Comparative meta transcriptomics uncovers the physiological responses of zooplankton to nutrient heterogeneity in their living environment

Muhammad Firman Nuruddin[1,2], Longjun Wu*[1,2]

[1]Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong SAR, China.

[2]Center for Ocean Research in Hong Kong and Macau, Hong Kong SAR, China.

*Corresponding author, E-mail: longjunwu@ust.hk

Abstract

Zooplankton plays a crucial role in coastal water ecosystems by linking energy transfer from primary producers to higher trophic level organisms and facilitating the biogeochemical cycle through their metabolism and vertical migration. However, our understanding of the zooplankton physiological response to nutrient variability in their living environment, particularly at the gene expression level, remains limited. This knowledge gap hinders our understanding of their physiological processes, which are essential for the biogeochemical cycle. To investigate the impact of coastal environmental nutrient heterogeneity on zooplankton community physiology, we conducted zooplankton samplings during the summer season in the western, southern, and eastern Coastal Hong Kong Waters. Meta transcriptomics analysis was employed to unravel the community composition and gene expression profiles of the sampled zooplankton. Our study revealed distinct patterns of differential gene expression associated with nitrogen and amino acid metabolism functions within the sampled zooplankton communities across the different sampling regions. Furthermore, we found several KEGG pathway members of gene modules which are classified as metabolism functions that highly correlated with nutrient concentrations. Importantly, these gene expression patterns corresponded to the observed heterogeneity in dissolved inorganic nitrogen and phosphorous concentration within coastal waters environment. To further validate our findings, we conducted reanalysis on the zooplankton model organism, *Daphnia magna*, transcriptome data under laboratory nitrogen and phosphorous imbalanced treatment which the result supported the gene ontology functions identified in the zooplankton community meta transcriptome. Our results contribute to a better understanding of zooplankton's physiological processes, especially highlighting the importance of incorporating gene expression analysis in studying zooplankton's role in the biogeochemical cycle.

Introduction

The pivotal contribution of zooplankton to aquatic ecosystems, encompassing their profound influence on the structure and functioning of food webs, as well as the intricate dynamics of biogeochemical cycles, has been extensively documented[1–5]. An important aspect of their contribution to nutrient cycling lies in the production of sinking fecal pellets, which effectively transport nutrients such as nitrogen, phosphorus, and iron from surface waters to deeper layers[6–10]. Through nutrient recycling via grazing and metabolism, zooplankton possess the capability to shape the availability of crucial elements, such as phosphorus and nitrogen, thereby influencing primary productivity [11]. This process plays a crucial role in both nutrient export and redistribution within aquatic systems, contributing to the overall functioning and stability of these environments.

Despite advancements in understanding zooplankton's ecological functions, our knowledge of the physiological processes occurring within zooplankton in their natural environments remains limited[12,13]. Comprehending these metabolic activities is crucial, as zooplankton are highly susceptible to various environmental factors, including temperature, salinity, nutrient concentration, and food availability[14–17]. Such understanding is essential for elucidating their role in nutrient cycling within these ecosystems.

Coastal waters exhibit considerable variability in nutrient concentrations due to upwelling events, riverine inputs, and anthropogenic influences. These fluctuations can significantly impact the physiological processes and overall health of zooplankton communities. However, traditional ecological approaches have offered limited insights into the molecular mechanisms governing these responses. Therefore, there is an urgent need to explore gene expression patterns and metabolic responses in zooplankton to enhance our understanding of their physiological processes and their contributions to biogeochemical cycles.

The coastal waters of Hong Kong present a unique setting for such investigations, characterized by complex environmental dynamics[18–20]. The western part is heavily influenced by the Pearl River discharge, which introduces significant amounts of freshwater and nutrients, exacerbating eutrophication, particularly during the wet summer months[21]. In contrast, the eastern part is primarily affected by oceanic shelf waters from the South China Sea, leading to lower nutrient levels due to dilution by oceanic currents[22]. Furthermore, the influence of northeast and southeast monsoon winds adds complexity to the hydrodynamics and biogeochemical processes

in these waters[19]. Thus, Hong Kong's coastal environment serves as an excellent model to investigate how nutrient heterogeneity affects zooplankton physiology at the community level.

Recent advancements in next-generation sequencing technologies, along with the growth of sequence databases and improved data analysis methodologies, have revolutionized our ability to study the combined effects of various environmental parameters on communities in their natural habitats [12,23–25]. One such method is metatranscriptomics, which allows for the examination of gene expression at the community level. This state-of-the-art approach integrates mRNA transcripts from all individuals in a sample, providing vital information about the physiological status of the community and enhancing our understanding of metabolic responses to environmental conditions [26].

In this study, we analyzed the gene expression profiles of zooplankton community samples collected in August from the western, southern, and eastern coastal waters of Hong Kong, each characterized by distinct nutrient concentration dynamics. By employing a comparative metatranscriptomics approach and conducting downstream analyses—including functional enrichment and weighted gene co-expression network analysis—we aim to elucidate the intricate relationships between zooplankton community gene expression and environmental heterogeneity. Our findings contribute to a deeper understanding of the physiological roles of zooplankton in biogeochemical cycles, ultimately enhancing our knowledge of these vital aquatic ecosystems.

Materials and Methods

Zooplankton field sampling and sample processing

We collected zooplankton samples during cruises using a yacht in Hong Kong coastal waters from the surface (2 m depth, neustonic layer) and near-bottom (approximately 1 m from the seafloor) layers using a custom-made plankton pump equipped with a 150 μm mesh sieve. The sampling sites were selected based on the long-term monitoring stations of the Environmental Protection Department Hong Kong Monitoring Program. These sites represented the western, transitional, and eastern regions of the Hong Kong coastal waters, which are influenced by the Pearl River Estuary, the meeting region between estuarine and oceanic waters, and the East China Sea shelf waters, respectively (Figure 1).

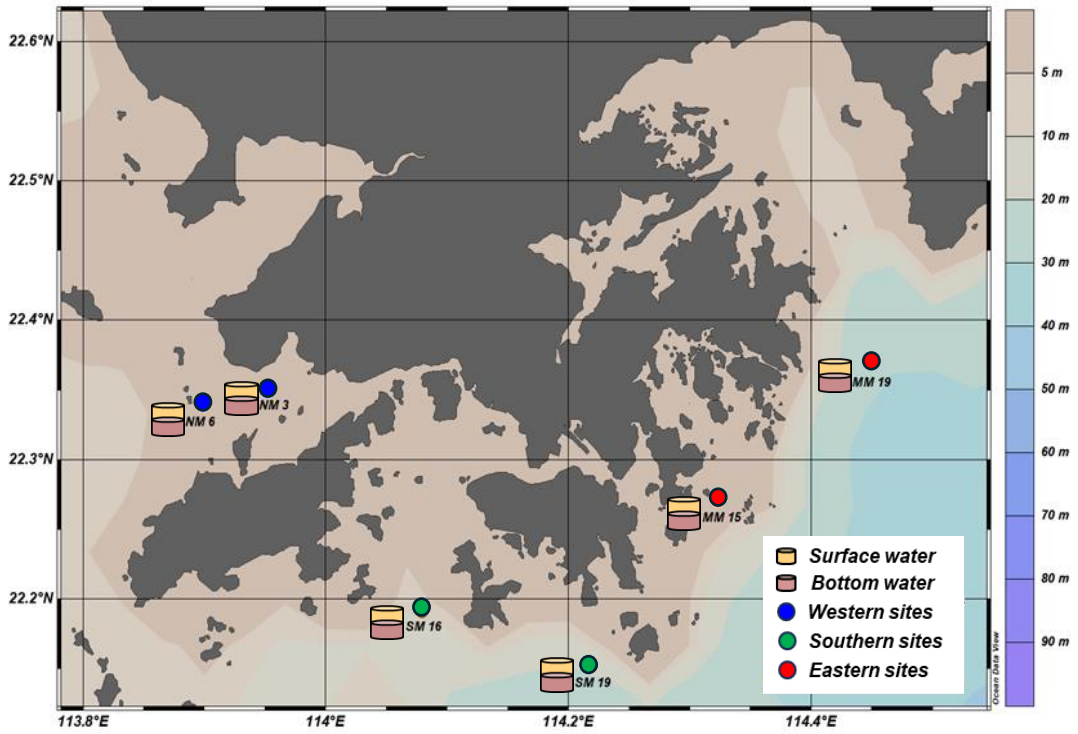

Figure 1. Illustrates the geographical locations of the six sampling stations in the coastal waters of Hong Kong. The sampling locations include MM 15 and MM 19 in the eastern region (marked in red), SM 16 and SM 19 in the southern region (marked in green), and NM 3 and NM 6 in the western region (marked in blue). In each sampling station we collect the samples from the surface (orange) and bottom (brown) water layer.

