NUTRIGENOMICS OF *DAPHNIA*: EXAMINING EFFECTS OF NUTRIENTS ON THE *DAPHNIA PULEX* TRANSCRIPTOME AND IDENTIFYING NUTRITIONAL PHENOTYPES

A Thesis Submitted to the Committee on Graduate Studies in Partial Fulfillment of the Requirement for the Degree of Doctor of Philosophy in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

© Copyright by Catriona Lucy Clare Jones 2022

Environmental and Life Sciences PhD Graduate Program

May 2022

ABSTRACT

Nutrigenomics of D*aphnia*: Examining effects of nutrients on the *Daphnia pulex* transcriptome and identifying nutritional phenotypes

Catriona Lucy Clare Jones

Organismal nutrition lies at the interface between biotic and abiotic factors in an ecosystem, dictating the transfer of energy and nutrients across trophic levels. Our ability to detect nutritional limitation in consumers is reliant on a priori knowledge of dietary history due to our inability to differentiate nutrient stress based on body-wide responses. Molecular physiological responses are increasingly being used to measure physiological stress with high levels of specificity due to the specific modes of action ecological stressors have on organismal molecular physiology. Because animal consumers respond to varying nutrient supplies by up- and down-regulating nutrient-specific metabolic pathways, we can quantify nutritional status by quantifying the expression of those pathways. Here I present an investigation into the use of transcriptomics to detect nutritional stress in the keystone aquatic herbivore, *Daphnia pulex*, I use RNAseq and quantitative PCR (qPCR) identify nutritional indicator genes. I found that nutritional status could be determined with 100% accuracy with just ten genes. Additionally, the functional annotation of those genes uncovered previously unidentified responses to dietary stress. Further testing and validation of the selected indicator genes is required however these findings have the potential to revolutionize our ability to measure and

monitor consumer nutritional stress.

ii

PREFACE: This thesis was carried out by myself, Catriona Jones, and in the majority represents my individual efforts. However, this work would not have been achieved without mentorship and collaboration and I use the word 'we' within data chapters to reflect this collaboration. Where data chapters have been published or submitted to journals, I have provided author lists and contributions in the same format as they were to the journal. Unpublished data chapters are prefaced with an acknowledgement of collaborator support.

Acknowledgements

This thesis would not have been possible without the collaboration, support, and mentorship of many people.

Firstly, I would like to thank my supervisor Dr. Paul Frost for his support throughout my PhD and for his patience, knowledge, motivation, and enthusiasm. I could not have asked for a better supervisor and mentor.

I would also like to thank my committee members, Dr. Aaron Shafer and Dr. Tereasa Crease for their technical advice, insightful comments, and encouragement.

I thank my fellow students and colleagues, former and current, in the Frost, Shafer and Xenopoulos labs at Trent University: Cody Starke, Emma McKnight, Katlin Doughty, Marcus Rice, Oghenemise Abirhire, Olivia Maillet, Sandra Klemet-N'guessan, Sarah D'Amario, Sarah Haworth, Sarah King, and Victoria Reimer for their practical support in the lab, helping care for my *Daphnia* cultures, and for being such a kind, supportive group of people.

I also thank Dr. Clay Prater, Dr. Nicole Wagner, and Dr. Nolan Pearce for their guidance and advice.

Thanks also go to the crew from office D118: Allison Kwok, Dr. Debbie Jenkins, Lauren Hooton, and Maegwin Bonar, for creating such a great working environment and informal support group!

Finally, I would like to thank my parents, Ken Jones and Priscilla Jones, my aunt Deborah Coleman, my brother Philip Jones, as well as Marg Moffat and Brenda

iv

Duncanson, and my chosen family, Badger, Cara, Lulu, Maxi, Millie, and Rook for their ongoing and unwavering support. Love you all.

Land Acknowledgement

I wish to acknowledge that the work for this thesis was carried out on the traditional territory of the Michi Saagiig Anishnaabeg. I am grateful to have had the opportunity to work on this land and thank the First Peoples for their care, stewardship, and teaching. May we honor those teachings. Miigwetch.

Contents

List of tables

Table 2.1. ANOVA table of MSGR and Ca:P ratio of daphniids grown under varied conditions of phosphorus [P] and calcium [Ca] limitations. Significant results in bold.

Table 3.1. The elemental ratios, algae species, daily dilution rates, N and P concentrations and culture media for 6 dietary treatments to which *Daphnia pulex* were exposed. *COMBO medium from (Kilham et al. 1998).

Table 3.2. ANOVA table of MSGR, C:P ratio, C:N ratio, N:P ratio and body Ca% (by dry weight) of *Daphnia pulex* grown under varied conditions of calcium [Ca], food quantity $[C]$, cyanobacterial diet (Cy) , nitrogen $[N]$, and phosphorus $[P]$ limitations. Significant results in bold.

Table 3.3. Global patterns of gene expression in *Daphnia pulex* exposed to 5 nutrient limited dietary treatments (low calcium, high cyanobacteria, low food/carbon, low nitrogen, and low phosphorus) showing total number of genes significantly differentially affected by nutritional limitation, total number of significantly DE genes unique to each treatment and percentage of total DE genes unique to each treatment.

Table 3.4. Outcome of 'leave one out' cross-validation within sparse partial least squares models for algorithm-produced subsets of nutritional indicator genes from *Daphnia pulex* exposed to five nutrient limited diets. All parameters have been averaged across the treatments. *MSEP = Mean square error of predictions.

Table 3.5: List of genes forming the 13 nutritional biomarkers identified from *Daphnia pulex* exposed to five nutrient limited diets (CA = low calcium, CY = high cyanobacteria, $CB = low carbon/food, NT = low nitrogen, and PH = low phosphorus). The biomarkers$

were selected with an elastic net regression and are shown here with KOG gene function and treatments under which the gene was up/down regulated.

Table 4.1: Details of qPCR primers designed for genes to be used as nutritional state indicators in *Daphnia pulex* exposed to six dietary treatments: high quantity of highquality food, low calcium, high cyanobacteria, low carbon/food quantity, low nitrogen, and low phosphorus. Metrics include melt temperature (Tm), limit of quantification (LOQ) and limit of detection (LOD).

Table 4.2. ANOVA table of MSGR, C:P ratio, C:N ratio, N:P ratio and body Ca% (by dry weight) of *Daphnia pulex* grown under varied conditions of calcium [Ca], food quantity [C], cyanobacterial diet (Cy), nitrogen [N], and phosphorus [P] limitations. Significant results in bold.

Table 4.3: Results of linear discriminant analysis to assess predictive capacity of five pairs of nutrient-specific indicator genes in *Daphnia pulex*. Nutrient-limited n=3, nutrient-replete n=15.

Table 4.4: Details of predicted gene function and potential links to biological processes for ten indicator genes selected for use in *Daphnia pulex* to predict nutritional status of animals fed on one of six diets: high quantities of high-quality food, low calcium, high cyanobacteria, low carbon/food, low nitrogen, and low phosphorus.

List of figures

Figure 2.1. A. Mean MSGR values $(\pm 1 \text{ S.D.})$ of daphniids grown under varied conditions of phosphorus [P] and calcium [Ca] limitations. **B**. Mean Ca:P ratios (± 1) S.D.) of daphniids grown under varied conditions of phosphorus and calcium limitations. Lowercase letters on both plots represent Tukey's post hoc comparisons.

Figure 2.2. Venn diagram showing proportions of genes affected by each nutritional treatment in *Daphnia pulex*; high calcium, low phosphorus (H600), low calcium, high phosphorus (L100) and low calcium, low phosphorus (L600). (total # genes not affected $= 19,941$.

Figure 2.3. A. Plot of 5 indicator genes selected through SLDA regression. For *Daphnia pulex* grown under the following conditions: high calcium, low phosphorus (H600), low calcium, high phosphorus (L100) and low calcium, low phosphorus (L600) (n=6). (total $#$ genes not affected = 19,941) (Model prediction accuracy = 100% , CV accuracy = 100% , ellipses = 95% confidence intervals). **B.** Heatmap of 10 differentially expressed genes selected through SLDA and their molecular functions (Red = up-regulated, Blue = downregulated, White = no change). (* = genes shown in figure 3a). For *D. pulex* grown under the following conditions: high calcium, low phosphorus (H600), low calcium, high phosphorus (L100) and low calcium, low phosphorus (L600) (n=6).

Figure 3.1. Mean mass-specific growth rates (MSGR), C:P ratio, C:N ratio, N:P ratio and body %Ca (by dry weight) of *Daphnia pulex* grown under 6 dietary treatments: high quantity of high-quality food (fast growth), low calcium, high cyanobacteria, low food quantity (carbon), low nitrogen and low phosphorus. **A.** Mean MSGR \pm 1 S.E. (n=10). **B.**

Mean C: P ratio \pm 1 S.E (by mol) (n = 5). **C.** Mean C:N ratio \pm 1 S.E. (by mol) (n = 10). **D.** Mean N:P ratio \pm 1 S.E. (n = 5). **E.** Mean Ca% \pm 1 S.E. (by dry weight) (n = 5).

Figure 3.2. Scatter plot of variable loadings from sparse partial least squares regression of four algorithm-selected subsets of potential nutritional indicator genes in *Daphnia pulex* exposed to five nutrient limited diets. Genes were selected by the following variable selection models. **A.** Elastic net regularised regression. **B.** Sparse partial least squares discriminant analysis **C.** KOG enrichment. **D.** 13 potential biomarkers.

Figure 3.3. Heatmap of mean log fold change from four algorithm-selected subsets of potential nutritional indicator genes in *Daphnia pulex* exposed to five nutrient limited diets. Genes were selected by the following variable selection models. **A.** Elastic net regularised regression. **B.** Sparse partial least squares discriminant analysis **C.** KOG enrichment. **D.** 13 potential biomarkers.

Figure 4.1. Mean mass-specific growth rates (MSGR), C:P ratio, C:N ratio, N:P ratio and body %Ca (by dry weight) of *Daphnia pulex* grown under 6 dietary treatments: high quantity of high-quality food (fast growth), low calcium, high cyanobacteria, low food quantity (carbon), low nitrogen and low phosphorus. **A.** Mean MSGR \pm 1 S.E. (fast growth $n = 9$, nutrient limited $n = 3$). **B.** Mean C: P ratio ± 1 S.E (by mol) (fast growth $n =$ 9, nutrient limited $n = 3$). **C.** Mean C:N ratio ± 1 S.E. (by mol) (fast growth $n = 9$, nutrient limited n =3). **D.** Mean N:P ratio \pm 1 S.E. (fast growth n = 9, nutrient limited n =3). **E.** Mean Ca% \pm 1 S.E. (by dry weight) (fast growth n = 9, nutrient limited n = 3).

Figure 4.2. Mean relative expression $(\Delta \Delta \text{Ct}) \pm 1$ s.e. (n=3) of ten nutritional indicator genes in *Daphnia pulex* exposed to six diets: high quantities of high-quality food (fast

growth), low calcium, low carbon/food, high cyanobacteria, low nitrogen, and low phosphorus.

Figure 4.3. Linear Discriminant Analysis (LDA) loadings plot of relative expression (ΔΔCt) (n=3) of ten nutritional indicator genes in *Daphnia pulex* exposed to six diets: high quantities of high-quality food (fast growth), low calcium, low carbon/food, high cyanobacteria, low nitrogen, and low phosphorus.

Chapter One - Introduction

Organismal nutrition is at the center of ecological interactions between consumers and their food. The acquisition and transfer of nutrients across trophic levels directly affects the movement of energy and matter through the food chain while indirectly affecting processes such as nutrient cycling and community structure (Lindeman 1942). Nutritional ecology seeks to understand and explain these interlinked effects using an integrative framework that combines knowledge from fields such as biogeochemistry, physiology, behavioural and evolutionary ecology (Raubenheimer et al. 2009). By connecting individual physiology with wider ecological processes, nutritional ecology allows us to use data collected on the nutritional status of an individual to help understand processes across scales from genes to whole ecosystems.

A major concern in Canada, and across the industrialized world, is the impact of anthropological stressors on ecosystem health and robustness (Freedman and Beauchamp 1998, Vinebrooke et al. 2004). Of particular interest within nutritional ecology is the effect that humans are having on nutrient dynamics in ecosystems (Millennium Ecosystem Assessment 2005). This could be through decreasing levels of nutrients as has been seen with the effects of logging and acid deposition leaching calcium out of lakes on the Canadian Shield and on the Scandinavian peninsula (Skjelkvåle 2001, Jeziorski et al. 2008, Jeziorski and Smol 2017). Alternatively, this could be through the external loading of excess nutrients such as phosphorus and nitrogen, which alter phytoplankton community structure and lead to toxic algal blooms (Davis et al. 2012, Bunnell et al. 2014, Verhamme et al. 2016). Frequently, these human-derived nutrient shifts occur concurrently, with levels of some nutrients increasing to harmful levels as others

dramatically decline (Crain et al. 2008). This mix of nutrient stresses makes it difficult to dissect each nutrients' effect on wider ecosystem processes and presents a unique challenge when it comes to mitigating and managing these effects.

Past efforts to measure and track the nutrient content of lakes and rivers through long term water quality monitoring projects have provided vital data on the functioning of freshwater ecosystems (Blanchfield et al. 2009, Arvola et al. 2011, Dodds et al. 2012) and have been key in highlighting the aforementioned declines in calcium and phosphorus across northern lakes (Skjelkvåle 2001, Eimers et al. 2009, Jeziorski and Smol 2017, Huser et al. 2018). However, these datasets provide an incomplete picture and have the distinct drawback that they take many years of repeated sampling to accumulate sufficient data to effectively track change (Magurran et al. 2010). While knowing the long-term trends in water chemistry is beneficial, a true picture of ecosystem health also requires understanding how these abiotic factors affect the organisms living in these ecosystems. The bottom-up controls exerted by changes in the available nutrients can drive rapid micro-evolutionary changes in the producers and primary consumers at the base of the food web (Rudman et al. 2015, Gibert and Yeakel 2019). A wide variety of factors such as the existing community structure, the length of time over which change occurs, and the presence of additional environmental stressors can interact in unpredictable ways meaning that lake ecosystems can respond in very different ways to very similar chemical parameters (Crain et al. 2008, Jackson et al. 2016). For example, two lakes might both exhibit similarly low calcium concentrations but one of these lakes may have historically low calcium with a plankton community that is well-adapted to these conditions, producing a relatively stable food web. The other lake meanwhile may

be historically high in calcium and rapid declines in calcium concentrations in response to anthropogenic acidification could drive the rapid extinction of maladapted plankton and destabilize the food web. Clearly measuring calcium concentrations by itself would not be sufficient to detect which ecosystem is under stress and requires immediate remediation efforts.

The complexities of identifying and monitoring changes to nutrient dynamics at an ecosystem level may be simplified by focusing instead on the nutritional status of individual organisms within that ecosystem and using these individual physiological responses to draw conclusions about the wider functioning of the ecosystem (Bae and Park 2014). In lakes, for example, herbivorous zooplankton such as *Daphnia* occupy a key position at the base of the food web as primary consumers (Miner et al. 2012). In addition to being under strong control from the bottom-up effects resulting from nutrient shifts, many of these zooplankton have a short lifespan, high population turnover and high levels of phenotypic plasticity in response to environmental change (Colbourne et al. 2011, Miner et al. 2012). This means that zooplankton are among the earliest responders to environmental change in lakes and might serve as an early warning system for events such as anthropogenic nutrient loading, nutrient depletion, and toxic algal blooms (Eggermont and Martens 2011).

To serve as an effective early warning system for environmental change (including changes to nutrient dynamics), the physiological responses of zooplankton to changes in nutrient supply must be well-characterized and stressor-specific (Wagner et al. 2013, Frost et al. 2014).

Nutritional ecology has traditionally relied on growth bioassays to examine the effects of poor nutrition on consumer physiology (e.g., (Hessen 1992, Martin-Creuzburg et al. 2009, Martin-Creuzburg et al. 2010, Prater et al. 2016, Schalicke et al. 2019a), with *Daphnia* being frequently used as a model organism in these experiments. These assays use food with known quantities of each limiting nutrient to detect the effects of these nutritional limitations on growth and life history (Kogan 1986). They typically require high levels of *a priori* knowledge of the consumer's nutritional history, growth and life history reaction norms and the nutritional content of the experimental food. Although these types of experiments have provided information on the physiological effects of poor nutrition (Boggs and Ross 1993, Frost and Elser 2002, Acharya et al. 2004, Frost et al. 2005, Prater et al. 2016), they are unable to identify the type and intensity of nutritional limitation in consumers with an unknown nutritional history and unknown baseline reaction norms. Without knowing the nutritional make-up of the diet that the consumer has been feeding on, there is no way to tell at a physiological level which nutrient is limiting because the effects of different nutrients on physiological and life history traits might be similar (Muller-Navarra 1995, Urabe and Sterner 2001). Meanwhile, nutritional limitation is measured in these assays by comparing traits such as growth rate and reproductive output with individuals in a control treatment (Kogan 1986). This approach cannot be used when measuring nutritional limitation in free-ranging populations where there is no control for comparison. This brings us to one of the main challenges currently facing nutritional ecologists – how to determine the nutritional status of consumers with an unknown nutritional history?

The ability to measure nutritional limitation in wild populations and/or consumers with unknown nutritional histories has far reaching implications, both in applied sciences and in the theoretical understanding of physiology and nutrition. These include the ability to track and regulate nutrition, immunity and overall performance in cells cultures used in food production (e.g., yeast) (Boer et al. 2003, Boer et al. 2010) and bioengineering (e.g., algae) (Bogaert et al. 2019, Loftus and Johnson 2019), and the animals and aquatic plants grown in aquaculture (Chandhini and Kumar 2019) more closely. In addition to these practical applications, the *in-situ* measurement of nutritional status can answer fundamental questions about the fate of elements in the ecosystem and the role of nutrition in consumer physiology.

A robust set of nutritional indicators needs to use tools and reflect traits which are uniquely affected by each ecologically relevant nutrient and can be measured as discrete values with no external baseline needed (Wagner et al. 2013). As many physiological and life history responses are not specific enough to differentiate each different nutritional limitation, one possibility is that these types of nutrient-specific responses can be found at the molecular level. There is a unique set of metabolic pathways associated with maintaining critical supplies each essential macronutrient in an organism's diet, through changes to ingestion, assimilation, use efficiency and excretion (Sterner and Elser 2002). These metabolic pathways in turn produce distinct changes to cellular pools of biological molecules, including proteins, metabolites, RNA and the ionome (Wagner et al. 2013). Measuring the levels of these pools of biological molecules (i.e., '-omics') should allow us to detect which element(s) are limiting and to what extent. Past work began by identifying potential molecular responses of interest in a wide variety of taxa. Using

proteomics, alkaline phosphatase activity has been linked to phosphorus limitation in *Daphnia* (Wagner and Frost 2012). An early application of transcriptomics in the yeast, *Saccharomyces cerevisiae,* identified 1,881 genes which responded to nutrient limitation with 484 of those gene expression responses being specific to a single nutrient limitation (Boer et al. 2003). Ongoing work into the proteome and ionome of salamanders has uncovered ontogenetic shifts in ionic composition, with life stage being the most significant driver of trace element composition (Prater et al. 2019), highlighting the shifts in nutritional requirements and status across an organism's lifespan. While these early studies demonstrate the potential value of different types of nutrient-sensitive molecular responses, their full development has been slow to materialize.

Omics-based techniques have the potential to be powerful indicators of nutritional status, however, each method presents its own methodological challenges. With metabolomics for example, the highly conserved nature of metabolites means that metabolomic indicators might be applied across multiple taxa with very little change to the assay technique (Wagner et al. 2013). However, this brings the disadvantage that food metabolites may be indistinguishable from the consumer's metabolites within the consumer organism's metabolome (Wagner et al. 2013). This necessitates additional steps within the sample preparation, such as clearing of the guts or targeting of specific tissues, which may not be possible for all organisms. Metabolomics has been successful for monitoring and tracking nutritional status of large herbivores (such as moose, (Fohringer et al. 2021)) where tissue-specific samples may easily be obtained. However, in smaller invertebrates such as *Daphnia*, the collection of specific tissues is timeconsuming or completely impossible. Additionally, using metabolomics to differentiate

environmental stress (such as food quantity and propranolol exposure) has been found to be most effective in adult *Daphnia* >8 days old (Jeong and Simpson 2019). *Daphnia* indicators therefore could focus on omics-techniques where whole-animal homogenization does not present a methodological issue and detection of stress is possible in younger individuals.

Proteomics in contrast avoids these issues due to the target proteins (such as protein complexes involved in oxidative phosphorylation) being highly species- and pathway-specific (Givskov et al. 1994, LaRoche et al. 1996). The highly specific nature of these proteins means that each protein requires a uniquely designed and targeted assay (Wagner et al. 2013). We are currently unable to measure and characterize an organism's full proteome in response to nutritional limitation, which makes characterization of novel proteins difficult and biases variable (protein) selection towards pre-existing knowledge.

Both of these methodological issues might be dealt with by using transcriptomics, which measures the mRNA present in a tissue or organism. There are sufficient genetic differences between consumers and their food to avoid the need for gut clearing and tissue specific assays (Wagner et al. 2013) particularly when consumers are in a different domain of life to their food (such as with herbivorous individuals like *Daphnia* that consume unicellular organisms such as algae and bacteria). Additionally, the existence of a published *Daphnia* genome (Colbourne et al. 2011) and associated genomic database (Colbourne et al. 2005) provides the necessary information to identify *Daphnia* specific transcripts assuming sufficient dietary differentiation. This means that in addition to allowing us to identify potential genes of interest for nutritional indicators, we also have a large dataset available which can help identify novel nutritional pathways and further our fundamental understanding of the physiology of poor nutrition.

Transcriptomics is not without its own drawbacks. One is the disconnect between transcription and translation, which may complicate conclusions drawn from transcriptomic data (Greenbaum et al. 2003, Tian et al. 2004, Feder and Walser 2005, Evans 2015). Not every molecule of transcribed mRNA in a cell is translated into a protein and some transcripts may be broken down into nucleotides without being translated (Feder and Walser 2005). Genes can also produce isoforms, where multiple mRNAs can be produced from a single locus while differing in function (Wang et al. 2008). While this disconnect can introduce a slight disconnect between amount of mRNA and amount of translated proteins, there remains a quantifiable link between the transcripts present in a cell and the level of activity within their associated molecular pathways. Gene expression (transcriptomic) studies also benefit from a built-in baseline for detecting change, in the form of housekeeping genes (Thellin et al. 1999, Vandesompele et al. 2002, Huggett et al. 2005). This in-built baseline allows us to detect change in individuals with an unknown nutritional history without the need for controls.

In this thesis, I present an investigation of the transcriptomic responses of *Daphnia pulex* to six ecologically relevant diets: low calcium, low carbon (or low food quantity), high cyanobacteria, low nitrogen, low phosphorus, and high quantities of highquality food. I discuss the ways in which these responses can be harnessed as molecular biomarkers of nutritional stress, identify potential indicator genes, and validate their capacity as biomarkers using a quantitative PCR assay. I also highlight the predicted functional annotations for these genes and explore how these functional annotations can

contribute to our fundamental understanding of the effects of poor nutrition on an animal consumer and its molecular physiology.

In chapter two, I address the problem of nutrient co-limitation and the effect of limiting supplies of two nutrients on the nutrient-specific patterns of gene expression, using next generation RNA sequencing, and phosphorus and calcium as the limiting nutrients. According to Liebig's Law of the Minimum, organisms may only be limited by a single nutrient at a time, that being the element found most limiting (Von Liebig 1840, Paris 1992). If Liebig's Law is incorrect then transcriptomic responses should be completely non-additive and we may see a completely different signature of transcriptomics responses for each diet. These would be distinct from the responses to individual nutrients and mean that the only way to proceed with transcriptomics-based nutritional indicators would be to characterize the unique responses of every possible combination of nutrients, an unwieldy and complex task. Here I show that co-limitation induces a patchwork of responses, many of which are non-additive and unpredictable, but that there remains a core of nutrient-specific responses that are comparable between singly- and co-limited animals. This gene expression data thus indicates that the *Daphnia* are able to detect and respond to multiple limitations at once and that those responses could be easily tracked and differentiated without the need for specific 'co-limited' markers.

