5-21-2020 10:00 AM

Amount and Sources of Population Variability in the Metabolome of a Crayfish Species

Cora Bilhorn, *The University of Western Ontario*

Supervisor: Yates, Adam G., *The University of Western Ontario*
Co-Supervisor: Brua, Robert B., *Environment and Climate Change Canada*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Geography

Abstract

My study investigated the amount of variation associated with region and stream scales in the metabolomes of northern crayfish (*Faxonius virilis*), collected from seven streams distributed across three ecoregions in Western Canada. Additionally, my study measured metabolomes of crayfish from the same seven streams after experiencing a common laboratory environment to separate the effects of environmental and genetic variation. Region and stream scales were found to be poor predictors of metabolomic variation among crayfish sampled in the field but exhibited increased predictive ability among crayfish exposed to the common environment, indicating crayfish from separate populations responded differently to the common environment. Furthermore, variation among the crayfish metabolomes did not decrease in the common environment, indicating the important influence of genetic variation. These findings show unstressed populations of the northern crayfish display similar metabolomes despite experiencing differing environmental conditions. Reference conditions derived for metabolomic-based bioassessment may thus be applicable across regions.

Keywords

Streams, Crayfish, *Faxonius virilis*, Western Canada, Ecoregions, Metabolomics, Environmental Variation, Genetic Variation

Summary for Lay Audience

Bioassessment uses biological responses to evaluate the health of an ecosystem. Stream bioassessment will benefit from the ability of metabolomics (the study of all small molecules in an organism) to detect sub-lethal organism stress through changes in the metabolite profile (i.e., the metabolome). However, for the metabolome to be integrated into bioassessment programs, the amount of natural variation among and within populations must be established, creating a baseline to which potentially stressed populations can be compared. For instance, the metabolome of an organism may be affected by its surrounding environment, such as the climate, topography, and geology characteristics described by different ecoregions. However, the metabolomes of a species taken from several different ecoregions may also vary because the different populations are genetically different. My study determined the amount of variation in the metabolome of the northern crayfish (*Faxonius virilis*) that could be attributed to the ecoregion or stream of origin and whether environment or genetics was the more important source of variation. I collected northern crayfish from seven streams distributed across three ecoregions in Western Canada and compared their metabolomes. As well, I kept crayfish from the same seven streams under similar environmental conditions in the laboratory for 16 days. I found that the ecoregion and stream that the crayfish originated from were poor predictors of the crayfish's metabolome among crayfish sampled directly from the streams. In contrast, ecoregion and stream of origin were better predictors of the metabolome among crayfish that had experienced the common environment, indicating crayfish of separate origins responded differently to the common environment. Furthermore, variation among the crayfish metabolomes did not decrease in the common environment, indicating that genetic variation was an important influence on the crayfish metabolome. These findings show that unstressed populations of the northern crayfish display similar metabolomes despite experiencing differing environmental conditions. The consistency in the northern crayfish metabolome across several ecoregions suggests that a single crayfish metabolome baseline could be used in bioassessment programs across the sampled ecoregions.

Co-Authorship Statement

This thesis contains content that will become one manuscript. Cora Bilhorn will be the first author because of her primary role in conducting the research. Dr. Robert B. Brua will be the second author because he co-supervised the research and provided expertise, equipment, and funding. Dr. Adam G. Yates will be the last author because he acted as the principal advisor and provided significant input into the design and analysis of the study.

Acknowledgments

Foremost, I would like to thank Dr. Adam Yates for all the patient guidance he provided me throughout this project. He has taught me an immense amount about scientific thinking, doing, and writing that I will draw on for the rest of my life. Besides being a great scientist, he is a thoughtful and caring individual, and I am so glad I have been able to work with him!

Likewise, I would like to thank Dr. Robert Brua for all the problem-solving energy and technical advice he provided throughout this project, as well as his immense patience as I learned about ^{1}NMR Spectroscopy. Also, his kind encouragement when I thought this project may not work out will never be forgotten.

I would also like to thank Jonathan Keating and Daryl Halliwell for their assistance with the field and lab work. Without their expertise, energy, and champion kick-netting, the data for this project would have never been obtained.

Many others helped with this project and deserve much thanks. Dr. Robert Bailey provided invaluable advice on the study design and statistics of this project. My lab mates Nolan Pearce and Lauren Banks provided abundant help in troubleshooting R and ArcGIS. Sarah Pomfret was tremendously helpful in explaining the steps of a metabolomics project while I was planning my project. Christopher Lucas and Kristen Painter were just great people to share an office with.

I must also thank my undergraduate research advisor, Dr. Jessica Orlofske, for her guidance. Without the encouragement and inspiration that I received from her, I never would have pursued a graduate degree, and I am so glad I did!

Finally, I would like to thank my family for their help and encouragement. I am very thankful to my aunt and uncle for their hospitality during my stay in Canada. Most importantly, I would like to thank my parents for their unwavering support, and my mother, especially, for regular calls to check that I was hard at work!

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1 Introduction

1.1 Stream Bioassessment

Streams are an important natural resource that support many human activities (Carpenter et al., 2011), and also serve as hotspots of biodiversity (Robinson et al., 2002). Unfortunately, streams are increasingly threatened by anthropogenic pollution and modification (Vorosmarty et al., 2010). Because of these threats, resource managers are commonly tasked with evaluating the health of streams to determine if interventions are needed (Norris & Hawkins, 2000). Consequently, many approaches to stream assessment have been developed.

The most basic stream assessment approach is the direct measurement of known contaminants and other water quality parameters (Connon et al., 2012; Ekman et al., 2017). However, this method fails to capture the full effect that contaminants may be having on stream biota for several reasons: 1) some contaminants have biological effects at levels below current detection capabilities; 2) periodic sampling may fail to capture events during which contaminant levels were temporarily higher; 3) there are a large and growing number of potential contaminants that are not routinely tested for, and; 4) some contaminants may only be problematic for biota when combined with other contaminants or natural stressors (Connon et al., 2012; Ekman et al., 2013; Skelton et al., 2014). Many of these limitations of directly measuring aquatic stressors are addressed by the practice of bioassessment.

Bioassessment uses changes in the structure or function of ecosystems and their resident biota to evaluate the impacts of human activities on the ecosystem (Stoddard et al., 2006). Biota can be examined at different levels of biological organization (i.e., cellular molecules, cells, tissues, organisms, populations, or communities) and, at each level, different measurements may serve as indicators of stress. Some of the most commonly used bioassessment indicators are indices summarizing fish, benthic macroinvertebrate or algal assemblages by describing the abundance, behavior, or diversity of taxa in a community. These indices assume that more sensitive taxa will be replaced with more tolerant taxa as

the ecosystem is exposed to more pollution (Connon et al., 2012). Some examples of community indices include taxa richness, number of sensitive taxa present, and functional feeding groups. Although in common use, community indices have several shortcomings as bioassessment indicators. Indeed, community indices can detect anthropogenic stress and are considered easy to use, but are unable to detect stress before it has had lethal effects on some species (Mayer et al., 1992). Moreover, these indices cannot determine the cause of stress (Connon et al., 2012) and can become expensive and difficult to implement because of the extensive field sampling and taxon identification required to compile the needed information (Bonada et al., 2006). Thus, there is an increasing need to develop a bioassessment indicator that addresses these shortcomings (Bonada et al., 2006; Pomfret et al., 2020).

1.2 Metabolomics in Stream Bioassessment

Metabolomics is a field of science that studies the metabolites (low molecular weight molecules such as sugars, amino acids, and small lipids) present in a tissue or organism. The concentrations of all metabolites considered collectively are referred to as the metabolome. The metabolome can indicate the physiological state of an organism and can be used to measure the effects of environmental factors on an organism's health (Viant, 2008). Metabolomics has been used in toxicology studies under laboratory conditions and has proven able to distinguish control organisms from organisms exposed to sub-lethal doses of many toxicants and environmental stressors (Lankadurai et al., 2013). In addition to laboratory applications, researchers have proposed that metabolomics could be used in bioassessment by detecting stress in organisms sampled directly from ecosystems (Bundy et al., 2009; Pomfret et al., 2020). Indeed, several studies have been conducted in which organisms collected from polluted sites and reference sites were found to have distinct metabolomes (e.g., Cappello et al., 2017; Fernandez-Cisnal et al., 2018; Gago-Tinoco et al., 2014). The metabolome may thus have potential as a bioassessment tool.

The metabolome has several potential advantages over conventional bioassessment indicators. First, the metabolome has the potential to detect organism stress early because changes at the molecular level must precede changes in individual fitness and subsequent population level effects (e.g., reduced reproduction and increased death rates) (Martyniuk,

2018; Miller, 2007). An example of early stress detection using metabolomics is a study by Taylor et al. (2018), who discovered metabolic biomarkers that could predict reduced reproductive fitness in *Daphnia magna*. Second, the metabolome has the potential to identify the cause of organism stress, because changes in the many different metabolites measured can create a 'fingerprint' unique to individual stressors (Bundy et al., 2009). Indeed, Jeppe et al. (2017) measured the levels of multiple contaminants in the field and were able to associate many of the individual contaminants with distinct metabolite changes in midge larva taken from the same sites. Third, changes in the metabolite concentrations of the metabolome may indicate the molecular mode of action of individual stressors (Bearden et al., 2012), making it easier to link the molecular effect of stressor exposure to eventual population-level effects. Fourth, measuring the metabolome of a stream organism would eliminate the need for detailed community taxonomic identification, which would make the metabolome an easier to use and a less expensive bioassessment indicator than the many community indices (Pomfret et al., 2020).

Despite its possible advantages, there are hurdles to using the metabolome as a bioassessment indicator. Of primary importance is defining a 'normal' or 'reference' metabolome, along with the amount of deviation from 'reference' that would justify concern (Bahamonde et al., 2016). Indeed, bioassessment, regardless of the indicator used, relies on having a 'reference condition' that represents the state of the indicator if no or few human impacts are present and against which measurements at a potentially impacted or "test" site can be compared (Hawkins et al., 2010). The Reference Condition Approach (RCA) provides a way to estimate population variability within the reference condition by using multiple reference sites (Bailey et al., 2004). By quantifying the variability of the bioassessment indicator among the reference sites, RCA provides a measure of the inherent 'background' variability that alterations caused by a stressor would have to exceed to be detected (Bailey et al., 2004). Thus, the metabolome will be a more sensitive indicator of stressors if among population variability in the reference metabolome is low. However, it is possible that population variability will be large, making it difficult to distinguish stressed metabolomes from unstressed metabolomes (Fig. 1). Despite the importance of measuring among population variability in the reference metabolome, to date all published field studies assessing biological effects of human activities using metabolomics, have

been conducted with the use of a single reference site (Cappello et al., 2017; Fernandez-Cisnal et al., 2018; Gago-Tinoco et al., 2014; Melvin et al., 2018; Watanabe et al., 2015). Consequently, there is a lack of understanding of among reference site variability and its sources in the reference metabolome.

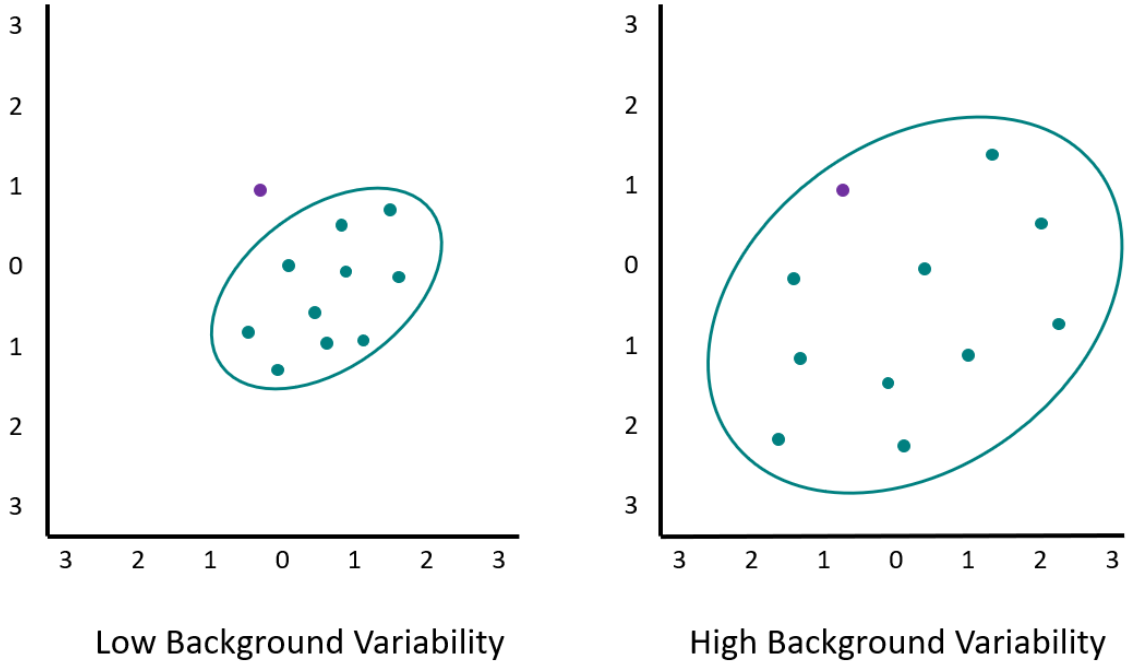


Figure 1. Representation of low and high variability in the reference condition. The ellipses, which represent the reference condition, demonstrate how low variability in the reference condition (plot on the left) allows the impacted population (purple dot) to be easily distinguished from reference populations (teal dots), while high variability in the reference condition (plot on the right) does not allow this distinction.

1.2.1 Variability in the Reference Metabolome

Variability in a reference metabolome is of two main types: within population and among population. Within population variability includes variability from such sources as sex, life stage, and season. For example, Hines et al. (2007) showed that the sexes within a species can have distinctly different metabolomes and that mixing the sexes makes it more difficult to use the metabolome to detect stressors. Moreover, Wu et al. (2017) found that life stage changed an organism's metabolic response to a stressor. Additionally, it has been shown that the metabolomes of organisms from the same site can change significantly from one season to the next (Melvin et al., 2018), although Aru et al. (2017) found that metabolomes from the same site remained similar if sampled within the same season from year to year.

Although there will be some within population variability caused by environmental patchiness and within population genetic differences, the findings of Aru et al. (2017) suggest that within population variability can largely be controlled for by stratifying sampling by variability causing factors, such as by only sampling organisms of the same sex and life stage at a specific time of year.