The zooplankton samples were collected by deploying a plankton pump for 15 minutes. The mesh size and sampling duration were chosen to match previous zooplankton surveys conducted in the Hong Kong coastal region and other comparable areas [27]. We did filtration using 10 μm nylon membrane to separate the zooplankton samples from the seawater. The zooplankton-containing filters were then preserved in Invitrogen™ RNAlater™ Stabilization Solution and stored at -80°C until processing for RNA extraction in the laboratory. Concurrently, water quality parameters including temperature, salinity, dissolved oxygen, and chlorophyll-a were measured at each sampling station using a multi-parameter CTD-DO prior to zooplankton sampling. Additionally, this study utilized complementary environmental data collected by the Environmental Protection Department, encompassing a range of parameters such as ammonia nitrogen (mmol/m³), chlorophyll-a (μg/L), dissolved oxygen (mmol/m³), E. coli (cfu/100mL),

nitrate nitrogen (mmol/m³), nitrite nitrogen (mmol/m³), orthophosphate phosphorus (mmol/m³), pH, phaeo-pigments (µg/L), Secchi disc depth (m), silica (mmol/m³), total nitrogen (mmol/m³), total phosphorus (mmol/m³), turbidity (NTU), unionized ammonia (mmol/m³), and volatile suspended solids (mmol/m³). These comprehensive environmental data were instrumental in further characterizing the heterogeneity of Hong Kong's coastal waters.

RNA extraction, library preparation, and sequencing

The collected zooplankton samples preserved in RNAlater were used for zooplankton total RNA extraction by using FastPure® Cell/Tissue Total RNA Isolation Kit V2 following the manufacturer's protocol. The bulk zooplankton pellets were added to Vazyme lysis reagent (Vazyme) and homogenized using RNase-free pestle. The FastPure® gDNA-Filter Column III and DNAse treatment were used to remove DNA contamination. Total RNAs were then cleaned and concentrated with FastPure® RNA Column III (Vazyme). For the purpose of RNA quantity and quality checking, Qubit, Biodrop, and Gel Electrophoresis were used prior to library creation and sequencing.

In the RNA sequencing step, each sample was processed by isolating mRNA from one microgram of total RNA through poly-T oligo-attached magnetic bead selection. Subsequently, the mRNAs were reverse-transcribed into cDNAs to construct RNA sequencing libraries with average insert sizes ranging from 250 to 300 bp, utilizing TruSeq™ RNA Sample Prep Kits (Illumina) in accordance with the manufacturer's guidelines. To ensure robustness and account for biological variability, two to three biological replicates were included for each sample collected from the Eastern, Western, and Southern Hong Kong Coastal Waters, resulting in a total of 34 libraries. The metatranscriptomes were subjected to pair-end sequencing, producing 150 bp reads using the HiSeq™ 2000 sequencing platform (Illumina) following standard protocols.

Meta transcriptome de novo assembly and transcript abundance

The raw data quality was evaluated using FastQC Version 0.11.5 [[28], followed by quality trimming of the raw reads for poor quality bases and RNA-Seq adapters using Trimmomatic. The resulting clean reads from all 34 libraries were utilized to construct the meta-assembly employing Trinity software with default setting [29]. The assembled sequenced then being used for open reading

frame prediction by using transdecoder against pfam database. Subsequently, the read alignment and transcript abundance estimation was performed by using salmon-based approach. In brief, the reads are first aligned to the assembled contigs using Bowtie2. The output BAM file is then used as the input for salmon tool. Normalization of transcript per million (TPM) was performed based on library sizes using the “calcNormFactors (method = "TMM")” function to ensure accurate quantification of gene expression levels across the samples [30].

Functional annotation

To obtain functional annotations for the assembled contigs, we utilized the Trinotate pipeline [31]. This process also needs protein sequence as one of the inputs for the functional annotations, consequently we used Transdecoder software packaged with Trinity to predict the protein translations from the contig. Subsequently, the unigenes obtained were queried against several public databases, including SwissProt, Pfam-A, and eggnog, with an E-value threshold of less than 1E-5 using the Trinotate pipeline. In this annotation process, BLAST 2.9.0+ was employed for gene assignments, using BLASTp for protein queries and BLASTx for nucleotide queries [32]. Both query sets were compared against the Uniprot/SwissProt database of reference proteins [33]. To identify conserved domains within the protein models, we utilized HMMER v.3.3 and the Pfam-A database.[34,35] The results were then loaded into Trinotate's SQL database. Trinotate used the top BLAST hits to populate the database with gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms.

Statistical analysis

Principal component analyses were performed on pre-processed log-transformed transcripts using R package. The arrows in the PCA plot represent the correlation between the first two principal components and environmental parameters. Subsequently we perform differential gene expression (DEG) analysis, by the Bioconductor R package edgeR[36,37]. Prior to the DEG analysis, transcript abundance normalization was performed using the TMM (Trimmed Mean of M-values) method[38]. The resulting p-values were adjusted for multiple testing using the Benjamini-Hochberg method. In this study, we considered genes with an adjusted p-value (Padj) less than 0.05 and an absolute log2-fold change greater than 2 as differentially expressed genes (DEGs) among the zooplankton datasets obtained from various sampling sites. To gain insights

into the functional profiles of the DEGs, we conducted gene ontology term enrichment analysis using the GOSeq R package[39]. The FDR approach was employed to correct the significant level of enrichment for multiple testing, and terms with a significant level below the threshold of 0.05 were considered significantly enriched by the DEGs.

To further explore the gene function and its association with environmental variables and zooplankton community relative abundances, we also performed a weighted gene co-expression network analysis (WGCNA) using the WGCNA package (version 1.72-1) in R[40]. This approach was modified from the pipeline described by Cohen et al.[41]. The expression matrix (TMM) of the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotated transcript was used as the input file for this analysis, with the aim of identifying gene modules with strong co-expression patterns. The WGCNA network construction and module detection were conducted using an unsigned type of topological overlap matrix (TOM) and the TOM-based dissimilarity measure (1-TOM). The soft threshold power β, based on the scale-free topology criterion, was set at 30, the minModuleSize at 70, and the mergeCutHeight at 0.4[42]. The goal was to identify co-expression modules of KEGG KO identifiers from normalized transcript per millions sums (TPM) of KEGG KOs over annotated transcripts. We calculated module eigengenes and assessed their correlations with environmental parameters and taxonomic groups TPM abundances. To identify metabolic pathways associated with a given module eigengene, we performed Mann-Whitney U tests to assess which pathways were more represented at the top of a ranked list of KEGG KO identifiers, ordered according to decreasing module membership, than expected by chance. Module membership of a pathway was defined as the Pearson correlation between module eigengene expression and KEGG KO identifier. Resulting p-values were adjusted for multiple testing with the Benjamini-Hochberg procedure, which controls the false discovery rate (FDR)[42,43].

Results and Discussion

Environmental heterogeneity in the Hong Kong coastal waters

Concurrent measurements of environmental parameters revealed differences in the vertical profiles of temperature, salinity, dissolved oxygen, and chlorophyll-a across the sampling sites (Figure 2 (a) to (c)). Generally, the characteristics of water parameters in the Eastern and Southern sites exhibit more similar vertical profiles, while distinct patterns were observed in the measurements taken at the Western sites. Regarding dissolved oxygen concentration and

chlorophyll-a, the Western sites of Hong Kong waters have the lowest values compared to the other regions (Figure 2 (b)). This result indicates that despite relatively narrow region, Hong Kong coastal waters pose a heterogeny environmental characteristic as this region located on the southern coast of China, lying southeast of the Pearl River estuary (PRE) and facing the northwestern part of the South China Sea (SCS). The western part of Hong Kong's coastal waters is primarily influenced by PRE river discharge which gives a supply of freshwater containing large amounts of nutrients. Meanwhile, the eastern part is dominantly affected by the oceanic shelf waters from the South China Sea, which make the nutrient levels there substantially lower than in western waters due to the dilution of oceanic currents [20,44].