In Chapter three, I address the issue of how to select a manageable number of robust and highly predictive indicator genes from a large RNA seq dataset. Previous investigations into *Daphnia* nutrigenomics used microarrays to measure transcriptomic responses (e.g., (Jeyasingh et al. 2011, Frisch et al. 2014, Chowdhury et al. 2015, Becker et al. 2018) where a limited number of genes were selected for probes based on functional annotation and known links to nutrient-specific metabolic pathways. Although these studies have provided an important basis for the study of nutrigenomics in *Daphnia*, selecting target genes based on current functional annotation risks biasing the study towards existing knowledge and overlooking potentially robust and responsive indicators because they have yet to be theoretically or empirically connected to nutrient limitation. This is especially true for a new model organism such as *Daphnia*, where the functional annotation of the genome is largely incomplete and the majority of annotations are only predicted functions (Ravindran et al. 2019). In this chapter, I demonstrate that the *Daphnia pulex* transcriptome provides a rich source of potential indicators, with around 47% of the total transcriptome being sensitive to at least one form of nutritional or dietary limitation. However, such a large volume of potential indicators presents its own issues, namely, how to reduce a list of over 15,000 potentially affected genes down to a more easily assayed list of potential indicators containing only the most responsive and predictive genes. I explored several dimension-reducing techniques including elastic net regularized regression (ENR) and sparse partial least squares discriminant analysis (sPLS-DA) and found that the ENR was the most successful at removing redundant variables and producing a shortlist of around 130 highly responsive genes. For future ecological applications, we chose to reduce this number further manually to a manageable assay of <20 highly predictive potential indicator genes. Using manual selection, I identify 11 genes which are capable of discriminating between the six aforementioned nutritional states with 100% accuracy.

In Chapter four, I used quantitative PCR (qPCR) to validate the responses of ten potential indicator genes, two per nutrient-limiting diet, against five nutrient-limiting diets and one nutrient-replete diet and explored two assaying approaches. The first approach considered each set of two nutrient-specific genes as discreet assays and tested their ability to discriminate low from high supplies of the nutrient in question. This approach was successful in discriminating low supplies of some nutrients, with the calcium-sensitive genes showing a particularly strong, calcium-specific response. However, the nutrient-specific patterns found in the RNAseq data for the phosphorus and cyanobacteria indicator genes was not reflected in the qPCR. Instead, a more nuanced suite of responses was found across all the nutrient-limited treatments, reducing the discriminating ability to 0%. I then combined all ten indicators into a single nutritional state assay where responses across all the genes were considered as a single 'barcode'. Using this approach proved to be far more accurate, with 100% discrimination accuracy across all treatments.

The findings I present in this thesis provide tantalizing evidence for the use of indicator genes in diagnosing nutritional stress in *Daphnia*. I have addressed several key methodological issues such as dealing with co-limitation and large datasets and I present an analytical pipeline for identifying and validating genetic biomarkers using RNA sequencing, bioinformatics, and qPCR. Nutrigenomics as a tool for long-term biomonitoring is still in its infancy but the data presented in this thesis provide the building blocks necessary to produce a library of robust, validated nutritional indicators for *Daphnia*. In addition to the potential indicator genes highlighted in Chapters three and four, I have also contributed a permanent genetic resource to the field in the form of RNA

sequence data deposited in the NIH sequence read archives and a database of \sim 130 nutrient-responsive genes, which will contribute both to our understanding of poor nutrition in *Daphnia* and to the field of *Daphnia* genomics. At a fundamental level, this work provides support for the hypothesis that consumers respond to nutritional stress through the up- and downregulation of nutrient specific pathways. Further, I demonstrate that patterns of differential gene expression can be used to detect these nutrient-induced changes. The determination of nutritional state without the need for information on dietary history will allow for more in-depth and complex studies into the effects of poor nutrition and the role of nutrition in consumer health. At an applied level, this will allow for the testing of hypotheses on the role of nutrition in consumer-environment interactions. It will enable the close tracking of nutrients through an ecosystem and allow researchers to identify the role nutrition plays in population and community dynamics. Finally, the ability to diagnose the nutritional state of free-roaming consumers will reduce the volume of data required to track long-term changes to nutritional dynamics, making management and mitigation decisions timelier.

Chapter Two: The complexity of co-limitation – Nutrigenomics reveal non-additive interactions of calcium and phosphorus on gene expression in *Daphnia pulex*

A VERSION OF THIS CHAPTER WAS PUBLISHED AS: Jones, C. L. C., Shafer, A. B. A., Kim, W. D., Prater, C., Wagner, N. D., & Frost, P. C. (2020). The complexity of co-limitation: nutrigenomics reveal non-additive interactions of calcium and phosphorus on gene expression in Daphnia pulex. *Proceedings of the Royal Society B-Biological Sciences, 287*(1941). doi:10.1098/rspb.2020.2302

Author contributions: P.C.F., C.P., A.B.A.S. and N.D.W. conceived this study; W.D.K., C.P. and N.D.W. carried out the laboratory work; P.C.F., C.L.C.J. and A.B.A.S. carried out data analysis and P.C.F., C.L.C.J., C.P., A.B.A.S. and N.D.W. wrote the manuscript

Abstract

Many lakes across Canada and northern Europe have been experiencing declines in ambient phosphorus (P) and calcium (Ca) supply for over 20 years. While these declines might create or exacerbate nutrient limitation in aquatic food webs, our ability to detect and quantify different types of nutrient-stress on zooplankton remains rudimentary. Here, we used growth bioassay experiments and whole transcriptome RNA-seq, collectively nutrigenomics, to examine the molecular phenotypes produced by single (P and Ca) and combined (CaxP)-limitation in the freshwater zooplankter, *Daphnia pulex*. We found that daphnids limited by P, Ca or CaxP grew slower and differed in their elemental composition. Our RNA-seq results revealed unique phenotypes in Ca-, P-, and CaxPstressed animals. The phenotype of CaxP-stressed *Daphnia* was intermediate to those found in Ca- and P-stressed animals, which provides evidence of nutritional co-limitation by both nutrients. Additional linear discriminant analysis found a significant separation between treatments based upon gene expression patterns with the expression patterns of just 5 genes needed to predict animal nutritional status with 92% accuracy. These data reveal how aquatic consumer phenotypes are altered by individual and co-limitation of two highly important nutritional elements and challenge universal application of Liebig's Law of the Minimum. This use of nutrigenomics demonstrates its potential to address many of the inherent complexities in studying interactions between multiple nutritional stressors in ecology and beyond.

Introduction

Calcium (Ca) concentrations in lakes have been declining for the past couple of decades in northern regions of Canada and Europe downwind of industrialized areas. These declines are connected to changes in soil chemistry after reduced acidification and regrowth of forests following logging (Watmough and Dillon 2002, Jeziorski et al. 2008, Jeziorski and Smol 2017, Weyhenmeyer et al. 2019). Lakes in these regions are also notoriously phosphorus (P)-deficient as a result of low P export from especially old and nutrient-poor geological substrates (Armstrong and Schindler 1971, Arvola et al. 2011, Huser et al. 2018). Given this, Ca and P both have the potential to create nutritional stress in zooplankton populations in these lake regions (Prater et al. 2016). Low Ca concentrations have been shown to limit growth and reproduction and increase mortality in Ca-rich *Daphnia* taxa (Cairns and Yan 2009, Arvola et al. 2011, Azan et al. 2015). Similarly, low P concentrations in lakes can produce P-limitation in the same Casensitive zooplankton (Prater et al. 2016, 2017) by elevating carbon (C):P ratios in phytoplankton food sources. Due to low ecosystem supplies of both elements, there remains uncertainty about which element, Ca, or P (or possibly both), proximally limit zooplankton in these lakes. Fully resolving this uncertainty remains difficult given limitations to our ability to detect and distinguish between more than two forms of limitation in zooplankton and other animal consumers (Wagner et al. 2013).

Nutritional limitation in animal taxa is created when dietary supplies of nutrients fail to match its metabolic needs for growth, reproduction, and maintenance. Differences in food nutrient content from some optimal mixture can create limitation by the nutrient in shortest supply (Sterner and Hessen 1994, Frost and Elser 2002, Sterner and Elser

2002). This perspective of a single limiting nutrient is consistent with Liebig's Law of the Minimum, which posits that an organism's growth rate should match the supply rate of the most limiting nutrient (Von Liebig 1840, Sperfeld et al. 2012). This framework rests on the assumptions that organisms have requirements for energy and multiple nutrients in varying proportions and that these resources are not substitutable. It follows that: 1) the nutrient provided in lowest relative supply limits organismal growth and 2) increased growth only results with greater supply of the limiting nutrient up to the point where another nutrient becomes limiting. In other words, an organism can only be growth limited by a single non-substitutable nutrient at a time and is not subject to co-limitation by multiple nutrients (Paris 1992).

This perspective may not fully capture the interactive effects of multiple nutrients on animal metabolism. For example, the single-limiting nutrient perspective assumes that all growth-related traits are intrinsically linked or require every nutrient in a set proportional mixture (Paris 1992). On the other hand, animal growth is known to involve multiple independent processes, which each may require different proportions of key dietary elements (Boggs and Ross 1993, Boggs 2009). For example, the production of somatic tissues may have different nutrient requirements than those needed for reproduction. This was observed in a larval lepidopteron, *Speyeria mormonia*, when its somatic growth was nitrogen-limited and, simultaneously, its egg production was primarily C-limited (Boggs and Ross 1993). Despite this possibility, detecting and describing co-limitation remains empirically difficult due to an over-reliance on evidence provided by diet-growth experiments.

One problem with diet-growth experiments is that slow growth produced by one or multiple nutrients, on its own, is indistinguishable, on a mass basis alone, from slow growth produced by any other single element. Consequently, simply measuring growth of animals consuming different diets of varying quality may not be useful in determining the nature of dietary limitation (Sperfeld et al. 2012). An approach that escapes these experimental constraints is to characterize an animal's nutritional phenotype. The nutritional phenotype can be viewed as a collection of molecular, biochemical, and physiological traits connected to animal metabolism and that respond to changes in dietary conditions (Leal et al. 2017). Consumers experiencing low intake of a dietary nutrient adjust their metabolism and physiology to increase the acquisition and use efficiency of the limiting nutrient (Boer et al. 2003, Schwarzenberger et al. 2010, Wagner et al. 2013). For example, a P-limited animal should increase its absorption and net retention of this element while eliminating other elements found in relative excess. At a molecular level, this could mean increasing the expression of genes underlying metabolic pathways that acquire or recycle the limiting nutrient (Sugiura et al. 2003, Jeyasingh et al. 2011, Wagner et al. 2013). While such changes would provide evidence of acute limitation by specific, dietary nutrients, how these nutritional phenotypes respond to the low supply of two or more nutrients has yet to be examined.

Here we examined the gene expression responses of *Daphnia pulex* to low supplies of Ca and P*.* We used whole transcriptome (i.e., RNA-seq) data to assess differential gene expression in this freshwater crustacean associated with limitation by these two nutrients. We expected that gene expression would differ between animals experiencing Ca- and P-limitation due to the important but distinct roles that each of

these nutrients have in *Daphnia* metabolism. These differences could allow us to assess whether animals experiencing low supplies of Ca and P together were exhibiting signs of co-limitation. A high degree of similarity between an individual form of stress (Ca or P) and the double stress (Ca and P) would indicate single limitation whereas a unique (possibly intermediate) response would indicate animals detect and respond to both forms of limitation simultaneously.

Methods and Materials

Culturing algae and Daphnia

We purchased *Scenedesmus obliquus* from the Canadian Physiological Culture Centre (strain 10, originally as *S. acuutus*) and grew semi-continuous cultures in 2 L culture flasks. We ensured the algal cells received constant aeration, high light intensity (>150 µmol s⁻¹ m⁻¹) and were diluted daily with differentially enriched P-media (Sterner 1993, Wagner and Frost 2012, Prater et al. 2016). We sampled algal cultures once daily and returned them to their original volume with fresh, autoclaved media. We centrifuged collected algal cells at 4066 g for 10 min and re-suspended in P, N, and Ca free COMBO media (Kilham et al. 1998). We determined P content of algae on subsamples that had been filtered onto pre-weighed GF/C glass fibre filters, which were then dried for at least 45 min at 60°C. After drying, we reweighed filters to determine algal dry mass. We measured algal P content after digestion with potassium persulfate using ascorbic acidmolybdate blue colorimetry (APHA 1992). We used these data on algal mass and P content to create algal food mixtures with different P-content (nominal molar C:P ratios of 100 and 600). Additionally, we saved samples of the mixed food and re-measured

them to confirm actual C:P ratios used in experiments. We verified the C content of postmixed algal samples with an elemental analyzer (Vario EL III, Elementar Inc., Mt Laurel, New Jersey).

We maintained the freshwater zooplankter, *Daphnia pulex* (clone # TCO SOM 1.1 AKA 'The Chosen One'), in continuous culture in the lab in P-free COMBO media for > 2 years prior to these experiments. The Chosen One clone was collected from a pond in Oregon in 2000. The isolate used in this study was obtained from the Yan lab at York University in Ontario, Canada and was maintained in continuous culture in the Frost lab in P-free COMBO media for > 2 years prior to these experiments. To generate experimental animals, we reared ten clonal sisters of *D. pulex* in jars containing 400 mL of P-free COMBO (Kilham et al. 1998). These brood mothers were fed high quantities (C:P ~ 80-100) of P-rich *S. obliquus* daily. For the experiments, we collected neonates (\leq 24 h old) released from the 2nd to 4th broods from brood mothers.

Experimental set-up

We collected neonate *Daphnia* from $2nd$ -4th broods and rinsed them twice with Ca-, N- and P-free COMBO (Kilham et al. 1998) prior to use in experiments to eliminate possible transfer of media-associated Ca. We subsequently placed 10 neonates in each experimental jar (n=26 jars for H100 and n=36 jars for H600, L100 and L600) containing 400 mL of N- and P-free COMBO media containing either 1.5 mg L^{-1} (low Ca) or 6 mg L^{-1} (high Ca) of calcium. The experimental codes used are as follows: H100 = high Ca (6mg L⁻¹) and C:P 100 food; H600 = high Ca and C:P 600 food; L100 = low Ca (1.5mg L^{-1}) and C:P 100 food; and L600 = low Ca and C:P 600 food. At the start of the

experiment, we saved 3 separate groups of 20 neonates to determine the initial mass of *D. pulex*. Finally, we saved COMBO samples from both Ca treatments for subsequent measurement of Ca concentrations.

We fed animals high quantities of food to prevent low food quantity effects. We initially fed neonates 2 mg C L^{-1} of either C:P 100 or C:P 600 at both Ca concentrations. On days 2 and 4, we fed the animals again with their prescribed P-diet at food concentrations of 2 and 4 mg C L^{-1} respectively. Throughout the experiment, we held the daphniids at 20°C in an environmental chamber. After six days of growth, we collected animals from replicate jars by first rinsing twice in N-, P-, and Ca-free media and then placing groups of animals from each replicate jar into separate containers designated for each type of analysis. Animals were saved for elemental analysis and RNA extraction.

Mass-specific growth rate (MSGR) and elemental analysis of Daphnia

We placed all ten *Daphnia* from ten replicate jars into pre-weighed aluminum cups and dried them at 60°C. After drying, we weighed animals on a microbalance and calculated the mass-specific growth rate (MSGR) using the following equation:

 $MSGR = [ln (B2) - ln (B1)]/time$

with B₁ and B₂ being the average initial and final mass per *Daphnia*, and time being the number of days of growth (Prater et al. 2016).

To measure the dry weight %P and %Ca of the *Daphnia,* we used 5 experimental replicates per element, each containing 10 individual animals. We placed the animals into aluminum cups (1 cup = 1 replicate) and dried at 60° C. After weighing dried samples on a microbalance, we analyzed the P content of animal bodies using the same molybdateblue method described above (APHA 1992). We took an additional 5 replicates per treatment, each consisting of 20 animals and analyzed *Daphnia* Ca body using a TRXF spectrometer with a read time of 90 s as described in (Prater et al. 2016).

Analysis of growth rate and elemental content data

To examine the effects of P and Ca on daphnid growth, %P, %Ca, and Ca:P body content, we used a two-factor analysis of variance (ANOVA) with type I errors using R (R Core Team 2019) before completing Tukey HSD tests using the lsmeans package (Lenth 2016) to determine differences among food treatments. We then plotted these data using the ggplot2 package (Wickham 2016).

RNA extraction and sequencing

For each treatment combination, we used 6 replicate 1.5 mL RNase DNase free microfuge tubes. Each replicate tube for the H100 treatment contained 10 individuals, pooled from a single jar whilst each of the nutrient-limited treatments contained 20 individuals, pooled from 2 jars. After removing excess COMBO, we added 100 μ L of RNAlater® to each tube and flash froze each sample in liquid nitrogen. We stored samples in a -80°C freezer until extraction. Before RNA extraction, we removed the RNAlater® and then followed the manufacturer's protocol for RNA extraction using the Qiagen RNeasy Minikit. We checked the quality and quantity of RNA using a NanoDrop™ spectrophotometer and a BioAnalyzer. Reverse transcription of the RNA samples, library preparation, and Illumina sequencing on a single lane (HiSeq 4000 PE 100) was done at the Genome Quebec facilities.

Analysis of RNAseq data

We assessed quality control of the raw Illumina RNAseq daphnid reads (available in the NIH sequence read archive, accession # PRJNA613855) using FastQC v. 0.11.5 (Andrews 2010).,We then mapped the raw read data to the *D. pulex* genome (acquired from ENSEMBL – V1.0.43) with HiSat2 (Kim et al. 2015) using default conditions and converted bam files using Samtools v1.5 (Li et al. 2009). We used featureCounts (Liao et al. 2014) to count the mapped reads to genomic features with the annotated version of the *D. pulex* genome (acquired from ENSEMBL, 2019 release – V1.0.43).

We completed differential gene expression (DGE) analysis using EdgeR (Robinson et al. 2010), on pseudo-transformed raw count data (+1 to every count data to remove 0s). We fit three sets of pairwise general linear models (GLMs) with a Poisson link function (model fit tested using the ggfortify package in R (Tang et al. 2016)) to each annotated gene, comparing each of our treatments to the H100 daphniids (R Core Team 2019). We corrected the 0.05 critical P values with the false discovery rate (FDR) method using the qvalue package in R (Storey et al. 2019) and used these P values to determine what genes were differentially expressed among treatments. We also built a Venn diagram to the list of DE genes from all treatments using the VennDiagram package (Chen 2018) to identify overlap between treatments and number of unique genes per treatment

We used the corrected GLM results to select the 1000 most DE genes and used this dataset to test if DEG can predict nutrient limitation with a stepwise linear discriminant regression (SLDA) using the caret (Kuhn 2020) and klaR (Weihs et al. 2005) R packages. We used the initial dataset of 1000 genes to fit two models, one with

10 predictors (genes) and one with five predictors. We decided upon these numbers of predictors based upon preliminary analysis into the predictive power of different fractions of the top 1000 dataset. We calculated model prediction accuracy (%) and 'leave one out' cross-validation accuracy (CV) for both models. For the 5-gene model, we used ggplot2 (Wickham 2016) to plot the top two linear discriminants with points clustered by treatment to a 95% confidence interval. We then used the pheatmap package (Kolde 2019) to plot the expression patterns of the genes from the 10-gene model, with the molecular and biological functions of these genes identified using the wFleaBase (Colbourne et al. 2005) and UniProt (Bateman et al. 2019) databases.

Results

Growth limitation and elemental composition

We found slower growth rates in *Daphnia* provided low Ca and food P conditions (Fig. 2.1, Table 2.1). The highest mass-specific growth rates were seen in *Daphnia* provided high Ca, high food P (H100) conditions, whereas lower but roughly equal growth rates were observed in animals under low Ca (L100) and low food P (H600) conditions and the lowest growth rates were observed in animals under low Ca, low P (L600) conditions (Fig. 2.1A, Table 2.1). Additionally, we found a change in this animal's body elemental composition as a result of our nutritional treatments (Fig. 2.1B, Table 2.1). The lowest body Ca:P was found in H100 whereas H600 animals exhibited the highest body Ca:P ratios with intermediate values seen for L100 and L600 *Daphnia*.
Global patterns of differential gene expression

The *Daphnia pulex* genome (TCO, 2011 Ensembl release) contains a total of 33,049 in a genome of 123Mb. We found a total of 13,108 genes (representing \approx 39% of the genome) to be significantly differentially expressed (DE) in animals experiencing one or more types of nutritional limitation Of these, we found that 3592 of these DE genes were unique to the H600 treatment, 582 were unique to the L100 treatment and 554 were unique to the L600 treatment. In addition to these uniquely expressed transcripts, we found that 1416 DE genes were shared between the H600 and L100 treatments, 2160 DE genes were shared between the H600 and L600 treatments and 797 DE genes were shared between the L100 and L600 treatments. Finally, we identified a further 4007 genes as being differentially expressed in all nutrient-limited treatments compared to the nonlimited animals (Fig. 2.2). Across all the significantly affected genes, the mean log fold change (LFC) for the H600 treatment was -0.02 with a median LFC of -0.005. The mean LFC for the L100 treatment was 0.05 with a median LFC of 0.18. Finally, the mean LFC for the L600 treatment was -0.25 with a median LFC of -0.035.

Nutritional phenotypes

Both the 5 and 10 gene SLDA models predicted nutritional status with 100% accuracy. However, when we performed cross-validation (CV) analysis, we found that the 10-gene model had a CV accuracy of 95.8% while the 5-gene model had a CV accuracy of 100%. Fitting the 5 gene SLDA revealed that the first two linear discriminants explained 99% of the variation (Fig. 2.3A). We also observed clear

separation among the 95% confidence interval ellipses for the control and all three treatments along both LD1 and LD2 (Fig. 2.3A).

We explored the molecular functions underpinning each of the nutritional phenotypes using a heatmap of SLDA selected genes (Fig. 2.3B).The H600 treatment primarily induced the up-regulation of genes associated with P metabolism and secondary metabolite biosynthesis and a slight down-regulation genes associated with insect cuticle proteins and acyltransferases (Fig. 2.3B). The L100 treatment induced the downregulation of genes associated with insect cuticle proteins, collagen structures and NTP (nucleotide triphosphate) binding (Fig 2.3B). The L600 treatment consisted of a unique pattern of gene expression which combined elements of both the H600 and L100 treatments' expression patterns (Fig. 2.3B).

Discussion

For more than 175 years, the law of the minimum has guided our understanding of nutritional physiology by its fundamental assumption of single resource limitation of organismal growth rate. We have provided novel molecular evidence that demonstrates the limits of this paradigm by identifying an intermediate co-limited phenotype possessing molecular characteristics of both singly limited phenotypes. This raises important questions about the ubiquitous application of Liebig's Law to understanding multiple nutrient limitation in animal consumers.

Breaking Laws through Molecular Phenotyping

25

Prior to any molecular characterization of phenotypes, we first examined the effects of our nutritional treatments on the broader-scale traits of MSGR and body Ca:P ratio in *D. pulex* (Fig. 2.1, Table 2.1). This demonstrated a similar magnitude of stress induced by each limiting nutrient and highlights the relative importance of both P and Ca for *Daphnia* growth. Additionally, the co-limited treatment (L600) showed slower growth than either of the singly limited treatments, providing an initial indication of sub-additive co-limitation of these elements. These results differ from predictions derived from Liebig's Law of the Minimum, which posits that growth is constrained by only one element at a time. Following Liebig's Law, we predicted that under low supplies of both Ca and P, the growth rate of the doubly limited nutritional phenotype would match that of the most limiting single nutrient but would be less than that of the secondary limiting nutrient.