Among population variability has two primary sources: the environment and genetics. Environmental differences among reference sites could be a key source of among population variability in the reference metabolome. Stream characteristics, such as temperature, salinity, pH, and dissolved oxygen, likely affect the metabolome (Simmons et al., 2015), as could channel morphology and stream flow (Martyniuk, 2018), all of which are shaped by the geology, topography, and climate of stream catchments (Allan, 2004). Studies that have measured the metabolomes of organisms taken from several field sites have often noted distinct metabolomes among sites (Fernandez-Cisnal et al., 2018; Gago-Tinoco et al., 2014; Watanabe et al., 2015) although sites with more similar characteristics typically produce more similar metabolomes (Gago-Tinoco et al., 2014). For example, Gidman et al. (2007) found that mussel metabolomes could be distinguished by their estuary of origin, but estuaries with the most similar pH and nutrient concentrations produced the most similar metabolomes. It follows that the metabolomes of organisms taken from streams that differ greatly in their environmental characteristics would be dissimilar, although there is a need for studies that test this hypothesis across large environmental gradients (Pomfret et al., 2020). However, to date this hypothesis has not been tested and thus, the amount of variability among metabolomes sampled from reference sites in landscapes with different environmental characteristics is not known.

The influence of landscape on among population variability is further complicated by the possibility of genetic variation among populations of a given species. As with landscape differences, genetic differences are expected to increase with increasing geographic distance as a result of genetic drift, genetic adaptation, or a combination of both (Orsini et al., 2013). Genetic differences among wild populations has rarely been considered in studies using metabolomics for bioassessment, but the influence of genetic differences could be significant and confound the interpretation of metabolomic data (Schoenfuss &

Wang, 2015). Indeed, genetic differentiation across a species' range was linked to among population variation in the metabolome of the two-spotted spider mite (*Tetranychus urticae*) from nine populations located throughout northern Europe. It was observed that metabolomes remained distinct among populations after being exposed to similar laboratory conditions for two generations (Van Petegem et al. 2016). Species that exhibit distinct metabolomes because of among population differences in genetics could pose problems for use in bioassessment protocols utilizing the metabolome. There is thus a need for metabolomic studies that examine genetic diversity in taxa being used for bioassessment and the extent to which these differences influence spatial variability in the metabolome (Bundy et al., 2009).

1.2.2 Crayfish as a Sentinel Taxa

A sentinel organism is any species that is well suited for indicating the impact that stressors may have on an environment (Lower & Kendall, 1990). Although the metabolome can indicate the physiological state of an organism, bioassessment seeks to determine the health of the ecosystem. Thus, measuring the metabolome of sentinel species provides the link that allows resource managers to assess the condition of an ecosystem based on the health of a single species (Miller, 2007). Besides their responsiveness to stressors of community-wide concern, the characteristics of an ideal sentinel organism include: 1) wide geographic and ecological distribution, 2) low mobility, 3) abundance, 4) ease of collection and identification, and 5) hardiness to a wide range of stressor exposure while showing detectable responses at even low levels of exposure (Johnson et al, 1993; Lower & Kendall, 1990). Because of their ability to fulfill these requirements, crayfish (order Decapoda; families Astacidae, Cambaridae, and Parastacidae) are a commonly used sentinel organism in freshwater ecosystems (Johnson et al., 1993).

Crayfish have been studied extensively and much is known about their physiology and environmental requirements and preferences (Kozak & Kuklina, 2016; Willis-Jones et al., 2016). Researchers have measured the physiological responses of crayfish to stressors such as metals (Kouba et al., 2010), hypoxia (Izral et al., 2019), acidity, salinity, and nutrients (Willis-Jones et al., 2016). Moreover, many crayfish species can tolerate highly polluted conditions (Del Ramo et al., 1987; Roldan & Shivers, 1987), but have been found to show

measurable, physiological effects across a wide range of stressor exposure levels (Izral et al., 2018; Johnson et al., 1993), making them a useful indicator of the effects of human activity. As well, crayfish are often considered keystone species in aquatic communities. Crayfish influence food webs because of their large total biomass in the aquatic environment and because their omnivorous habits link them to many other community members at different trophic levels (Willis-Jones et al., 2016). Crayfish can also act as ecosystem engineers because of their ability to disrupt sediment through tail flipping and burrowing (Johnson et al., 2010; Statzner et al., 2003), and the consumption of macrophytes (Rodriguez et al., 2003).

Crayfish are globally distributed in freshwater environments, with native populations in all continents, except Africa and Antarctica (Souty-Gosset & Fetzner, 2016). The cosmopolitan distribution of crayfish enables them to be sampled at many locations and the results to be compared (Lower & Kendall, 1990). Individual crayfish typically have small home ranges and will only be influenced by the conditions close to the site where they were collected (Kouba et al., 2010). For instance, Bubb et al. (2002) found that *Pacifastacus leniusculus* had a median linear range of 23.8 m (25% quartile = 10 m, 75% quartile = 46 m) over 127 days of observation, Barbaresi et al. (2004) found that *Procambarus clarkii* had a mean linear range of 32.8 ± 23.6 m for males and 39 ± 7.6 m for females over a 10 day study, and, in a study spanning over a year, Hazlett et al. (1974) found that *Faxonius virilis* (formerly *Orconectes virilis*) had a mean linear range of 33 m for males and 32 m for females, but a mode of less than 5 m. Thus, stressor effects indicated by the crayfish metabolome can be associated with the collection site, an assumption that may not be warranted with more mobile species (e.g., fish). Crayfish also have a long life relative to many other invertebrates, spanning from a couple years to decades (Vogt, 2012). The long lifespan allows crayfish to potentially bioaccumulate pollutants, which can be informative for bioassessment. Also, unlike many other aquatic invertebrates that emerge to a terrestrial life stage, crayfish spend their entire life as an aquatic organism and can be sampled in any season to assess the aquatic environment through their metabolome. Moreover, because crayfish are often abundant (Willis-Jones et al., 2016), they can, in many circumstances, be sampled without significantly depleting individual populations. The ubiquity of crayfish also makes them easy to collect in a variety of habitats, although

some species can be difficult to differentiate from each other. Finally, crayfish also have a relatively large body size enabling dissection of individual tissues, which, coupled with their easily identified sexual organs and reproductive status, make them a particularly useful sentinel taxa for metabolomics-based assessments.

1.3 Research Questions

As part of investigating the suitability of the crayfish metabolome for use as a bioassessment tool, the goal of this study was to determine the amount and sources of among population variability inherent in the metabolomes of crayfish collected from reference condition streams in distinct regions. This goal was met by asking two questions:

1) What is the relative amount of variability in the metabolomes of crayfish associated with the region and stream scales?

In response to this question, this study made its first prediction.

1) A majority of the total variability among the crayfish metabolomes will be attributable to regional and stream scales, with the regional scale explaining the greatest amount of variability and stream scale explaining the second greatest amount of variability.

This prediction assumes that factors that vary at the region and stream scales are important influences on the metabolome. The influence of the landscape and site of origin on the metabolome has been suggested for both environmental and genetic factors (Bundy et al., 2009; Martyniuk, 2018). If these factors explain much of the variability in the metabolome, metabolomes should group by region and stream of origin (Fig. 2A). In contrast, if these factors account for little of the variability among metabolomes, there may be no discernable pattern in a plot of the metabolomes identified by their region and stream of origin (Fig. 2B).

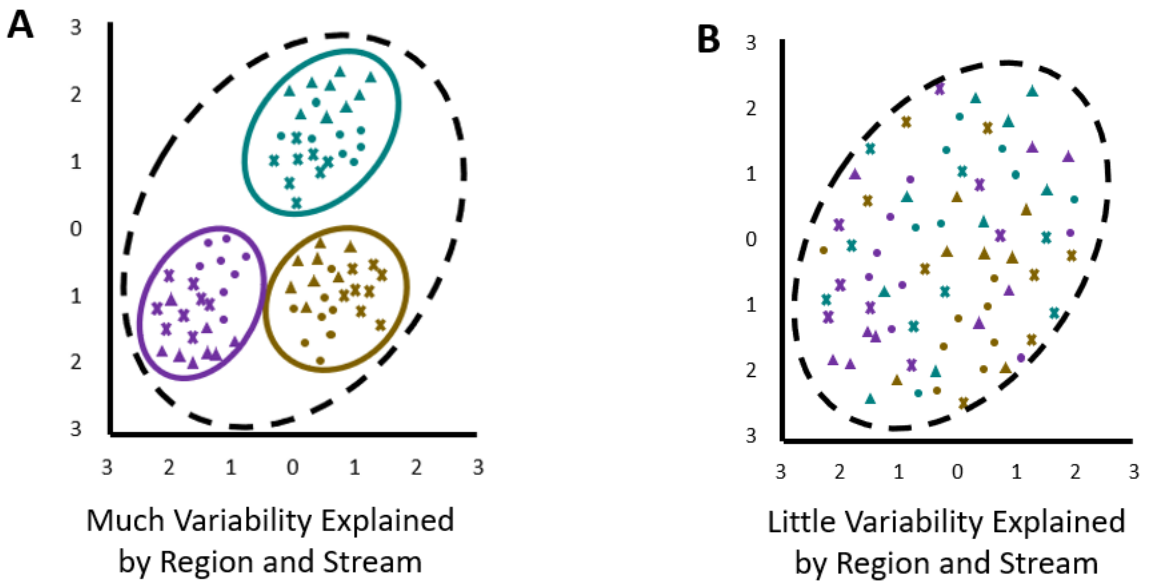


Figure 2. Two extremes of how metabolomes may group based on region and stream of origin. The hypothetical plots show crayfish metabolomes from three regions (purple, teal, and brown), with three streams (circle, triangle, and cross) in each region. The dashed line represents the total variability. In Figure 2A, the metabolomes can be easily grouped by their region and stream of origin. In Figure 2B, however, no pattern based on the regions or streams can be discerned. Although the total variability is the same in both plots, the proportion of variability explained by region and stream is greater in 2A than in 2B.

Because metabolome variability at the region and stream scales could have different sources, this study asked a second question.

2) Is the variability among crayfish metabolomes from different regions and streams caused largely by environmental differences among the locations or genetic differences among the crayfish at those locations?

In response to this question, this study made its second prediction.

2) Metabolome differences among crayfish populations will be caused largely by the environmental differences among the regions and streams that the crayfish inhabit.

Isolation of environmental and genetic influences on variability in the crayfish metabolome can be achieved by removing environmental differences. By placing crayfish originating in separate regions and streams under similar environmental conditions (i.e., a common environment experiment) and allowing them time to adjust to the new environment, the direct effect of environmental differences on the metabolome would be reduced. The degree to which variability among the metabolomes was reduced after completion of the common environment experiment would indicate the relative contribution of

environmental factors to total variability (Fig. 3). A small reduction in variability would indicate that genetic factors contribute more to the variability among the metabolomes, while a large reduction would indicate that environmental factors are more influential. This study predicted that environmental influences would be more influential on the crayfish metabolome because there were many known environmental differences among the regions being sampled and no known research indicating important genetic differences among crayfish in these regions.

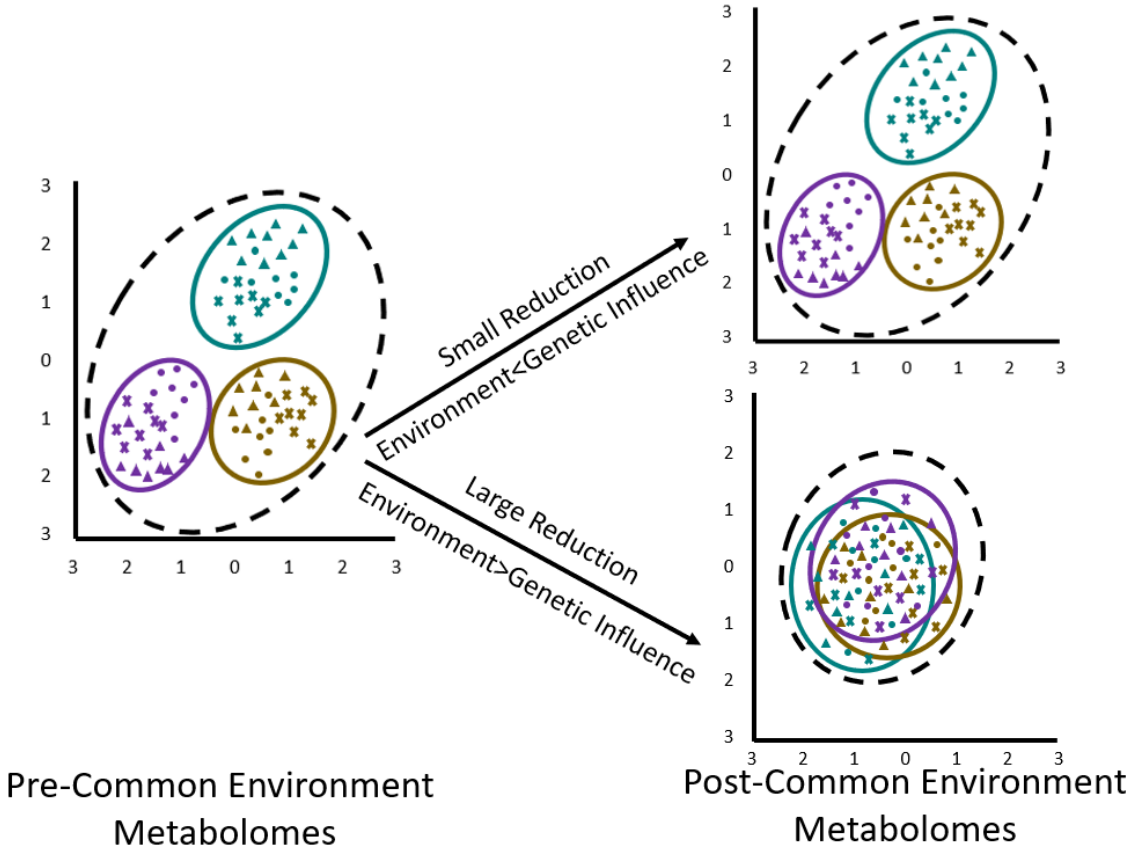


Figure 3. Two circumstances that variability among metabolomes could change in response to the common environment experiment. The hypothetical plots show crayfish metabolomes from three different regions (purple, teal, and brown) with three streams (circle, triangle, and cross) in each region. The dashed line represents total variability. If there is little reduction in the total variability over the course of the common environment experiment (top plot), then the environmental influence on the metabolome is less than the genetic influence. However, if there is a large reduction in the total variability (bottom plot), then the environmental influence on the metabolome is greater than the genetic influence.

Chapter 2

2 Methods

2.1 Study Design

2.1.1 Field Study

The proportionate influence of region and stream on the crayfish metabolome was determined with a hierarchical study design that had two or three streams within each of three ecoregions, for a total of seven streams (Fig. 4A). Separate regions were chosen based on ecoregions, which are areas with distinctive ecological characteristics, such as in climate, physiography, soil, or vegetation (Ecological Stratification Working Group [ESWG], 1995). The selected ecoregions were the Mixed-Grasslands ecoregion, the Mid-Boreal Uplands ecoregion, and the transitional area between the Lake of the Woods ecoregion and the Inter-Lake Plain ecoregion. The selected regions were located within the provinces of Saskatchewan and Manitoba because these areas of Canada have only one native species of crayfish, *Faxonius virilis* (Taylor et al., 2007), helping to ensure that only one species was collected. Streams in each region were selected to represent minimally disturbed reference conditions.