To further understand the concentrations of nitrogen and phosphorus in the study regions, we have collected marine water quality data from the Hong Kong Environmental Protection Department as shown in the Figure 2 (d). The results indicate that the eastern sampling sites, both in the surface and bottom waters, exhibit the lowest levels of nitrogen and phosphorus concentrations, as well as the lowest ratio of nitrogen to phosphorus (N:P). In contrast, the western sites tend to have significantly higher nutrient levels (nitrogen and phosphorus) and N:P ratios. The southern sites, located in the transition zone, show high N:P ratios in the surface water but low ratios in the bottom level, with relatively lower concentrations of nitrogen and phosphorus compared to the western sites. Based on this result, nitrogen (N) is potentially the limiting nutrient in the eastern Hong Kong coastal waters, while phosphorus (P) appears to be the major limiting nutrient in the western and southern Hong Kong waters. These data in line with the previous study which explained that there were clear spatial variations of the nutrient concentration in this region during summer season with mostly P limitation happened in the Southern and Western meanwhile co limitation of N and P in the Eastern sites [45].

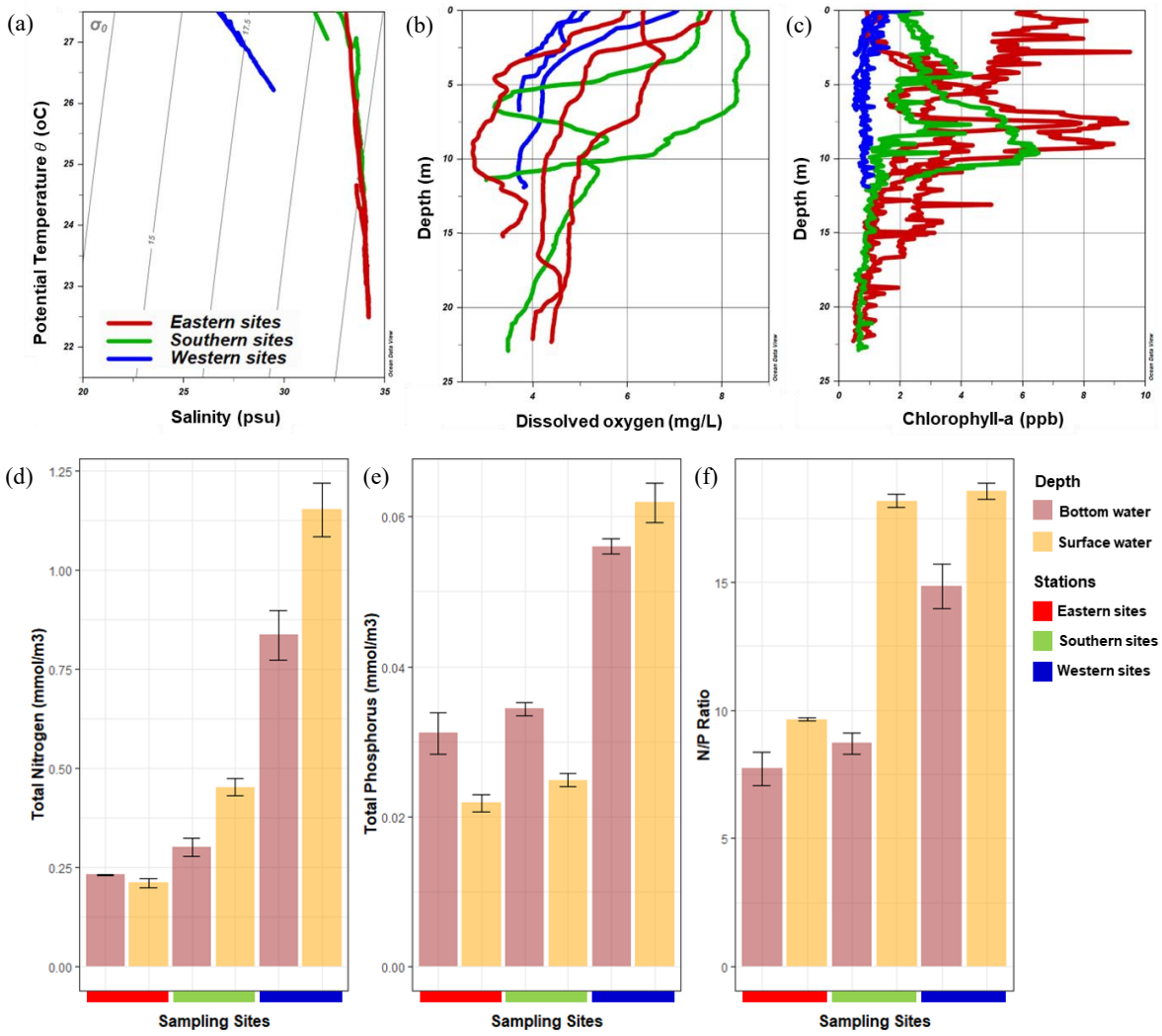

Figure 2 Vertical profile of the (a) potential temperature and salinity, (b) Dissolved Oxygen, (c) chlorophyll-a concentration. (d) Total nitrogen (mmol/m^3), (e) phosphorous (mmol/m^3), and (f) the ratio of nitrogen over phosphorous concentration from the Eastern (red), Southern (green), and Western (blue) Hong Kong sites. The color in the bar represents bottom (brown) and surface (orange) water layers the depth in which the samples were collected.

Exploratory analysis of the zooplankton gene expression

In order to investigate the clustering patterns of zooplankton community gene expression profiles, we initiated our analysis by conducting principal component analysis (PCA). The PCA computation unveiled that the first two principal components collectively capture 50.84% of the

total variance present in the zooplankton gene expression dataset. As illustrated in Figure 3, the PCA plot vividly illustrates discernible clustering patterns among the samples. Notably, samples collected from the western sites, both at the surface and bottom levels, exhibit a cohesive grouping within one cluster, whereas samples from the eastern and southern sites tend to group in a separate component displaying a more dispersed arrangement. This clustering phenomenon resonates with the spatial heterogeneity observed in the environmental parameters across the Hong Kong coastal waters, indicating that the gene expression profiles of the zooplankton community may mirror the spatial disparities within their habitat.

The clustering pattern of zooplankton community gene expression identified in our study stands in stark contrast to a previous investigation on population-level transcriptome analysis of Copepods (*Temora longicornis*) conducted in the Belgian part of the North Sea (BPNS), where no distinct spatial variation was evident[46]. The heightened environmental heterogeneity within the Hong Kong coastal waters, alongside the inherent differences among zooplankton community members, likely underlie the observed pattern dissimilarity at the genetic level. This discrepancy aligns with findings from a prior study on *Neocalanus flemingeri* in the northern Gulf of Alaska, where clustering patterns were discerned in response to regional heterogeneity among sampling sites [47]. The intricate interplay between environmental factors and genetic expression patterns in zooplankton communities underscores the need for further exploration to unravel the underlying mechanisms governing their spatial dynamics and ecological responses.