Applying Liebig's Law at the molecular level, we would predict that only one set of metabolic pathways (associated with the most limiting nutrient) should be up- or down-regulated at a time. However, our results provide evidence that metabolic pathways associated with two different nutrients can be up- or down-regulated simultaneously and, that under co-limitation, an additive phenotype is expressed with traits of both Ca- and Plimitation present (Figs. 2.3A and 2.3B). This molecular evidence suggests that organisms can respond metabolically and physiologically to limiting quantities of two separate nutrients and that animals in our experiment were co-limited by low supplies of Ca and P. Our results demonstrate the need to re-examine the application of Liebig's Law to animal nutrition with aid from a molecular lens to better understand its limitations and its caveats.

Each molecular phenotype in our study was comprised of thousands of individual genes, many of which have no current functional annotation and are indicative of complex multiple pathway responses to each nutritional treatment. By identifying predicted functions of a small subset of these genes, selected based upon their ability to explain 99% of variation between treatments, we lay the groundwork to begin understanding the complex interactions and specific metabolic pathways that help *Daphnia* mitigate the effects of poor elemental nutrition (Fig. 2.3B). We saw the clearest example of this in the P-limited (H600) phenotype, which was characterized by 6-fold up-regulation of two genes associated with P metabolism and secondary metabolite synthesis. This is in line with previous findings on the effects of low P on *Daphnia* physiology, which show that in order to meet their P requirements, *Daphnia* increase their P uptake- and use-efficiency (DeMott et al. 1998, Frisch et al. 2014). The Ca-limited (L100) phenotype was characterized by down-regulation of 2 genes connected to insect cuticle protein and extracellular collagen structures. This is in line with our current understanding of *Daphnia* response to low ambient Ca*,* which is primarily a structural element and limits growth by preventing sufficient hardening of the carapace following a molt (Cairns and Yan 2009). It therefore follows that at a molecular level, the down regulation of these genes translates to a decrease in the compounds that bind with Ca to form structural features. The co-limited (L600) treatment meanwhile forms an intermediate phenotype comprised of molecular features from both singly limited phenotypes; we found up-regulation of P metabolism and secondary metabolite production associated with low P, and the downregulation of cuticle protein and collagen structures associated with low Ca.

Nutritional phenotyping has clear practical applications for diagnosing nutrient limitation in the field. One persistent challenge in studying animal nutrition has been to determine the type and strength of nutritional limitation and in differentiating singular nutrient limitation from co-limitation (Wagner et al. 2013). As we demonstrate with this study, a powerful solution to this problem is found in tools provided by the growing field of nutrigenomics (O'Sullivan et al. 2011, Wagner et al. 2013, Frost et al. 2014, Rozenberg et al. 2015), which examines transcriptome responses of nutrient-stressed animal consumers. As well as providing evidence of co-limitation by two nutrients, we also demonstrated that gene expression patterns can predict nutritional state to a high level of accuracy with a small number of genes (5-10), providing an initial proof of principle for the application of molecular nutritional indicators.

The production of an intermediate co-limited phenotype has other important ramifications for the way in which we understand animal nutrition and its effects on their physiological processes. Previous assumptions that nutritional limitation in animals meets assumptions based on Liebig's Law have informed the building of animal growth models based on mass balance principles (Sterner 1997, Frost and Elser 2002). These models are used to estimate threshold elemental (TER) ratios, where limitation sharply shifts from one nutritional element to another (Anderson and Hessen 2005, Frost et al. 2006). Our results indicate the need to reconsider the formulation of TER models to allow for and assess the conditions that create elemental co-limitation. Given the role of these models in understanding growth limitation of animal consumers and their contributions to nutrient cycling in ecosystems, it is clearly important that we more fully reconcile our

observation of co-limitation at a molecular level with whole animal models of elemental nutrition.

Although providing initial compelling evidence of nutrient co-limitation in *Daphnia*, it may be worth considering whether our results are specific to the two elements under study. *Daphnia* absorb P and many other nutrients obtained from the digestion of ingested food whereas they acquire most of their Ca directly from the water column (Hessen et al. 2000). It may be that co-limitation was only possible because these elements (Ca and P) have largely separate uptake and usage pathways. To better understand the specificity of co-limitation, there should be additional experiments using combinations of nutrients in the same matrix (i.e., food) and with similar acquisition pathways such as nitrogen and P. There are also additional elements and complex biomolecules that are key nutritional components and a host of biotic and abiotic factors (e.g., predation, temperature and genotypic/ontogenetic variation) that have the potential to affect nutritional phenotypes and mask or exacerbate the nutrient-stressed signal (Wagner et al. 2013). Future application of nutrigenomics-based indicators should identify sets of responsive genes for environmentally relevant stressors. Following that, intensive experimental validation is needed to determine whether these indicator genes provide accurate and specific information of specific types of nutrient stress in animals exposed to these confounding biotic and abiotic conditions. Overall, our study highlights the importance of incorporating molecular and genetics tools into ecological studies and provides proof of concept for the application of nutrigenomics in nutritional ecology.

Figures and Tables

Table 2.1. ANOVA table of MSGR and Ca:P ratio of daphniids grown under varied conditions of phosphorus [P] and calcium [Ca] limitations. Significant results in bold.

Figure 2.1. A. Mean MSGR values $(\pm 1 \text{ S.D.})$ of daphniids grown under varied conditions of phosphorus [P] and calcium [Ca] limitations. **B**. Mean Ca:P ratios (± 1) S.D.) of daphniids grown under varied conditions of phosphorus and calcium limitations. Lowercase letters on both plots represent Tukey's post hoc comparisons.

Figure 2.2. Venn diagram showing proportions of genes affected by each nutritional treatment in *Daphnia pulex*; high calcium, low phosphorus (H600), low calcium, high phosphorus (L100) and low calcium, low phosphorus (L600). (total # genes not affected $= 19,941$

Figure 2.3. A. Plot of 5 indicator genes selected through SLDA regression. For *Daphnia pulex* grown under the following conditions: high calcium, low phosphorus (H600), low calcium, high phosphorus (L100) and low calcium, low phosphorus (L600) (n=6). (total # genes not affected = 19,941) (Model prediction accuracy = 100% , CV accuracy = 100% , ellipses = 95% confidence intervals). **B.** Heatmap of 10 differentially expressed genes selected through SLDA and their molecular functions (Red = up-regulated, Blue = $down$ regulated, White = no change). (* = genes shown in figure 3a). For *D. pulex* grown under the following conditions: high calcium, low phosphorus (H600), low calcium, high phosphorus (L100) and low calcium, low phosphorus (L600) (n=6).

Chapter Three: Characterizing nutritional phenotypes using experimental nutrigenomics: Is there nutrient-specificity to different types of dietary stress?

UNDER REVIEW AT: Molecular Ecology Resources

AS: Characterizing nutritional phenotypes using experimental nutrigenomics: Is there nutrient-specificity to different types of dietary stress?

Catriona L. C. Jones* , Aaron B. A. Shafer, and Paul C. Frost

SUBMITTED: 14th January 2022

Author contributions: C.L.C.J., A.B.A.S. and P.C.F. conceived this study; C.L.C.J. carried out the laboratory work; C.L.C.J., A.B.A.S. and P.C.F. carried out data analysis and C.L.C.J., A.B.A.S. and P.C.F. wrote the manuscript

Abstract

The ability to directly measure and monitor poor nutrition in individual animals and ecological communities is hampered by methodological limitations. In this study, we use nutrigenomics to identify nutritional biomarkers in the freshwater zooplankter, *Daphnia pulex,* a ubiquitous primary consumer in lakes and sentinel of environmental change. We grew animals in 6 ecologically relevant nutritional treatments: nutrient replete, low carbon (food), low phosphorus, low nitrogen, low calcium, and high Cyanobacteria. We extracted RNA for transcriptome sequencing to identify genes that were highly nutrient responsive and capable of predicting nutritional status with a high degree of accuracy. We selected a list of 125 candidate genes, which were subsequently pruned to 13 highly predictive potential biomarkers. Using a nearest neighbours' classification algorithm, we demonstrate that these potential biomarkers are capable of classifying our samples into the correct nutritional group with 100% accuracy. The functional annotation of the selected biomarkers revealed some highly specific nutritional pathways and supported our hypothesis that animal responses to poor nutrition are highly nutrient specific and not simply different presentations of slow growth or energy-limitation. This is a key step in uncovering the causes and consequences of nutritional limitation in animal consumers and their responses to small- and large-scale changes in biogeochemical cycles.

Introduction

Nutritional ecology aims to understand the environmental controls and ecological effects of organismal nutrition (Scriber and Slansky 1981). The study of environmental nutrition requires the ability to assess the presence and intensity of nutritional stress on organisms across the foodweb. Assessing nutrition in animal consumers has typically relied on indirect evidence from controlled experiments and models that infer the nutritional state of consumers (Boggs and Ross 1993, Boersma 2000, Carrillo et al. 2001, Becker and Boersma 2003, Frost et al. 2006, Cook et al. 2016, Pérez-Fuentetaja and Goodberry 2016, Khattak et al. 2018). While these approaches have provided useful information regarding animal nutrition in diverse ecosystems, there are limitations to these types of assessments. For example, while growth bioassays can indicate whether an animal is limited by one nutrient or another, they can be confounded by the low supply of other potentially limiting nutrients (Urabe et al. 1997). More direct assessments of the nutritional status would be useful to avoid these types of issues when determining the nutrition of individual consumers having unknown dietary histories (Wagner et al. 2013).

Biochemical indicators of nutritional stress, also referred to as biomarkers, have been proposed as a source of more direct knowledge of consumer nutrition. For example, alkaline phosphatase activity (APA) was identified as a possible biomarker of phosphorus (P) limitation in a lake zooplankter (McCarthy et al. 2006), but the activity of this enzyme was subsequently shown to covary with other nutritional stressors (Wagner and Frost 2012). While work on metabolomics has demonstrated the possible value of a multi-indicator profile approach (Wagner et al. 2014, Wagner et al. 2017, Alfaro and Young 2018, Buoso et al. 2021), these metabolite-based approaches may be confounded

by ingested food and have yet to be validated against multiple nutritional and environmental stresses. There remains a need for techniques that directly measure the type and intensity of nutritional limitation in consumers and that are free of confounding effects of multiple nutrients and other logistical limitations (Wagner et al. 2013).

One key issue complicating the nutritional indicator approach is that many ecologically relevant macronutrients, when in low supply, have similar effects on the physiology and life history of animal consumers. For example, poor nutrition generally reduces animal growth rates and reproductive output regardless of the limiting nutrient (Becker and Boersma 2003). The similarity of animal responses to low supplies of different nutrients means that these indicators are not informative when differentiating one type of nutritional limitation from another in consumers of unknown nutritional history. In addition, being able to reliably measure changes in these traits requires a baseline (i.e., mean growth rate and reproductive output of well-nourished individuals) from which to define significant increases or decreases. One set of emerging techniques that may overcome some of these limitations is based on gene expression and involves quantitatively tracking nutrient-specific molecular changes to metabolic and ingestionacquisition-excretion pathways.

Patterns of gene expression associated with nutritional metabolism could be a rich source of nutritional indicators for animal consumers. When a consumer feeds on poor quality food (i.e., deficient in one or more essential nutrients), they adjust their metabolic and nutrient acquisition-assimilation-excretion pathways to maintain broad nutritional homeostasis and meet their nutritional requirements (Frost et al. 2005). These effects are seen through physiological changes in ingestion, excretion and assimilation and result

from molecular and cellular level responses (such as differential gene expression), which manage the ebb and flow of metabolites and proteins (Jones et al. 2020). Previous work has found that around 30-40% of an organism's transcriptome, comprising thousands of genes, may be differentially affected by nutritional limitation (Boer et al. 2003, Jones et al. 2020) with many of those genes likely associated with previously unknown and uninvestigated nutritional pathways. Developing RNA-based nutritional biomarkers could allow for more wide-scale application of nutritional assessments in ecology. The use of internal housekeeping genes to normalise expression levels in gene expression assays such as quantitative PCR (qPCR) would also mean that there are built-in baselines (Vandesompele et al. 2002) that would remove the need for external controls or *a priori* knowledge of nutritional history.

In this study, we grew the freshwater zooplankton, *Daphnia pulex,* on six different diets (high abundance of high-quality food (fast growing), low food abundance, low phosphorus, low nitrogen, low calcium, and cyanobacterial) and used highthroughput sequencing (HTS) of RNA and differential gene expression analysis (DGE) with the high abundance of high-quality food diet as the calibrator, or baseline, to identify nutrient-specific patterns of gene expression. We then used 3 variable selection methods to identify and select subsets of significantly affected genes with the potential to act as biomarkers of each specific type of nutritional stress. The predictive power of these genes was then compared using a common sparse partial least squares regression (sPLS), which was followed by a more detailed examination of classification accuracy of selected biomarker subsets. The identification of robust and nutrient-specific biomarkers, as done here, will allow for unprecedented characterisation of nutritional state in animal

populations and aid in the long-term monitoring and management of our freshwater resources.

Materials and Methods

Culturing algae and Daphnia

Scenedesmus obliquus (high abundance of high-quality food, low P, low N, low food abundance and low Ca treatments) and *Romeria leopoliensis* (cyanobacterial diet) were purchased from the Canadian Physiological Culture Centre (*S. obliquus* strain 10, originally as *S. acuutus; R. leopoliensis* strain 102, originally as *Synechoccocus leopoliensis*) and were grown in 2 L culture flasks in COMBO medium (Kilham et al. 1998) for *S. obliquus* and BG-11 (adapted from Rippka et al (1979) by Wagner and Frost (Wagner and Frost 2012)) for *R. leopoliensis*. Algal cells received constant aeration, high light intensity (>150 µmol s⁻¹ m⁻¹) and were diluted daily with differentially enriched P and N-media (Table S1; adapted from Sterner et al. (1993), Wagner and Frost (2012), Prater et al. (2016)). *R. leopoliensis* was cultured initially in BG-11 medium and was diluted with increasing amounts of COMBO; days $1-2 = 10\%$ COMBO: 90% BG-11, days 3-4 = 20% COMBO: 80% BG-11 until 100% COMBO. Algal cultures were diluted daily (see Table S1) and resuspended according to the methods described in Jones et al. (2020). Details of treatment conditions can be found in Table 1.

The freshwater zooplankter, *Daphnia pulex* (clone # TCO SOM 1.1 AKA 'The Chosen One'), was collected from a pond in Oregon in 2000. The isolate used in this study was obtained from the Yan lab at York University in Ontario, Canada and was maintained in continuous culture in the Frost lab in P-free COMBO media for > 2 years

prior to these experiments, kept at 18-20 °C and fed on ad libitum *S. obliquus*. We collected *D. pulex* neonates (<24 hrs old) from the 3rd clutch of brood mothers that had been grown in COMBO medium (Kilham et al. 1998), which were prepared, assigned to jars, and fed according to the methods described in Jones et al. (Jones et al. 2020).

Preparation of low nutrient treatments

We prepared low Ca *Daphnia* COMBO by modifying the Kilham (1998) formulation to adjust the amount of $CaCl₂•2H₂0$ to the nominal concentrations given in Table 3.1. We verified actual concentrations of Ca media for all experiments using total reflection X-ray fluorescence spectroscopy (TXRF) on the PICOFOX (S2, Billerica, MA) (Prater et al. 2016). We grew nutrient sufficient, low P and low N algae in growth media adjusted according to amounts in Table 3.1. We estimated the biomass using subsamples of centrifuged algal food suspensions that had been pipetted onto pre-weighed GF/C glass fibre filters. These subsamples were then dried for at least 45 min at 60°C. After drying, filters were re-weighed to determine algal dry mass. We estimated algal C content as 50% of algal dry mass and calculated feeding amounts needed to produce 0.125 mg L^{-1} of C for the low food treatment and 3 mg L^{-1} of C for the high food treatment. We saved algal samples and verified the C and N content of all algal samples with an elemental analyser (Vario EL III, Elementar Inc., Mt Laurel, New Jersey). We determined algal P content of all algal samples using potassium persulfate digestion followed by ascorbic acid-molybdate blue colorimetry (APHA 1992). We used data on algal mass and P content to create algal food mixtures with high and low P-content (nominal molar C:P

ratios of 100 and 800). Samples of the mixed food were saved and re-measured to confirm actual C:P ratios used in experiments.

Mass specific growth rate (MSGR) and elemental analysis of Daphnia

Mass-specific growth rate (MSGR), body C:N ratio, body C:P ratio and body % Ca by dry weight was measured from animals in each treatment. Mass-specific growth rate was calculated using the equation: $\frac{MSGR = [\ln(B_2) - \ln(B_1)]}{\lim_{h \to 0}$ $\frac{\ln(B_2)-\ln(B_1)}{\lim e}$ where B_1 and B_2 are the average initial and final mass of *Daphnia*, and time is the number of days of growth (i.e., six). To measure the body C:N ratio of *Daphnia*, we collected 10 experimental replicates, each containing 10 individual animals. We placed these groups of animals into preweighed tin cups (1 cup = 1 replicate), dried animals at 60° C and reweighed each sample before analysing on an elemental analyser (Vario EL III, Elementar Inc., Mt Laurel, New Jersey). To measure the dry weight %P of the *Daphnia,* we used 5 of the experimental replicates from the MSGR analysis per element, each containing 10 individual animals. These animals were then prepared and measured using the methods described in Jones et al. (2020). We analysed *Daphnia* dry weight %Ca by placing all ten *Daphnia* from ten replicate jars into pre-weighed aluminium cups and then drying them at 60°C. After drying, we analysed Ca content using a Bruker S2 TRXF spectrometer as described in Prater et al. (2016).

Statistical analysis of growth and elemental data

To examine the effects of low supplies of food, N, P, Ca, and cyanobacterial food on daphnid MSGR, body C:P ratio, body C:N ratio, body N:P ratio, and body % Ca

content, we used a two-factor analysis of variance (ANOVA) with type II errors to account for unbalanced design due to loss of some samples. Analyses were performed using R (R Core Team 2019) and the car package (Fox et al. 2013).

RNA extraction and sequencing

A single jar of ten *Daphnia* was used for each treatment combination and replicate (n=8). The 10 animals per replicate were place in 1.5 mL RNase DNase free microfuge tubes. After removing excess COMBO, we added 100 μ L of RNAlater[®] to each tube and flash froze each sample in liquid nitrogen. Samples were stored in a -80°C freezer until extraction. Before RNA extraction, we removed the RNAlater® and then followed the manufacturer's protocol for RNA extraction using the Qiagen RNeasy Minikit. Reverse transcription of the RNA samples, library preparation, and Illumina sequencing on a single lane (HiSeq 4000 PE 100) was completed at the Genome Quebec facilities.

Analysis of RNAseq data

We used the bioinformatics pipeline described in Jones et al. (2020) to determine quality control of the raw Illumina RNAseq daphniid reads (available in the NIH sequence read archive, accession # PRJNA784271), map the RNAseq reads to the annotated *D. pulex* genome (acquired from ENSEMBL – V1.0.45) using Hi-Sat2 (Kim et al. 2015) and count the mapped reads to genomic features. We then conducted differential gene expression analysis and fit a generalised linear models with a Poisson link function to each of the 33,049 genes using the method described in Jones et al.

(2020). Using an alpha threshold of 0.05 for the false discovery rate (FDR) corrected P values, we excluded genes that were not significantly affected by any of the nutritionally limited treatments from further analysis, which left us with a list of 15,653 potential indicator genes.

We used the log2 fold change (LFC) for each of the 15,653 genes with three variable selection techniques to identify differentially expressed gene relative to the fast growth treatment that were highly correlated with nutritional treatment and therefore able to predict nutritional state with a high degree of accuracy. The following 3 techniques were used: 1) elastic net regularised regression (ENR), 2) a multi-dimensional discriminant model, sparse partial least squares discriminant analysis (sPLS-DA), and 3) KOG enrichment analysis based on functional groups. The ENR was fit using the GLMnet package in R (Friedman et al. 2010) and a Y matrix with 5 parameters each representing a different nutritional treatment (C:P ratio for P, C:N ratio for N, mg C/L for C, mg Ca/L for Ca and proportion of diet made up of *R. leopoliensis* for cyanobacteria). The parameters were then normalised using the equation $(x - \bar{x})$ /s where $x =$ data value, \bar{x} = mean of dataset and s = standard deviation of dataset, bringing the variance to between 0 and 1 in order to meet the assumptions of the model. The sPLS-DA was fit using the mixOmics package in R (Le Cao et al. 2016, Rohart et al. 2017); this model assumes sparsity in the data (i.e., that the significant majority of observed variance is driven by just a few predictors). Finally, we used the WFleaBase (Colbourne et al. 2005) and UniProt (Bateman et al. 2019) databases to assign a KOG (eukaryotic orthologous group) functional identifier. Genes with no KOG identifier were excluded. We then used

the KOGMWU package in R (Matz 2019) to determine which KOG classes were significantly enriched with up- or down-regulated genes.

Once the models were fit, variables (genes) were then ranked according to their overall contribution to the model using variable importance in projection, coefficient values (ENR), VIP values (sPLS-DA), or K-enrichment values (KOG). For the ENR, we selected all coefficients (genes) that were not reduced to 0 by the model. For the SPLS-DA model, all genes with a positive variable importance value were used. The KOG enrichment analysis uses a Mann-Whitney U test to group the genes into KOG classes and rank the classes according to their significance. We used a cut-off value of ≥ 80 for the delta ranking and used all genes which fell into that KOG class to create the data frame. Finally, we calculated mean square error (MSE) for ENR measured as the mean distance between each predicted point and the fitted regression line. MSE here is given as the LFC of the count data, squared.

The resulting three subsets of genes were then compared using a common sparse partial least squares (sPLS) regression in the mixOmics package version 6.3.2 (Le Cao et al. 2016, Rohart et al. 2017). We used the same Y matrix as described for the ENR model to meet the assumptions of the sPLS regression. The models were validated using 'leave one out' cross-validation. \mathbb{R}^2 , mean square error of predictions (MSEP), and \mathbb{Q}^2 values were obtained from the fitted models and used to compare model accuracy. These values were then averaged across treatments. Finally, the sPLS projections for each subset of genes were plotted in a clustered scatter plot to determine class separation and the LFC for each individual replicate of each gene against the treatment matrix were plotted in a heatmap using the pheatmap package in R (Kolde 2019) to identify hotspots of nutrient

specific responses. For the purposes of interpreting our sPLS models, we used the following criteria: an \mathbb{R}^2 value of > 0.7 indicates a reliably fitted model (Peng et al. 2016), a Q^2 of >0.4 indicates a model that remains reliably fitted after cross-validation to a suitable level for a biological system where additional noise is to be expected (Westerhuis et al. 2008) and a difference of <0.3 between the R^2 and Q^2 model indicates a model which retains predictive capacity after cross-validation (Veerasamy et al. 2011).

After assessing the performance of each set of model-selected genes using sPLS regression, we manually reduced the ENR-selected genes to identify a smaller set of potential biomarkers. We used the gene set selected by the ENR as this proved to be the best performing model based on correlation coefficients between individual genes and the nutritional treatment. We used PCA loadings and Euclidian distance trees to manually prune the ENR-selected genes down to a shorter list of potential biomarkers composed of genes that retained high prediction accuracy. This final subset of 13 highly differentially expressed nutrient-correlated genes was tested within the sPLS model as with previous subsets and plotted in a clustered scatterplot and heatmap. Finally, we fit a nonparametric nearest neighbours (k-NN) model using the caret (Kuhn 2020) and class packages (Venables and Ripley 2002) on the ENR and biomarkers gene sets to estimate their balanced accuracy of prediction. All analysis was performed in R v.4.1.1.

Results

Mass-specific growth rate and elemental content

Mass-specific growth rate (MSGR) was significantly reduced in all nutritionally limited treatments compared to the fast growth (FG) animals (Fig. 3.1A, Table 3.2). Body C:P ratio was significantly affected by P only and increased in response to the low P treatment (Fig. 3.1B, Table 3.2) whereas body C:N ratio increased in animals consuming low P and low food (Fig. 3.1C, Table 3.2). Body N:P ratios increased in response to low P, low food, and the cyanobacterial diet (Fig. 3.1D, Table 3.2). Three diets had significantly higher body %Ca: cyanobacteria, low food, and low P (Fig. 3.1E, Table 3.2).