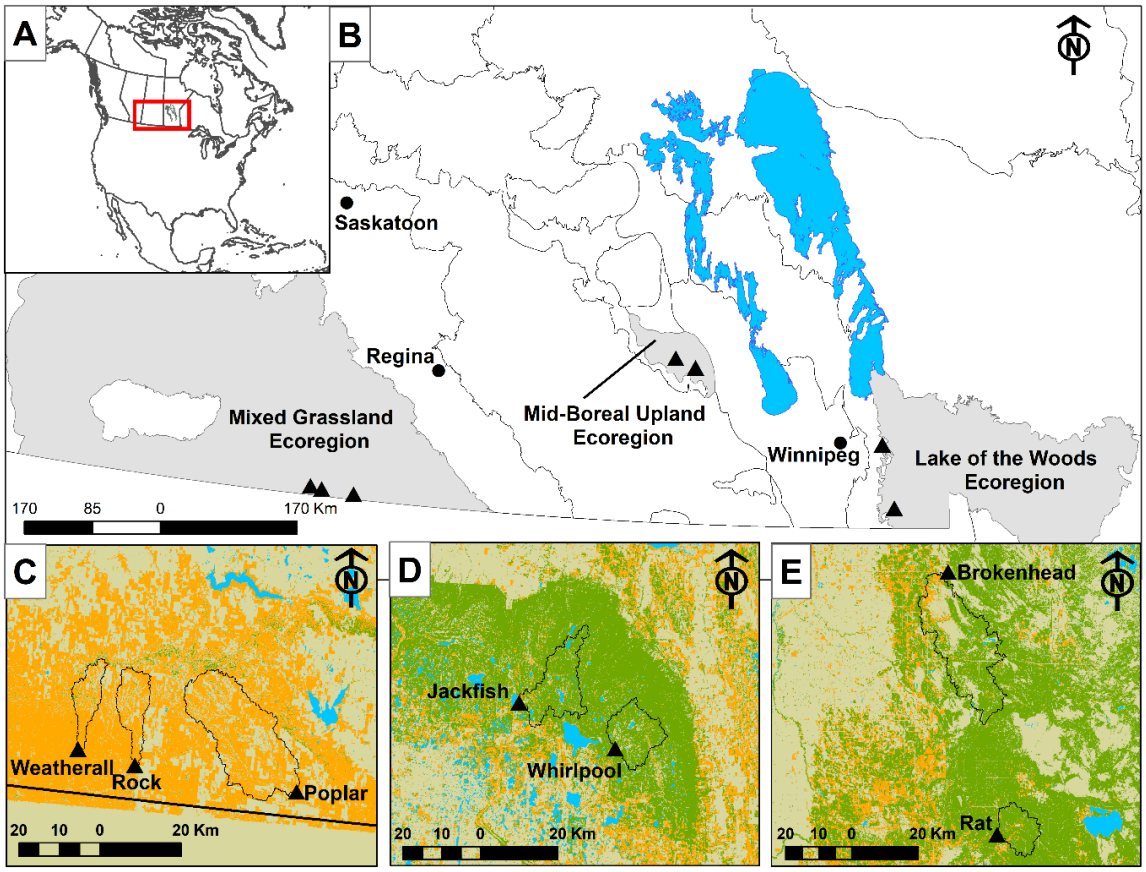


Figure 4. Maps of the study locations: the study location in North America (Panel A), the three sampled ecoregions (Panel B) and the sampling sites with upstream catchments and land cover for the Mixed Grassland (Panel C), the Mid-Boreal Upland (Panel D), and the Lake of the Woods (Panel E). Yellow = grassland, green = forest, blue = water, and tan = all other land covers.

The Mixed Grasslands ecoregion of Southern Saskatchewan has warm summers (mean temperature of 16°C), relatively mild winters (mean temperature of -10°C), and low precipitation (250-350 mm per year), especially in late summer (ESWG, 1995). Its geology is formed largely by Cretaceous sediments, covered with loamy glacial deposits and lacustrine sediments and some sandy eolian deposits. There are areas of high soil salinity (ESWG, 1995). Its topography is flat to undulating, and its native vegetation consists predominately of grasses and sagebrush (ESWG, 1995).

Within the Mixed Grasslands ecoregion three streams, Poplar River, Rock Creek, and Weatherall Creek, were selected (Fig. 4B; Fig. 5A, B, and C; Table 1). Sampling sites on Rock and Weatherall Creeks were located within the East Block of Grasslands National Park, whereas the Poplar River site was located outside the park. Catchments upstream of

the three sampling sites are characterized by sedimentary bedrock and chernozemic soils, although there are small areas of solonetzic soils in the Poplar River Catchment. The majority of land cover in all three catchments (64-69%) is grassland with smaller amounts of cropland (20-4%) and shrubland (21-12%).

The Mid-Boreal Upland ecoregion has short, cool summers (mean temperature of 13°C to 15.5°C), cold winters (mean temperature of -13.5°C to -16°C), and precipitation of 400-550 mm per year (ESWG, 1995). Its geology is dominated by Cretaceous shales covered by loamy to clayey glacial and lacustrine deposits with some coarse fluvioglacial deposits. It has many lakes and poorly drained fens, as well as forests of aspen, poplar, and conifers (ESWG, 1995). The Mid-Boreal Upland ecoregion includes the land covered by Riding Mountain National Park in Manitoba, and Jackfish Creek and the Whirlpool River (Fig. 4C; Fig. 5D and E; Table 1) were selected for crayfish sampling within the park. Catchments upstream of the sampling sites on these streams have sedimentary bedrock and luvisolic soils, and the dominant land cover is forest (77-79%), followed distantly by wetland (8-4%) or shrubland (6-7%).

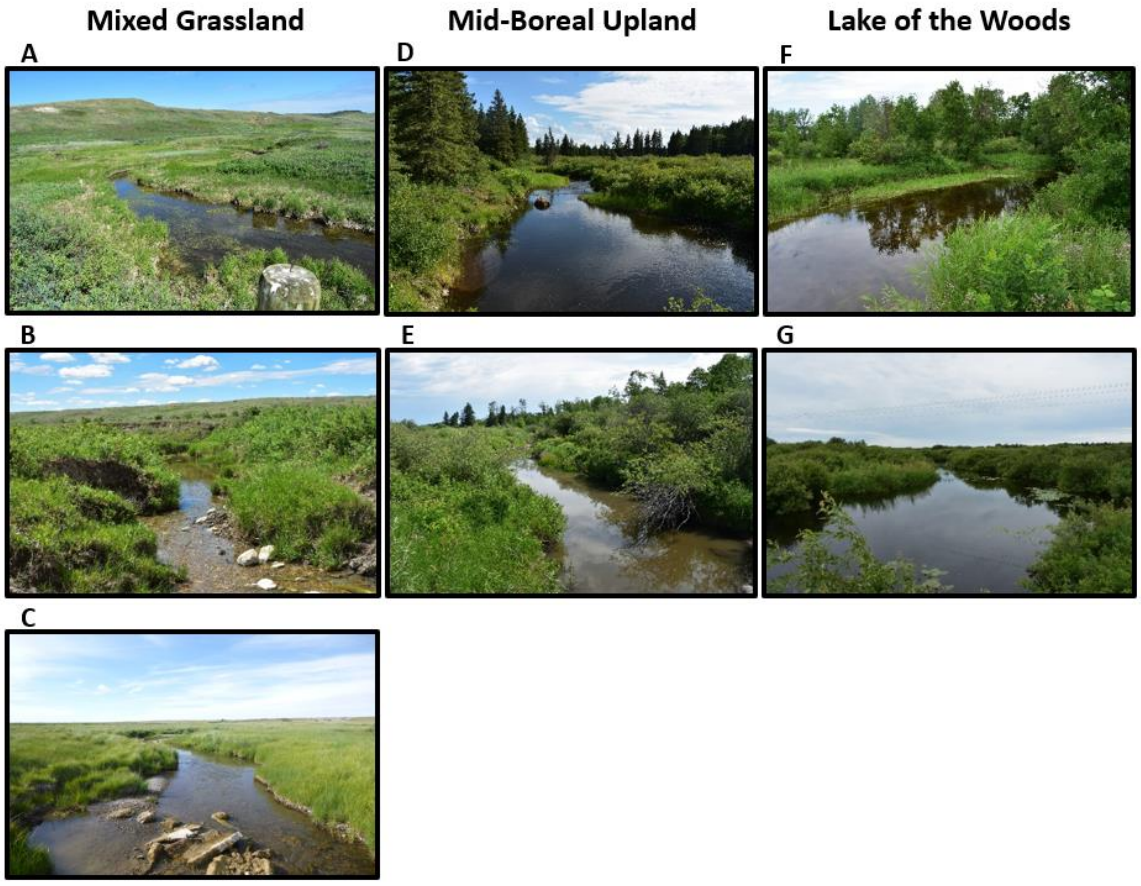


Figure 5. Pictures of the seven sampled streams. In the Mixed Grassland, A = Rock Creek, B = Weatherall Creek, and C = Poplar River. In the Mid-Boreal Upland, D = Whirlpool River and E = Jackfish Creek. In the Lake of the Woods, F = Brokenhead River and G = Rat River.

In southeastern Manitoba, the landscape transitions from the Lake of the Woods ecoregion in the east to the Inter-Lake Plain ecoregion in the west. Both ecoregions have warm summers (mean temperature of 15 °C and 16 °C, respectively), cold winters (mean temperature of -13 °C and -12.5 °C, respectively), and ample precipitation (500-700 mm and 450-700 mm, respectively) (ESWG, 1995). The Lake of the Woods ecoregion is underlain by acidic, crystalline bedrock that commonly protrudes or is thinly covered with brunisolic soil, although there are also areas covered by fluvioglacial outwash or lacustrine clay. The topography is broadly undulating, and the vegetation cover is dominated by birch, aspen, and conifers (ESWG, 1995). Geology in the Inter-Lake Plain ecoregion consists of limestone bedrock covered with calcareous glacial till in the north and loamy, sandy, or clayey lacustrine deposits in the south. The topography is very flat, and the native

vegetation cover consists of oak and aspen groves and rough fescue grasslands (ESWG, 1995).

Sites on the Brokenhead River and Rat River (Fig. 4D; Fig. 5F and G; Table 1) within Sandilands Provincial Forest were selected for crayfish sampling. The catchment upstream of the sampling site on the Brokenhead River has nearly equal areas designated as Lake of the Woods and Inter-Lake Plain. The geology of this catchment is roughly split between sedimentary and intrusive bedrock, and, while nearly half of the soil is organic, the other half is a mix of chernozemic, brunisolic, and luvisolic soils. The catchment upstream of the sampling site on the Rat River is entirely within the Lake of the Woods ecoregion. It has mostly volcanic bedrock, with smaller portions of sedimentary and intrusive bedrock. The soils are brunisolic. Forest is the majority land cover in both catchments, although it is more dominant in the Rat watershed (87%) than in the Brokenhead watershed (56%). Wetlands are an important land cover in the Brokenhead catchment (27%), but not in the Rat catchment (3%).

Table 1. Summary of site location and climate of the seven study streams and associated catchment areas. Mean max. and mean min. temperatures refers to the mean of the monthly maximum or minimum temperatures. The watershed areas were calculated using Arc-GIS (see Data Collection), and the climate data are a mean of data collected from 1971 to 2000[1].

Region	Stream	Latitude	Longitude	Area (km^2)	Mean Max. Temp. (C)[1]	Mean Min. Temp. (C)[1]	Mean Precip. (mm)[1]
Mixed Grassland	Poplar	49.0536°	-105.9455°	488	10.4	-3.1	354
	Rock	49.0718°	-106.5296°	121	10.2	-3.4	348
	Weatherall	49.0924°	-106.7376°	94	10.4	-3.4	347
Mid-Boreal Upland	Jackfish	50.7524°	-100.2305°	212	6.8	-5.0	516
	Whirlpool	50.6594°	-99.8533°	160	6.2	-5.2	552
Lake of the Woods	Brokenhead	49.8850°	-96.3669°	575	8.1	-3.4	566
	Rat	49.2103°	-96.1480°	137	8.2	-3.5	578

[1] (Meteorological Service of Canada, n. d.)

2.1.2 Common Environment Experiment

To address the second question of this study regarding the relative influence of environment and genetics on the crayfish metabolome, a common environment experiment was performed. Crayfish from the seven streams were kept under similar environmental conditions for a period of 16 days, although these periods were staggered for crayfish from different regions. The 16-day time period was considered an appropriate amount of time

for the crayfish to recover from travel and adjust to the common environment conditions because previous studies have shown that two weeks is sufficient to influence the crayfish metabolome (Izral et al., 2018). Crayfish were housed in individual 1.4 litre aquaria with a plastic flowerpot provided for shelter and aerated water purified by reverse osmosis. Water temperature was 20.6 (± 0.4) °C, pH was 7.50 (± 0.20), and dissolved oxygen was 8.02 (± 0.51) mg/L. Crayfish were fed *ad libitum*, with a commercial crayfish pellet. The aquariums were cleaned every 3-5 days. Any incidences of unusual behavior or ecdysis were recorded.

2.2 Data Collection

2.2.1 Environmental Data Collection

The Arc Hydro extension in Arc-GIS 10.5.1 was used to delineate catchment boundaries upstream of the selected sampling sites for each stream. Delineations were based on digital elevation models (DEMs) with a 1 arc second (28-26 m^2 cell size) resolution and stream network layers, obtained from the ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) Global Digital Elevation Model V003 2013-11-30 at Nasa Earth Data (https://earthdata.nasa.gov/) and the Government of Canada's National Hydro Network (https://open.canada.ca/data/en/dataset/a4b190fe-e090-4e6d-881e- b87956c07977), respectively. The climate of the delineated catchments was collected from data layers compiled by the Meteorological Service of Canada (n.d.), and the geology of the catchments was collected from data layers created by the Geological Survey of Canada (Wheeler et al., 1996). Agriculture and Agri Food Canada's (AAFC) Soil Landscapes of Canada version 3.2 (http://sis.agr.gc.ca/cansis/nsdb/slc/v3.2/index.html) was used to calculate the proportions of soil types present in the catchments, and the AAFC's Annual Crop Inventory 2018 (https://open.canada.ca/data/en/dataset/1f2ad87e-6103-4ead-bdd5-147c33fa11e6) was used to calculate the proportions of land cover present.