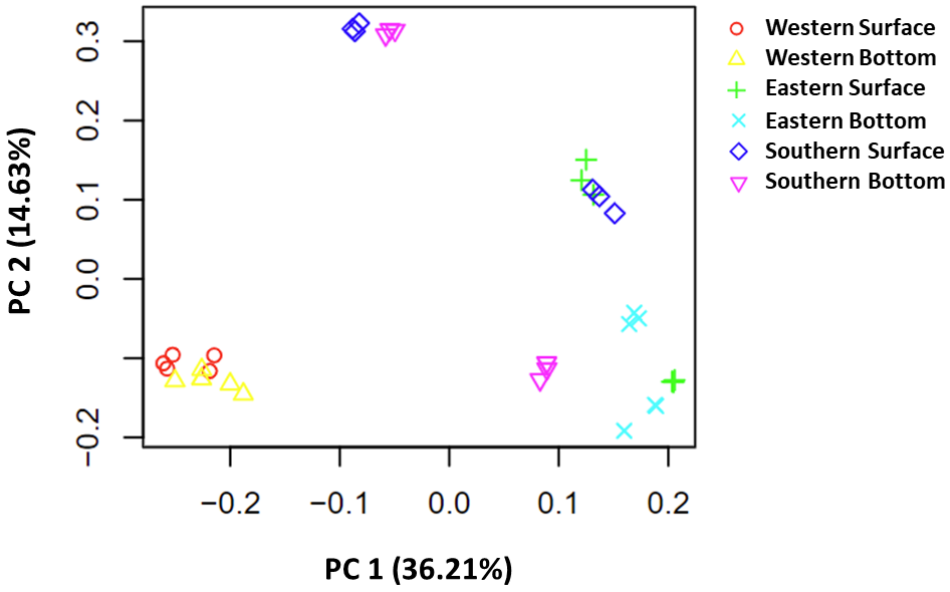

Figure 3 Principal component analysis of the zooplankton community transcript abundance. The symbol inside the plot represents the samples which consists of western, eastern, southern surface and bottom, respectively.

Zooplankton community gene expression pattern in response to heterogeneity of the nutrient concentration in Hong Kong Coastal waters

We perform gene ontology (GO) enrichment analysis to the zooplankton community differentially expressed genes (DEG) to further identify the significantly overrepresented biological processes, molecular functions, and cellular components within the zooplankton community's gene expression profiles. This analysis involved three comparisons of the sample's scenarios: eastern versus western surface waters, eastern versus southern surface waters, and western versus southern surface waters. These scenarios were designed to elucidate the zooplankton's responses to nitrogen and phosphorus limitation and to determine whether nutrient concentrations or ratios (N:P) are more important in affecting zooplankton physiological responses. Overall, our findings reveal a pattern of enrichment of the gene expression associated with nutrients (nitrogen and phosphate), amino acid, and oxygen metabolism related functions. These patterns corresponded to the observed heterogeneity in nutrient concentration within the zooplankton living environment.

Figure 4 showcases the outcomes of the first comparison scenario, contrasting samples obtained from the surface waters of the Western and Eastern study sites. The western sites exhibit a relative phosphorus limitation, characterized by elevated absolute concentrations of nitrogen (N) and phosphorus (P), whereas the eastern sites manifest nitrogen limitation, marked by diminished absolute levels of both N and P. The enrichment analysis uncovers several significantly enriched cellular component gene ontologies related to ribosomal subunits and the cytoplasm. Furthermore, molecular function gene ontologies associated with structural molecule activity and the structural constituents of ribosomes are notably enriched. Among the top 30 most significantly enriched GO terms, biological processes encompass functions associated with the metabolism and biosynthesis of phosphate, organonitrogen compounds, cellular nitrogen compounds, amides, macromolecules, and translation.

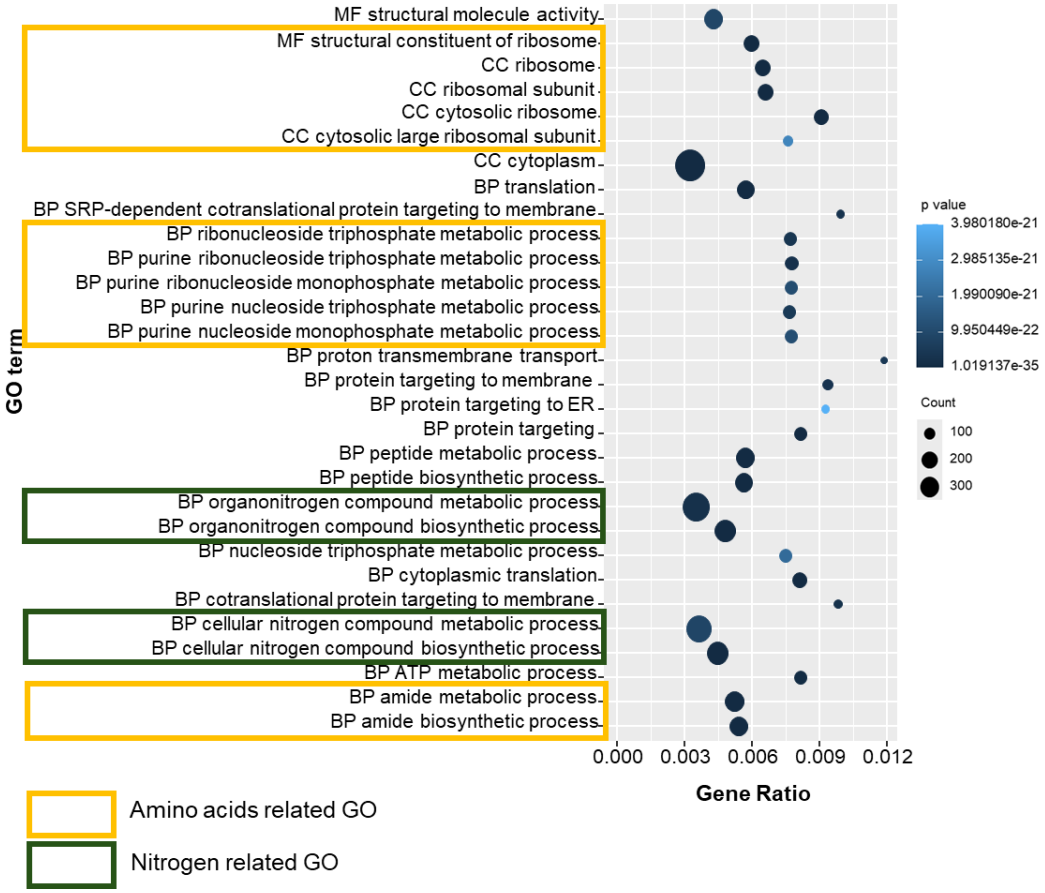

Figure 4 The top 30 most significant zooplankton gene ontology (GO) enrichment results from the comparison of samples collected from the surface of Western and Eastern sites.

BP represents biological process, CC cellular component, and MF means molecular function.

The zooplankton GO enrichment analysis results presented in Figure 5 compare samples from the Western and Southern surface water study sites, representing the second set of comparison scenarios. The western surface exhibits phosphorus limitation, characterized by elevated concentrations of both nitrogen (N) and phosphorus (P), whereas the Southern surface water experiences phosphorus limitation with relatively lower absolute levels of N and P. Overall, the GO enrichment patterns observed in this second scenario are similar to the first comparison scenario (Figure 5). Significantly enriched cellular component gene ontologies include ribosomal subunits, protein complexes, and mitochondria. Enriched biological processes encompass metabolism and biosynthesis of peptides, organonitrogen compounds, cellular nitrogen compounds, amides, macromolecules, and translation. At the molecular function GO level, structural molecule activity and structural constituents of ribosomes are also notably enriched.

The GO enrichment profiles derived from the first two comparison scenario further adding our comprehension of how zooplankton respond physiologically toward nutrient imbalance in their natural habitat, more specifically coastal water environments. Previous studies have indicated that differentially expressed Gene Ontologies (GOs) in Copepods are associated with functions related to ribosomal activity, lipid synthesis, glycolysis, and translation[46–48]. However, none of these studies have identified functions linked to nitrogen and amide biosynthesis. Zooplankton rely on nitrogen and phosphorus obtained through their prey to fuel essential biological processes such as cellular construction, growth, amino acid and protein synthesis, nucleic acid production, and overall survival[17]. The capacity of zooplankton to retain and utilize these nutrients is critical for supporting their metabolic activities and facilitating nutrient cycling within aquatic ecosystems. Understanding the intricate interplay between nutrient acquisition, utilization, and metabolic processes in zooplankton is fundamental to elucidating their ecological roles and the broader functioning of marine ecosystems.