Global gene expression profiles

The sequence data passed initial QA/QC assessments with between 9 and 13 million reads per sample (Table S3.1). In total, 15,653 genes were significantly affected by at least one low nutrient treatment. This translated to about 47% of the entire transcriptome being sensitive to one or more type of nutrient limitation. The total number of significantly affected genes per treatment and the number of uniquely significantly affected genes per treatment varied some among the types of dietary stress (Table 3.3). We found the highest number of affected genes in the animals fed Cyanobacteria, while those consuming N-poor food had the fewest affected genes. Animals consuming Cyanobacteria food were found to have the highest % of uniquely affected genes (1.91% of total genes in *Daphnia*'s genome) whereas animals consuming N-poor food had the lowest % of uniquely affected genes (0.088%) .

Variable selection

With the ENR, we identified a total of 127 genes with a non-zero coefficient and this produced a model with an overall mean square error (MSE) of 3.58 LFC^2 . Using these 127 genes, we found there was a moderate level of separation between treatments in multidimensional space with slight overlaps between the calcium and fast growth diets, and the phosphorus and fast growth diets (Fig. 3.2A), indicating largely differentiated phenotypes for each limiting nutrient. The heatmap (Fig. 3.3A) showed little change in the majority of the genes (see yellow) however each nutrient showed distinct regions of strong up- and down-regulation, indicating differentiated nutritional phenotypes (see red and blue).

For the sPLS-DA, we identified a single gene set of 19 genes of which only three were also identified by the ENR. There was no clear separation between treatments, with the calcium and cyanobacterial treatments overlapping in one group and the remaining treatments overlapping in a second group (Fig. 3.2B). There was a distinct phosphorus phenotype that was characterised predominantly by down-regulation of many genes (Fig. 3.3B – see blue). As the patterns of differential expression for the other treatments were fairly homogenous, the majority of genes selected by the SPLS-DA model showed very little treatment-specific differentiation (Fig. 3.3B).

The KOG enrichment dataset identified 67 genes. Of these, only two were shared with the ENR identified set and one was shared with the SPLS-DA-selected set. There was little separation between the treatments in the PCA with only the low food treatment separating cleanly (Fig. 3.2C). The majority of genes selected by the KOG method were downregulated in animals experiencing nutrient stress compared to FG animals. The calcium and nitrogen treatments had largely homogenous expression with no clearly defined phenotypes (Fig. 3.3C).

Predictive capacity of selected variables

The highest performing model on all sPLS parameters (R^2, Q^2) and MSEP) was the ENR (Table 3.4), with slight overfitting but an overall high R^2 and Q^2 . The KOG enrichment analysis resulted in an overfit model (Q^2 is 25% lower than R^2) while the SPLS-DA analysis resulted in a slightly underfit model (Q^2 is 5% higher than R^2). Because ENR was the best performing model, the correlation coefficients from this set of genes were used to select a smaller set of potential biomarkers. We found that a set of 13 biomarkers, when tested against the other 3 gene subsets, out-performed both the KOG and SPLS-DA selected genes but not the full set of ENR-selected genes (Table 3.4).

In our scatter plot of loadings from the smaller set of genes, there was complete separation between all treatments (Fig. 3.2D). The heatmap for this set of thirteen genes showed unique patterns of expression for all nutritional treatments, in addition to unique 'barcodes' of expression across all 13 genes. This indicates that there are indicator phenotypes for all 5 growth-limiting nutrients that could be comprised of a relatively small subset of all DE genes (Fig 3.3D).

Balanced accuracy of ENR and potential biomarkers

The balanced accuracy of classification for each treatment and overall accuracy for the model was calculated for both the ENR-selected set of genes and the 13 potential biomarkers using a k-NN algorithm. The ENR-selected set of genes had an overall

accuracy of 100%, with a balanced accuracy of 100% for all the treatments. We found this 100% accuracy of classification was retained when the larger gene set was reduced to the 13 potential biomarkers.

Discussion

The ability to directly quantify poor nutrition in individual animals would have very high utility in ecology and other animal sciences. Here we used RNA-seq to identify potential biomarkers for five types of nutritional stress in the freshwater zooplankter, *Daphnia pulex*. This represents a key step forward in developing new approaches to tracking and quantifying the intensity of nutritional limitation in zooplankton and other aquatic animals. If applied to wild populations, *in-situ* measurements of nutritional status would allow us to more broadly describe and explain, with higher levels of precision, the type of nutritional limitation acting on animal consumers in nature, how this limitation affects ecological interactions and how it responds to small- and large-scale changes in biogeochemical cycles.

Characterizing coarse physiological responses to poor nutrition

The molecular characterization of nutrient-specific phenotypes is premised on experimental animals responding specifically to a particular type of nutritional stress rather than exhibiting a general growth-limited response. In all nutrient-limited treatments, we found a significant reduction of growth rates (i.e., MSGR), which indicates that we produced limitation in the most general sense. Each of the treatments applied here has previously been shown to produce moderately nutrient-limited growth of *Daphnia*. While it is not possible to identify *a priori* which nutritional element is limiting *Daphnia* in any experiment (Sterner and Schulz 1998), our ability to control animal diets through manipulation of algal and animal media creates a very high likelihood that animals experienced the type of limitation associated with each treatment. The Cyanobacteria treatment may be the hardest to parse out as the specific nutrient limiting animals consuming this diet in our experiment may have been one or more fatty acids, sterols, or some other lipid-associated biochemical limitation (Martin-Creuzburg et al. 2005, Wacker and Martin-Creuzburg 2007, Schwarzenberger et al. 2010, Koussoroplis et al. 2017).

Transcriptomic responses to poor nutrition

The transcriptome showed a high level of sensitivity to nutrient limitation with almost half (~48%) of the transcriptome significantly differentially expressed in response to one or more nutritional treatments. Compared to previous sequencing-based nutrigenomics studies, this represents a much higher nutrient-sensitive fraction of the transcriptome than has previously been identified. For instance, Boer et al (2003) found ~31% of the *Saccharomyces cerevisiae* (a yeast) transcriptome to be sensitive to carbon, nitrogen, phosphorus, or sulphur limitation. With a comparable study on *D. pulex*, Jones et al. (2020) found ~39% of the transcriptome to be sensitive to calcium and phosphorus limitation. Our high number of significantly affected genes is probably due, at least in part, to the higher number of limiting nutrients used in this study. It is also possible that the *Daphnia* transcriptome has more regions dedicated to addressing nutrient stress and growth limitation than the yeast. As a metazoan, *Daphnia* must coordinate growth

between multiple differentiated tissues and organs, which could require a higher level of genomic control and coordination (Puig and Tjian 2006).

It is already well-documented that *Daphnia* are phenotypically plastic and have an ecoresponsive genome that responds rapidly to environmental stress (Miner et al. 2012, Lind et al. 2015, Yamamichi et al. 2015). Studies into *Daphnia's* plastic genome have primarily focused on anti-predator responses (Rozenberg et al. 2015, Hales et al. 2017) and ecotoxicology (Kim 2015, Fuertes 2019, Campos 2021). However, our results add to a growing body of evidence that the genome of *Daphnia* enables plastic responses at the molecular level to a wide variety of biotic and abiotic stressors. This broad scope of responses was also demonstrated by Orsini et al. (2018) after exposing *D. magna* to 12 non-nutrient stressors and sequencing their mRNA. They found that only 44% of the significantly affected genes have orthologs outside of the *Daphnia* lineage (and potentially outside the Crustacea), in comparison to other metazoans where up to 67% of genes have orthologs in other lineages. Meanwhile, 72% of affected genes in *Daphnia* had no orthologs outside of the crustaceans (compared to 25% for other metazoans). Our results suggest that this lineage-specific adaptive architecture might also be present for nutrient stress in *Daphnia*.

From transcriptome to biomarkers

Isolating robust biomarkers from a high volume of sequence data requires reducing the large number of significantly affected genes to a smaller set of informative genes. Statistical techniques based on log fold change have the advantage of using existing expression levels, which theoretically should remain largely unchanged in

subsequent assays (e.g., qPCR, microarray, etc (Griffith et al. 2010, Shi and He 2014, Wu et al. 2014)). However, there is a need for statistical approaches that can address the complexities of genomic data including: a) the large number of response variables (high dimensions), which each have a relatively small contribution to the overall fit of the model and b) the high levels of correlation seen between many response variables.

We used three variable selection techniques, two statistical and one functional, to identify highly responsive genes and then tested the predictive capabilities of these sets of selected genes using sparse partial least squares regression (sPLS). Elastic net regression (ENR) has been developed to address both issues by eliminating genes with a low input (i.e, those with low or zero counts) and automatically clustering highly correlated features into groups for selection (Zou and Hastie 2005). This leads to a better, more highly predictive model fit than the other models used here. Furthermore, this high level of predictive capacity was retained when the list of genes was pruned to 13 potential indicators, which is compelling evidence that the use of transcriptome-derived data will be useful in diagnosing types of nutritional stresses. Although sparse partial least squares discriminant analysis (sPLS-DA) is capable of addressing the first issue (high dimensions) (Cao et al. 2011), it has been found to not perform well in heavily correlated datasets (Tinnevelt et al. 2020). As *Daphnia* responses to environmental stress are driven by modules of co-expressed genes (Asselman et al. 2018, Orsini et al. 2018), sPLS-DA is likely not an appropriate model for this application.

Functional annotation has been the basis of target gene selection for a number of *Daphnia* transcriptomics studies (Jeyasingh et al. 2011, Altshuler et al. 2015, Christjani et al. 2016). It is important to consider gene function when selecting relevant biomarkers as

this can reduce the risk of selecting biomarkers that turn out to be sensitive to other environmental or biological factors. However, this method may not be ideal for non- or new model organisms whose genomes have yet to be fully annotated. Because much of the *D. pulex* genome remains uncharacterized (Ravindran et al. 2019), focusing purely on functional annotation risks biasing variable selection away from highly responsive, nutrient-specific genes simply because their function has yet to be determined. Indeed, identifying conditions under which uncharacterized genes are differentially expressed may well aid in ultimately uncovering their function (Williams and Auwerx 2015). Almost half (6/13) the genes that we identified as potential biomarkers (using ENR) are currently uncharacterised in terms of function (Table 3.5). Yet these genes appear capable of discriminating between treatments with a high level of accuracy, which further highlights the importance of full transcriptome sequencing when studying transcriptomic responses to stress.

For those genes that did have a functional annotation (Table 3.5), there were some compelling links between gene function and nutrient stress pathways. For example, DAPPUDRAFT_320940 codes for Acyl-CoA synthetase. Klumpen et al. (2021) found that Acyl-CoA synthetase is up-regulated in *D. pulex* experiencing starvation conditions. In our dataset, this gene was strongly upregulated under low food quantity conditions (Table 3.5, Fig. 3.3D) and to a lesser extent under phosphorus limitation. DAPPUDRAFT 251853 is part of the phosphate transport pathway, which has been linked with post-digestive absorption of P from the gut in humans (Sabbagh et al. 2011). In our dataset, this gene was strongly down-regulated in the cyanobacterial treatment and up-regulated in the calcium, carbon, and phosphorus limited treatments (Table 3.5, Fig.

3.3D), with the phosphorus treatment showing the highest up-regulation, indicating that this pathway has a similar role in *Daphnia* and humans. DAPPUDRAFT_324898 codes for insect cuticle protein, which is one of a group of proteins that play a key role in moulting and exoskeleton formation (Charles 2010). This gene was heavily downregulated under all conditions of nutrient stress (Table 3.5, Fig. 3.2D), which is consistent with findings from Hessen and Rukke (2000) that the moulting process in *Daphnia* is tightly connected to animal growth rates. What is less clear is the reason for the up regulation of chitinase (DAPPUDRAFT_300583, Table 3.5) in response to the calcium, cyanobacterial, and phosphorus treatments. Chitinase is responsible for the digestion of the old cuticle during moulting (Giraudo et al. 2017) thus its up regulation would seem to suggest an up regulation in moulting, contrary to the patterns exhibited in the insect cuticle protein genes. This demonstrates the need for more intensive studies into the functional annotation of the *Daphnia* genome and consideration of whether annotating based on generic orthologous groups is suitable for an organism like *Daphnia* whose genome is so highly conserved within the genus. These links between gene function and nutritional limitation provide compelling evidence that nutrient-limited animals in our study not only grew slower, but they were also able to detect which nutrient was in short supply and actively adjust specific metabolic pathways in order to address these nutrient imbalances. Additionally, the presence of unique phenotypes for each treatment confirms that each diet type produced a distinct type of growth limitation, rather than simply causing starvation or energy limitation by way of a single, generalized metabolic mechanism.

The use of RNA-seq to identify targets for qPCR assays has yielded environmental biomarkers for plant (Ferrandez-Ayela et al. 2016) and other animal (Akbarzadeh et al. 2020) taxa. Our library of DE genes generated by the ENR analysis can thus be viewed as a promising source of raw materials for further development and refinement of nutritional profiling with more targeted assays such as qPCR. While our findings demonstrate the feasibility of developing sets of gene expression-based biomarkers of nutrition stress, they raise some practical questions about their further development and implementation. First, what is the smallest number of predictors (genes in this case) needed for accurate and robust classification of an animal's nutritional status? Frost et al. (2014) used modelled datasets and classification algorithms to explore the process and minimal requirements of nutritional profiling. They found that a limited set (-10) of indicators could be powerful enough to differentiate among four phenotypes, but the minimum number also depended on the relative magnitude of effect sizes (for example, fold change in gene expression) and variance in response variables. The somewhat larger set (13) of genes found here to yield high classification accuracy may reflect our classification of six nutritional treatments, our relatively small effect sizes, and the higher level of variance present in the RNA-seq database. However, based on the analysis in Frost et al. (2014), it seems feasible that an even smaller subset of genes could be used to differentiate the six types of nutrition studied here provided that these responses can be measured with more precision. The selection of an appropriate number of indicator genes must therefore strike a balance between making the assay manageable and cost-effective, and having sufficient genes to ensure accuracy in the event of unexpected expression patterns from one or more genes.

The development of nutritional indicators in animals has primarily focused on increasing yield in agriculture (Spence 2010) and aquaculture (Chandhini and Kumar 2019) by reducing dietary deficiencies related to biomolecules such as lipids and proteins. Similar molecular biomarkers of environmental stress have also been successfully deployed for a variety of species and applications from studying the effect of xenobiotic toxins on moon jellies (Schroth et al. 2005) to tracking immune system function in Pacific cod (Mao et al. 2015). Here we demonstrate this approach can be used to identify nutritional state of an animal consumer associated with growth limitation by a specific macronutrient with no information on their past diet. The use of nutritional indicators to diagnose the nutritional status in individuals with an unknown nutritional history in more natural settings, if applied more broadly, promises unprecedented insight into the role of dietary stress in ecological interactions.

Figures and Tables

Table 3.1. The elemental ratios, algae species, daily dilution rates, N and P

concentrations and culture media for 6 dietary treatments to which *Daphnia pulex* were

exposed. *COMBO medium from (Kilham et al. 1998).

Table 3.2. ANOVA table of MSGR, C:P ratio, C:N ratio, N:P ratio and body Ca% (by dry weight) of *Daphnia pulex* grown under varied conditions of calcium [Ca], food quantity [C], cyanobacterial diet (Cy), nitrogen [N], and phosphorus [P] limitations. Significant results in bold.

Trait	Treatment	DF	${\bf F}$	\overline{P}
MSGR	Ca	1,39	50.568	< 0.0001
	$\mathbf{C}\mathbf{y}$	1,39	102.841	< 0.0001
	$\mathbf C$	1,39	191.667	< 0.0001
	N	1,19	87.955	0.0001
	${\bf P}$	1,19	144.995	< 0.0001
C: P	Ca	1,43	1.2471	>0.1
	$\mathbf{C}\mathbf{y}$	1,43	9.9881	>0.01
	C	1,43	2.7145	< 0.1
	$\mathbf N$	1,43	0.1093	>0.1
	${\bf P}$	1,43	11.6825	< 0.005
C: N	Ca	1,55	1.291	>0.1
	Cy	1,55	1.2884	< 0.1
	$\mathbf C$	1,55	13.7497	>0.0005
	$\mathbf N$	1,55	1.3509	>0.1
	${\bf P}$	1,55	41.6449	< 0.0001
N: P	Ca	1,41	0.3424	>0.1
	$\mathbf{C}\mathbf{y}$	1,41	12.2481	< 0.005
	$\mathbf C$	1,41	4.12	< 0.05
	$\mathbf N$	1,41	0.2699	>0.1
	${\bf P}$	1,41	7.9571	0.01
Body Ca%	Ca	1,38	0.767	>0.1
	$\mathbf{C}\mathbf{y}$	1,38	11.5193	< 0.005
	$\mathbf C$	1,38	57.3044	< 0.0001
	$\mathbf N$	1,38	0.6113	>0.1
	${\bf P}$	1,38	28.7979	< 0.0001

Table 3.3. Global patterns of gene expression in *Daphnia pulex* exposed to 5 nutrient limited dietary treatments (low calcium, high cyanobacteria, low food/carbon, low nitrogen, and low phosphorus) showing total number of genes significantly differentially affected by nutritional limitation, total number of significantly DE genes unique to each treatment and percentage of total DE genes unique to each treatment.

Table 3.4. Outcome of 'leave one out' cross-validation within sparse partial least squares models for algorithm-produced subsets of nutritional indicator genes from *Daphnia pulex* exposed to five nutrient limited diets. All parameters have been averaged across the treatments. *MSEP = Mean square error of predictions.

Table 3.5: List of genes forming the 13 nutritional biomarkers identified from *Daphnia pulex* exposed to five nutrient limited diets (CA = low calcium, CY = high cyanobacteria, $CB = low carbon/food, NT = low nitrogen, and PH = low phosphorus). The biomarkers$ were selected with an elastic net regression and are shown here with KOG gene function and treatments under which the gene was up/down regulated.

Figure 3.1. Mean mass-specific growth rates (MSGR), C:P ratio, C:N ratio, N:P ratio and body %Ca (by dry weight) of *Daphnia pulex* grown under 6 dietary treatments: high quantity of high-quality food (fast growth), low calcium, high cyanobacteria, low food quantity (carbon), low nitrogen and low phosphorus. **A.** Mean MSGR \pm 1 S.E. (n=10). **B.** Mean C: P ratio \pm S.E (by mol) (n = 5). **C.** Mean C:N ratio \pm 1 S.E. (by mol) (n = 10). **D.** Mean N:P ratio \pm 1 S.E. (n = 5). **E.** Mean Ca% \pm 1 S.E. (by dry weight) (n = 5).

Figure 3.2. Scatter plot of variable loadings from sparse partial least squares regression of four algorithm-selected subsets of potential nutritional indicator genes in *Daphnia pulex* exposed to five nutrient limited diets. Genes were selected by the following variable selection models. **A.** Elastic net regularised regression. **B.** Sparse partial least squares discriminant analysis **C.** KOG enrichment. **D.** 13 potential biomarkers.

Figure 3.3. Heatmap of mean log fold change from four algorithm-selected subsets of potential nutritional indicator genes in *Daphnia pulex* exposed to five nutrient limited diets. Genes were selected by the following variable selection models. **A.** Elastic net regularised regression. **B.** Sparse partial least squares discriminant analysis **C.** KOG enrichment. **D.** 13 potential biomarkers.

Chapter Four: Well-fed or nearly dead? Using quantitative PCR to detect dietary stress in *Daphnia pulex*

Acknowledgement: I would like to thank Paul Frost, Aaron Shafer, and Teresa Crease for their collaboration and support in conceiving this study, analyzing data, and writing the manuscript.

Abstract

The ability to detect and track the molecular effects of environmental stress on a variety of taxa has benefited from the development of increasingly sophisticated molecular techniques. This has generated an unprecedented insight into the interactions between organisms and their environment and allowed for the rapid detection of, and response to, negatively impacted populations and communities. In this study, a qPCR assay was used to examine responsiveness of ten genes identified as potential indicators of nutritional state in the freshwater zooplankton *Daphnia pulex.* Animals were grown in 6 ecologically relevant nutritional treatments: nutrient replete, low carbon (food), low phosphorus, low nitrogen, low calcium, and high cyanobacteria. Ten nutrient sensitive genes, two per limiting nutrient, and two reference genes were selected from an RNA sequencing dataset to design and validate qPCR primers. Using the nutrient-specific pairs of genes for single nutrient limitation assays showed mixed results with high discrimination success for some treatments (calcium and nitrogen) but not others (carbon, phosphorus, and cyanobacteria). However, when all ten genes were used as a single nutritional state assay, they were able to discriminate between our six nutritional states with 100% accuracy. These results present a compelling proof of concept for the use of gene-based nutritional biomarkers, paving the way for genetic nutritional state biomarkers to join other biomarkers in the biomonitoring toolkit.

Introduction

Environmental stressors, such as limiting nutrient supplies, drought, and pollutants, can have profound effects on ecosystem structure, function and diversity (Freedman 2015). Tracking the presence and effects of environmental stressors through ecosystems presents a methodological challenge as stressor effects can be confounded, or masked, by interactions with other biotic and abiotic factors making these effects hard to detect until they reach a critical level (Maltby 1999). High throughput sequencing assays allow the development of new approaches to measuring stress at a molecular level that can be highly specific to different kinds of environmental stress (e.g., Jemec et al 2010). Transcriptomics tracks the relative expression of genes and is being increasingly used to track organismal responses to pollutants and toxins. This approach could also be used to assess and monitor animal responses to nutritional stress (Poynton et al. 2007, Poynton et al. 2011, Campos et al. 2021). Both toxins and nutrients have specific modes of action associated with metabolic responses, meaning that specific, sensitive pathways can be targeted using quantitative assays (e.g., qPCR, RNAseq). Indeed, the development of toxicology biomarkers has contributed greatly to our understanding of the mechanism of action of many of these toxins (Poynton et al. 2007). Using this approach for nutritional limitation would not only bolster our ability to track and monitor poor nutrition in wild animals, but it would also help us more fully understand the molecular effects of limitation by these nutrients.

Given the physiological and morphological responses of invertebrates to nutritional stress, it is likely that different types of nutritional stress produce different and predictable effects on gene expression (Boer et al. 2003, Wagner et al. 2013, Frost et al.

2014). There are relatively few types of potentially important nutritional stressors acting on zooplankton in lake ecosystems (Sterner and Schulz 1998). Crustaceans, particularly zooplankton, such as *Daphnia,* can require high levels of calcium to build and harden their carapace (Porcella et al. 1969) and are also incapable of storing calcium between molts, meaning that they require a consistent supply of calcium to replenish what is lost during moulting (Alstad et al. 1999, Hessen et al. 2000). Because the carapace forms the first line of defense between the *Daphnia* and their external environment, an insufficiently hardened carapace can increase mortality risk, particularly from predators (Riessen et al. 2012, Rice et al. 2021).

Another type of nutritional stress that may be common in lakes is low food abundance. Low quantities of food can create energy and carbon limitation if *Daphnia* are unable to meet their structural and metabolic requirements for C (Anderson and Hessen 2005, Bukovinszky et al. 2012, Lukas and Wacker 2014). The consumption of Cyanobacteria has also been identified as a potential source of nutritional limitation as these microscopic cyanobacterial producers are naturally low in PUFAs (polyunsaturated fatty acids), especially EPA (eicosapentaenoic acid) (Wacker and Martin-Creuzburg 2007). *Daphnia* with a primarily cyanobacterial diet must therefore adjust their allocation of essential fatty acids away from egg production and reduce their overall fitness (Wacker and Martin-Creuzburg 2007). Many cyanobacteria also produce protease inhibitors, presenting a further dietary challenge for consumers (Schwarzenberger et al. 2010).