At each stream, a YSI probe was used to measure pH and specific conductivity. Water samples were collected in a well-mixed area of the stream at approximately 60% depth from the surface in polyethylene bottles: 2 L for nutrients other than ammonium (dissolved organic carbon, nitrate/nitrite, particulate organic nitrogen, soluble reactive phosphorus,

total dissolved nitrogen, total dissolved phosphorus, and total phosphorus), 500 ml for metals, and 250 ml for ammonium. The water sampled for ammonium analysis was acidified with 1 mL of sulfuric acid immediately after collection. All water samples were kept in the dark after collection and refrigerated from the evening of the day of collection until they were analyzed. Nutrients were analyzed at Environment and Climate Change Canada's Saskatoon – National Lab for Environmental Testing (Appendix A.1). Total nitrogen was calculated from the results by summing the concentrations of particulate organic nitrogen and total dissolved nitrogen. Water samples for analysis of dissolved metals were shipped with ice to Environment and Climate Change Canada's Burlington – National Lab for Environmental Testing (Appendix A.2).

2.2.2 Crayfish Tissue Collection

At each of the seven streams, at least eight crayfish were collected and immediately dissected to obtain the tail muscle tissue (Table 2). An additional minimum of eight crayfish were collected for the common environment experiment and placed in containers with aerated stream water before being transported to the National Hydrology Centre in Saskatoon. Crayfish were collected at each stream site over two to four days, using both dip nets and minnow traps baited with dog food. Only female crayfish that were not bearing eggs or young were collected because sex (Hines et al., 2007), and reproductive status (Martyniuk, 2018), may affect the metabolome. Collected crayfish had carapace lengths between 17.9 and 38 mm, so that the crayfish would be roughly matched by age. For both the field crayfish dissected at the stream site and the common environment crayfish dissected at the end of the experiment, the crayfish were first weighed, and the length of their carapace, from the tip of the rostrum to the end of the carapace, was measured. Dissection instruments were cleaned with ethanol and allowed to air dry between dissections to avoid cross-contamination of samples. Immediately after dissection, the tail tissue was placed in a cryo-vial and flash frozen in a dewar of liquid nitrogen. The tissue samples were kept in the dewar until they could be transferred to a -80°C freezer. Eight crayfish died before the end of the common environment experiment: three from Rock, one from Weatherall, and four from Jackfish.

Table 2. Sampling dates and the number of crayfish collected from each stream for the field study and the common environment experiment.

Region	Stream	Sampling Dates	Field Crayfish	Common Environment Crayfish
Mixed Grassland	Poplar	June 10-13, 2019	9	8
	Rock	June 11-13, 2019	10	8
	Weatherall	June 12-14, 2019	15	8
Mid-Boreal Upland	Jackfish	July 9-12, 2019	8	10
	Whirlpool	July 10-12, 2019	8	9
Lake of the Woods	Brokenhead	July 17-18, 2019	12	13
	Rat	July 16-18, 2019	8	8
		Total:	70	64

2.3 Sample Preparation and Analysis

Crayfish tail muscle samples were freeze-dried for 48-hours and ground in a chilled Precellys homogenizer. To extract the polar metabolites using a methanol:chloroform:water (2:2:1.8) extraction (Viant, 2007), 0.60 ml of ice-cold methanol and 0.27 ml of ice-cold Milli-Q water were added to 10 mg of powdered tissue. The solution was vortexed for 15 seconds three times over 2 minutes and centrifuged at 13,400 g at 4°C for 10 minutes. The supernatant was removed, and 0.60 ml of ice-cold chloroform and 0.27 ml of ice-cold Milli-Q water were added. This solution was vortexed for 60 seconds and left on ice for 10 minutes to begin partitioning. To complete partitioning, the solution was centrifuged at 13,400 g at 4°C for 10 minutes. The methanol fraction was removed, placed in a separate Eppendorf tube, and evaporated until dry in a Speedvac Concentrator (ThermoScientific). The evaporated samples were stored in a -80°C freezer until their metabolite content was measured with proton nuclear magnetic resonance (^{1}H NMR) spectroscopy. At that time, each sample was resuspended in 0.75 ml of 100 mM sodium phosphate buffer (pH = 7.0), made with deuterium (D_2O) and containing 3 mM sodium azide (NaN_3) as a preservative and 1 mM trimethylsilylpropanoic acid (TMSP) as an internal standard. After being vortexed for 10 seconds, 0.60 ml of the resuspended sample was transferred to a 5 mm NMR tube.

Spectra of the crayfish tail muscle polar metabolomes were obtained on a Bruker Avance 500 MHz spectrometer, at 298 K, using a 5 mm Bruker TCI cryoprobe. The spectrometer was operated at a frequency of 500.27 MHz and was locked to D_2O. The spectrometer was tuned and shimmed before each spectral acquisition. Spectra were obtained using an

excitation sculpting pulse program (Bollard et al., 2005) with 60° pulse, a relaxation delay of 2 seconds, a spectral width of 9,615.38 Hz, and 128 scans. The obtained spectra were modified by 0.3 Hz line broadening and zero-filling to 32,768 points. Each spectrum was then Fourier transformed, phased, baseline corrected with a polynomial function, and referenced to TMSP.

2.4 Data Analysis

2.4.1 Analysis of Environmental Data

To examine environmental variation among the seven streams, three separate principal component analyses (PCAs) were performed: one on land cover in the upstream catchments, one on stream nutrients, and one on dissolved metals. Land cover data (AAFC, 2018) was reclassed into broader categories for the purposes of this investigation by combining specific crop covers and fallow land into a single "cropland" category, different types of forest (e.g., broadleaf and coniferous) into a single "forest" category, and grassland and pasture into a single "grassland" category. These three broader categories plus wetland and shrubland described over 93% of the land cover in each catchment; all other categories comprised <6% of the land cover and were dropped from the analysis. A Pearson correlation matrix among the five remaining land cover variables was calculated. Shrubland and grassland had a high ($r = 0.91$) positive correlation, and both had a high negative correlation ($r \geq 0.89$) with forest. Therefore, shrubland and grassland were removed from the analysis, and the PCA was performed with forest, cropland, and wetland as variables. Land cover variables were ArcSine/Square root transformed in Excel prior to performing the PCA.

For the nutrients PCA, nitrate/nitrite was not used because it was below the detection limit (10 ug/l) at all the streams. Particulate organic nitrogen (PON) and total dissolved nitrogen (TDN) were excluded because they were highly correlated ($r = 0.91$ and 0.89, respectively) with their sum, total nitrogen (TN). Total dissolved phosphorus (TDP) and soluble reactive phosphorus (SRP) were excluded because they were highly correlated with total phosphorus (TP) ($r = 0.96$). Thus, the nutrient variables retained for the PCA were

ammonia as nitrogen (NH_3), dissolved organic carbon (DOC), total nitrogen (TN), and total phosphorus (TP). All nutrient concentrations were log (x+1) transformed prior to the PCA.

For the metals PCA, all dissolved metals except calcium, magnesium, sodium, and iron were excluded because of low concentrations, low variability, or a lack of biological justification. Specific conductivity was not used in the metals PCA because it was highly correlated (r = 0.95) with dissolved sodium. Thus, the variables used in the metals PCA were pH and dissolved calcium, magnesium, sodium, and iron. pH has known effects on the crayfish exoskeleton, and therefore growth and moulting (Willis-Jones et al., 2016), while calcium is an essential component of the exoskeleton (McLay & van den Brink, 2016). Magnesium is important, along with calcium, for determining water hardness, a parameter that can influence a crayfish's response to pH (Berrill et al., 1985). Sodium was included because the high concentrations in some of the sampled streams could affect a crayfish's osmotic regulation. Finally, iron was included because it could potentially be a limiting nutrient that affects stream productivity and the amount of food available to resident crayfish. Prior to being used in the PCA, the dissolved metal data was log (x+1) transformed and, along with the pH data, normalized to put the metal concentrations and pH values on a common scale.

All PCA's were performed in R 3.6.0 (R Core Team, 2018) using the function "prcomp" (R Core Team, 2018), which uses singular value decomposition. The argument "scale" was set to "False" so that the data was not scaled to have unit variance prior to the analysis. The number of significant PC axes was determined with the broken stick model, using the R function "bsDimension" (Coombes & Wang, 2019).

2.4.2 Analysis of Spectral Data

Field and common environment crayfish datasets were individually analyzed using the following analyses. MATLAB was used to segment each spectrum into 0.01 ppm bins, scale the bins by the TMSP peak, and remove the bins containing the water peak (4.705 to 4.805 ppm, inclusive) and TMSP peak (0 to 0.195 ppm, inclusive). MetaboAnalyst 4.0 (Chong et al., 2019) was used to data filter the spectral bins with the interquartile range and to normalize the spectral bins with auto-scaling.

In R 3.6.0 (R Core Team, 2018), the spectral bins were used to calculate Euclidean distance matrices for both the field and the common environment datasets, and the distances were ordinated with nonmetric multidimensional scaling (nMDS) to visualize the similarity among the crayfish metabolomes within each dataset. The metabolomes of three common environment crayfish (one from each of Rock, Weatherall, and Poplar) appeared highly dissimilar from the metabolomes of the other crayfish and the data from these crayfish were removed from the dataset. Lab records indicated that the metabolite profile of these crayfish may have been affected by special circumstances: one crayfish had a very small hepatopancreas when dissected, indicating physiological stress; one crayfish had white muscle tissue, indicating a microsporidium infection; and the tail muscle tissue of one crayfish failed to fully freeze-dry initially, requiring that it be freeze-dried a second time. After removing the data from these crayfish, the Euclidean distance matrix for the common environment crayfish was rerun and re-visualized with a nMDS plot.

Separate PERMANOVA's were run on both the field and common environment distance matrices, using a design file that identified "Region" as a fixed factor and "Stream" as a random factor nested within "Region". The PERMANOVA's were run with 9999 permutations using Type I sums of squares partitioning method, and "Permutation of residuals under a reduced model" as the permutation method. An alpha of 0.1 was used to decide if the effects of region and stream were significant, and the estimates of the components of variation were used to find the proportion of the total variability that was attributable to region and stream. The estimates of the components of variation were compared between the field crayfish and the common environment crayfish to determine how the total variability and the division of the total variability among region, stream, and residuals changed between the field and the common environment experiment. The change in total variability between the field crayfish and the common environment crayfish was used to deduce the relative contribution of environment and genetics to total variability.

Partial least squares discriminant analysis (PLS-DA) in MetaboAnalyst 4.0 (Chong et al., 2018) was used to discover the bins, representing metabolites, that were most responsible for regional and stream differences among the field and common environment crayfish metabolomes. Bins assigned the highest variable importance in projection (VIP) scores in

MetaboAnalyst 4.0 (Chong et al., 2018) were examined in Chenomx Profiler v8.5 to find the metabolites responsible for the peaks within those bins. Spectra for the average crayfish metabolome from each stream and region for the field and common environment crayfish were created by plotting the average value for each bin in R 3.6.0 (R Core Team, 2018). Where possible, the peaks of metabolites with high VIP scores were highlighted on the average spectra.

Chapter 3

3 Results

3.1 Characterization of Stream Environments

GIS analysis of land cover in the catchments upstream of the selected sampling sites found that grassland was the dominant land cover in the Mixed Grassland catchments, whereas forest was the majority land cover in the Mid-Boreal Upland and the Lake of the Woods catchments (Table 3). For the PCA of land cover in the upstream catchments, the broken-stick model revealed that only the first axis, explaining 92% of the variability among watersheds, was important. This axis was a gradient of forest cover (Forest loading = 0.941). The watersheds of the three streams in the Mixed Grassland were distributed towards the negative end of this axis (axis scores: Poplar = -0.62; Rock = -0.61; Weatherall = -0.49). The catchments of streams in the Mid-Boreal Upland (axis scores: Jackfish = 0.42; Whirlpool = 0.48) and Lake of the Woods (axis scores: Brokenhead = 0.28; Rat = 0.54) were distributed towards the positive end of the axis (Table 3).

Table 3. Classified land cover (%) in the upstream catchments of the sampled streams.

Region	Streams	Shrubland	Wetland	Grassland	Cropland	Forest
Mixed Grassland	Poplar	12.2	1.0	64.4	20.1	1.0
	Rock	18.5	0.6	68.4	10.8	0.7
	Weatherall	21.3	0.3	69.5	4.3	3.1
	Mean ± 1SD	17.3 ± 4.7	0.6 ± 0.4	67.4 ± 2.7	11.7 ± 7.9	1.6 ± 1.3
Mid-Boreal Upland	Jackfish	6.4	3.5	6.2	0.4	77.3
	Whirlpool	6.9	8.5	3.0	0.0	78.7
	Mean ± 1SD	6.6 ± 0.3	6.0 ± 3.5	4.6 ± 2.2	0.2 ± 0.3	78.0 ± 1.0
Lake of the Woods	Brokenhead	5.7	26.6	8.5	0.4	56.3
	Rat	1.5	2.9	5.4	1.0	87.5
	Mean ± 1SD	3.6 ± 3.0	14.8 ± 16.8	6.9 ± 2.2	0.7 ± 0.4	71.9 ± 22.1

According to regional means, stream water nutrient concentrations in the Mixed Grassland catchments were between 5% and 87% lower than in the Mid-Boreal Upland or Lake of the Woods (Table 4). As well, the Mixed Grassland and the Lake of the Woods tended to be less variable in their nutrient concentrations than the Mid-Boreal Upland. For instance, total nitrogen concentrations ranged from 333-514 μg/L in the Mixed Grassland and 929-

980 μg/L in the Lake of the Woods, but from 486-1590 μg/L in the Mid-Boreal Upland. However, all streams were consistent in having nitrate/nitrite concentrations below the detection limit (10 μg/L). Jackfish Creek was notable in having the highest nutrient concentrations, except in DOC, of all seven streams.

PCA of NH_3, DOC, TP, and TN concentrations resulted in one interpretable axis according to the broken-stick model, and this axis explained 72% of the variability. Although TP made the most important contribution to this axis (TP loading: -0.634), the other nutrient forms also made important contributions of the same direction (TN = -0.500, DOC = -0.488, and NH_3 = -0.333). Thus, axis 1 represented a gradient of nutrient enrichment, separating streams with higher amounts of nutrients from streams with lower concentrations. Jackfish, followed by the Rat (axis scores of -0.822 and -0.433, respectively), fell at the negative end of the axis with high nutrient concentrations, whereas the Mixed Grassland streams (axis scores: Poplar = 0.336, Rock = 0.472, and Weatherall = 0.290) fell at the positive end with low nutrient concentrations. Whirlpool and Brokenhead (axis scores of 0.251 and -0.095, respectively) fell between these two extremes.

Table 4. Nutrient concentrations (μg/L) in sampled streams. DOC = dissolved organic carbon, SRP = soluble reactive phosphorus, TDP = total dissolved phosphorus, TP = total phosphorus, NH_3 = ammonia as nitrogen, NO_3/NO_2 = nitrate/nitrite, PON = particulate organic nitrogen, TDN = total dissolved nitrogen, TN = total nitrogen, and DL = detection limit.