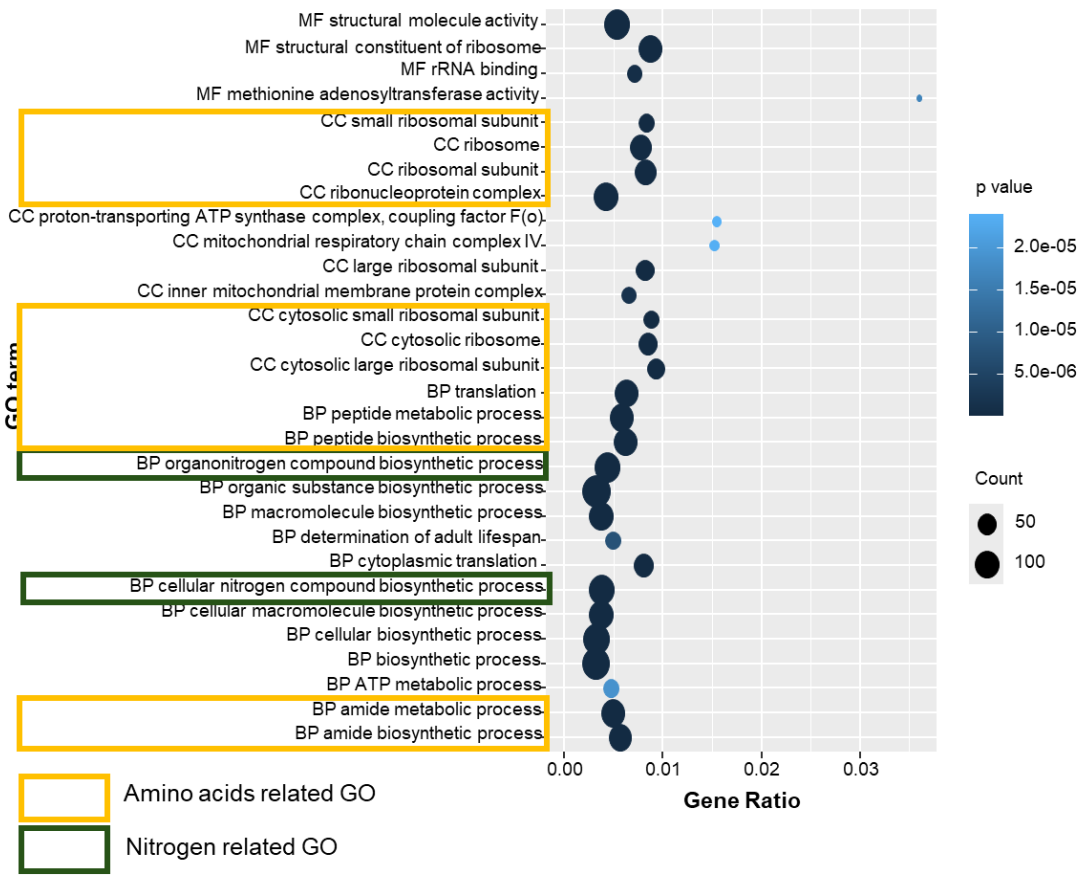

Figure 5 The top 30 most significant zooplankton gene ontology (GO) enrichment results from the comparison of samples collected from the surface of Western and Southern sites. BP represents biological process, CC cellular component, and MF means molecular function.

In the third comparison scenario, we conducted an enrichment analysis by contrasting samples from the surface layer of Eastern and Southern Hong Kong coastal waters (Figure 6). These regions represent distinct nutrient limitations, nitrogen and phosphorus limitations, respectively, with similar absolute nutrient values (low nitrogen and phosphorus). A unique pattern of gene ontology enrichment profiles emerged in this scenario, deviating from the previous two scenarios. The cellular component gene ontologies related to respiratory organelles, such as mitochondria, respiratory chain, cytochrome complex, and respirasome[49,50], were mostly enriched, rather than those related to ribosomes. Furthermore, molecular function and biological processes related to respiration, rather than nutrient metabolism as observed in the prior scenarios, were enriched. These results collectively indicate that the absolute concentrations of nutrient levels

(phosphorus and nitrogen) exert a more dominant influence on zooplankton gene expression profiles than nutrient ratios. This finding aligned with the previous study which has demonstrated the crucial role of certain nutrients in supporting the fundamental biological structures and core functions of organisms. For instance, nitrogen is a vital element for the synthesis of nucleic acids and amino acids, while phosphorus is a key component of rRNA, which is essential for ribosome biogenesis and, consequently, cellular and organismal growth [51].

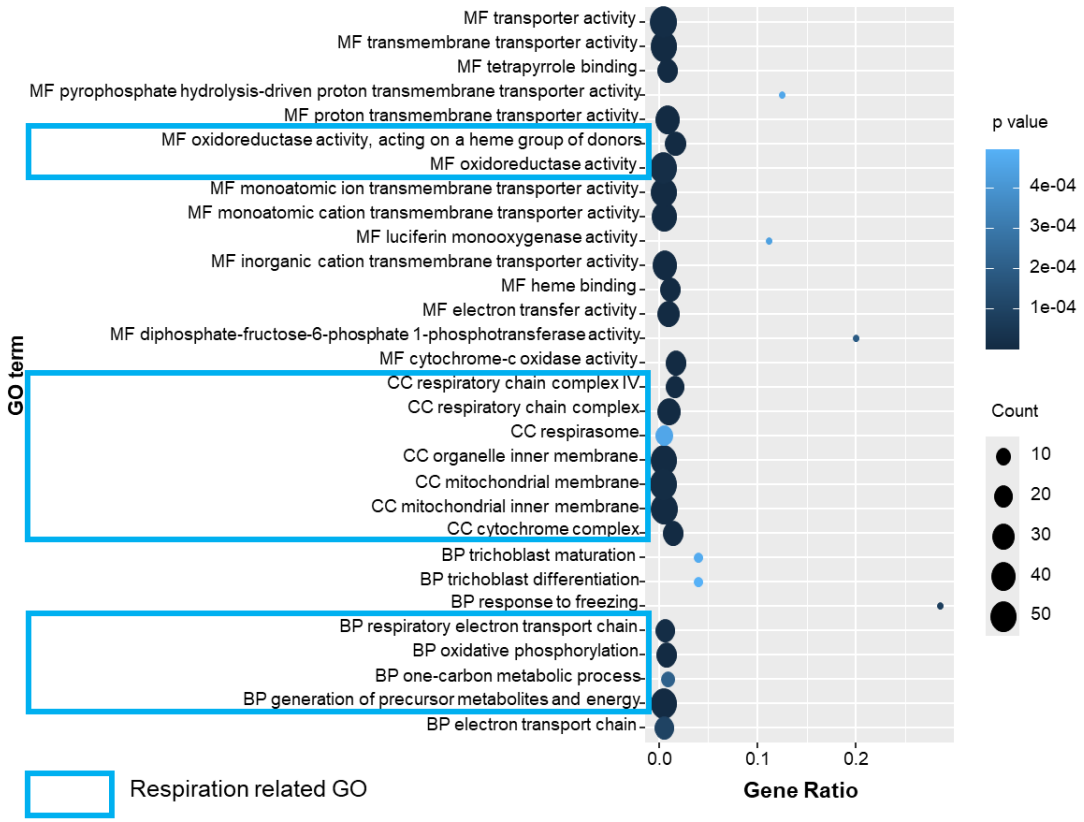

Figure 6 The top 30 most significant zooplankton gene ontology (GO) enrichment results from the comparison of samples collected from the surface of Southern and Eastern sites. BP represents biological process, CC cellular component, and MF means molecular function.

The dissolved oxygen concentration varies across the sampling regions, with Figure 2 (c) highlighting this discrepancy, showcasing the Southern sites with the highest dissolved oxygen levels compared to the Eastern and Western sites. These oxygen level variations trigger diverse physiological responses in zooplankton respiration processes. This is evident in the top thirty gene ontology (GO) enrichment results displayed in Figure 6, where all these top thirty GO terms are

associated with functions related to respiration. This outcome provides deeper insights into the intricate mechanisms through which zooplankton contribute to the carbon cycle via the respiration process, warranting further investigation[3].