While energy and carbon are required to fuel anabolism of organic molecules, nitrogen is required in animal diets for protein synthesis in growth and reproduction

(Sterner and Elser 2002). The effects of nitrogen limitation in *Daphnia* tend to be less severe than the effects of phosphorus limitation due to the daphniids' high P content relative to their N content (Sterner 1993, Sterner et al. 1993). However, a low nitrogen diet can still significantly reduce growth rate and reproductive output. Nitrogen-limited green algae tends to be lower in essential amino acids (Wagner et al. 2017) and harder for *Daphnia* to digest (Xu et al. 2021). Under a low nitrogen diet, the necessary building blocks for protein synthesis, both in the form of nitrogen and free amino acids, are in low supply or lacking altogether. There is also a significant drain on energy and resources due to the increase in digestive enzymes needed to obtain sufficient dietary nitrogen. Additionally, Xu et al (2021) found that at a transcriptomic level, methionine metabolism is up-regulated to offset the effects of nitrogen limitation but that this has the effect of suppressing reproduction.

Finally, the effects of phosphorus limitation on *Daphnia* have been well studied in the field of ecological stoichiometry (Scavia and Gardner 1982, Urabe et al. 1997, Becker and Boersma 2005, DeMott and Pape 2005, Ferrao-Filho et al. 2007, Chowdhury et al. 2015, Elser et al. 2016, Prater et al. 2016, Khattak et al. 2018, Jones et al. 2020), where the Growth Rate Hypothesis (GRH) was developed specifically to address the well characterized linkages between growth rate, ribosomal RNA production, and phosphorus content (Elser et al. 2003). DeMott et al. (1998) observed that *Daphnia* fed P-poor diets exhibited lower growth rates, lower birth rates, and lower offspring mass. Moreover, the specific P content of *Daphnia* and their offspring was reduced by up to 50% under Pdeficiency. In order to maintain internal P balance, they found *Daphnia* increased both

assimilation efficiency of P and the excretion of C, and reduced the assimilation efficiency of C.

When consumers adjust metabolic pathways in order to better meet nutritional requirements under nutritional limitation, these adjustments are triggered through the upand down-regulation of pathway specific genes (Boer et al. 2003, Jeyasingh et al. 2011). By measuring the expression of the genes controlling these processes, we can build a picture of the metabolic adjustments a consumer is making in response to poor nutrition (Wagner et al. 2013). For the purposes of this study, we have categorized potential effects of poor nutrition on metabolic pathways that are specific to a type of nutritional stress into three general categories: nutrient uptake, maintenance/basal metabolism, and nutrient acquisition/sequestration. With the first, poor nutrition could increase gene expression connected to metabolic pathways associated with increasing the animal's uptake of the limiting nutrient. This uptake could be through increased digestion and absorption efficiency of the dietary nutrient found in lowest supply (Muyssen et al. 2009). With the second group, there could be up or down regulation of pathways associated with maintenance of physiological processes in the face of low nutrient supplies. For example, low food abundance could result in adjusting the carbon budget and decreasing metabolism under C limitation (or food starvation) (Lukas and Wacker 2014) with subsequent effects on gene expression. Finally, there could be altered gene expression connected to maximizing the limiting supply through either increasing use efficiency or decreasing excretion of the limiting nutrient. For example, under N limitation, animal may suppress anabolic pathways, which would require changes in gene expression associated with these metabolic adjustments (Acquisti et al. 2009).

We examined the responses of ten indicator genes, two per limiting diet. These genes were selected based upon their potential to signal limitation of different dietary nutrients (Jones et al, *Chapter three*), which was assessed using Poisson regression models. We also considered each gene's predicted functional annotation (where available) and whether there was connection to predicted affected pathways that are likely connected to nutrient limitation. We selected two reference genes from an RNAseq dataset based on stability of expression across all nutrient-limiting conditions (Jones et al, *Chapter three*). We used a qPCR assay to examine the expression of these genes in animals exposed to six ecologically relevant diets .

Methods

Culturing algae and Daphnia

Two algal species were used, *Scenedesmus obliquus* (high abundance of high quality, low P, low N, low food abundance and low Ca treatments) and *Romeria leopoliensis* (cyanobacterial diet). These were purchased from the Canadian Physiological Culture Centre (*S. obliquus* strain 10, originally as *S. acuutus*; *R. leopoliensis* strain 102, originally as *Synechoccocus leopoliensis*) and grown in 2 L culture flasks in COMBO medium (Kilham et al. 1998) for *S. obliquus* and BG-11 (adapted from Rippka et al (1979) by Wagner and Frost (2012)) for *R. leopoliensis.* Algal cells received constant aeration, high light intensity $(>150 \mu mol s-1 m-1)$ and were diluted daily with differentially enriched P and N-media following methods described in Jones et al (*Chapter 3*; , and Prater et al. (2016). This approach allowed us to produce 5 algal-based treatments (high abundance of high-quality food, low P, low N, low food abundance, and cyanobacteria) in addition to one medium-based treatment (low Ca). The

freshwater zooplankter, *Daphnia pulex* (clone # TCO SOM 1.1 AKA 'The Chosen One') was maintained in continuous culture in the Frost lab in P-free COMBO media for > 2 years prior to these experiments, kept at 18-20 °C and fed on ad libitum *S. obliquus*. The Chosen One clone was collected from a pond in Oregon in 2000. The isolate used in this study was obtained from the Yan lab at York University in Ontario, Canada We collected *D. pulex* neonates (<24 hrs old) from the 3rd clutch of brood mothers that had been grown in COMBO medium (Kilham et al. 1998), which were fed using methods described in Jones et al. (2020).

Preparation of low nutrient treatments

We prepared low Ca COMBO by modifying the Kilham et al. (1998) formulation to adjust the amount of CaCl2•2H20 to the nominal concentrations of 1.5 mg/L (low Ca) and 10 mg/L (remaining treatments). We verified actual concentrations of Ca media for all experiments using total reflection X-ray fluorescence spectroscopy (TXRF) on the PICOFOX (S2, Billerica, MA) (Prater et al. 2016). We grew nutrient sufficient, low P and low N algae in growth media adjusted according to amounts described in Jones et al. (*Chapter 3*). We centrifuged subsamples of algal food suspensions that had been pipetted onto pre-weighed GF/C glass fibre filters in order to estimate algal biomass. These subsamples were then dried for at least 45 min at 60°C. After drying, filters were reweighed to determine algal dry mass. We estimated algal C content as 50% of algal dry mass and calculated feeding amounts needed to produce 0.125 mg L-1 of C for the low food treatment and 3 mg L-1 of C for the remaining treatments. We saved algal samples and verified the C and N content of all algal samples with an elemental analyzer (Vario

EL III, Elementar Inc., Mt Laurel, New Jersey). We determined algal P content of all algal samples using potassium persulfate digestion followed by ascorbic acid-molybdate blue colorimetry (APHA 1992). We used the data on algal dry mass and P content to create algal food mixtures with different nutritional ratios: C:P 100 for all treatments excluding low P, C:P 800 for the low P treatment, C: N 8 for all treatments excluding low N, and C:N 16 for the low N treatment.

Experimental set-up

We ran three experiments in total, with each lasting six days. The first experiment had two limiting treatments: low calcium and low nitrogen. The second experiment also had two limiting treatments: low food quantity and low phosphorus. The third experiment had one limiting treatment: cyanobacterial diet. The fast growth treatment was the same for all three experiments and fed animals high quantities of high quality *S. obliquus* in high Ca COMBO medium (Kilham et al. 1998).

For all experiments, we used 15 replicate jars per treatment, with each jar representing a discreet experimental unit. Each jar contained 400ml of COMBO medium and ten >24-hour old *Daphnia* neonates. We fed the *Daphnia* on days 1 and 4 of the experiment, with feeding amounts calculated according to Jones et al. (*Chapter 3*). On day 6 of the experiment, we pooled all ten animals from each jar to form a single experimental replicate and haphazardly assigned jars to the following analysis: three replicate jars were used to calculate mass-specific growth rate (MSGR) with those animals then being used for P analysis ($n=3$); three replicate jars were used for C: N analysis ($n=3$); three replicate jars were used for body % Ca analysis ($n=3$). For RNA

extraction, we used three jars for three replicates for the fast growth phenotype $(n=3)$ and for the nutrient limited treatments, to maximize yields from the smaller animals, we pooled two jars into a single replicate of 20 animals and used six jars total to create three experimental replicates (n=3). Because all three experiments included a fast growth treatment, there were nine replicates for the fast growth phenotype for all analyses. For the MSGR and elemental analysis, all 9 replicates were analyzed. However, for the qPCR portion, we used a Nanodrop™ spectrophotometer to select the three highest RNA quality/yield samples for the assay based upon A260 and A280 readings and total yield.

MSGR and elemental analysis of Daphnia

Animals from each treatment were measured for mass-specific growth rate (MSGR), body C:N ratio, body C:P ratio and body % Ca by dry weight. We calculated MSGR using the equation: $MSGR = (\lfloor ln(B2) - ln(B1) \rfloor)/time$ where B1 and B2 are the average initial and final mass of *Daphnia*, and time is the number of days of growth (i.e., six). To measure the body C:N ratio of *Daphnia*, we collected three experimental replicates, each containing 10 individuals. We placed these groups of animals into preweighed tin cups (1 cup = 1 replicate), dried animals at 60° C for at least 12 hours and reweighed each sample before measuring on an elemental analyzer (Vario EL III, Elementar Inc., Mt Laurel, New Jersey). To measure the dry weight %P of the *Daphnia*, we used three of the experimental replicates from the MSGR analysis per treatment, each containing 10 individuals. These animals were then prepared and measured using the methods described in Jones et al. (2020). We analyzed *Daphnia* dry weight %Ca by placing all ten *Daphnia* from three replicate jars into pre-weighed aluminium cups and

then drying them at 60°C. After drying, we analyzed Ca content using a Bruker S2 TRXF spectrometer as described in Prater et al. (2016).

Statistical analysis of growth and elemental data

To examine the effects of low food abundance, N-limitation, P-limitation, Ca-limitation, and cyanobacterial diet on *Daphnia* MSGR, body C:P ratio, body C:N ratio, body N:P ratio, and body % Ca content, we used a two-factor analysis of variance (ANOVA) with type II errors to account for unbalanced design due to loss of some samples from the fast growth calcium analysis and the nitrogen and calcium limited CN analysis. The analyses were done using R version 4.1.1. (R Core Team 2019) and the car package (Fox et al. 2013).

RNA extraction and cDNA preparation

For each treatment combination, a single jar of 10 *Daphnia* was used for each replicate (n=5). We placed these 10 animals in a 1.5 mL RNase DNase free microfuge tubes. After removing excess COMBO, we added 100 µL of RNAlater® to each tube and flash froze each sample in liquid nitrogen. Samples were stored in a -80°C freezer until extraction. Before RNA extraction, we removed the RNAlater®. We used a partial kitbased extraction combined with a Trizol-chloroform extraction in order to overcome the methodological issues of extracting genomic material from a calcified, chitinous invertebrate (Athanasio et al. 2016). We used the wash buffers and spin columns from a PureLink™ RNA mini kit with a trizol-chloroform phase separation and additional oncolumn DNase treatment prior to washing and elution. We then quantified the RNA

concentration using a NanoDrop 8000 spectrophotometer. We converted the RNA to cDNA using the Applied Biosystems High-Capacity RNA-to-cDNA™ kit. Following the cDNA conversion, we ran PCR with an 18S (Crease and Colbourne 1998) primer pair on all samples to check the integrity of the extracted product, which was visualized on a 2% agarose gel.

Primer design and qPCR

We selected potential indicator genes from an RNA seq dataset produced by Jones et al. (*Chapter three*). This dataset (available in the NIH sequence read archive, accession # PRJNA784271) was produced by growing animals in the six nutritional treatments, followed by high throughput RNA sequencing. After identifying differentially expressed genes (Jones et al, Chapter 3), we used the WFleabase (Colbourne et al. 2005) and Uniprot (Bateman et al. 2019) databases to assign predicted functional annotations to the potential genes. We screened genes with significant, nutrient-specific differential expression using a Poisson regression and functional annotation to identify five genes per limiting treatment which were significantly differentially expressed in that treatment only and whose function could be tied to nutrient-limited molecular processes where possible. This list of genes was different to the list selected in Chapters 2 and 3, with no overlap between the two. Additionally, we selected five potential housekeeping genes based upon the literature and stability of their expression across all treatments using the RNAseq dataset from chapter 3. We obtained sequences and *Daphnia* exon-intron junctions (*Daphnia pulex* genome assembly v1.0) for each gene from the JGI genome portal (Grigoriev et al. 2012, Nordberg et al. 2014) and used them to design exon-exon junction

qPCR primers. We used the Primer3plus (Untergasser et al. 2007) online primer design tool to design two primers per gene, which were then checked for self-dimers, heterodimers and secondary structures using the IDT Oligo Analyzer 3.1 (Integrated DNA Technologies, Inc.) and ran on BLAST (Altschul et al. 1990) to check for specificity. Each primer pair was first run on a control sample using standard PCR to check for appropriate binding, amplification, and product size before running standard curves and gels to obtain the limit of detection (LOD) and limit of quantification (LOQ). Following this initial screening, we selected primers for two genes per nutritional treatment and two reference genes for the final qPCR assay. We randomly selected one fast growth sample to act as an experiment-wide reference sample, three more fast growth samples were selected based on A280 and A260 values to be run as experimental samples. All experimental samples and reference samples were amplified in triplicate. Details of qPCR primers, including sequences, LOD and LOQ can be found in Table 4.1. We used the QuantStudio 3 real-time PCR system to run standard curves and comparative ΔΔCt assays, using SYBR Green PowerUp™ mastermix and the following cycling conditions: 95 \degree C for 30s; (95 \degree C for 5s; annealing temp for 30s) x40; 4 \degree C for ∞ . We ran melting curves for each primer set before running the qPCR product on 2% agarose TBE gels with ethidium bromide to confirm target-specific amplification.

Statistical analysis of qPCR data

We first assessed the raw Ct data for outliers by calculating standard deviation for each set of three analytical replicates. If the standard deviation was above 0.2 then the most extreme outlier was removed prior to analysis (Table S4.1). We then assessed the

stability of each housekeeping gene using the Ct values of each housekeeping gene and the refFinder algorithm (Xie et al. 2012). RefFinder uses four different stability algorithms (BestKeeper (Pfaffl et al. 2004), NormFinder (Andersen et al. 2004), Genorm (Vandesompele et al. 2002), and the comparative delta-Ct method (Silver et al. 2006)) computes the geometric mean from these four algorithms to give a Comprehensive gene stability value (Table S4.2). We then used the 2^{\wedge} - $\Delta\Delta$ Ct method to calculate relative gene expression using the following equations:

- 4.1 Δ Ct experimental = Ct target gene Ct housekeeping gene
- 4.2 Δ Ct reference= Ct target gene Ct housekeeping gene
- 4.3 $\Delta \Delta \text{C}t = \Delta \text{C}t$ experimental $-\Delta \text{C}t$ reference
- 4.4 2^{\wedge} - $\Delta \Delta$ Ct = 2 \wedge (- $\Delta \Delta$ Ct)

Where Ct is the threshold cycle, or the number of cycles at which the fluorescent signal is detectable above the background level and $\rm Ct$ housekeeping gene is the geometric mean Ct of two housekeeping genes.

In order to visualize the expression patterns, we plotted a bar graph of 2^{\wedge} - $\Delta\Delta$ Ct values for each gene in each treatment using ggplot2 (Wickham 2016). We then divided the dataset into the five groups of two nutrient specific genes (Table 2) with two treatment levels; the low supply treatment level comprised the three nutrient limited samples and the high supply treatment comprised all 15 remaining observations. We then fit six linear discriminant analyses (LDA) to the 2^{\wedge} - $\Delta\Delta$ Ct values using the MASS package in R (Venables and Ripley 2002). We fit five LDAs to each of the nutrientspecific datasets to test the capacity of each pair of indicator genes to predict limitation by the specific nutrient and fit box and whisker plots using the linear discriminant

coefficients from the first linear discriminant. We then fit an LDA to the full set of 10 indicator genes in order to test the predictive capacity of the genes when used in a single, combined assay. For the 10 gene model, we used the class predictions to make a variable loadings plot in ggplot2 (Wickham 2016).

Results

MSGR and elemental content

All nutrient limited animals grew significantly slower compared to well-fed animals (Fig. 4.1A, Table 4.2). The cyanobacterial diet produced the slowest growth, with a 61% decrease in growth rate compared to well-fed animals. The low food treatment had the second greatest decrease with a 42% reduction in growth, closely followed by the low phosphorus treatment with a 36% reduction in growth. Finally, the low nitrogen and low calcium treatments both reduced growth by around 9%. Body C:P ratio was significantly decreased under the cyanobacterial and low food quantity treatments and significantly increased under the low P treatment (Fig. 4.1B, Table 4.2). The body C:N ratio was significantly decreased under the low food and cyanobacterial treatments but not significantly affected by any other treatments (Fig. 4.1C, Table 4.2). The body N:P ratio meanwhile was only significantly affected by one diet, increasing significantly under the low P treatment (Fig. 4.1D, Table 4.2). Finally, body Ca% was significantly higher in animals consuming two of the diets, low food, and low phosphorus (Fig. 4.1E, Table 4.2).

Indicator genes as nutrient specific biomarkers

Firstly, the selected housekeeping genes (DAPPUDRAFT_59577 [ligand-gated ion channel] and DAPPUDRAFT__92597 [Oligoribonuclease]) showed stable and consistent expression, with refFinder rankings of 1.189 and 1.414 respectively across all plates (Table S4.2). We found expression patterns ($\Delta \Delta Ct$) that were different among indicator genes across all nutrient limited treatments (Figure 4.2). The vast majority of the selected genes were down-regulated relative to the reference condition (fast growth phenotype) with DAPPUDRAFT_320529 (AMP-dependent synthetase) and DAPPUDRAFT_238038 (Phosphatidylinositol transfer protein SEC14) being the only genes to show mainly up-regulation under limiting conditions. With the exception of DAPPUDRAFT_241140, which had very little response to any of the treatments, all the other genes showed different patterns of expression across treatments with no more than two treatments at a time showing similar levels of expression (Fig. 4.2). The fast growth treatment showed some responsiveness to around half of the genes, showing up- and down-regulation compared to the control genes. This could be demonstrating the possibility that the fast growth animals are adjusting pathways which allow them to exploit plentiful resources (Fig. 4.2). Alternatively, this may simply represent the inherent stochasticity of gene expression (Kaern et al. 2005).

Using linear discriminant analysis, we tested the ability of each primer set (2 per treatment) to predict high or low quantities of the target nutrient (or presence in diet in the case of the cyanobacterial treatment) (Table 4.3). Each individual two gene assay had n=3 low nutrient observations and n=15 high nutrient observations. The nitrogen sensitive (DAPPUDRAFT_315119 and DAPPUDRAFT_300433) genes had the best

overall discriminant success with a 100% accuracy in discriminating low supplies of nitrogen and 93% accuracy in discriminating high supplies of nitrogen. Both the calcium sensitive genes (DAPPUDRAFT_127554 and DAPPUDRAFT_36257) and the carbon (DAPPUDRAFT_313603 and DAPPUDRAFT_320529) sensitive genes had equal success in discriminating low nutrient conditions with 66.67% accuracy, but the calcium genes had slightly higher success than the carbon genes in discriminating high supplies of nutrient (100% vs 93%). Meanwhile, the phosphorus (DAPPUDRAFT_41601 and DAPPUDRAFT_241140) sensitive genes were able to successfully discriminate 33.34% of low phosphorus individuals and 100% of high phosphorus individuals. The cyanobacteria (DAPPUDRAFT_238038 and DAPPUDRAFT_252775) sensitive genes had the lowest discrimination accuracy, with 0% of the cyanobacteria-fed individuals correctly classified and 100% of the green algae-fed individuals correctly classified.

Indicator genes as a single nutritional-state assay

When the 10 genes were taken as a single assay and used to predict one of six nutritional states, they were able to discriminate the five nutrient-limited and one nutrient-replete states with 100% accuracy. When the linear discriminates of this single model were plotted as a scatter plot (Fig. 4.3), there is a clear separation between the majority treatments, with only a slight overlap between the low food and cyanobacteria individuals, and all three replicate samples cluster very closely within their groups. All five nutrient-limited diets separate from the fast growth diet along linear discriminant 1 with 90% of the between-treatment variance explained by that axis. The low phosphorus and low calcium diets additionally separate from the fast growth diet along linear

discriminant 2 with just over 5% of the between-treatment variance explained by that axis.

Discussion

In this study we show how to measure and track nutritional status in the freshwater zooplankton, *Daphnia pulex*. Consumer nutritional state is an important driver of ecosystem processes, affecting nutrient cycling, population dynamics and community interactions (Frost et al. 2002, Raubenheimer et al. 2009, Wagner et al. 2013). Successfully measuring the nutritional state of animals in nature would give us a new and deeper insight into these processes. Moreover, developing single nutrient-specific assays would allow us to parse the individual and interactive roles of multiple key nutrients in a wide range of ecological interactions (Wagner et al. 2013).

MSGR and elemental content

The molecular changes being quantified in this study underpin physiological and life history responses to poor nutrition. We used MSGR as a first order response to confirm whether our nutrient-limited animals were expressing a nutrient-limited phenotype. We found significant reductions in growth rate associated with all five of the nutrient-limited treatments, indicating that all the nutrient-limited animals experienced some form of nutritional limitation. Although we cannot use these data to specifically identify the nutrient which is in short supply, the observed reductions in growth are consistent with the type of limitation that should have occurred as a result of our manipulating the diet or growth media (Sterner and Schulz 1998). Body C:P, C:N, N:P

and Ca% also cannot be used alone to determine which nutrient is in short supply as they all, to some extent, scale with body size and nutrient supply (Prater et al. 2016, Wagner et al. 2017). However, taken with a priori knowledge of diet, these data can tell us how consumers alter the nutritional make up of their body in response to poor nutrition and how different nutrient limitations can affect nutritional homeostasis. C:P ratio for example significantly increased under the low P treatment only, indicating that our P treatment was sufficiently severe enough to weaken the homeostatic control of their body P. Conversely, C:N ratio was only significantly altered in the smallest (Cyanobacteria and low food) animals and not by a low N diet, implying that although the N limitation was severe enough to reduce growth, it was not severe enough for the animals to be pushed away from strict homeostatic control of their body C:N ratio.

Nutritional state indicators

The housekeeping genes selected (DAPPUDRAFT_59577 [ligand-gated ion channel] and DAPPUDRAFT__92597 [Oligoribonuclease]) showed consistently stable expression, with Comprehensive gene stability rankings in line with the most highly ranked housekeeping genes from a range of recent housekeeping gene validation studies (Panina et al. 2018, Kohsler et al. 2020, Cherubini et al. 2021, Giri and Sundar 2022, Gupta et al. 2022). We therefore deemed these genes appropriate for normalization of experimental samples. It is however worth noting that identifying reference genes which are stable across multiple environmental and dietary stressors may not be feasible moving forward as more layers of complexity are added to the validation assays for these primers. It is therefore worth identifying a larger library of reference genes which can be

specifically tailored to each nutrient-specific primer set. We tested our indicator genes with two different assay types: the first was a nutrient-by-nutrient approach where a small number of genes (two in this case) were used to target limitation by a single nutrient of interest. This approach has the advantage of being more efficient and cost effective if researchers are only concerned with one or two nutrients. However, developing indicators that can work in a single state assay can be fraught with complications. Firstly, this approach relies on unambiguous monotonic responses which are sensitive to a single nutrient only across a variety of assay techniques. Our indicator genes were developed from RNA sequence data and were selected based on their responsiveness to a single nutrient. However, the qPCR assays presented here tell a more complex story. Some of the genes, including those for nitrogen and calcium, show very distinct patterns across both the RNAseq and the qPCR analyses with the calcium limiting condition inducing up-regulation in both genes while all other forms of limitation induced down-regulation. Others however, such as the cyanobacteria and phosphorus genes, deviated from the single nutrient response observed in the RNAseq data and revealed far more nuanced expression patterns across multiple nutritional states with this more sensitive qPCR assay, which clouds the expected nutrient-specific response. These deviations primarily took the form of differences in magnitude of gene expression (fold change) and presence of significant differential expression where previously no response was detected. These results show definite promise for the measurement of calcium, carbon, and nitrogen limitation using individual genes, with an overall classification accuracy of 94% for the calcium and nitrogen genes and 88.89% for the low carbon genes. Further development and testing will be needed to identify phosphorus-, and cyanobacteria-specific genes

(different from the ones studied here) which retain tightly nutrient-specific responses across different assays. In addition, more low nutrient samples should be run to determine whether the discrimination accuracy can be improved for the low calcium and low carbon genes or whether new genes need to be identified.