	Stream	DOC	SRP	TDP	TP	NH_3	NO_3/NO_2	PON	TDN	TN
Mixed Grassland	Poplar	4340	3	15.1	28.7	19	<10 (DL)	56	352	408
	Rock	4770	4	16.9	29.1	8	<10 (DL)	38	295	333
	Weatherall	7730	1	14.6	26.1	9	<10 (DL)	34	480	514
	Mean	5613	3	15.5	28.0	12	NA	43	376	418
	± 1SD	± 1507	± 1	± 1.0	± 1.3	± 5		± 10	± 77	± 74
Mid-Boreal Upland	Jackfish	17000	13	35.4	180.0	33	<10 (DL)	666	924	1590
	Whirlpool	8470	3	7.5	17.0	26	<10 (DL)	20	466	486
	Mean	12735	8	21.5	98.5	30	NA	343	695	1038
	± 1SD	± 4265	± 5	± 14.0	± 81.5	± 3.5		± 323	± 229	± 552
Lake of the Woods	Brokenhead	21400	3	13.0	20.3	20	<10 (DL)	36	893	929
	Rat	18500	10	19.6	70.0	24	<10 (DL)	189	791	980
	Mean	19950	7	16.3	45.2	22	NA	113	842	955
	± 1SD	± 1450	± 4	± 3.3	± 24.9	± 2		± 77	± 51	± 26

Regional means showed that specific conductivity (SPC) was over 1 ½ -times greater in the Mixed Grassland than in the Mid-Boreal Upland or the Lake of the Woods (Table 5). As well, magnesium was also over 1 ½ - times greater, and sodium was over 10 times greater in the Mixed Grassland. Within the Mixed Grassland, sodium concentrations were lowest in the easternmost stream (Poplar), at 52.8 mg/L, and highest in the westernmost stream (Weatherall), at 220 mg/L. pH was also higher by at least 0.58 in the Mixed Grassland compared to the other regions. In contrast, calcium was over a third lower in the Mixed Grassland than in the other two regions. Iron was more than twice as high in the Rat than in any other stream but did not show any regional trends.

The PCA of dissolved metals and pH resulted in one interpretable axis according to the broken-stick model. This axis explained 68% of variability among the streams, with pH (loading = -0.523), sodium (loading = -0.493), magnesium (loading = -0.475), and calcium (loading = 0.441) making important contributions. Thus, the axis represented a gradient with high pH, sodium, and magnesium on one end and high calcium on the other end. Streams from the Mixed Grassland (axis scores: Poplar = -1.84, Rock = -1.24, and Weatherall = -2.46) fell at the negative end of the axis, whereas streams from the Mid-Boreal Upland (axis scores: Jackfish = 0.94 and Whirlpool = 1.10) and the Lake of the Woods (axis scores: Brokenhead = 0.95 and Rat = 2.56) fell towards the positive end with high calcium.

Table 5. Specific conductivity (ms/cm), pH, and concentrations of dissolved metals (mg/L) in the water of the sampled streams. SPC = specific conductivity.

Region	Stream	SPC	pH	Calcium	Magnesium	Sodium	Iron
Mixed Grassland	Poplar	786	8.58	42.0	54.7	52.8	0.054
	Rock	937	8.43	34.7	28.4	151.0	0.085
	Weatherall	1393	8.61	35.9	52.3	220.0	0.050
	Mean	1039	8.54	37.5	45.1	141.3	0.063
	± 1SD	± 316	± 0.10	± 3.9	± 14.5	± 84.0	± 0.019
Mid-Boreal Upland	Jackfish	726	7.98	87.6	32.0	22.9	0.081
	Whirlpool	524	7.91	68.0	26.2	3.3	0.039
	Mean	625	7.95	77.8	29.1	13.1	0.060
	± 1SD	± 143	± 0.05	± 13.9	± 4.1	± 13.9	± 0.030
Lake of the Woods	Brokenhead	482	8.21	67.3	19.1	5.3	0.049
	Rat	410	7.71	61.4	15.9	2.9	0.173
	Mean	446	7.96	64.3	17.5	4.1	0.111
	± 1SD	± 51	± 0.35	± 4.2	± 2.3	± 1.7	± 0.087

3.2 Field Study: Regional and Stream Sources of Variability

Among the crayfish used in the field study, crayfish from the Mixed Grassland had a significantly lower mean mass (3.4 ± 1.2 g) than crayfish from the Mid-Boreal Upland (8.2 ± 2.5 g; $p < 0.001$), or the Lake of the Woods (7.7 ± 2.2 g; $p < 0.001$) (Fig. 6A). Similarly, mean carapace length of crayfish from the Mixed Grassland (23 ± 3 mm) was significantly shorter than the mean carapace length of crayfish from the Mid-Boreal Upland (31 ± 3 mm; $p < 0.001$), and the Lake of the Woods (30.5 ± 3; $p < 0.001$) (Fig. 6B). There were no significant differences in mass ($p = 0.45$) or carapace length ($p = 0.90$) between the Mid-Boreal Upland crayfish and the Lake of the Woods crayfish.

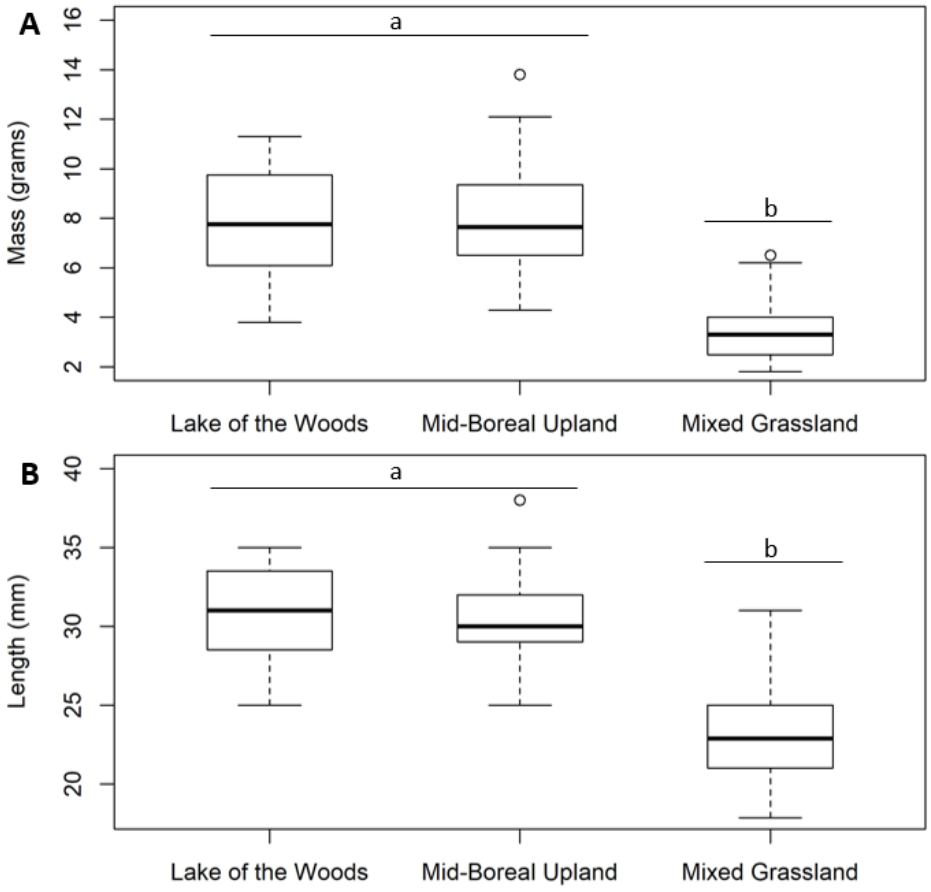


Figure 6. Mass (A) and carapace length (B) of crayfish from each region used in the field study. Boxplots show median (dark horizontal bar) flanked by the interquartile range (25th to 75th percentiles), with whiskers that extend to the closer of either the minimum or maximum value or 1.5x the interquartile range. Values higher or lower than 1.5x the interquartile range are shown as hollow circles. Significant differences are designated by different letters over the boxplots.

An nMDS ordination of the field crayfish metabolomes revealed no interpretable pattern based on the region or stream of origin (Fig. 7). The metabolomes from the seven streams had similar distributions, except for metabolomes from the Whirlpool River, which spread from edge to edge of the plot. The PERMANOVA analyzing the field crayfish metabolomes found an effect of region ($p = 0.089$) and of stream within region ($p = 0.002$) on the crayfish metabolome. The estimates of the components of variation revealed that, out of a total variation of 747, region accounted for 5.0% (36.7) and stream within region accounted for an additional 7.6% (56.5). The remaining 87% (654) of variation was among individual crayfish from a given stream within a given region.

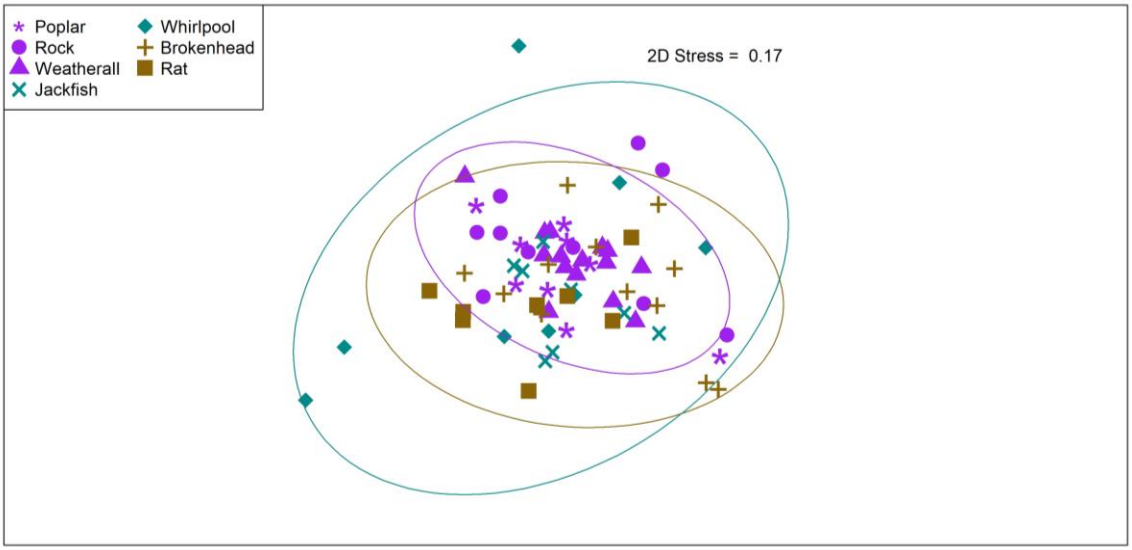


Figure 7. nMDS ordination based on Euclidean distance showing the relative similarities between the field crayfish metabolomes. The markers are colour-coded by region: Mixed Grassland (purple), Mid-Boreal Upland (teal), and Lake of the Woods (brown). Ellipses show the 95% confidence intervals.

PLS-DA identified three metabolites that primarily drove regional differences among the field crayfish (Appendix B.1). The amino acid, glycine, was most abundant in crayfish from the Mixed Grassland and least abundant in crayfish from the Lake of the Woods (Fig. 8A). Methionine was most abundant in crayfish from the Mid-Boreal Upland, and arginine was most abundant in crayfish from the Lake of the Woods, with both amino acid metabolites being least abundant in crayfish from the Mixed Grassland.

PLS-DA identified several additional metabolites that differentiated the field crayfish metabolomes by stream (Appendix B.2). Although abundant in all the metabolomes from the Mixed Grassland region, glycine was particularly abundant in crayfish from Weatherall (Fig. 9). Glycine was least abundant in crayfish from the Rat. Malonate was most abundant in crayfish from Whirlpool and Rock, and least abundant in crayfish from Jackfish. Whirlpool crayfish also had the highest abundance of methionine, whereas Weatherall crayfish had the lowest. Lactate was most abundant in crayfish from Rat, followed by crayfish from Weatherall and Poplar, but least abundant in crayfish from Rock.

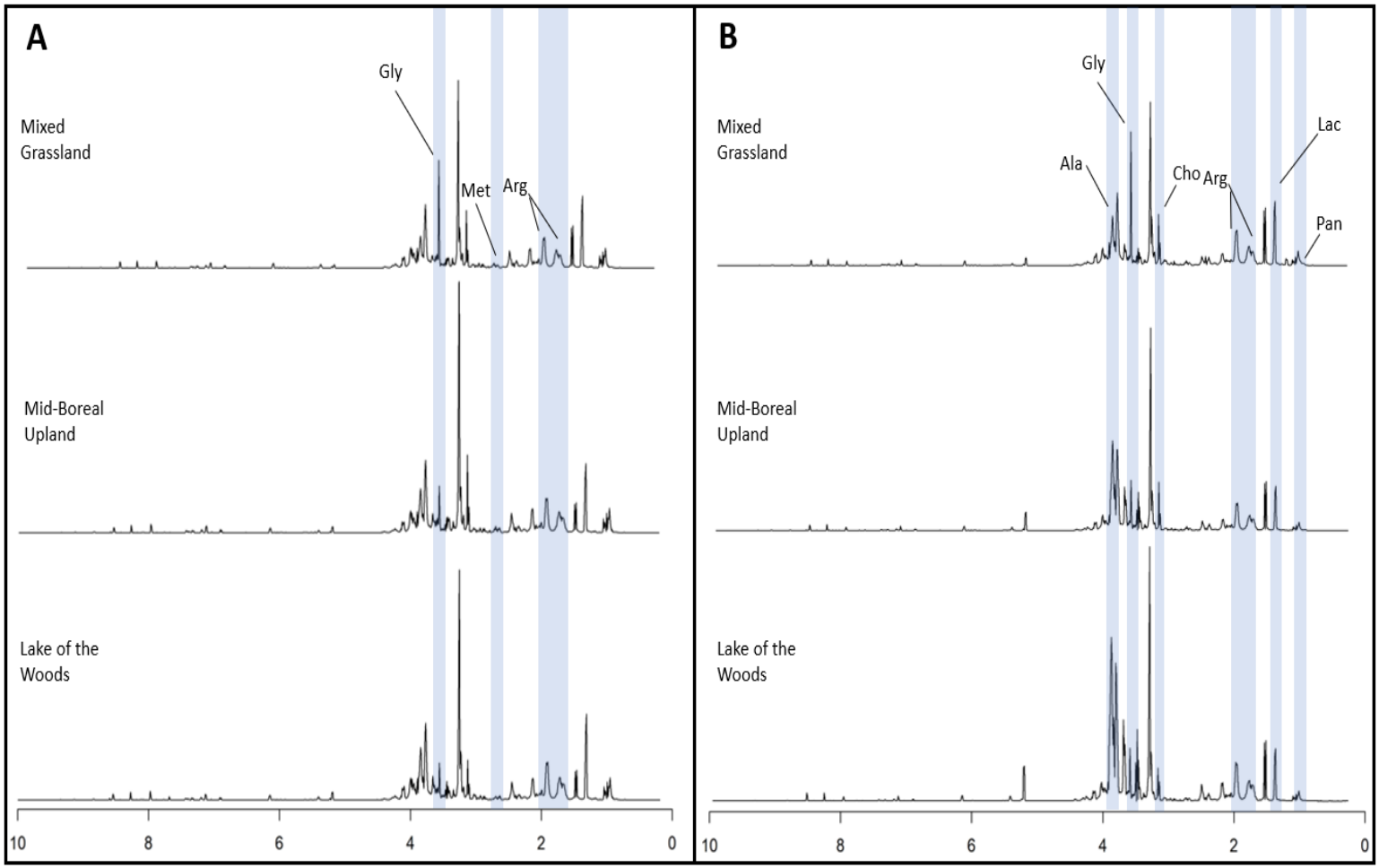


Figure 8. Average metabolomes of the crayfish collected from each region for the field crayfish (A) and the common environment crayfish (B). The spectral peaks of metabolites that are important for regional differences are highlighted: Ala = Alanine, Arg = Arginine, Cho = Choline, Gly = Glycine, Lac = Lactate, Met = Methionine, Pan = Pantothenate.