Weighted gene co-expression profiles and its association with environmental parameters

The relationship between the KEGG-annotated transcripts of the zooplankton community and environmental factors was elucidated through a gene network analysis of meta-transcriptomic data. Correlation analyses between module eigengenes and the environmental parameters unveiled several significant associations as depicted in Figure 7. The darkolivegreen gene modules exhibited clear positive correlations with nutrient concentrations, including ammonia nitrogen, nitrite nitrogen, nitrate nitrogen, total nitrogen, total phosphorus, silica, and orthophosphate phosphorus ($p < 0.05$). In contrast, the greenyellow module showed a negative association with nutrient levels. These patterns suggest a physiological response of the zooplankton community to fluctuations in nutrient availability, further supporting the previous pattern revealed by GO enrichment analysis. Additionally, the floralwhite module was positively correlated with *E. coli* concentration, while the grey module was negatively associated with temperature and dissolved oxygen levels ($p < 0.01$). Conversely, there are no clear associations observed between environmental variables and several other modules, such as lightsteelblue1, brown4, and darkorange2.

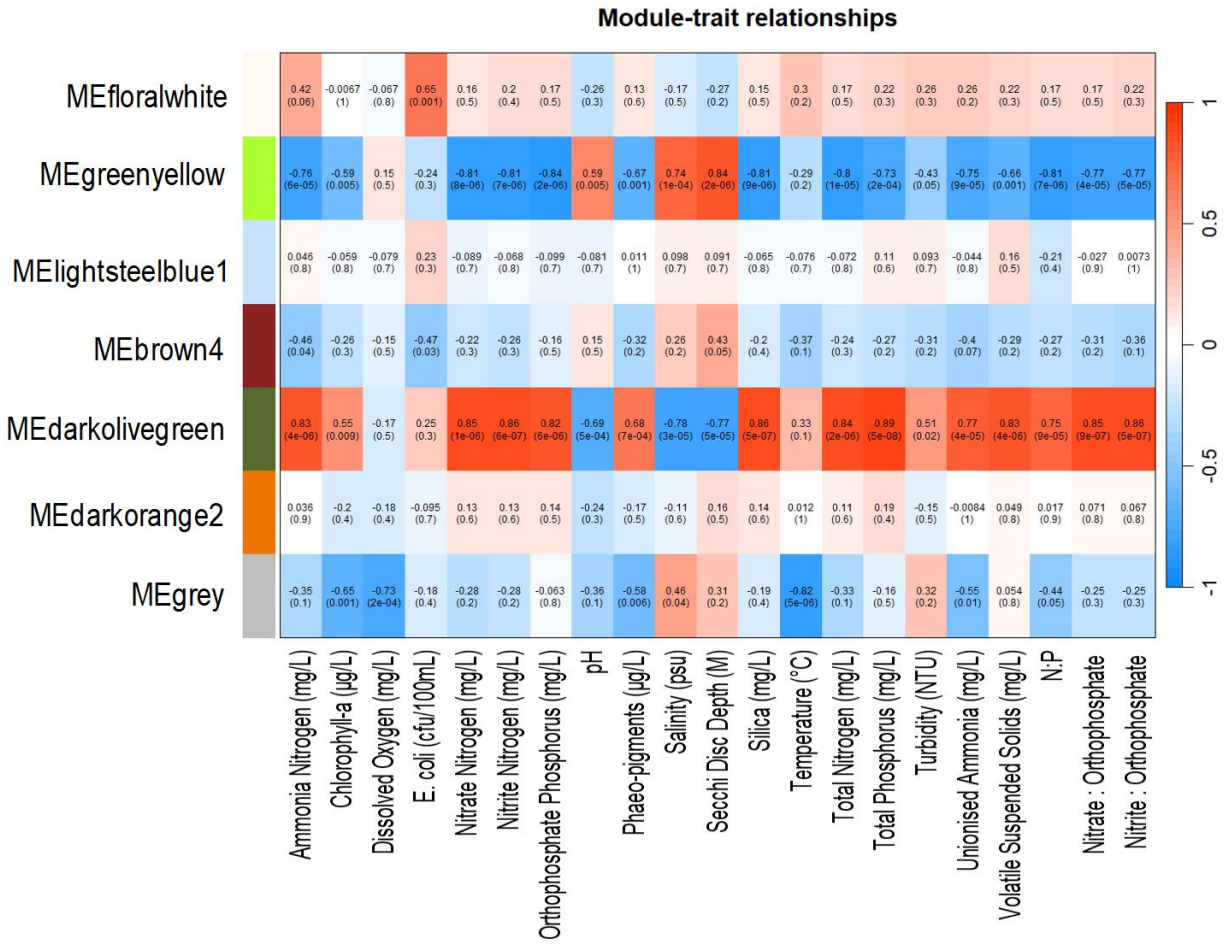

Figure 7 Correlation between module eigengene expression and environmental parameters. The upper part number and color gradients represents Pearson's correlation coefficients while the below part number indicate the statistical significance (inside the bracket).

Subsequent analysis using Mann-Whitney U test was performed to further identify the key KEGG members within the gene modules that exhibited clear associations with nutrients. We found that KEGG pathway related to several metabolism processes such as amino acids, energy, carbohydrate, glycan, cofactors and vitamins, terpenoid and polyketides, and xenobiotics degradations are the key members of darkolivegreen module (Figure 8). This result consistent with the previous study on Copepods population transcriptome collected from BPNS region which also found that the KEGG pathways annotated sequences involved the metabolic pathways and glycolytic metabolism[46,48].

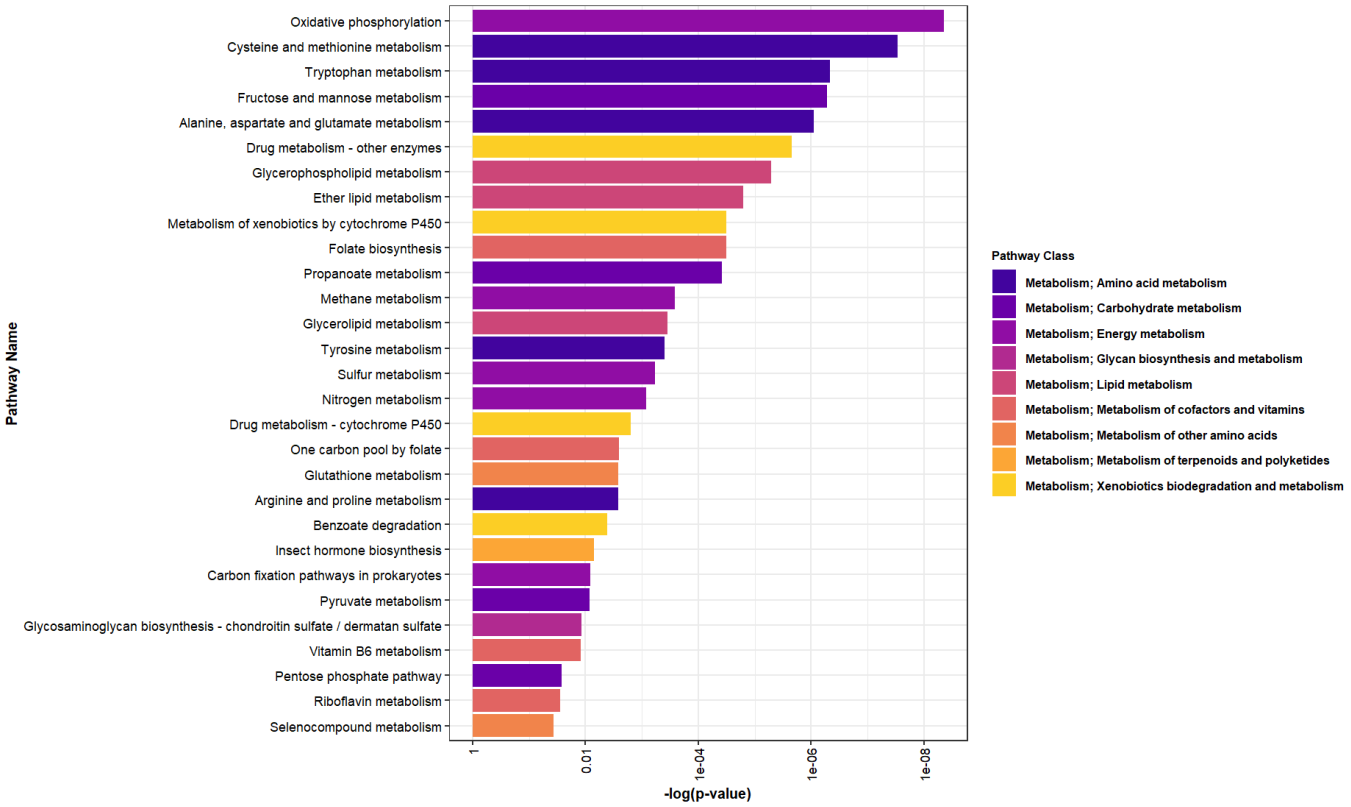

Figure 1 KEGG Pathways members of darkolivegreen gene modules result from Mann Whitney U Test. The color denoting the pathway class categories.