The second approach we used was a full nutritional state assay, using all ten indicator genes at once with a 'barcode' approach, where responses across all ten genes provide information on the prevailing nutritional state. Although this approach involves a larger and therefore more expensive assay, it also yielded a far greater success rate (100% classification accuracy for all 6 diets, including fast growth animals). These indicator genes will require further validation and testing prior to deployment, with the next step being to assess responses across a gradient of limitation to determine 'dose dependency' and testing stability of expression in response to other environmental stressors, such as temperature and salinity (Wagner et al. 2013, Frost et al. 2014). However, the results presented here represent a first order confirmation that nutritional profiling using transcriptomics can yield highly nutrient specific results and can be used to classify animal nutritional state.

Functional lessons from the transcriptome

While functional annotation plays an important role in selecting nutrient- and other stress-responsive genes, the responses of these genes also have the potential to inform functional annotations and contribute to our understanding of the effects of poor nutrition on consumers.

For example, both calcium indicator genes are tentatively associated with the moulting process (Table 3), one (NADP) with the synthesis of moulting hormones and the other (succinate dehydrogenase) with the Krebs cycle as a potential precursor to moulting. The nutrient-specific expression patterns of these two genes (Fig. 1, DAPPUDRAFT_127554 and DAPPUDRAFT_36257) show a clear down-regulation of DAPPUDRAFT_127554 (NADP) under the other growth limiting diets but up-regulation under calcium limitation and fast growth. Similarly, DAPPUDRAFT_36257 (succinate dehydrogenase) showed a clear down-regulation under three of the growth limiting diets, no response to the cyanobacterial diet and up-regulation under low calcium. This supports the currently held consensus that calcium affects moulting processes (Porcella et al. 1969, Alstad et al. 1999, Hessen et al. 2000, Pérez-Fuentetaja and Goodberry 2016), and provides support to the hypothesis that phosphorus and carbon are also key components of the moulting cycle (Hessen and Rukke 2000). It also demonstrates a previously unidentified link between other forms of growth limitation (low nitrogen and cyanobacterial) and changes to the moulting cycle. Finally, DAPPUDRAFT_36257 (succinate dehydrogenase) was up regulated in the fast-growing animals and was by far the greatest response to any of the 10 genes in the fast growth phenotype. This provides compelling evidence that not only are the *Daphnia* making metabolic adjustments to deal with the effects of poor nutrition but that they are also making metabolic adjustments to better exploit plentiful resources.

The low carbon genes meanwhile were associated with production of juvenile hormone (JH), which in turn controls the production of male neonates as the population shifts from parthenogenic to sexual reproduction under times of environmental stress

(DAPPUDRAFT_313603, ionotropic glutamate receptor) and regulates energy balance and metabolism (DAPPUDRAFT_320529, AMP-dependent synthetase). Interestingly, it appears that all the growth limiting treatments showed down-regulation of the ionotropic glutamate receptor (IGR), with particularly strong responses from the low food (carbon) and cyanobacterial treatments. Miyakawa et al. (2015) hypothesized a possible antagonistic relationship between the expression of IGR and the expression of JH. The down-regulation of IGR under conditions (low food availability) known to induce a switch from parthenogenesis to sexual reproduction (Ebert 2005) that we see in our data would appear to support this hypothesis. Additionally, the production of sexual offspring has primarily been associated with high population density, high temperatures and intermittent environments (such as seasonal drying of ponds) (Ebert 2005). Our data show that the strongest down-regulation of IGR actually occurred under the cyanobacterial treatment, suggesting that *Daphnia* populations may also use sexual reproduction and diapause as an adaptive response to cyanobacterial blooms.

For the nitrogen limited gene, DAPPUDRAFT_315119, there is no current functional annotation but we found that it is clearly sensitive to nutritional limitation (Fig. 4.2), particularly from calcium, nitrogen, and phosphorus. This gene, which is 229 bp in length, appears to have no current orthologs outside of the crustaceans, with a BLAST search only finding alignments within the *Daphnia* genus. Additionally, none of the *Daphnia* alignments identified in the BLAST search had any functional annotation other than "predicted" or "hypothetical" protein. Given the highly genus-specific stress architecture identified by Orsini et al (2018), this gene potentially represents part of this architecture.

These findings demonstrate that patterns of differential gene expression in a selected set of indicator genes are highly nutrient specific and able to discriminate between 6 nutritional states with 100% accuracy when used together in a nutritional state assay. This has implications for the study of organismal nutrition and associated ecological processes in allowing us to detect nutritional status in animals with an unknown nutritional history. Future nutritional ecology studies, particularly those based in wild or semi-wild environments, will be able to incorporate consumer responses in addition to abiotic drivers and have greater certainty in assessing the type and intensity of nutritional limitation.

Figures and tables

Table 4.1: Details of qPCR primers designed for genes to be used as nutritional state indicators in *Daphnia pulex* exposed to six dietary treatments: high quantity of high-quality food, low calcium, high cyanobacteria, low carbon/food quantity, low nitrogen, and low phosphorus. Metrics include melt temperature (Tm), limit of quantification (LOQ) and limit of detection (LOD).

transfer

protein

91

92

Table 4.2. ANOVA table of MSGR, C:P ratio, C:N ratio, N:P ratio and body Ca% (by dry weight) of *Daphnia pulex* grown under varied conditions of calcium [Ca], food quantity [C], cyanobacterial diet (Cy), nitrogen [N], and phosphorus [P] limitations. Significant results in bold.

Trait	Treatment	DF	\boldsymbol{F}	$\mathbf P$
MSGR	Ca	1,18	7.409	< 0.05
	$\mathbf C$	1,18	308.739	< 0.0001
	Cy	1,18	152.349	< 0.0001
	$\mathbf N$	1,18	7.676	< 0.05
	${\bf P}$	1,18	107.951	< 0.0001
C: P	Ca	1,18	1.432	>0.1
	\mathcal{C}	1,18	15.341	< 0.005
	Cy	1,18	20.608	0.001
	${\bf N}$	1,18	0.18	>01
	${\bf P}$	1,18	19.512	0.001
C: N	Ca	1,18	2.209	>0.1
	C	1,18	6.204	< 0.05
	Cy	1,18	6.8364	< 0.05
	${\bf N}$	1,18	1.508	>0.1
	${\bf P}$	1,18	1.918	>0.1
N:P	Ca	1,18	0.199	>0.1
	C	1,18	0.5672	>0.1
	Cy	1,18	1.488	>0.1
	\overline{N}	1,18	0.619	>0.1
	${\bf P}$	1,18	25.909	0.01
Body Ca%	Ca	1,18	0.149	>0.1
	$\mathbf C$	1,18	0.8223	>0.1
	$\mathbf{C}\mathbf{y}$	1,18	17.464	0.001
	\overline{N}	1,18	1.871	>0.1
	${\bf P}$	1,18	4.773	< 0.05

Table 4.3: Results of linear discriminant analysis to assess predictive capacity of five pairs of nutrient-specific indicator genes in *Daphnia pulex*. Nutrient-limited n=3, nutrient-replete n=15.

Table 4.4: Details of predicted gene function and potential links to biological processes for ten indicator genes selected for use in *Daphnia pulex* to predict nutritional status of animals fed on one of six diets: high quantities of high-quality food, low calcium, high cyanobacteria, low carbon/food, low nitrogen, and low phosphorus.

Figure 4.1. Mean mass-specific growth rates (MSGR), C:P ratio, C:N ratio, N:P ratio and body %Ca (by dry weight) of *Daphnia pulex* grown under 6 dietary treatments: high quantity of high-quality food (fast growth), low calcium, high cyanobacteria, low food quantity (carbon), low nitrogen and low phosphorus. **A.** Mean MSGR \pm 1 S.E. (fast growth $n = 9$, nutrient limited $n = 3$). **B.** Mean C: P ratio ± 1 S.E (by mol) (fast growth $n =$ 9, nutrient limited $n = 3$). **C.** Mean C:N ratio ± 1 S.E. (by mol) (fast growth $n = 9$, nutrient limited n =3). **D.** Mean N:P ratio \pm 1 S.E. (fast growth n = 9, nutrient limited n =3). **E.** Mean Ca% \pm 1 S.E. (by dry weight) (fast growth n = 9, nutrient limited n = 3).

Figure 4.2. Mean fold change $(2^{\wedge} - \Delta \Delta \text{Ct}) \pm 1$ s.e. (n=3) of ten nutritional indicator genes in *Daphnia pulex* exposed to six diets: high quantities of high-quality food (fast growth), low calcium, low carbon/food, high cyanobacteria, low nitrogen, and low phosphorus.

Figure 4.3. Linear Discriminant Analysis loadings plot of relative expression (2^(- ΔΔCt)) (n=3) of ten nutritional indicator genes in *Daphnia pulex* exposed to six diets: high quantities of high-quality food (fast growth), low calcium, low carbon/food, high cyanobacteria, low nitrogen, and low phosphorus.

Conclusion

In this thesis, I investigated patterns of gene expression in *Daphnia* in response to different forms of nutritional stress. The goal of carrying out this research was to assess whether nutrient-specific patterns of gene expression can be used as indicators of nutritional stress in *Daphnia* and to identify a select number of highly responsive genes to be used as biomarkers in future nutritional state assays. In the preceding three chapters, I addressed three key methodological and theoretical issues with regards to deploying transcriptomic nutritional state indicators: 1) differentiating singular limitation from colimitation, 2) variable reduction and selection from a large RNAseq dataset, and 3) validation of selected genes with qPCR. Having addressed these issues and identified a pipeline for the identification and validation of nutritional biomarkers, I present several suggested directions for future work in this area.

In Chapter two, I addressed the issue of co-limitation, specifically whether *Daphnia* can detect and respond to limiting supplies of more than one nutrient at once, and whether this co-limitation can be differentiated in the transcriptome. I used phosphorus and calcium as the two nutrients and deployed a fully factoral experimental design and sequenced RNA from these animals. I found that clear and distinct nutrient-limited phenotypes could be identified from *Daphnia* RNA sequences. Moreover, the co-limited animals showed a patchwork of responses, some of which were non-additive combinations unique to the co-limited animals, but overall, the co-limited animals presented a blended phenotype with distinct and recognizable features of the singly limited phenotypes. These findings are of great importance to both fundamental and applied research into consumer nutrition as they confirm that *Daphnia* (and therefore

potentially other herbivore consumers) can detect and respond to limiting supplies of more than one nutrient. This largely contradicts the Law of the Minimum (Von Liebig 1840, Paris 1992, van der Ploeg et al. 1999) for animal consumers and paves the way for future studies that will more closely examine the mechanisms by which consumers detect and react to low supplies of nutrients. Additionally, these findings provide evidence that genetic biomarkers of growth limitation by a single nutrient or limiting diet can also be used to detect co-limitation, making them far more useful in applied studies and ecosystem management scenarios than if specific biomarkers were needed for every possible combination of nutrients/diets. It is possible that the blended phenotype produced by the co-limited animals was due to the fact that phosphorus and calcium have such different uptake and acquisition pathways, with phosphorus being obtained from the diet and taken up post-digestion through the gut while calcium is taken up as cations directly from the water column and largely bypass the digestive system. Previous studies examining and confirming zooplankton dietary co-limitation at a physiological or molecular level have mostly focused on co-limitation from complex biomolecules (such as low PUFAs as a result of cyanobacterial diet) with or without a single dietary nutrient (such as P) (Martin-Creuzburg et al. 2009, Martin-Creuzburg et al. 2010, Schalicke et al. 2019a, Schalicke et al. 2019b). Consequently, while my findings are supported by a welldeveloped body of literature, there has been little focus on co-limitation by nutrients sharing very similar intake and acquisition pathways. A key next step in validating potential biomarkers would therefore be to study more mixtures of dietary nutrients, such as phosphorus and nitrogen or carbon.

In Chapter three, I addressed the issue of selecting a few highly predictive genes from a large RNAseq dataset. Perhaps the most compelling finding in this chapter was the need for human intervention at the final stages of the selection process. I explored multiple machine learning and variable reduction algorithms optimized specifically for large, sparse datasets such as mine and found that elastic net regularized regression (ENR) was by far the most successful at eliminating redundant variables from my dataset and selecting a reduced subset (~ 130) from the main dataset of $\sim 33,050$ genes, compared to other algorithms such as sparse partial least squares regression and KOG enrichment analysis. However, 130 genes still represent a potentially unwieldy number of genes to routinely assay in subsequent studies. This is particularly true for applied, ecology-driven studies where willingness to employ molecular techniques is highly dependent on cost and ease of deployment. Although custom microarrays are available, which have the potential to detect thousands of genes on a single chip, these microarrays have been found to have a lower dynamic range than RNAseq and must frequently be validated using qPCR (Zhao et al. 2014). In order to reduce the number of potential indicators down to a manageable assay of 10-20 genes, it was necessary to manually prune the list of 130 using PCA loadings and Euclidian distance trees. The algorithms I used for dimension reduction proved to be quite powerful at identifying and excluding redundant variables that did not contain unique information. However, once many of the redundant variables are removed, they were unsuccessful at further trimming the dataset to include only the most highly nutrient-sensitive, predictive genes. A recent review (Arowolo et al. 2021) on variable selection from RNAseq data found that removal of redundant variables and selection of highly predictive variables are best done in two steps with two models,

one optimized for dimension reduction and one optimized for feature selection, supporting my findings that ENR performs well at dimension reduction but that different models (PCA and Euclidian trees) are necessary for further feature selection. Ultimately, it was necessary to develop my own pipeline for dimension reduction and variable selection because my objectives and downstream applications were so different from the majority of much of the published literature in this area. Many of the studies I found were focused on clinical applications of genetic biomarkers (e.g., (Wang et al. 2017, Ching et al. 2018, Pouyan and Kostka 2018, Kobak and Berens 2019, Petegrosso et al. 2020, Arowolo et al. 2021)), where binary outcomes are more common (i.e., presence or absence of disease) than multiple outcomes (i.e., six distinct nutritional phenotypes). A more tailored approach was therefore needed for my dataset. An additional complication was the release of a new version of the *D. pulex* genome (Ye et al. 2017), sequenced from the PA42 clone. Because the TCO clone has been found to have a >90% reduction in nucleotide diversity, leading to multiple deleterious genomic features, the original TCO sequence (Colbourne et al. 2011) is believed to over-estimate the true *D. pulex* genome size by around 7000 genes. I made the choice to use the sequenced genome closest to my experimental clone, which is the over-estimated 2011 release, however future studies should prioritize using the 2017 PA42 genome release as it is more representative of the broader *Daphnia pulex* population.

In Chapter four, I validated ten potential indicator genes and two potential reference genes using qPCR. I tested two approaches to nutritional profiling: the first was a nutrient-by-nutrient approach where I used each pair of nutrient-sensitive genes as a discreet assay to detect limitation by a single nutrient. The second was a full nutritional

state assay where I used expression patterns of all ten nutrient sensitive genes as unique barcodes to discriminate samples into one of six possible nutritional states. The ten genes used in this chapter were manually selected from the RNAseq dataset in Chapter three using the results from Poisson distributed linear models. The first finding of note was the difference in expression patterns between the RNAseq data and the qPCR data. In the RNAseq dataset, each of my selected genes was sensitive to one nutrient while the qPCR assay revealed a more nuanced set of responses with most of the selected genes responding in phenotypes, both nutrient-limited and fast growth. This unexpected suite of responses found by qPCR reduced the predictive capability of the discreet gene-by-gene assays, with only the calcium sensitive genes able to discriminate high from low supplies with 100% accuracy. In contrast, the expression patterns generated by all ten genes were distinct enough that nutritional state (including well fed, fast growing) could be discriminated with 100% accuracy. This represents a convincing first order confirmation that transcriptomics is a powerful tool for detecting nutritional stress in *Daphnia* in both lab and field experiments, although application in the field has yet to be demonstrated..

Transcriptomic biomarkers of stress are already being investigated and deployed for a variety of applications including monitoring the health of crops (Ferrandez-Ayela et al. 2016) and livestock (Xu et al. 2017) in agriculture and aquaculture (Overturf and Hardy 2001, Tupac-Yupanqui et al. 2013, Chandhini and Kumar 2019), tracking the effects of climate change and heat stress on corals (Morgan et al. 2001, Kenkel et al. 2014, Dixon et al. 2015), sea turtles (Tedeschi et al. 2015), and other marine organisms such as oysters (Farcy et al. 2009), and as early warning systems of exposure to toxins (Snell et al. 2003) such as cadmium (Roh et al. 2006, Liu et al. 2008) and other heavy metals (Benhamed et

al. 2016). In order to use nutritional stress biomarkers in the same way, my selected genes will require extensive validation. This goes both for the larger library identified in Chapter three, which requires primers to be designed and tested and for the genes highlighted in Chapter four, for which working primers have been designed.

The validation of potential genetic biomarkers includes assessing the stability of their expression under other environmental stressors, such as temperature, anoxia, and predation; testing responsiveness in a variety of different genotypes; assessing the effect of mixtures and co-limitation; and finally, testing in a field setting where environmental conditions are variable and the animal's nutritional history is unknown. Starting with a larger library of potential genes means that we can retain a working library while discarding genes that may prove to be sensitive to other environmental stressors, or behave unpredictably in mixtures, or even potentially be too highly conserved within the one genotype that I used. In addition to furthering the development of nutritional indicators, I have also contributed two large RNA sequence datasets to the NIH sequence read archives. These sequences form a permanent genetic resource to be utilized by other researchers, not just for bioindicators but for a wide-range of studies including functional annotation of nutrient-sensitive genes and isoform analysis to investigate the differentiation of nutritional and growth phenotypes.

This work represents a step forward in our understanding of the effects of poor nutrition on zooplankton physiology in both fundamental and applied contexts. First, these results contribute to our understanding of the metabolic and molecular processes that consumers adjust/deploy in response to poor nutrition. For example, my finding that pathways relating to ecdysis are not just altered by calcium limitation but by all five

growth limiting treatments. This finding challenges the paradigm that *Daphnia* molt on a consistent schedule regardless of nutrient availability (Porcella et al. 1969) and raises further questions about the role of ecdysis in somatic growth. Secondly, my findings could help contribute towards further functional annotation of the *Daphnia* genome. The current genome annotation is largely incomplete and made up primarily of predicted functions based on gene ontology with other sequenced taxa (Ravindran et al. 2019). Knowing how the expression of each gene changes in relation to nutrient supply will help researchers begin to identify functional groups of genes and aid in targeting genes for functional assays such as knockout studies. Finally, being able to measure the nutritional state of consumers with an unknown nutritional history could transform the study of ecological interactions in the long-term monitoring of aquatic ecosystems. Most knowledge of animal nutrition *in situ* is inferred from measurement of food and does not involve sampling animals directly. The direct measurement of consumer nutritional state could allow us to directly measure nutrient flow and track trophic interactions in natural foodwebs, opening the door for further studies into the interactions between nutrition, nutrient cycling, population growth, and community interactions. In addition, our ability to monitor nutrients and their fluxes in freshwater and marine ecosystems relies on collecting years of data in order to establish a baseline for that specific system and then detect any significant deviations from that baseline. Measuring nutritional limitation/stress in the organisms rather than nutrient concentrations in the water means that we can potentially detect changes to nutrient dynamics in a single sampling point. Daphniid responses in particular are so useful because their ecoresponsive genome responds quickly to environmental stress and facilitates rapid local adaptation to

environmental conditions (Boersma et al. 1999, DeClerck et al. 2001, Schwarzenberger et al. 2012). This means that if a *Daphnia* population is showing signs of nutritional limitation, this likely reflects a new, rapid, or extreme change warranting further investigation and management. Scaling the application of nutritional indicators up to a population level is not without its challenges. In addition to the complex interactions between nutritional and environmental stressors present at an ecosystem level, wild populations themselves are inherently more complex than experimental populations grown in the lab. Wild populations contain a mix of ages, life stages, and sexes meaning that the responses of the nutritional indicator must be stable enough that ontogenic noise does not cloud the response. Wild populations also contain more mutations, as well as polyploid hybrids and, depending on the environmental conditions, a mixture of haploid and diploid individuals. All these factors have the potential to add noise to the overall transcriptomic stress signature. However, once these issues are addressed and nutritional indicators have been included as part of a larger biomonitoring toolkit, they have the potential to revolutionize how we monitor and manage ecosystems in the face of increasing anthropogenic change.

References

- Acharya, K., M. Kyle, and J. J. Elser. 2004. Effects of stoichiometric dietary mixing on *Daphnia* growth and reproduction. Oecologia **138**:333-340.
- Ackrell, B. A. C., M. K. Johnson, R. P. Gunsalus, and G. Cecchini. 1992. Structure and function of succinate dehydrogenase and fumarate reductase. Chemistry and. Biochemistry of Flavoenzymes, ed. F. Müller **3**:229-297.
- Acquisti, C., S. Kumar, and J. J. Elser. 2009. Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus. Proceedings of the Royal Society B-Biological Sciences **276**:2605-2610.
- Akbarzadeh, A., A. L. S. Houde, B. G. Sutherland, O. P. Gunther, and K. M. Miller. 2020. Identification of Hypoxia-Specific Biomarkers in Salmonids Using RNA-Sequencing and Validation Using High-Throughput qPCR. G3-Genes Genomes Genetics **10**:3321-3336.
- Alfaro, A. C., and T. Young. 2018. Showcasing metabolomic applications in aquaculture: areview. Reviews in Aquaculture **10**:135-152.
- Alstad, N. E. W., L. Skardal, and D. O. Hessen. 1999. The effect of calcium concentration on the calcification of *Daphnia magna*. Limnology and Oceanography **44**:2011-2017.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology **215**:403-410.
- Altshuler, I., A. M. McLeod, J. K. Colbourne, N. D. Yan, and M. E. Cristescu. 2015. Synergistic interactions of biotic and abiotic environmental stressors on gene expression. Genome **58**:99-109.
- Andersen, C. L., J. L. Jensen, and T. F. Orntoft. 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 Research **64**:5245-5250.
- Anderson, T. R., and D. O. Hessen. 2005. Threshold elemental ratios for carbon versus phosphorus limitation in *Daphnia*. Freshwater Biology **50**:2063-2075.
- Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data., [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)..
- APHA. 1992. Standard Methods for the Examination of Water and Wastewater. 18th Edition, American Public Health Association (APHA), American Water Works Association (AWWA) and Water Pollution Control Federation (WPCF), Washington DC.
- Armstrong, F. A. J., and D. W. Schindler. 1971. Preliminary Chemical Characterization of Waters in the Experimental Lakes Area, Northwestern Ontario. Journal of the Fisheries Research Board of Canada **28(2)**:171-187.
- Arowolo, M. O., M. O. Adebiyi, and A. A. Adebiyi. 2021. Enhanced Dimensionality Reduction Methods for Classifying Malaria Vector Dataset using Decision Tree. Sains Malaysiana **50**:2579-2589.
- Arvola, L., M. Jarvinen, and T. Tulonen. 2011. Long-term trends and regional differences of phytoplankton in large Finnish lakes. Hydrobiologia **660**:125-134.
- Asselman, J., M. E. Pfrender, J. A. Lopez, J. R. Shaw, and K. A. C. De Schamphelaere. 2018. Gene Coexpression Networks Drive and Predict Reproductive Effects in

Daphnia in Response to Environmental Disturbances. Environmental Science & Technology **52**:317-326.