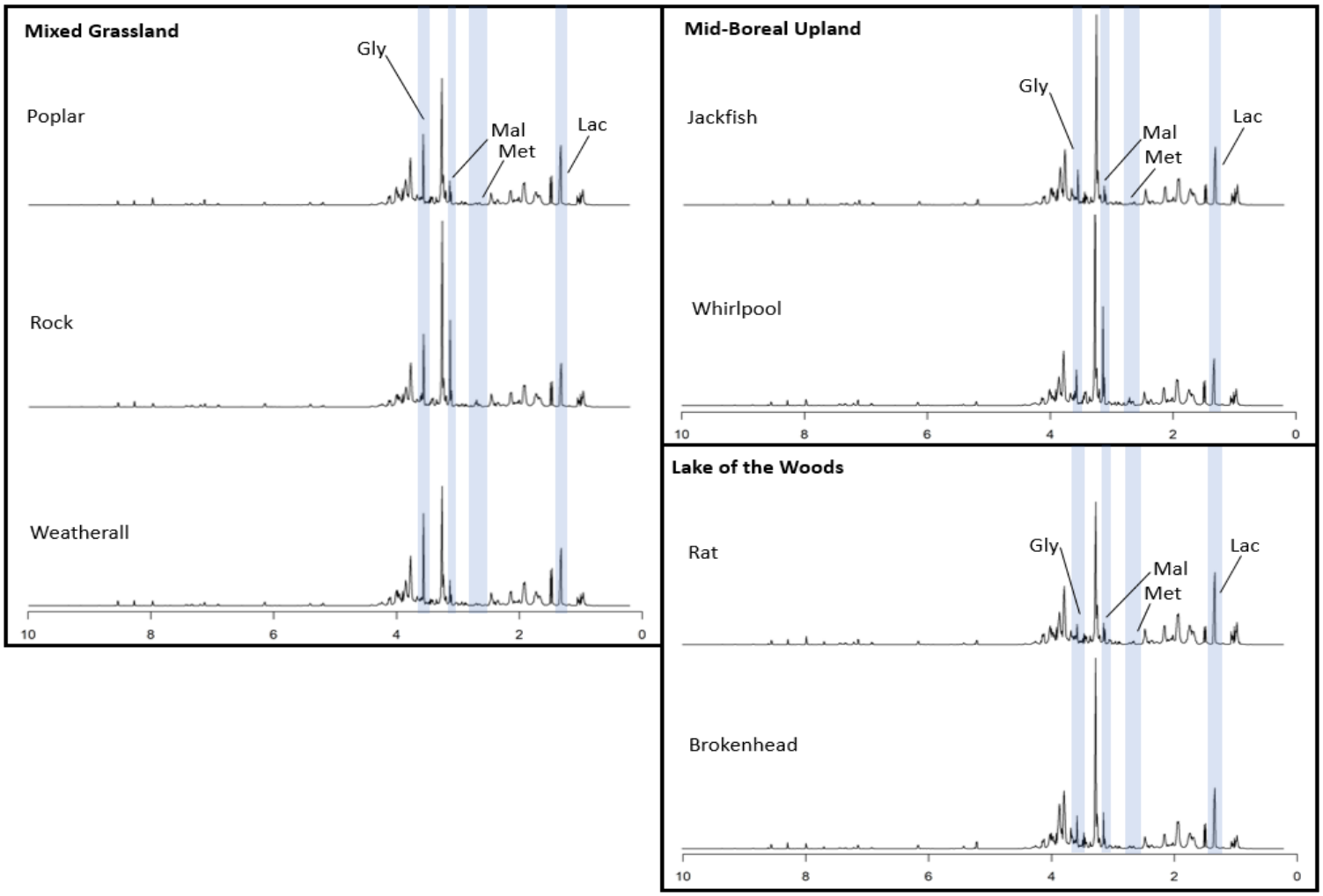


Figure 9. Average metabolomes of the field crayfish collected from each stream. The spectral peaks of metabolites that are important for differences by stream are highlighted. Gly = Glycine, Lac = Lactate, Mal = Malonate, Met = Methionine

3.3 Common Environment Experiment: Environmental and Genetic Sources of Variability

Among the common environment crayfish, the mean masses of the crayfish from all three regions were significantly different from each other, with Lake of the Woods crayfish having greater mass (9.4 ± 2.4 g) than crayfish from both the Mid-Boreal Upland (5.6 ± 2.0 g; $p < 0.001$) and the Mixed Grassland (3.4 ± 1.1 g; $p < 0.001$), and crayfish from the Mid-Boreal Upland having greater mass than crayfish from the Mixed Grassland ($p = 0.002$) (Fig. 10A). Likewise, Lake of the Woods crayfish had longer carapaces (33 ± 2 mm) than both Mid-Boreal Upland crayfish (28 ± 4 mm; $p < 0.001$) and Mixed Grassland crayfish (24 ± 3 mm; $p < 0.001$), and Mid-Boreal Upland Crayfish had longer carapaces than Mixed Grassland crayfish ($p < 0.001$) (Fig. 10B).

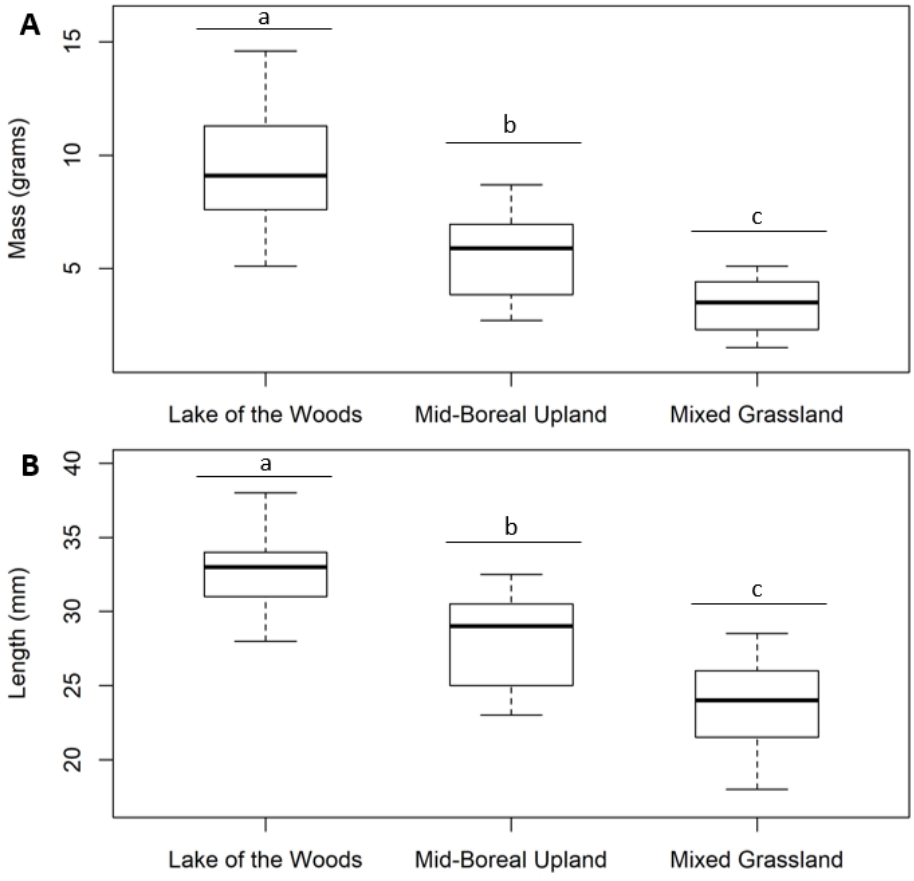


Figure 10. Mass (A) and carapace length (B) of the crayfish from each region used in the common environment experiment. The boxplots show the median flanked by the interquartile range (25th to 75th percentiles), with whiskers that extend to the closer of either the minimum or maximum value or 1.5x the interquartile range. Significant differences are designated by different letters over the boxplots.

An nMDS ordination of the common environment crayfish metabolomes revealed groups based on the region and stream of origin (Fig. 11). Metabolomes from the Mixed Grassland grouped in the centre-right of the plot, whereas metabolomes from the Mid-Boreal Upland appeared largely in the lower centre-left of the plot, with the exceptions of a Jackfish metabolome in the far left of the plot and a Whirlpool metabolome in the far right. Metabolomes from the Lake of the Woods spread from the left to the upper centre of the plot. More separation among streams was also apparent, such as with the metabolomes from Weatherall and Poplar, which separated into two adjacent, but non-overlapping groups.

The PERMANOVA on the common environment crayfish metabolomes found significant effects of both region ($p = 0.023$) and stream within region ($p = 0.0001$). The estimates of the components of variation summed to 784, a 5% increase from the field crayfish. However, 20% more of this variability was explained by region and stream, with 15% (114) explained by region and 18% (141) explained by stream within region. The remaining 67% (528) of variation was among individual crayfish from a given stream within a given region.

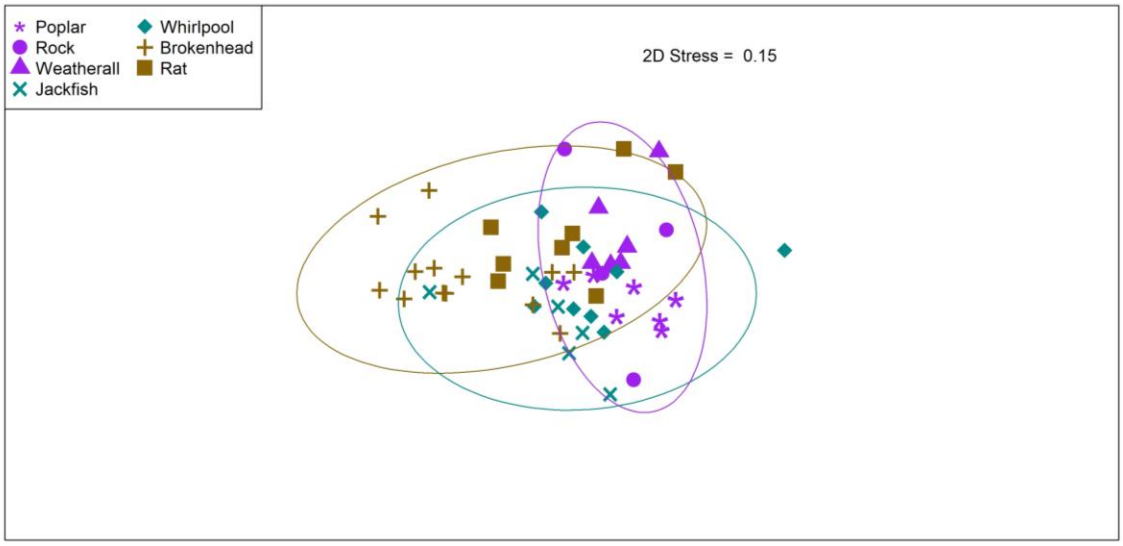


Figure 11. nMDS ordination based on Euclidean distance showing the relative similarities between the common environment crayfish metabolomes. The markers are colour-coded by region: Mixed Grassland (purple), Mid-Boreal Upland (teal), and Lake of the Woods (brown). Ellipse represent 95% confidence intervals.

PLS-DA identified multiple metabolites that drove regional differences among the common environment crayfish (Appendix B.3). As among the field crayfish, glycine was most abundant in the crayfish from the Mixed Grassland and least abundant in the crayfish from the Lake of the Woods (Fig. 8B). This pattern was followed by most of the other metabolites – choline, arginine, lactate, and pantothenate – that contributed to differentiation of the metabolomes by region. One exception was alanine, which was most abundant for crayfish from the Lake of the Woods and least abundant in crayfish from the Mixed Grassland.

PLS-DA also found that the metabolites carnitine, trehalose, and methionine were important for explaining differences by stream among the metabolomes of the common environment crayfish (Appendix B.4). Carnitine was most abundant in the crayfish from Poplar, followed by the other Mixed Grassland streams, and least abundant in crayfish from Brokenhead (Fig. 12). Methionine was most abundant in crayfish from Rock, followed by Poplar and Whirlpool, and least abundant in Brokenhead. The most apparent peak changes on the spectra from stream to stream were caused by changes in the abundance of trehalose, which had multiple peaks throughout the spectra. The abundance of trehalose was greatest in the crayfish from Brokenhead, followed by Jackfish and Whirlpool, and lowest in crayfish from streams in the Mixed Grasslands.