Validation through Daphnia magna transcriptome analysis

Previous study conducted by Zu et. al, have tried to elucidate the transcriptome response of zooplankton (*Daphnia magna*) towards nutrient limited (nitrogen and phosphorous algae feeding treatment) food under laboratory settings experiments. To further validate our gene expressions profiles findings with the transcriptome study under simulated nutrient conditions environment, we conducted GO enrichment analysis on the that transcriptome data [52] which the result is shown in Figure 9. We compare the gene ontology enrichment of the samples from phosphorous limited food treatment against the control, nitrogen limited food treatment against the control, and between nitrogen and phosphorous limited treatment. Among those comparisons, we found that the GO enrichment results from the control vs phosphorous limited food showing the enrichment profiles which align with the GO enrichment functions identified in the zooplankton community meta transcriptome (Figure 4 and 5). Specifically, those enriched GO include several

cellular components such as ribosome and ribonucleoprotein complex, molecular function related to structural constituent of ribosome, and biological processes such as peptide, cellular nitrogen, amide, and organonitrogen biosynthesis process. Moreover, several GOs related to chitin-based cuticle also show enrichment pattern.

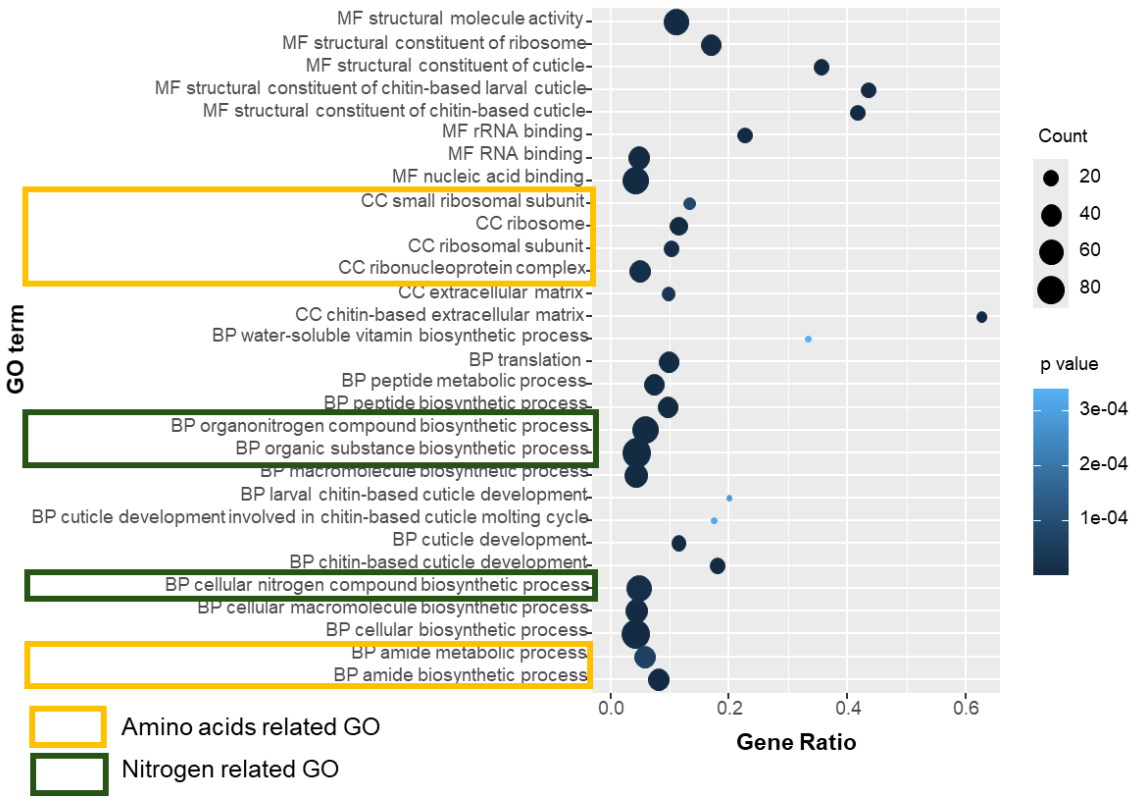

Figure 2 Top thirty gene ontology enrichment results from *Daphnia magna* control group compared to phosphorous limited food treatment. BP represents biological process, CC cellular component, and MF means molecular function.

Numerous studies have been conducted to investigate the phenotypic and physiological responses of zooplankton to environmental changes, particularly nutrient imbalances. However, these studies have predominantly focused on analyzing individual zooplankton species rather than exploring gene expression patterns at the community level. Our study breaks new ground by delving into the collective gene expressions within the zooplankton community, shedding light on their functional profiles in response to fluctuations in nutrient concentrations within their habitat. These findings are in alignment with controlled laboratory experiments simulating nutrient imbalances. While our study uncovers intriguing patterns in Gene Ontology (GOs) and Kyoto

Encyclopedia of Genes and Genomes (KEGG) pathways, we acknowledge certain limitations in our analysis. The absence of phytoplankton Carbon:Nitrogen:Phosphorous (C:N:P) ratio data poses a challenge in providing definitive evidence of nutrient imbalances, given zooplankton's dependence on primary producers for essential nutrients. Furthermore, the incomplete taxonomy annotation of zooplankton based on metatranscriptomics due to the eukaryotes database limitations hinders our ability to accurately discern species-specific responses towards environmental heterogeneity within the metatranscriptome data. Addressing these limitations will be crucial for advancing our understanding of zooplankton responses to changing environmental conditions and further elucidating their intricate gene expressions dynamics within aquatic ecosystems.

Conclusion

In our metatranscriptomic study, we have uncovered a distinct pattern of differential gene expression enrichment in zooplankton, specifically associated with nutrient, amino acid biosynthesis, and respiration functions. The gene ontology enrichment profiles, particularly those related to nutrient and respiration processes, along with the identified pathway members of gene modules linked to metabolism functions, align closely with the observed variations in nutrient concentrations and dissolved oxygen levels in their natural habitat. These findings underscore the significant impact of nutrient and oxygen availability on zooplankton's metabolic activities.

To validate our results, we conducted a reanalysis of transcriptome data from the model organism *Daphnia magna* under laboratory nitrogen and phosphorous imbalanced treatments. The outcomes of this analysis supported the gene ontology functions identified in the zooplankton community metatranscriptome, emphasizing the importance of gene expression analysis in understanding the role of zooplankton in the biogeochemical cycle.

The complex interplay between the transcriptional profiles of the zooplankton community and environmental conditions emphasizes the intricate ecological mechanisms that govern their responses. By elucidating the molecular underpinnings of zooplankton responses to environmental heterogeneity, our study provides valuable insights into the physiological response strategies employed by these organisms in heterogenous coastal waters environments. This deeper understanding also contributes to a better understanding of zooplankton’s physiological processes, especially the importance of incorporating gene expression analysis when studying zooplankton community’s role in the biogeochemical cycle.

Acknowledgements

We would like to express our sincere gratitude for the funding support from Center for Ocean Research in Hong Kong and Macau (CORE), the joint research centre for ocean research between Qingdao National Laboratory for Marine Science and Technology (QNLM) and Hong Kong University of Science and Technology (HKUST).

References

1. Ducklow, H. W., Steinberg, D. K. & Buesseler, K. O. Upper ocean carbon export and the biological pump. *Oceanography* **14**, 50–58 (2001).

2. Steinberg, D. Zooplankton biogeochemical cycles. *Marine Plankton: A practical guide to ecology, methodology, and taxonomy* **1**, (2017).

3. Steinberg, D. K. & Landry, M. R. Zooplankton and the ocean carbon cycle. *Ann Rev Mar Sci* **9**, 413–444 (2017).

4. Valdés, V. P., Fernandez, C., Molina, V., Escribano, R. & Joux, F. Dissolved compounds excreted by copepods reshape the active marine bacterioplankton community composition. *Front Mar Sci* **4**, 343 (2017).