- Athanasio, C. G., J. K. Chipman, M. R. Viant, and L. Mirbahai. 2016. Optimisation of DNA extraction from the crustacean *Daphnia*. PeerJ:e2004 **4**.
- Azan, S. S. E., S. E. Arnott, and N. D. Yan. 2015. A review of the effects of *Bythotrephes longimanus* and calcium decline on zooplankton communities - can interactive effects be predicted? Environmental Reviews **23**:395-413.
- Bae, M. J., and Y. S. Park. 2014. Biological early warning system based on the responses of aquatic organisms to disturbances: A review. Science of the Total Environment **466**:635-649.
- Balla, T. 2013. Phosphoinositides: Tiny lipids with a giant impact on cell regulation. Physiological Reviews **93**:1019-1137.
- Bateman, A., M. J. Martin, S. Orchard, M. Magrane, E. Alpi, B. Bely, and e. al. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research **47**:D506-D515.
- Becker, C., and M. Boersma. 2003. Resource quality effects on life histories of *Daphnia*. Limnology and Oceanography **48**:700-706.
- Becker, C., and M. Boersma. 2005. Differential effects of phosphorus and fatty acids on *Daphnia magna* growth and reproduction. Limnology and Oceanography **50**:388- 397.
- Becker, D., Y. Reydelet, J. A. Lopez, C. Jackson, J. K. Colbourne, S. Hawat, M. Hippler, B. Zeis, and R. J. Paul. 2018. The transcriptomic and proteomic responses of

Daphnia pulex to changes in temperature and food supply comprise environmentspecific and clone-specific elements. Bmc Genomics **19**.

- Benhamed, S., F. A. Guardiola, S. Martinez, M. J. Martinez-Sanchez, C. Perez-Sirvent, M. Mars, and M. A. Esteban. 2016. Exposure of the gilthead seabream (*Sparus aurata*) to sediments contaminated with heavy metals down-regulates the gene expression of stress biomarkers. Toxicology Reports **3**:364-372.
- Blanchfield, P. J., M. J. Paterson, J. A. Shearer, and D. W. Schindler. 2009. Johnson and Vallentyne's legacy: 40 years of aquatic research at the Experimental Lakes Area INTRODUCTION. Canadian Journal of Fisheries and Aquatic Sciences **66**:1831- 1836.
- Boer, V. M., C. A. Crutchfield, P. H. Bradley, D. Botstein, and J. D. Rabinowitz. 2010. Growth-limiting Intracellular Metabolites in Yeast Growing under Diverse Nutrient Limitations. Molecular Biology of the Cell **21**:198-211.
- Boer, V. M., J. H. de Winde, J. T. Pronk, and M. D. W. Piper. 2003. The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. Journal of Biological Chemistry **278**:3265-3274.
- Boersma, M. 2000. The nutritional quality of P-limited algae for *Daphnia*. Limnology and Oceanography **45**:1157-1161.
- Boersma, M., L. De Meester, and P. Spaak. 1999. Environmental stress and local adaptation in *Daphnia magna*. Limnology and Oceanography **44**:393-402.
- Bogaert, K. A., E. Perez, J. Rumin, A. Giltay, M. Carone, N. Coosemans, M. Radoux, G. Eppe, R. D. Levine, F. Remacle, and C. Remacle. 2019. Metabolic, Physiological,

and Transcriptomics Analysis of Batch Cultures of the Green Microalga *Chlamydomonas* Grown on Different Acetate Concentrations. Cells **8**.

- Boggs, C. L. 2009. Understanding insect life histories and senescence through a resource allocation lens. Functional Ecology **23**:27-37.
- Boggs, C. L., and C. L. Ross. 1993. The effect of adult food limitation on life-history traits in S*peyeria mormonia* (Lepidoptera, nymphalida). Ecology **74**:433-441.
- Bukovinszky, T., A. M. Verschoor, N. R. Helmsing, T. M. Bezemer, E. S. Bakker, M. Vos, and L. N. d. S. Domis. 2012. The Good, the Bad and the Plenty: Interactive effects of food quality and quantity on the growth of different *Daphnia* species. Plos One **7**.
- Bunnell, D. B., R. P. Barbiero, S. A. Ludsin, C. P. Madenjian, G. J. Warren, D. M. Dolan, T. O. Brenden, R. Briland, O. T. Gorman, J. X. He, T. H. Johengen, B. F. Lantry, B. M. Lesht, T. F. Nalepa, S. C. Riley, C. M. Riseng, T. J. Treska, I. Tsehaye, M. G. Walsh, D. M. Warner, and B. C. Weidel. 2014. Changing Ecosystem Dynamics in the Laurentian Great Lakes: Bottom-Up and Top-Down Regulation. Bioscience **64**:26-39.
- Buoso, S., N. Tomasi, M. Arkoun, A. Maillard, L. Jing, F. Marroni, S. Pluchon, R. Pinton, and L. Zanin. 2021. Transcriptomic and metabolomic profiles of *Zea mays* fed with urea and ammonium. Physiologia Plantarum **173**:935-953.
- Cairns, A., and N. Yan. 2009. A review of the influence of low ambient calcium concentrations on freshwater daphniids, gammarids, and crayfish. Environmental Reviews **17**:67-79.
- Campos, B., B. Pina, and C. Barata. 2021. *Daphnia magna* gut-specific transcriptomic responses to feeding inhibiting chemicals and food limitation. Environmental Toxicology and Chemistry **40**:2510-2520.
- Cao, K. A. L., S. Boitard, and P. Besse. 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. Bmc Bioinformatics **12**.
- Carrillo, P., M. Villar-Argaiz, and J. M. Medina-Sanchez. 2001. Relationship between N : P ratio and growth rate during the life cycle of calanoid copepods: An *in situ* measurement. Journal of Plankton Research **23**:537-547.
- Chandhini, S., and V. J. R. Kumar. 2019. Transcriptomics in aquaculture: current status and applications. Reviews in Aquaculture **11**:1379-1397.
- Charles, J. P. 2010. The regulation of expression of insect cuticle protein genes. Insect Biochemistry and Molecular Biology **40**:205-213.
- Chen, H. 2018. VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20. [https://CRAN.R-project.org/package=VennDiagram](https://cran.r-project.org/package=VennDiagram)
- Cherubini, A., F. Rusconi, and L. Lazzari. 2021. Identification of the best housekeeping gene for RT-qPCR analysis of human pancreatic organoids. Plos One **16**.
- Ching, T., X. Zhu, and L. X. Garmire. 2018. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. Plos Computational Biology **14**.
- Chowdhury, P. R., D. Frisch, D. Becker, J. A. Lopez, L. J. Weider, J. K. Colbourne, and P. D. Jeyasingh. 2015. Differential transcriptomic responses of ancient and modern *Daphnia* genotypes to phosphorus supply. Mol Ecol **24**:123-135.
- Christjani, M., P. Fink, and E. von Elert. 2016. Phenotypic plasticity in three *Daphnia* genotypes in response to predator kairomone: evidence for an involvement of chitin deacetylases. Journal of Experimental Biology **219**:1697-1704.
- Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Caceres, L. Carmel, C. Casola, J. H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Frohlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V. Koonin, E. M. Zdobnov, I. V. Grigoriev, M. Lynch, and J. L. Boore. 2011. The ecoresponsive genome of *Daphnia pulex*. Science **331**:555-561.
- Colbourne, J. K., V. R. Singan, and D. G. Gilbert. 2005. wFleaBase: The *Daphnia* genome database. Bmc Bioinformatics **6:45**.
- Cook, S. C., M. D. Eubanks, R. E. Gold, and S. T. Behmer. 2016. Summer and fall ants have different physiological responses to food macronutrient content. Journal of Insect Physiology **87**:35-44.
- Crain, C. M., K. Kroeker, and B. S. Halpern. 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters **11**:1304-1315.
- Crease, T. J., and J. K. Colbourne. 1998. The unusually long small-subunit ribosomal RNA of the crustacean, *Daphnia pulex:* Sequence and predicted secondary structure. Journal of Molecular Evolution **46**:307-313.
- Davis, T. W., F. Koch, M. A. Marcoval, S. W. Wilhelm, and C. J. Gobler. 2012. Mesozooplankton and microzooplankton grazing during cyanobacterial blooms in the western basin of Lake Erie. Harmful Algae **15**:26-35.
- DeClerck, S., C. Cousyn, and L. De Meester. 2001. Evidence for local adaptation in neighbouring *Daphnia* populations: a laboratory transplant experiment. Freshwater Biology **46**:187-198.
- DeMott, W. R., R. D. Gulati, and K. Siewertsen. 1998. Effects of phosphorus-deficient diets on the carbon and phosphorus balance of *Daphnia magna*. Limnology and Oceanography **43**:1147-1161.
- DeMott, W. R., and B. J. Pape. 2005. Stoichiometry in an ecological context: testing for links between *Daphnia* P-content, growth rate and habitat preference. Oecologia **142**:20-27.
- Dixon, G. B., S. W. Davies, G. A. Aglyamova, E. Meyer, L. K. Bay, and M. V. Matz. 2015. Genomic determinants of coral heat tolerance across latitudes. Science **348**:1460-1462.
- Dodds, W. K., C. T. Robinson, E. E. Gaiser, G. J. A. Hansen, H. Powell, J. M. Smith, N. B. Morse, S. L. Johnson, S. V. Gregory, T. Bell, T. K. Kratz, and W. H. McDowell. 2012. Surprises and Insights from Long-Term Aquatic Data Sets and Experiments. Bioscience **62**:709-721.
- Ebert, D. 2005. Chapter 2, Introduction to *Daphnia* Biology. Ecology, Epidemiology, and Evolution of Parasitism in Daphnia Bethesda (MD): National Center for Biotechnology Information (US).
- Eggermont, H., and K. Martens. 2011. Preface: Cladocera crustaceans: sentinels of environmental change. Hydrobiologia **676**:1-7.
- Eimers, M. C., S. A. Watmough, A. M. Paterson, P. J. Dillon, and H. X. Yao. 2009. Long-term declines in phosphorus export from forested catchments in southcentral Ontario. Canadian Journal of Fisheries and Aquatic Sciences **66**:1682- 1692.
- Elser, J. J., K. Acharya, M. Kyle, J. Cotner, W. Makino, T. Markow, T. Watts, S. Hobbie, W. Fagan, J. Schade, J. Hood, and R. W. Sterner. 2003. Growth ratestoichiometry couplings in diverse biota. Ecology Letters **6**:936-943.
- Elser, J. J., M. Kyle, J. Learned, M. L. McCrackin, A. Peace, and L. Steger. 2016. Life on the stoichiometric knife-edge: effects of high and low food C:P ratio on growth, feeding, and respiration in three *Daphnia* species. Inland Waters **6**:136-146.
- Evans, T. G. 2015. Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. Journal of Experimental Biology **218**:1925-1935.
- Farcy, E., C. Voiseux, J. M. Lebel, and B. Fievet. 2009. Transcriptional expression levels of cell stress marker genes in the Pacific oyster *Crassostrea gigas* exposed to acute thermal stress. Cell Stress & Chaperones **14**:371-380.
- Feder, M. E., and J. C. Walser. 2005. The biological limitations of transcriptomics in elucidating stress and stress responses. Journal of Evolutionary Biology **18**:901- 910.
- Ferrandez-Ayela, A., A. B. Sanchez-Garcia, C. Martinez-Andujar, Z. Kevei, M. L. Gifford, A. J. Thompson, and e. al. 2016. Identification of novel stress-responsive biomarkers from gene expression datasets in tomato roots. Functional Plant Biology **43**:783-796.
- Ferrao-Filho, A. d. S., A. J. Tessier, and W. R. DeMott. 2007. Sensitivity of herbivorous zooplankton to phosphorus-deficient diets: Testing stoichiometric theory and the growth rate hypothesis. Limnology and Oceanography **52**:407-415.
- Fohringer, C., I. Dudka, R. Spitzer, F. Stenbacka, O. Rzhepishevska, J. Cromsigt, G. Grobner, G. Ericsson, and N. J. Singh. 2021. Integrating omics to characterize eco-physiological adaptations: How moose diet and metabolism differ across biogeographic zones. Ecology and Evolution **11**:3159-3183.
- Fox, J., M. Friendly, and S. Weisberg. 2013. Hypothesis Tests for Multivariate Linear Models Using the car Package. R Journal **5**:39-52.
- Freedman, B. 2015. Ecological Effects of Environmental Stressors. Oxford University Press.
- Freedman, B., and S. Beauchamp. 1998. Implications of atmospheric change for biodiversity of aquatic ecosystems in Canada. Environmental Monitoring and Assessment **49**:271-280.
- Fridolfsson, E., E. Lindehoff, C. Legrand, and S. Hylander. 2018. Thiamin (vitamin B-1) content in phytoplankton and zooplankton in the presence of filamentous cyanobacteria. Limnology and Oceanography **63**:2423-2435.
- Friedman, J. H., T. Hastie, and R. Tibshirani. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software **33**:1 - 22.
- Frisch, D., P. K. Morton, P. R. Chowdhury, B. W. Culver, J. K. Colbourne, L. J. Weider, and P. D. Jeyasingh. 2014. A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. Ecol Lett **17**:360-368.
- Frost, P. C., J. P. Benstead, W. F. Cross, H. Hillebrand, J. H. Larson, M. A. Xenopoulos, and T. Yoshida. 2006. Threshold elemental ratios of carbon and phosphorus in aquatic consumers. Ecology Letters **9**:774-779.
- Frost, P. C., and J. J. Elser. 2002. Growth responses of littoral mayflies to the phosphorus content of their food. Ecology Letters **5**:232-240.
- Frost, P. C., M. A. Evans-White, Z. V. Finkel, T. C. Jensen, and V. Matzek. 2005. Are you what you eat? Physiological constraints on organismal stoichiometry in an elementally imbalanced world. Oikos **109**:18-28.
- Frost, P. C., K. Song, and N. D. Wagner. 2014. A Beginner's Guide to Nutritional Profiling in Physiology and Ecology. Integrative and Comparative Biology **54**:873-879.
- Frost, P. C., R. S. Stelzer, G. A. Lamberti, and J. J. Elser. 2002. Ecological stoichiometry of trophic interactions in the benthos: understanding the role of $C : N : P$ ratios in lentic and lotic habitats. Journal of the North American Benthological Society **21**:515-528.
- Gibert, J. P., and J. D. Yeakel. 2019. Eco-Evolutionary Origins of Diverse Abundance, Biomass, and Trophic Structures in Food Webs. Frontiers in Ecology and Evolution **7**.
- Giraudo, M., M. Douville, G. Cottin, and M. Houde. 2017. Transcriptomic, cellular and life-history responses of *Daphnia magna* chronically exposed to benzotriazoles: Endocrine-disrupting potential and molting effects. PLOS One.
- Giri, A., and I. K. Sundar. 2022. Evaluation of stable reference genes for qPCR normalization in circadian studies related to lung inflammation and injury in mouse model. Scientific Reports **12**.
- Givskov, M., L. Eberl, and S. Molin. 1994. Responses to nutrient starvation in *Pseudomonas-putida* kt2442 - 2-dimensional electrophoretic analysis of starvation-induced and stress-induced proteins. Journal of Bacteriology **176**:4816- 4824.
- Greenbaum, D., C. Colangelo, K. Williams, and M. Gerstein. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biology **4**.
- Griffith, M., O. L. Griffith, J. Mwenifumbo, R. Goya, A. S. Morrissy, R. D. Morin, R. Corbett, M. J. Tang, Y. C. Hou, T. J. Pugh, G. Robertson, S. Chittaranjan, A. Ally, J. K. Asano, S. Y. Chan, H. Y. I. Li, H. McDonald, K. Teague, Y. J. Zhao, T. Zeng, A. Delaney, M. Hirst, G. B. Morin, S. J. M. Jones, I. T. Tai, and M. A. Marra. 2010. Alternative expression analysis by RNA sequencing. Nature Methods **7**:843-U108.
- Grigoriev, I. V., H. Nordberg, I. Shabalov, A. Aerts, M. Cantor, and e. al. 2012. The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Research.
- Gupta, D. G., N. Varma, A. Kumar, S. Naseem, M. U. S. Sachdeva, P. Bose, J. Binota, M. Gupta, P. Sonam, P. Rana, P. Malhotra, and S. Varma. 2022. Identification and validation of suitable housekeeping genes for gene expression studies in BCR-ABL1 positive B-lineage acute lymphoblastic leukemia. Molecular Biology Reports.
- Hales, N. R., D. R. Schield, A. L. Andrew, D. C. Card, M. R. Walsh, and T. A. Castoe. 2017. Contrasting gene expression programs correspond with predator-induced phenotypic plasticity within and across generations in *Daphnia*. Molecular Ecology **26**:5003-5015.
- Hessen, D. O. 1992. Nutrient element limitation of zooplankton production. American Naturalist **140**:799-814.
- Hessen, D. O., N. E. W. Alstad, and L. Skardal. 2000. Calcium limitation in *Daphnia magna*. Journal of Plankton Research **22**:553-568.
- Hessen, D. O., and N. A. Rukke. 2000. The costs of moulting in *Daphnia*; mineral regulation of carbon budgets. Freshwater Biology **45**:169-178.
- Huggett, J., K. Dheda, S. Bustin, and A. Zumla. 2005. Real-time RT-PCR normalisation; strategies and considerations. Genes and Immunity **6**:279-284.
- Huser, B. J., M. N. Futter, R. Wang, and J. Folster. 2018. Persistent and widespread longterm phosphorus declines in Boreal lakes in Sweden. Science of the Total Environment **613**:240-249.
- Jackson, M. C., C. J. G. Loewen, R. D. Vinebrooke, and C. T. Chimimba. 2016. Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. Global Change Biology **22**:180-189.
- Jeong, T. Y., and M. J. Simpson. 2019. *Daphnia magna* metabolic profiling as a promising water quality parameter for the biological early warning system. Water Research **166**.
- Jeyasingh, P. D., A. Ragavendran, S. Paland, J. A. Lopez, R. W. Sterner, and J. K. Colbourne. 2011. How do consumers deal with stoichiometric constraints? Lessons from functional genomics using *Daphnia pulex*. Mol Ecol **20**:2341-2352.
- Jeziorski, A., and J. P. Smol. 2017. The ecological impacts of lakewater calcium decline on softwater boreal ecosystems. Environmental Reviews **25**:245-253.
- Jeziorski, A., N. D. Yan, A. M. Paterson, A. M. DeSellas, M. A. Turner, D. S. Jeffries, B. Keller, R. C. Weeber, D. K. McNicol, M. E. Palmer, K. McIver, K. Arseneau, B. K. Ginn, B. F. Cumming, and J. P. Smol. 2008. The Widespread Threat of Calcium Decline in Fresh Waters. Science **322**:1374-1377.
- Jones, C. L. C., A. B. A. Shafer, W. D. Kim, C. Prater, N. D. Wagner, and P. C. Frost. 2020. The complexity of co-limitation: nutrigenomics reveal non-additive interactions of calcium and phosphorus on gene expression in *Daphnia pulex*. Proceedings of the Royal Society B-Biological Sciences **287**.
- Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins. 2005. Stochasticity in gene expression: From theories to phenotypes. Nature Reviews Genetics **6**:451-464.
- Kenkel, C. D., C. Sheridan, M. C. Leal, R. Bhagooli, K. D. Castillo, N. Kurata, E. McGinty, T. L. Goulet, and M. V. Matz. 2014. Diagnostic gene expression biomarkers of coral thermal stress. Molecular Ecology Resources **14**:667-678.
- Kessler, D. 2006. Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes. Fems Microbiology Reviews **30**:825-840.
- Khattak, H. K., C. Prater, N. D. Wagner, and P. C. Frost. 2018. The threshold elemental ratio of carbon and phosphorus of *Daphnia magna* and its connection to animal growth. Scientific Reports **8: 9673**.
- Kilham, S. S., D. A. Kreeger, S. G. Lynn, C. E. Goulden, and L. Herrera. 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. Hydrobiologia **377**:147-159.
- Kim, D., B. Landmead, and S. L. Salzberg. 2015. HISAT: a fast spliced aligner with low memory requirements. Nature Methods **12**:357-U121.
- Klumpen, E., N. Hoffschroeer, A. Schwalb, U. Gigengack, M. Koch, R. J. Paul, and B. Zeis. 2021. Metabolic adjustments during starvation in *Daphnia pulex*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology **255**.
- Kobak, D., and P. Berens. 2019. The art of using t-SNE for single-cell transcriptomics. Nature Communications **10**.
- Kogan, M. 1986. Bioassays for Measuring Quality of Insect Food. Pages 155-189 *in* J. R. Miller, Miller, T.A. (eds), editor. Insect-Plant Interactions. Springer series in experimental entomology. Springer, New York, NY.
- Kohsler, M., D. Leitsch, N. Muller, and J. Walochnik. 2020. Validation of reference genes for the normalization of RT-qPCR gene expression in Acanthamoeba spp. Scientific Reports **10**.
- Kolde, R. 2019. pheatmap: Pretty Heatmaps. R package version 1.0.12.
- Koussoroplis, A. M., A. Schwarzenberger, and A. Wacker. 2017. Diet quality determines lipase gene expression and lipase/esterase activity in *Daphnia pulex*. Biol Open **6**:210-216.
- Kuhn, M. 2020. caret: Classification and Regression Training., R package version 6.0-85.
- LaRoche, J., P. W. Boyd, R. M. L. McKay, and R. J. Geider. 1996. Flavodoxin as an in situ marker for iron stress in phytoplankton. Nature **382**:802-805.
- Le Cao, K., F. Rohart, I. Gonzalez, S. Dejean, B. Gautier, F. Bartolo, and e. al. 2016. mixOmics: Omics Data Integration Project. R package version 6.1.1. [https://CRAN.R-project.org/package=mixOmics.](https://cran.r-project.org/package=mixOmics)
- Leal, M. C., O. Seehausen, and B. Matthews. 2017. The Ecology and Evolution of Stoichiometric Phenotypes. Trends in Ecology & Evolution **32**:108-117.
- Lemos, M. F. 2021. Biomarker Studies in Stress Biology: From the Gene to Population, from the Organism to the Application. Biology (Basel) **10(12)**.
- Lemos, M. F. L., A. M. V. M. Soares, A. C. Correia, and A. C. Esteves. 2010. Proteins in ecotoxicology–how, why and why not? Proteomics **10**:873-887.
- Lenth, R. V. 2016. Least-Squares Means: The R Package lsmeans. Journal of Statistical Software **69**:1-33.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and P. Genome Project Data. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078-2079.
- Liao, Y., G. K. Smyth, and W. Shi. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**:923-930.
- Lind, M. I., K. Yarlett, J. Reger, M. J. Carter, and A. P. Beckerman. 2015. The alignment between phenotypic plasticity, the major axis of genetic variation and the response to selection. Proceedings of the Royal Society B-Biological Sciences **282**:9.

Lindeman, R. L. 1942. The trophic-dynamic aspect of ecology. Ecology **23**:399-418.