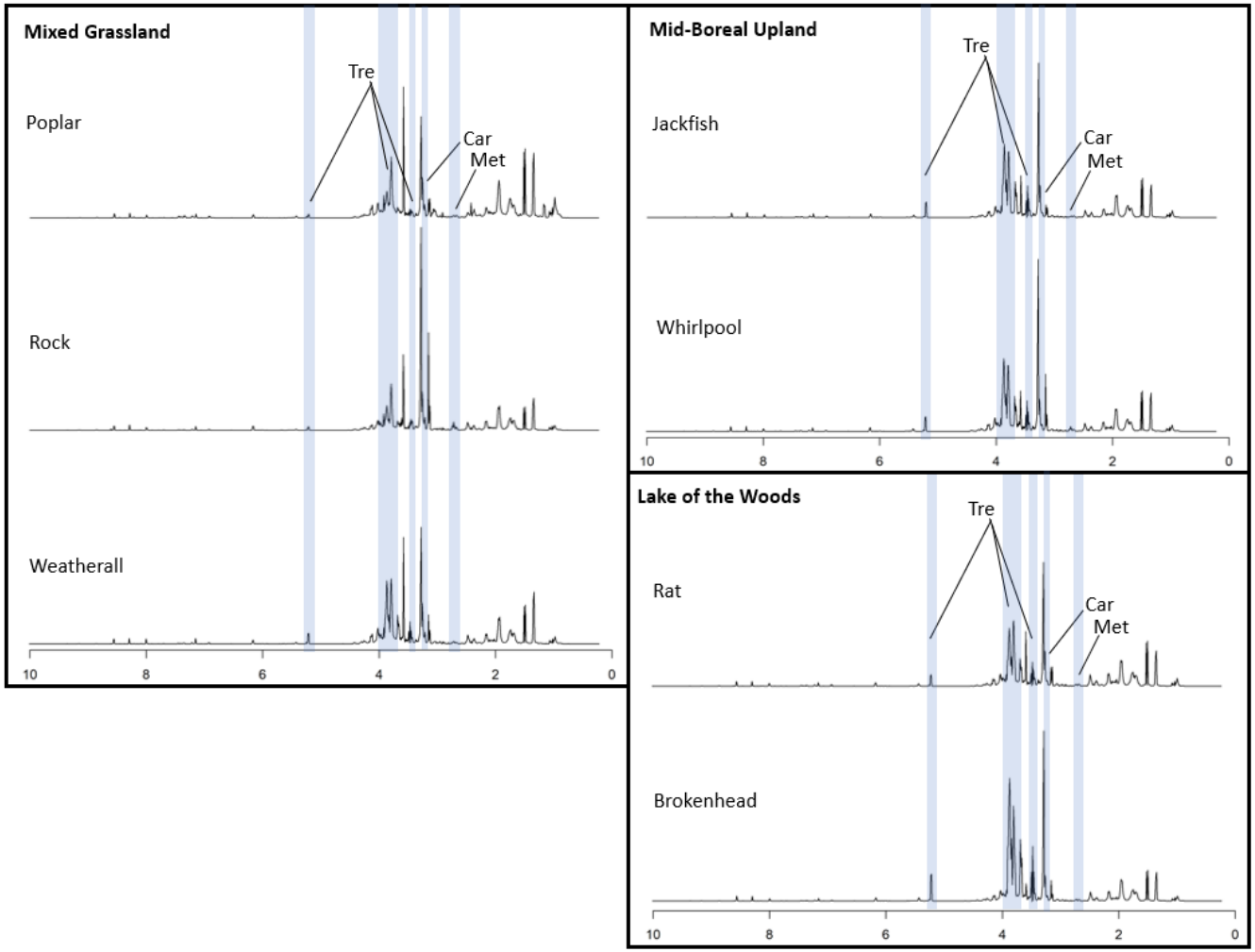


Figure 12. Average metabolomes of the common environment crayfish collected from each stream. The spectral peaks of metabolites that are important for differences by stream are highlighted. Car = Carnitine, Met = Methionine, Tre = Trehalose

4 Discussion

4.1 Regional and Stream Influences on the Metabolome

Contrary to my predictions, region and stream scales explained little of the variability among the metabolomes of the field crayfish. A possible explanation for this finding is that the range of environmental variation incorporated into my study was insufficient to generate physiological changes in my study species and thus result in substantive among region or stream differences in the metabolomes. Indeed, the streams in my study were chosen to represent reference conditions and likely did not exhibit the more extreme environmental conditions observed in streams impacted by human activities and associated with changes in the metabolome in previous studies (Fernandez-Cisnal et al., 2018; Gago-Tinoco et al., 2014). For instance, although I found that streams in the Mixed Grasslands had lower calcium concentrations than streams in the other two regions, the calcium concentrations for all streams in this study were far above (34.4 to 87.6 mg/L) the 8 mg/L threshold below which *Faxonius virilis* populations have been found to decline (Edwards et al., 2015). Thus, although calcium is essential to crayfish for the hardening of their exoskeleton (McLay & van den Brink, 2016), crayfish in my study likely exhibited little functional difference in response to the differences in calcium concentrations among the sampled streams. However, pH and sodium concentrations did show differences among the sampled streams that would be more likely to generate observable effects in the crayfish metabolome.

Streams in the Mixed Grassland were different from streams in the other two regions in having increased pH. Indeed, the Mixed Grassland streams had a pH range (8.43 to 8.61) starting 0.2 above the maximum of the other regions (7.71 to 8.21). pH differences, such as that exhibited by the streams in my study, have been demonstrated to impact the metabolomes of crustaceans. For example, a study on the shore crab (*Carcinus maenas*) showed that crabs kept in seawater with a pH of 8.08 for two weeks had significantly more glycine in the metabolome of their leg muscle tissue than crabs kept in seawater with a pH of 7.40 for the same amount of time (Hammer et al., 2012). My study appears to mirror this finding, as crayfish from the Mixed Grassland streams showed significantly higher abundances of glycine than crayfish from the other two regions.

Streams in the Mixed Grassland also stood apart in having far higher sodium concentrations (52.8 to 220 mg/L) than streams in the other two regions (2.9 to 22.9 mg/L). Although no studies have investigated the effect of high concentrations of sodium on the crayfish metabolome, Styrishave et al. (1995) used changing circadian rhythms in heart rate to detect stress in crayfish exposed to sodium chloride concentrations increasing from 5.3 to 1400 mg/L over 24 days. The Styrishave et al. (1995) study showed that crayfish death from mercury and copper exposure was proceeded by severe disruptions in circadian rhythms, and that crayfish began showing circadian rhythm disruptions at sodium chloride concentrations of 42 mg/L, indicating a stress response. Thus, it is likely that metabolic changes would also be detectable in crayfish facing increases in salinity, and it is surprising that there were not greater metabolic differences between the Mixed Grassland crayfish and the crayfish from the other two regions. Yet, excepting the marked increase in glycine in the Mixed Grassland crayfish, it appeared that the metabolism of the Mixed Grassland crayfish was not greatly disrupted by the high sodium concentrations in their environment.

An alternative to a lack of environmental variation explaining the low predictive ability of region and stream is that the crayfish populations have adapted to their respective environments, either through epigenetic regulation of gene expression or changes in allele frequencies. (Epi)genetic adaptions may allow crayfish living in different environments to operate close to the same metabolic optimum and thus have metabolomes that appear similar (Morgan et al., 2007). Environmental adaptation would be advantageous to crayfish populations colonizing new environments that they find physiologically challenging. Although the colonizing crayfish may initially acclimate with fast-acting, plastic responses, such as metabolic changes, these changes may be stressful to an organism, creating selective pressure in favour of those crayfish whose (epi)genetics allow their metabolism to operate closer to homeostasis in the new environment (Loria et al., 2019). Over time, the entire population may exhibit (epi)genetic changes that allow their metabolomes to operate optimally (Morgan et al., 2007). Thus, the weak regional differences among the field crayfish metabolomes in this study may reflect that all the crayfish were (epi)genetically well-adapted to their specific environments and metabolizing efficiently.

There is also evidence that genetic differentiation, possibly representing genetic adaptation to different environments, can arise rapidly among crayfish populations. For example, a study of *F. virilis* populations in the Midwestern United States found strong intraspecific genetic differentiation among crayfish from different streams (Fetzner et al., 1998), and similar results have been found in studies of European crayfish (Gouin et al., 2006; Gross et al., 2013; Vorburger et al., 2014). Further, Barnett et al. (2020) found genetic differentiation among crayfish sampled up-stream and down-stream of dams that had existed for only 40 to 110 years, indicating that genetic divergence among populations can occur quickly in evolutionary terms. Although not yet observed in crayfish, a study with *Daphnia magna* has shown that epigenetics can also lead to heritable environmental adaptations (Jeremais et al., 2018). In this study, a generation of the freshwater species *D. magna* responded to being raised in saline water with methylation patterns on the DNA strand that were retained by their descendants, raised in fresh water, three generations later (Jeremais et al., 2018). Although my study did not collect the data needed to determine if (epi)genetic adaptions had occurred among the crayfish populations, the metabolic similarities among crayfish from different environments suggests that environmental adaptation among wild populations should be investigated in future studies.

Despite the metabolic similarities among the field crayfish, metabolites were found that varied by region and stream. For instance, glycine was abundant in all the crayfish metabolomes from the Mixed Grassland. Interestingly, glycine has been identified as an amino acid that aids in osmotic regulation in crayfish and increases with the salinity of the external environment (Fujimori & Abe, 2002; Okuma & Abe, 1994). Moreover, as mentioned previously, glycine can also increase with increasing pH (Hammer et al., 2012). This linkage between osmotic regulation and acid-base regulation may be tied to the fact that the membrane proteins in crustaceans that transport sodium and chloride ions simultaneously transport hydrogen and bicarbonate ions (Wheatly & Henry, 1992). It is appropriate, therefore, that glycine was most abundant in the crayfish from the Mixed Grassland where water salinity and pH were the highest of the three regions. The correspondence between glycine abundance, salinity, and pH was also observed at the stream scale, as the stream within the Mixed Grassland with the highest salinity and pH, Weatherall Creek, also had the crayfish with the most abundant glycine levels.

Several other metabolites also varied by region and stream, although it is largely uncertain why these metabolites were more abundant in certain regions or streams than others. Crayfish in the Mid-Boreal Upland had the highest abundances of methionine, which is stored by crustaceans as a metabolic reserve in preparation for moulting (Maity et al., 2012). Crayfish from the Lake of the Woods had the highest abundances of arginine, which is stored in high concentrations in the muscle tissues of many invertebrates (Hird et al., 1986) and participates in a major energy storage pathway (Viant et al., 2001). Two additional metabolites, malonate and lactate, were important in distinguishing the crayfish metabolomes by stream. Malonate hinders the citric acid cycle by inhibiting the enzyme succinate dehydrogenase (Long et al., 1984). Lactate increases with a transition from aerobic to anaerobic metabolism (Gade, 1984) and has been noted to increase with several different types of stress, such as hypoxia (Bonvillain et al., 2012), light (Fanjul-Moles et al., 1998), temperature (Malev et al., 2010), and intense exercise (Gade, 1984). The differences in these metabolites were likely due to environmental or life history factors not measured in my study, and future studies should collect a wider array of environmental data to help understand the underlying causes of variation in these metabolites.

As a corollary of the small percentage of metabolome variability explained by region and stream, the majority of the metabolome variability was residual. The sources of the residual variability can only be speculated, but a likely source is within stream variability among individual crayfish, caused by differences in 1) social hierarchy position, 2) life-history stage, 3) disease state, and/or 4) genes. First, crayfish engage in agonistic behaviour that establishes intraspecific dominance hierarchies (Goessmann et al., 2000), and access to resources, such as food and shelter (Bergman & Moore, 2003; Martin III & Moore, 2008). Shelter, especially, is a resource that crayfish commonly fight over (Bergman & Moore, 2003; Fero et al., 2007), because it provides protection from conspecifics, predators, and environmental changes, particularly during moulting (Hazlett et al., 1974). Crayfish that lose bouts may be forced into areas of the stream where there are fewer or lower quality food resources and/or higher risk of predation. This inequality may cause crayfish metabolomes from the same stream to reflect varying energy stores and stress responses.

Second, crayfish undergo physiological cycles as part of their life history that likely cause fluctuations in their metabolome. Despite my attempts to minimize differences in life history by only collecting female crayfish of a certain size, additional physiological differences that were difficult to control for were likely responsible for some metabolic variability. In particular, crayfish are subject to a nonsynchronous moulting cycle that is under hormonal control and cyclically alters their feeding and movement behaviour (McLay & van den Brink, 2016). At sexual maturity, Cambarid crayfish, such as *Faxonius virilis*, alternately moult into sexually active and sexually inactive forms. Although this process is largely synchronous within a population, crayfish of different reproductive forms may coexist in the same stream (Guiasu & Dunham, 1998). There is also some evidence that the formation of the crayfish dominance hierarchy itself may alter the metabolome, with more dominant crayfish expressing higher abundances of serotonin (Herberholz et al., 2001). Although these physiological processes likely caused considerable within population variability, it is also possible that they are responsible for some among population variability. For instance, for either environmental or hard-wired reasons, many of the crayfish collected in the Mid-Boreal Upland may have been preparing to moult, which would explain their higher abundances of methionine.

Third, disease state is likely to be reflected in the metabolome, and crayfish can host a wide array of parasites and pathogens (Longshaw, 2016). In my study, a crayfish was harvested with starkly white muscle tissue, an indication of a microsporidian fungal infection (France & Graham, 1985). In silkworms (*Bombyx mori*), infection with microsporidia increases the production of many proteins needed for basic metabolism (Tang et al., 2020), and thus may have impacted the metabolomes of the infected crayfish in my study. Another disease that can infect crayfish, white spot syndrome virus, was investigated in white leg shrimp (*Litopenaeus vannamei*) and found to distinctly change the hepatopancreas metabolome, increasing glucose and decreasing lysine and tyrosine (Liu et al., 2015). Further studies will be needed to determine the degree to which disease can cause variability among the metabolomes within a population.

Fourth and last, within-population genetic variability could also contribute to variability among the crayfish metabolomes from a single stream. Indeed, Mathews et al. (2008)

demonstrated that *F. virilis* retrieved from the same stream can show enough genetic differentiation to be considered separate species and argued that *F. virilis* should be considered a species complex containing cryptic species. There was some potential evidence of such a response in the metabolomes of the common environment crayfish from Brokenhead River, which separated into two very distinct groups on the nMDS plot (Fig. 11), with a small group of four near the centre of the plot and the remainder at the far left of the plot. It is possible that these crayfish from the same stream reacted so differently to the common environment experiment because of distinct genetic backgrounds. Altogether, within-stream diversity in crayfish microhabitat, physiological state, and genetics could account for the residual variability among the crayfish metabolomes observed in my study.

4.2 Environmental and Genetic Influences on the Metabolome

The amount of variability among the crayfish metabolomes at the end of the common environment experiment was similar to the amount of variability among the field crayfish metabolomes, contradicting my prediction that the common environment experiment would cause a substantial decrease in variability. As well, the common environment experiment caused the crayfish metabolomes to become more distinct by region and stream (Fig. 13) and the proportion of variability explained by the region and stream scales to increase, contradicting my prediction that the crayfish metabolomes would become more similar in the common environment.