5. Vanni, M. J. Nutrient cycling by animals in freshwater ecosystems. *Annu Rev Ecol Syst* **33**, 341–370 (2002).

6. Corner, E. D. S. & Cowey, C. B. Biochemical studies on the production of marine zooplankton. *Biological Reviews* **43**, 393–426 (1968).

7. James, M. R. Role of zooplankton in the nitrogen cycle off the west coast of the South Island, New Zealand, winter 1987. *N Z J Mar Freshwater Res* **23**, 507–518 (1989).

8. Nuester, J., Shema, S., Vermont, A., Fields, D. M. & Twining, B. S. The regeneration of highly bioavailable iron by meso-and microzooplankton. *Limnol Oceanogr* **59**, 1399–1409 (2014).

9. Urabe, J. N and P cycling coupled by grazers' activities: food quality and nutrient release by zooplankton. *Ecology* **74**, 2337–2350 (1993).

10. Urabe, J., Nakanishi, M. & Kawabata, K. Contribution of metazoan plankton to the cycling of nitrogen and phosphorus in Lake Biwa. *Limnol Oceanogr* **40**, 232–241 (1995).

11. Schenone, L. *et al.* Modelling Key Variables for Understanding the Effects of Grazing and Nutrient Recycling by Zooplankton on the Freshwater Microbial Loop. *Freshw Biol* **66**, 2322–2337 (2021).

12. Lenz, P. H., Lieberman, B., Cieslak, M. C., Roncalli, V. & Hartline, D. K. Transcriptomics and metatranscriptomics in zooplankton: wave of the future? *J Plankton Res* **43**, 3–9 (2021).

13. Richardson, A. J. In hot water: zooplankton and climate change. *ICES Journal of Marine Science* **65**, 279–295 (2008).

14. Kwiatkowski, L., Aumont, O. & Bopp, L. Consistent trophic amplification of marine biomass declines under climate change. *Glob Chang Biol* **25**, 218–229 (2019).

15. Niu, Y., Wei, H., Zhang, Y. & Su, J. Transcriptome response of a marine copepod in response to environmentally-relevant concentrations of saxitoxin. *Mar Pollut Bull* **205**, 116546 (2024).

16. Li, Y. *et al.* Eutrophication and predation mediate zooplankton diversity and network structure. *Limnol Oceanogr* **67**, S133–S145 (2022).

17. Thomas, P. K., Kunze, C., Van de Waal, D. B., Hillebrand, H. & Striebel, M. Elemental and biochemical nutrient limitation of zooplankton: A meta-analysis. *Ecol Lett* **25**, 2776–2792 (2022).

18. Lai, R. W. S. *et al.* Hong Kong's marine environments: History, challenges and opportunities. *Reg Stud Mar Sci* **8**, 259–273 (2016).

19. Yin, K. Monsoonal influence on seasonal variations in nutrients and phytoplankton biomass in coastal waters of Hong Kong in the vicinity of the Pearl River estuary. *Mar Ecol Prog Ser* **245**, 111–122 (2002).

20. Lee, J. H. W., Harrison, P. J., Kuang, C. & Yin, K. Eutrophication dynamics in Hong Kong coastal waters: Physical and biological interactions. in *The Environment in Asia Pacific Harbours* 187–206 (2006). doi:10.1007/1-4020-3655-8_13.

21. Cai, Z., Liu, G., Liu, Z. & Gan, J. Spatiotemporal variability of water exchanges in the Pearl River Estuary by interactive multiscale currents. *Estuar Coast Shelf Sci* **265**, 107730 (2022).

22. Tao, W., Niu, L., Dong, Y., Fu, T. & Lou, Q. Nutrient pollution and its dynamic source-sink pattern in the pearl river estuary (South China). *Front Mar Sci* **8**, 713907 (2021).

23. Marchetti, A. *et al.* Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proceedings of the National Academy of Sciences* **109**, E317–E325 (2012).

24. Mojib, N., Thimma, M., Kumaran, M. K., Sougrat, R. & Irigoien, X. Comparative Metatranscriptomics Reveals Decline of a Neustonic Planktonic Population. *Limnol Oceanogr* **62**, 299–310 (2016).

25. Yadav, R. K., Bragalini, C., Fraissinet-Tachet, L., Marmeisse, R. & Luis, P. Metatranscriptomics of soil eukaryotic communities. *Microbial Environmental Genomics (MEG)* 273–287 (2016).

26. Shakya, M., Lo, C.-C. & Chain, P. S. G. Advances and challenges in metatranscriptomic analysis. *Front Genet* **10**, 904 (2019).

27. Wong, E., Yau, C. & Chan, K. Y. K. Seasonal and spatial dynamics of mesozooplankton community in a subtropical embayment. *Reg Stud Mar Sci* **56**, 102724 (2022).

28. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Preprint at (2010).

29. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).

30. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).

31. Bryant, D. M. *et al.* A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep* **18**, 762–776 (2017).

32. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, W20–W25 (2004).

33. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).

34. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29–W37 (2011).

35. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419 (2021).

36. Chen, Y., McCarthy, D., Robinson, M. & Smyth, G. K. edgeR: differential expression analysis of digital gene expression data User's Guide. *Bioconductor User's Guide* (2014).

37. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics* **26**, 139–140 (2010).

38. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, 1–9 (2010).

39. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, 1–12 (2010).

40. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

41. Cohen, N. R. *et al.* Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean. *Nat Microbiol* **6**, 173–186 (2021).

42. Perneel, M., Lagaisse, R., Mortelmans, J., Maere, S. & Hablützel, P. I. Seasonal Metabolic Dynamics of Microeukaryotic Plankton: A Year-long Metatranscriptomic Study in a Temperate Sea. *bioRxiv* 2021–2024 (2024).

43. Wright, R. M., Aglyamova, G. V, Meyer, E. & Matz, M. V. Gene expression associated with white syndromes in a reef building coral, Acropora hyacinthus. *BMC Genomics* **16**, 371 (2015).

44. Gan, J. *et al.* Assessing ecosystem response to phosphorus and nitrogen limitation in the Pearl River plume using the Regional Ocean Modeling System (ROMS). *J Geophys Res Oceans* **119**, 8858–8877 (2014).

45. Xu, J. *et al.* Temporal and spatial variations in nutrient stoichiometry and regulation of phytoplankton biomass in Hong Kong waters: influence of the Pearl River outflow and sewage inputs. *Mar Pollut Bull* **57**, 335–348 (2008).

46. Semmouri, I. *et al.* Spatio-temporal patterns in the gene expression of the calanoid copepod Temora longicornis in the Belgian part of the North Sea. *Mar Environ Res* **160**, 105037 (2020).

47. Roncalli, V., Cieslak, M., Germano, M., Hopcroft, R. R. & Lenz, P. H. Regional Heterogeneity Impacts Gene Expression in the Subarctic Zooplankter Neocalanus Flemingeri in the Northern Gulf of Alaska. *Commun Biol* **2**, (2019).

48. Semmouri, I., De Schamphelaere, K. A. C., Mees, J., Janssen, C. R. & Asselman, J. Evaluating the potential of direct RNA nanopore sequencing: Metatranscriptomics highlights possible seasonal

differences in a marine pelagic crustacean zooplankton community. *Mar Environ Res* **153**, 104836 (2020).

49. Wu, M. *et al.* Research journey of respirasome. *Protein Cell* **11**, 318–338 (2020).

50. Wikström, M., Pecorilla, C. & Sharma, V. Chapter Two - The mitochondrial respiratory chain. in *The Enzymes* (eds. Kaguni, L. S. & Tamanoi, F.) vol. 54 15–36 (Academic Press, 2023).

51. JEYASINGH, P. D. *et al.* How do consumers deal with stoichiometric constraints? Lessons from functional genomics using Daphnia pulex. *Mol Ecol* **20**, 2341–2352 (2011).

52. Xu, Z., Li, Y., Li, M. & Liu, H. Transcriptomic response of Daphnia magna to nitrogen-or phosphorus-limited diet. *Ecol Evol* **11**, 11009–11019 (2021).