- Liu, W., Y. S. Yang, D. Francis, H. J. Rogers, P. Li, and Q. Zhang. 2008. Cadmium stress alters gene expression of DNA mismatch repair related genes in Arabidopsis seedlings. Chemosphere **73**:1138-1144.
- Loftus, S. E., and Z. I. Johnson. 2019. Reused cultivation water accumulates dissolved organic carbon and uniquely influences different marine microalgae. Frontiers in Bioengineering and Biotechnology **7**.
- Lukas, M., and A. Wacker. 2014. *Daphnia's* dilemma: adjustment of carbon budgets in the face of food and cholesterol limitation. Journal of Experimental Biology **217**:1079-1086.
- MA. 2005. Millennium Ecosystem Assessment (Ch.12. Nutrient cycling). *Ecosystems and Human Well-being: Synthesis* . Island Press, Washington DC.
- Magurran, A. E., S. R. Baillie, S. T. Buckland, J. M. Dick, D. A. Elston, E. M. Scott, R. I. Smith, P. J. Somerfield, and A. D. Watt. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends in Ecology & Evolution **25**:574-582.
- Maltby, L. 1999. Studying stress: The importance of organism-level responses. Ecological Applications **9**:431-440.
- Mao, M.-G., X. Li, A. Perálvarez-Marín, J.-L. Jiang, Z.-Q. Jiang, S.-H. Wen, and H.-Q. Lü. 2015. Transcriptomic analysis and biomarkers (Rag1 and Igμ) for probing the immune system development in Pacific cod, *Gadus macrocephalus*. Fish & Shellfish Immunology **44**:622-632.
- Martin-Creuzburg, D., E. Sperfeld, and A. Wacker. 2009. Colimitation of a freshwater herbivore by sterols and polyunsaturated fatty acids. Proceedings of the Royal Society B-Biological Sciences **276**:1805-1814.
- Martin-Creuzburg, D., A. Wacker, and T. Basen. 2010. Interactions between limiting nutrients: Consequences for somatic and population growth of *Daphnia magna*. Limnology and Oceanography **55**:2597-2607.
- Martin-Creuzburg, D., A. Wacker, and E. von Elert. 2005. Life history consequences of sterol availability in the aquatic keystone species *Daphnia*. Oecologia **144**:362- 372.
- Matz, M., V. 2019. KOGMWU: Functional Summary and Meta-Analysis of Gene Expression Data. .
- McCarthy, V., I. Donohue, and K. Irvine. 2006. Field evidence for stoichiometric relationships between zooplankton and N and P availability in a shallow calcareous lake. Freshwater Biology **51**:1589-1604.
- Miner, B. E., L. De Meester, M. E. Pfrender, W. Lampert, and N. G. Hairston. 2012. Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proceedings of the Royal Society B-Biological Sciences **279**:1873-1882.
- Miyakawa, H., M. Sato, J. K. Colbourne, and T. Iguchi. 2015. Ionotropic Glutamate Receptors Mediate Inducible Defense in the Water Flea *Daphnia pulex*. Plos One **10**.
- Morgan, M. B., D. L. Vogelien, and T. W. Snell. 2001. Assessing coral stress responses using molecular biomarkers of gene transcription. Environmental Toxicology and Chemistry **20**:537-543.
- Muller-Navarra, D. C. 1995. Biochemical versus mineral limitation in *Daphnia*. Limnology and Oceanography **40**:1209-1214.
- Muyssen, B. T. A., K. A. C. de Schamphelaere, and C. R. Janssen. 2009. Calcium accumulation and regulation in *Daphnia magna*: Links with feeding, growth and reproduction. Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology **152**:53-57.
- Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, and e. al. 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates., Nucleic acids research.
- O'Sullivan, A., M. J. Gibney, and L. Brennan. 2011. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. American Journal of Clinical Nutrition **93**:314-321.
- Orsini, L., J. B. Brown, O. S. Solari, D. Li, S. He, R. Podicheti, and e. al. 2018. Early transcriptional response pathways in *Daphnia magna* are coordinated in networks of crustacean-specific genes. Molecular Ecology **27**:886-897.
- Overturf, K., and R. W. Hardy. 2001. Myosin expression levels in trout muscle: a new method for monitoring specific growth rates for rainbow trout *Oncorhynchus mykiss* (Walbaum)on varied planes of nutrition. Aquaculture Research **32**:315- 322.
- Panina, Y., A. Germond, S. Masui, and T. M. Watanabe. 2018. Validation of Common Housekeeping Genes as Reference for qPCR Gene Expression Analysis During iPS Reprogramming Process. Scientific Reports **8**.
- Paris, Q. 1992. The return of Von Liebig's Law of the Minimum. Agronomy Journal **84**:1040-1046.
- Peng, S. T., L. Dai, and T. T. Guo. 2016. The application of sparse partial least squares regression in electricity consumption of Yunnan province. Proceedings of the 2016 International Conference on Computer Science and Electronic Technology **48**:299-303.
- Petegrosso, R., Z. L. Li, and R. Kuang. 2020. Machine learning and statistical methods for clustering single-cell RNA-sequencing data. Briefings in Bioinformatics **21**:1209-1223.
- Pfaffl, M. W., A. Tichopad, C. Prgomet, and T. P. Neuvians. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. Biotechnology Letters **26**:509-515.
- Porcella, D. B., C. E. Rixford, and J. V. Slater. 1969. Moulting and calcification in *Daphnia magna*. Physiological Zoology **42**:148-&.
- Pouyan, M. B., and D. Kostka. 2018. Random forest based similarity learning for single cell RNA sequencing data. Bioinformatics **34**:79-88.
- Poynton, H. C., J. M. Lazorchak, C. A. Impellitteri, M. E. Smith, K. Rogers, M. Patra, K. A. Hammer, H. J. Allen, and C. D. Vulpe. 2011. Differential gene expression in *Daphnia magna* suggests distinct modes of action and bioavailability for ZnO nanoparticles and Zn ions. Environmental Science & Technology **45**:762-768.
- Poynton, H. C., J. R. Varshavsky, B. Chang, G. Cavigiolio, S. Chan, P. S. Holman, A. V. Loguinov, D. J. Bauer, K. Komachi, E. C. Theil, E. J. Perkins, O. Hughes, and C. D. Vulpe. 2007. *Daphnia magna* ecotoxicogenomics provides mechanistic insights into metal toxicity. Environmental Science & Technology **41**:1044-1050.
- Prater, C., D. E. Scott, S. L. Lance, S. O. Nunziata, R. Sherman, N. Tomczyk, K. A. Capps, and P. D. Jeyasingh. 2019. Understanding variation in salamander ionomes: A nutrient balance approach. Freshwater Biology **64**:294-305.
- Prater, C., N. D. Wagner, and P. C. Frost. 2016. Effects of calcium and phosphorus limitation on the nutritional ecophysiology of *Daphnia*. Limnology and Oceanography **61**:268-278.
- Prater, C., N. D. Wagner, and P. C. Frost. 2017. Interactive effects of genotype and food quality on consumer growth rate and elemental content. Ecology **98**:1399-1408.
- Puig, O., and R. Tjian. 2006. Nutrient availability and growth: regulation of insulin signaling by dFOXO/FOXO1. Cell Cycle **5(5):503-5.**
- Pérez-Fuentetaja, A., and F. Goodberry. 2016. *Daphnia's* challenge: survival and reproduction when calcium and food are limiting. Journal of Plankton Research **38**:1379-1388.
- R Core Team. 2019. R: A language and environment for statistical computing.*in* V. R Foundation for Statistical Computing, Austria. URL [https://www.R-project.org/.](https://www.r-project.org/) editor. R Foundation for Statistical Computing, Vienna, Austria. URL [https://www.R-project.org/.](https://www.r-project.org/)
- Raubenheimer, D., S. J. Simpson, and D. Mayntz. 2009. Nutrition, ecology and nutritional ecology: toward an integrated framework. Functional Ecology **23**:4-16.
- Ravindran, S. P., J. Luneburg, L. Gottschlich, V. Tams, and M. Cordellier. 2019. *Daphnia* stressor database: Taking advantage of a decade of *Daphnia* '-omics' data for gene annotation. Scientific Reports **9**.
- Rice, M. J., C. L. C. Jones, C. W. E. Starke, and P. C. Frost. 2021. Calcium stress in *Daphnia pulicaria* and exposure to predator-derived cues: Making a bad situation worse? Freshwater Science **40**:449-462.
- Riessen, H. P., R. D. Linley, I. Altshuler, M. Rabus, T. Soellradl, H. Clausen-Schaumann, C. Laforsch, and N. D. Yan. 2012. Changes in water chemistry can disable plankton prey defenses. Proceedings of the National Academy of Sciences of the United States of America **109**:15377-15382.
- Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier. 1979. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. Journal of General Microbiology **111**:1-61.
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139-140.
- Roh, J. Y., S. Y. Park, and J. Choi. 2006. Cadmium toxicity monitoring using stress related gene expressions in *Caenorhabditis elegans*. Molecular & Cellular Toxicology **2**:54-59.
- Rohart, F., B. Gautier, A. Singh, and K. A. Le Cao. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. Plos Computational Biology **13**.
- Rozenberg, A., M. Parida, F. Leese, L. C. Weiss, R. Tollrian, and J. R. Manak. 2015. Transcriptional profiling of predator-induced phenotypic plasticity in *Daphnia pulex*. Frontiers in Zoology **12**.
- Rudman, S. M., M. A. Rodriguez-Cabal, A. Stier, T. Sato, J. Heayyside, R. W. Ei-Sabaawi, and G. M. Crutsinger. 2015. Adaptive genetic variation mediates bottom-up and top-down control in an aquatic ecosystem. Proceedings of the Royal Society B-Biological Sciences **282**:125-132.
- Sabbagh, Y., H. Giral, Y. Caldas, M. Levi, and S. C. Schiavi. 2011. Intestinal Phosphate Transport. Advances in Chronic Kidney Disease **18**:85-90.
- Scavia, D., and W. S. Gardner. 1982. Kinetics of nitrogen and phosphorus release in varying food supplies by *Daphnia magna*. Hydrobiologia **96**:105-111.
- Schalicke, S., L. Y. Sobisch, D. Martin-Creuzburg, and A. Wacker. 2019a. Food quantity-quality co-limitation: Interactive effects of dietary carbon and essential lipid supply on population growth of a freshwater rotifer. Freshwater Biology **64**:903-912.
- Schalicke, S., J. Teubner, D. Martin-Creuzburg, and A. Wacker. 2019b. Fitness response variation within and among consumer species can be co-mediated by food quantity and biochemical quality. Scientific Reports **9**.
- Schroth, W., A. Ender, and B. Schierwater. 2005. Molecular biomarkers and adaptation to environmental stress in moon jelly (*Aurelia* spp*.*). Marine Biotechnology **7**:449-461.
- Schwarzenberger, A., C. J. Kuster, and E. Elert. 2012. Molecular mechanisms of tolerance to cyanobacterial protease inhibitors revealed by clonal differences in *Daphnia magna*. Molecular Ecology **21**:4898-4911.
- Schwarzenberger, A., A. Zitt, P. Kroth, S. Mueller, and E. Von Elert. 2010. Gene expression and activity of digestive proteases in *Daphnia*: effects of cyanobacterial protease inhibitors. BMC Physiology **10**:6.
- Scriber, J. M., and F. Slansky. 1981. The nutritional ecology of immature insects. Annual Review of Entomology **26**:183-211.
- Sengupta, N., D. C. Reardon, P. D. Gerard, and W. S. Baldwin. 2017. Exchange of polar lipids from adults to neonates in *Daphnia magna*: Perturbations in sphingomyelin allocation by dietary lipids and environmental toxicants. Plos One **12**.
- Shi, Y., and M. X. He. 2014. Differential gene expression identified by RNA-Seq and qPCR in two sizes of pearl oyster (*Pinctada fucata*). Gene **538**:313-322.
- Shindo, S., S. Numazawa, and T. Yoshida. 2007. A physiological role of AMP-activated protein kinase in phenobarbital-mediated constitutive androstane receptor activation and CYP2B induction. Biochemical Journal **401**:735-741.
- Silver, N., S. Best, J. Jiang, and S. L. Thein. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. Bmc Molecular Biology **7**.
- Skjelkvåle, B. L., Stoddard, J.L. & Andersen, T. 2001. Trends in Surface Water Acidification in Europe and North America (1989–1998) | SpringerLink. Water, Air & Soil Pollution.
- Snell, T. W., S. E. Brogdon, and M. B. Morgan. 2003. Gene expression profiling in ecotoxicology. Ecotoxicology **12**:475-483.
- Spence, J. T. 2010. Nutrigenomics and Agriculture: A Perspective. Journal of Nutrigenetics and Nutrigenomics **3**:290-295.
- Sperfeld, E., D. Martin-Creuzburg, and A. Wacker. 2012. Multiple resource limitation theory applied to herbivorous consumers: Liebig's minimum rule vs. interactive co-limitation. Ecology Letters **15**:142-150.
- Sterner, R. W. 1993. *Daphnia* growth on varying quality of *Scenedesmus* Mineral limitation of zooplankton. Ecology **74**:2351-2360.
- Sterner, R. W. 1997. Modelling interactions of food quality and quantity in homeostatic consumers. Freshwater Biology **38**:473-481.
- Sterner, R. W., and J. J. Elser. 2002. Ecological Stoichiometry:The Biology of Elements from Molecules to the Biosphere. Princeton University Press.
- Sterner, R. W., D. D. Hagemeier, and W. L. Smith. 1993. Phytoplankton nutrient limitation and food quality for *Daphnia*. Limnology and Oceanography **38**:857- 871.
- Sterner, R. W., and D. O. Hessen. 1994. Algal nutrient limitation and the nutrition of aquatic herbivores. Annual Review of Ecology and Systematics **25**:1-29.
- Sterner, R. W., and K. L. Schulz. 1998. Zooplankton nutrition: recent progress and a reality check. Aquatic Ecology **32**:261-279.
- Storey, J., A. Bass, A. Dabney, and D. Robinson. 2019. qvalue: Q-value estimation for false discovery rate control. R package version 2.18.0, <http://github.com/jdstorey/> qvalue.
- Sugiura, S. H., N. K. McDaniel, and R. P. Ferraris. 2003. In vivo fractional P-i absorption and NaPi-II mRNA expression in rainbow trout are upregulated by dietary P restriction. American Journal of Physiology-Regulatory Integrative and Comparative Physiology **285**:R770-R781.
- Tams, V., J. Nickel, A. Ehring, and M. Cordellier. 2018. Reproduction associated gene expression under predation risk in *Daphnia* - a comparative transcriptomic approach.
- Tang, Y., M. Horikoshi, and W. X. Li. 2016. ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages. R Journal **8**:474-485.
- Tedeschi, J. N., W. J. Kennington, O. Berry, S. Whiting, M. Meekan, and N. J. Mitchell. 2015. Increased expression of Hsp70 and Hsp90 mRNA as biomarkers of thermal stress in loggerhead turtle embryos (*Caretta caretta*). Journal of Thermal Biology **47**:42-50.
- Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, and E. Heinen. 1999. Housekeeping genes as internal standards: use and limits. Journal of Biotechnology **75**:291-295.
- Tian, Q., S. B. Stepaniants, M. Mao, L. Weng, M. C. Feetham, M. J. Doyle, E. C. Yi, H. Y. Dai, V. Thorsson, J. Eng, D. Goodlett, J. P. Berger, B. Gunter, P. S. Linseley, R. B. Stoughton, R. Aebersold, S. J. Collins, W. A. Hanlon, and L. E. Hood. 2004. Integrated genomic and proteomic analyses of gene expression in mammalian cells. Molecular & Cellular Proteomics **3**:960-969.
- Tinnevelt, G. H., U. F. H. Engelke, R. A. Wevers, S. Veenhuis, M. A. Willemsen, K. L. M. Coene, and e. al. 2020. Variable Selection in Untargeted Metabolomics and the Danger of Sparsity. Metabolites **10**.
- Tom, M., C. Manfrin, P. G. Giulianini, and A. Pallavicini. 2013. Crustacean oxireductases protein sequences derived from a functional genomic project potentially involved in ecdysteroid hormones metabolism - A starting point for function examination. General and Comparative Endocrinology **194**:71-80.
- Tupac-Yupanqui, I., L. Vasquez, M. Villarroel, and S. Dunner. 2013. Fatty acid food source affects expression of genes involved in the stress response in tilapia (*Oreochromis niloticus*). Journal of Biological Research-Thessaloniki **19**:30-37.
- Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. M. Leunissen. 2007. Primer3Plus, an enhanced web interface to Primer3 Nucleic Acids Research.
- Urabe, J., J. Clasen, and R. W. Sterner. 1997. Phosphorus limitation of *Daphnia* growth: Is it real? Limnology and Oceanography **42**:1436-1443.
- Urabe, J., and R. W. Sterner. 2001. Contrasting effects of different types of resource depletion on life‐history traits in *Daphnia*. Functional Ecology **15**:165-174.
- van der Ploeg, R. R., W. Bohm, and M. B. Kirkham. 1999. On the origin of the theory of mineral nutrition of plants and the law of the minimum. Soil Science Society of America Journal **63**:1055-1062.
- Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology **3**.
- Veerasamy, R., H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese, and R. K. and Agrawal. 2011. Validation of QSAR Models - Strategies and Importance. International Journal of Drug Design and Discovery **2**.
- Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S Fourth Edition edition. Springer, New York.
- Verhamme, E. M., T. M. Redder, D. A. Schlea, J. Grush, J. F. Bratton, and J. V. DePinto. 2016. Development of the Western Lake Erie Ecosystem Model (WLEEM): Application to connect phosphorus loads to cyanobacteria biomass. Journal of Great Lakes Research **42**:1193-1205.
- Vinebrooke, R. D., K. L. Cottingham, J. Norberg, M. Scheffer, S. I. Dodson, S. C. Maberly, and U. Sommer. 2004. Impacts of multiple stressors on biodiversity and ecosystem functioning: the role of species co-tolerance. Oikos **104**:451-457.
- Von Liebig, J. 1840. Die organische chemie in ihre anwendung auf agricultur und physiologie F. Vieweg und Sohn, Braunschweig.
- Wacker, A., and D. Martin-Creuzburg. 2007. Allocation of essential lipids in *Daphnia magna* during exposure to poor food quality. Functional Ecology **21**:738-747.
- Wagner, N. D., and P. C. Frost. 2012. Responses of alkaline phosphatase activity in *Daphnia* to poor nutrition. Oecologia **170**:1-10.
- Wagner, N. D., H. Hillebrand, A. Wacker, and P. C. Frost. 2013. Nutritional indicators and their uses in ecology. Ecology Letters **16**:535-544.
- Wagner, N. D., B. P. Lankadurai, M. J. Simpson, A. J. Simpson, and P. C. Frost. 2014. Metabolomic Differentiation of Nutritional Stress in an Aquatic Invertebrate. Physiological and Biochemical Zoology **88**:43-52.
- Wagner, N. D., Z. Yang, A. B. Scott, and P. C. Frost. 2017. Effects of algal food quality on free amino acid metabolism of *Daphnia*. Aquatic Sciences **79**:127-137.
- Wang, B., J. J. Zhu, E. Pierson, D. Ramazzotti, and S. Batzoglou. 2017. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nature Methods **14**:414-+.
- Wang, E. T., R. Sandberg, S. J. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, and C. B. Burge. 2008. Alternative isoform regulation in human tissue transcriptomes. Nature **456**:470-476.
- Watmough, S. A., and P. J. Dillon. 2002. The impact of acid deposition and forest harvesting on lakes and their forested catchments in south central Ontario: a critical loads approach. Hydrology and Earth System Sciences **6**:833-848.
- Weihs, C., U. Ligges, K. Luebke, and N. Raabe. 2005. klaR Analyzing German Business Cycles. Pages 335-343 Data Analysis and Decision Support. Springer-Verlag, Berlin.
- Westerhuis, J. A., H. C. J. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. J. van Velzen, and e. al. 2008. Assessment of PLSDA cross validation. Metabolomics **4**:81-89.
- Weyhenmeyer, G. A., J. Hartmann, D. O. Hessen, J. Kopacek, J. Hejzlar, S. Jacquet, S. K. Hamilton, P. Verburg, T. H. Leach, M. Schmid, G. Flaim, T. Noges, P. Noges, V. C. Wentzky, M. Rogora, J. A. Rusak, S. Kosten, A. M. Paterson, K. Teubner, S. N. Higgins, G. Lawrence, K. Kangur, I. Kokorite, L. Cerasino, C. Funk, R. Harvey, F. Moatar, H. A. de Wit, and T. Zechmeister. 2019. Widespread diminishing anthropogenic effects on calcium in freshwaters. Scientific Reports **9: 10450**:10.
- Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
- Williams, E. G., and J. Auwerx. 2015. The Convergence of Systems and Reductionist Approaches in Complex Trait Analysis. Cell **162**:23-32.
- Wu, A. R., N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke, and S. R. Quake. 2014. Quantitative assessment of single-cell RNA-sequencing methods. Nature Methods **11**:41-46.
- Xie, F. L., P. Xiao, D. L. Chen, L. Xu, and B. H. Zhang. 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Molecular Biology **80**:75-84.
- Xu, Q., Y. C. Wang, R. Liu, L. F. Brito, L. Kang, Y. Yu, D. S. Wang, H. J. Wu, and A. Liu. 2017. Differential gene expression in the peripheral blood of Chinese Sanhe cattle exposed to severe cold stress. Genetics and Molecular Research **16**.
- Xu, Z. M., Y. D. Li, M. Li, and H. B. Liu. 2021. Transcriptomic response of *Daphnia magna* to nitrogen- or phosphorus-limited diet. Ecology and Evolution **11**:11009- 11019.
- Yamamichi, M., C. L. Meunier, A. Peace, C. Prater, and M. A. Rua. 2015. Rapid evolution of a consumer stoichiometric trait destabilizes consumer-producer dynamics. Oikos **124**:960-969.
- Ye, Z., S. Xu, K. Spitze, J. Asselman, X. Jiang, M. S. Ackerman, J. Lopez, B. Harker, R. T. Raborn, W. K. Thomas, J. Ramsdell, M. E. Pfrender, and M. Lynch. 2017. A New Reference Genome Assembly for the Microcrustacean *Daphnia pulex*. G3 Genes|Genomes|Genetics **7**:2160-1836 (Electronic).
- Zhao, S. R., W. P. Fung-Leung, A. Bittner, K. Ngo, and X. J. Liu. 2014. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLOS One **9**.
- Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B-Statistical Methodology **67**:301- 320.

Appendix

List of tables

Table S3.1. Average number of reads per treatment from Illumina Hi-Seq HTS for *Daphnia pulex* exposed to one of six dietary treatments (high quantities of high-quality food, low calcium, high cyanobacteria, low carbon/food, low nitrogen, and low phosphorus) \pm 1S.D.

Table S4.1. Raw threshold cycle (Ct) values from qPCR on all samples, including target genes and housekeeping (HK) genes. The sample codes used are as follows: $Ca = low$ calcium, $LN = low$ nitrogen, $LP = low$ phosphorus, $FA = cyanobacteria$, $HCN/HFA/HPF$ $=$ fast growth, LF $=$ low food. Standard deviation for all triplicate analytical replicates run, if S.D. >0.2 then the most extreme outlier was removed prior to analysis.

Table S4.2. . Stability values of housekeeping genes from qPCR on *Daphnia pulex* under six different dietary treatments. Comprehensive ranking calculated with refFinder algorithm (Xie et al. 2012). See chapter 3 methods for full explanation of ranking system. Plate reference refers to each of the individual 96-well plates that were run for the qPCR where the letter represents a code for the target gene and the number represents the set of samples being run $(1 = low$ calcium, low nitrogen, and low phosphorus, as well as normalizer sample and negative controls and $2 =$ low carbon, cyanobacteria, and fast growth, as well as normalizer sample and negative controls).

Table S3.1. Average number of reads per treatment from Illumina Hi-Seq HTS for *Daphnia pulex* exposed to one of six dietary treatments (high quantities of high-quality food (fast growth), low calcium, high cyanobacteria, low carbon/food, low nitrogen, and low phosphorus) \pm 1S.D.

Treatment	# of samples	Average # of reads per sample \pm 1 s.d.
Fast growth	5	$9,059,056 \pm 1,419,302$
Low calcium	5	$13,000,258 \pm 5,988,704$
High cyanobacteria	5	$12,120,211 \pm 3,790,871$
Low carbon	5	$10,525,737 \pm 1,958,366$
Low nitrogen	5	$11,455,234 \pm 3,096,146$
Low phosphorus	5	$10,873,581 \pm 4,340,115$

Table S4.1. Raw threshold cycle (Ct) values from qPCR on all samples, including target genes and housekeeping (HK) genes. The sample codes used are as follows: $Ca = low$ calcium, $LN = low$ nitrogen, $LP = low$ phosphorus, $FA = cyanobacteria$, $HCN/HFA/HPF$ $=$ fast growth, LF $=$ low food. Standard deviation for all triplicate analytical replicates run, if S.D. >0.2 then the most extreme outlier was removed prior to analysis.

170

Table S4.2. Stability values of housekeeping genes from qPCR on *Daphnia pulex* under six different dietary treatments. Comprehensive ranking calculated with refFinder algorithm (Xie et al. 2012). See chapter 3 methods for full explanation of ranking system. Plate reference refers to each of the individual 96-well plates that were run for the qPCR where the letter represents a code for the target gene and the number represents the set of samples being run $(1 = low$ calcium, low nitrogen, and low phosphorus, as well as normalizer sample and negative controls and $2 =$ low carbon, cyanobacteria, and fast growth, as well as normalizer sample and negative controls).