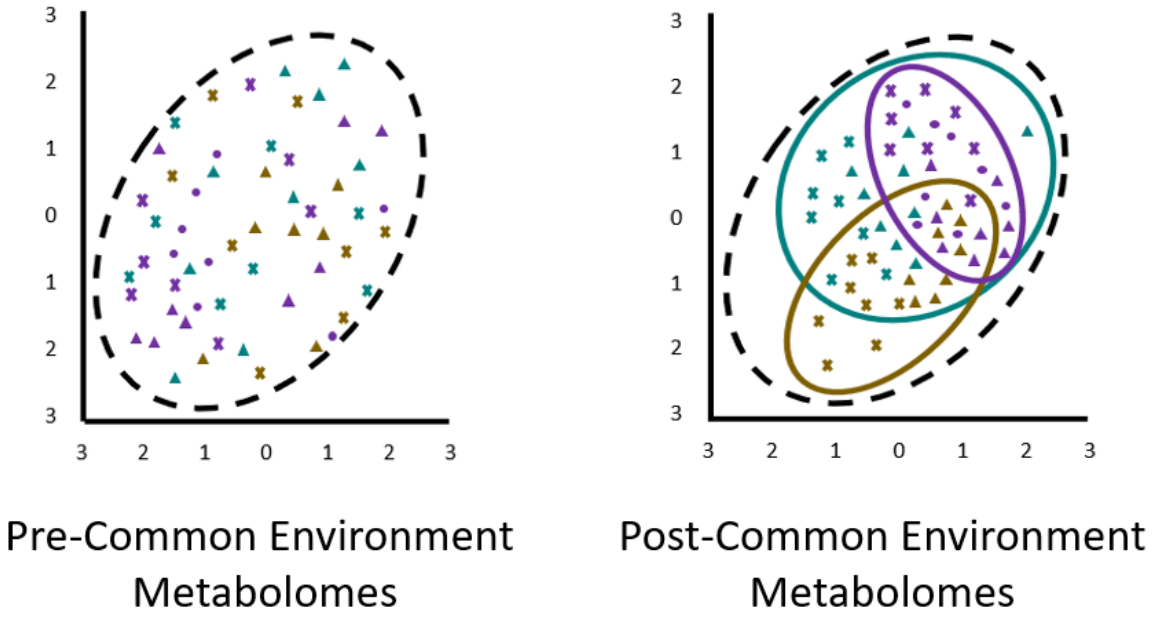


Figure 13. Representation of how the variability among crayfish metabolomes changed during the common environment experiment. Total variability remained the same, but the metabolomes formed more distinct groups by region and stream. The variability among the pre-common environment metabolomes was inferred from the variability measured among the metabolomes of the field crayfish.

The common environment experiment was designed to remove variability associated with the environment in order to isolate variability from genetic sources or sources that mimic genetics by being heritable and/or slow to respond to environmental changes, such as some epigenetic traits (Richards, 2006). Accordingly, I posited that, were metabolomic variability to be the same in the common environment as in the field, this would indicate among population variability was driven by genetics as opposed to environmental conditions (Fig. 3). My findings, in part, support this hypothesis because I did find the total variability associated with the crayfish metabolomes to be comparable between the common environment and the field populations. However, although the total variability was essentially the same, the partitioning of the variability among region, stream, and residuals changed, such that the amount of variability explained by region and stream was significantly greater, and the residual smaller, than observed for the field populations.

The decline in the amount of residual variability in the common environment indicates that the common environment was successful in removing sources of environmental variability that caused differences in crayfish metabolomes either among or within streams. Indeed,

given the small amount of variability explained by region or stream factors in the field crayfish, it is likely that the common environment experiment primarily reduced within population variability from sources associated with social hierarchy. For example, in the field, crayfish with similar social positions in different populations may share many key stressor experiences (e.g., predation risk, antagonistic interactions) and thus have more in common metabolically with each other than crayfish of the same population that occupy a different position in the social hierarchy. However, competition for resources and perceived threat of predation would have been the same for all crayfish in the common environment, eliminating associated stress responses that were likely influencing metabolomic variability among field crayfish. Thus, the removal of social stressors could have reduced commonalities between crayfish of different populations allowing population-specific attributes, such as (epi)genetic adaptations, to be more readily expressed in the metabolome. The hypothesis of elimination of social hierarchy stressors is supported by my observation that crayfish metabolomes were more similar within populations in the common environment than in the field.

In contrast, the increase in metabolomic variability explained by region and stream factors among the common environment crayfish suggests that the common environment accentuated among population differences, possibly through an interaction between crayfish (epi)genetics and the common environment. I propose two hypotheses as to the possible mechanisms behind the increased divergence of crayfish metabolomes by region and stream in the common environment. First, the common environment experiment may have produced different levels and types of stress for crayfish adapted to the environmental conditions associated with the sampled regions and streams. For instance, although the common environment was substantially different from all the stream environments in having a lower pH and conductivity, it was most different for crayfish from the Mixed Grassland, where pH and conductivity were highest. Likewise, crayfish from the other regions and streams may have found the common environment stressful to a different degree and for other reasons. Thus, the metabolomes of crayfish from different populations may have diverged because the common environment experiment removed them from their various optimal environments and placed them in a single, foreign environment, resulting in various, potentially permanent, states of stress for members of different populations.

Studies similar to my common environment experiment have found evidence for local adaptation. For example, in a common garden experiment with northern rock cress (*Arabidopsis lyrata* ssp. *petraea*), Kunin et al. (2009) concluded that adaptation to the native climate was responsible for differences they observed among rock cress metabolomes. Moreover, a follow-up study used reciprocal-transplant, common garden experiments in four different locations throughout the northern rock cress range to show that seeds from distant populations typically grew into less fit plants than the seeds from local populations, indicating that local adaptation had occurred and individuals were stressed by the novel common garden environment (Vergeer & Kunin, 2013). However, unlike my study, the studies on northern rock cress did not compare the metabolomes among populations growing in their native environments and therefore could not address whether metabolomes were more similar among populations in their native environments or in a common environment. Although I was unable to find past studies comparing metabolomes between populations in the field and in a common environment, such comparisons have been made by studies investigating the epigenome and phenotype (e.g., Clark et al., 2018; Gaspar et al., 2019; Groot et al., 2018; Shi et al., 2019). Indeed, Groot et al. (2018) found that the inflorescence height of its subject species, the plant *Scabiosa columbaria*, was significantly different between populations when plants from two populations were grown in a common garden, but that there was no difference between populations when the plants grew in their native environments. Likewise, my study observed greater similarity among the metabolomes of crayfish in their native environments than in the common environment. Although this finding is suggestive of local adaptation among crayfish, additional studies, such as a reciprocal transplant study, will be needed to test the hypothesis that local adaptation has occurred.

Second, the endpoint of my common environment experiment may have caught the crayfish populations at different stages of acclimation to their new environment. In other words, the crayfish may have been capable of acclimating to the stress of the common environment eventually, but their different adaptations caused them to adjust in different ways and at different rates. Because I sampled the crayfish metabolomes at only a single point in time, it is uncertain if the crayfish had fully acclimated to the common environment at the end of 16 days or if the process was still ongoing. It is possible, for instance, that a

common environment experiment of longer duration may have shown that the crayfish metabolomes diverged even more or that they became more similar again after their initial divergence.

My common environment experiment was run for 16 days based on previous studies that had found metabolic effects in crayfish after two weeks of exposure to different environmental conditions (Izral et al., 2018). However, common garden experiments from studies focusing on epigenetic gene regulation have used time periods of several months to years and found that full environmental acclimation can be a very gradual process (Clark et al., 2018; Gaspar et al., 2019; Groot et al., 2018) and may not be complete within a single organism's life span (Shi et al., 2019). The relatively short duration of my common environment experiment compared to the crayfish life span may mean that much epigenetic gene regulation that may have eventually changed, did not have time to adjust. Thus, the effects I could observe in this study were likely restricted to metabolic and epigenetic mechanisms that could adjust quickly to the common environment. In future studies, common environment experiments of longer duration that sample crayfish at multiple time points should be used to clarify how crayfish metabolomes acclimate to a common environment.

4.3 Implications for Bioassessment

My study indicated that the crayfish metabolome was conserved across environmental gradients, and this finding has implications for using the crayfish metabolome as a bioassessment tool. In particular, my findings apply to regional applications of the Reference Condition Approach (RCA), whereby the reference condition is ascertained by finding the range of values for a biological indicator at multiple sites that are considered to be in the best available ecological condition. Thus, a key assumption of the RCA is that test sites will not be significantly different from reference sites in the absence of stressors (Bowman & Somers, 1999). A corollary to this assumption is that reference sites used in RCA must consist of a single population with similar biotic conditions. This requirement often necessitates that regionally specific reference conditions be established as many biological indicators that rely on descriptions of taxonomic structure (e.g., benthic invertebrate community composition) vary substantially from region to region in

association with biogeographic gradients. The need to establish unique reference conditions is a significant impediment to global implementation bioassessment protocols and cross regional comparison of bioassessment datasets, as development of reference conditions requires significant expenditure of resources, and reference sites may be absent from regions where human activity is pervasive.

My findings that the metabolome of the crayfish, *Faxonius virilis* is comparable among reference sites in different regions supports hypotheses about the portability of the metabolome as an indicator. Indeed, Pomfret et al. (2020) hypothesized that metabolomes would be similar among populations of the same species across large spatial scales, removing the need for establishment of regionally specific reference conditions, because the metabolome reflects organism function as it reflects the rate of metabolic processes. As such, it has been argued that the metabolome is a functional indicator. Functional indicators detect changes in the patterns or rates of processes that are relevant to the health of an ecosystem, such as primary production or decomposition (Bunn & Davies, 2000; Gessner & Chauvet, 2002; Palmer & Febria, 2012; Von Schiller et al., 2017). Because the results of these processes are largely similar regardless of geography, functional indicators are predicted to be less dependent on regionally specific references (Gessner & Chauvet, 2002). However, as this hypothesis has received limited testing, our findings provide some of the first evidence that reference conditions derived from functional indicators, and the metabolome more specifically, can be applied across regions that span environmental gradients.

Although my findings indicate that metabolomes of *Faxonius virillis* from reference sites have the potential to be treated as a single reference condition across much of Canada's prairie provinces, my study was not designed to test the ability of the crayfish metabolome to distinguish between impacted and reference streams. Specifically, because my study did not include crayfish metabolomes from impacted streams, I could not determine whether the total variability among crayfish metabolomes from reference streams was sufficiently small to readily detect deviations from reference conditions caused by increasing exposure to anthropogenic stressors. Thus, future studies are needed to contextualize the utility of the reference condition "space" my study has described by testing the extent to which

crayfish metabolomes from impacted streams may fall within its boundaries. An inability of the crayfish metabolome to distinguish between reference streams and impacted streams seems unlikely given previous studies that have been able to clearly separate organisms from reference sites and impacted sites by their metabolomes (Cappello et al., 2017; Fernandez-Cisnal et al., 2018; Gago-Tinoco et al., 2014; Melvin et al., 2018; Skelton et al., 2014; Watanabe et al., 2015). However, given that these past studies used only a single reference site, thus leaving the reference space undefined, it is critical that this knowledge gap be addressed. Therefore, future studies should include crayfish metabolomes from several impacted streams along with crayfish metabolomes from several reference streams to determine how separate these two groups are and the relative amount of variability within each group. As well, additional studies should include crayfish metabolomes from streams impacted by different degrees of a stressor and streams impacted by different types of stressors to examine if the degree or type of stress affects the ability to distinguish impacted streams from reference streams.

My findings indicated that crayfish can adapt to relatively high levels of some stressors, possibly through (epi)genetic adaptations. This result raises the question of whether the metabolomes of these crayfish populations would be sensitive to human impact associated with stressors that the crayfish have adapted too. For example, the limited impact of the high levels of sodium observed in the Mixed Grassland streams on the metabolomes of the resident crayfish suggests that these populations may not be good indicators of the impacts of road salt (NaCl), which has been shown to raise sodium levels in nearby waterways (Cooper et al., 2014; Rosenberry et al., 1997). As a result, it may be important for bioassessment studies to identify the key stressors of interest and ensure that reference populations are selected that come from streams that are not unusually, but naturally, rich in those stressors to be assessed.

Another way in which the crayfish reference metabolome needs to be examined to ensure its applicability as a bioassessment tool is in its consistency over time. Although my study found that the crayfish metabolomes from different regions and streams were largely similar, this pattern could change from year to year. For instance, a very dry year may accentuate differences among streams. Crayfish from streams in the Mixed Grassland,

where evaporation is high, may be especially likely to diverge from crayfish in other streams in a dry year, as solute concentrations may increase to even more extreme levels. Such interannual variability could increase the range of variation among reference sites and, furthermore, may decrease the amount of similarity among regions and/or streams. Future studies should thus measure crayfish metabolomes from the same reference streams used in my study over several years to determine how well the reference space is conserved over time.

Additionally, future studies should try to identify the sources of the residual variability among crayfish metabolomes observed in my study by investigating the effects of social hierarchy, life history stage, disease, and genetics on the crayfish metabolome. The relative influence of these sources of variability on the crayfish metabolome will determine whether controlling for them will enhance the power of the metabolome as a bioassessment tool. First, to investigate the effect that a crayfish's social hierarchy position has on its metabolome and the noise this creates for the detection of other stressors, a study combining a stressor and intraspecific competition could be completed. Crayfish could be divided into several groups of three to five crayfish, with half of the groups exposed to an environmentally relevant concentration of a stressor (e.g., a metal contaminant) and the other half of groups left unexposed. The crayfish of each group could be watched to determine the social hierarchy that develops, as has been done in previous studies (Fero et al., 2007; Martin III & Moore, 2008). After an exposure period sufficient for the stressor to impact the metabolome and for social hierarchies to become established, the metabolomes of each crayfish could be measured to determine if the crayfish metabolomes group more clearly based on contaminant exposure or social position. If the social position of the crayfish does not obscure the contaminant exposure groups, then controlling for social position in bioassessment studies is likely unnecessary.

Second, a similar study could be used to investigate the effects of life-history stage, particularly the moulting cycle, on the crayfish metabolome. Crayfish of the same age, preferably of the same brood, could be divided into two groups. All environmental conditions would be kept the same for both groups except that one group would be exposed to environmentally relevant concentrations of a contaminant, such as a metal, that crayfish

tolerate well over extended periods of time (Kouba et al., 2010). By harvesting crayfish from both groups at regular intervals, changes in the metabolome over the five stages of the moult cycle (McLay & van den Brink, 2016) could be measured and compared between groups. Morphological observations and hemolymph concentrations of ecdysteroids, the hormones that control the moulting cycle, could be used to determine which stage of the moult cycle each crayfish was in at the time of harvest (Willig & Keller, 1973; Jegla et al., 1983). Thus, the crayfish metabolomes could be compared to see if they group more distinctly by contaminant exposure or by moulting cycle stage.

Third, to investigate the effects of disease on the crayfish metabolome, crayfish infected with a known disease, such as the microsporidium fungal infection observed in my study, would need to be obtained. Measuring the metabolome of these crayfish could identify a "metabolite fingerprint" characteristic of the disease that could be used to flag crayfish infected with the disease in future studies. In addition, the metabolomes of the infected crayfish could be measured after exposure to a contaminant or without exposure to determine if contaminant exposure or the infection was the greater influence on the metabolome. Such investigations would have to be repeated for different diseases. However, insights on how the crayfish metabolome would react to certain diseases may be drawn from studies in related crustaceans, and research on the metabolome's reaction to disease is becoming more common, especially in aquaculture research (Afaro & Young, 2018).

Fourth, genetic variability, both among and within crayfish populations, needs be quantified and its relationship to the metabolome established. Considering that cryptic species have been found within *F. virilis* and have been theorized to be widespread among North American crayfish (Mathews et al., 2008), genetics may be an important source of variability among crayfish metabolomes. Additional tissue samples were taken from the crayfish used in my study and could be used to determine their genetic relationships. This genetic information could then be combined with the metabolomic information on each crayfish to determine if genetic variability aligns with metabolomic variability and if it explains some of the residual variability observed in my study. It would be interesting to know, for example, if genetics was responsible for the metabolomes of crayfish from the

Brokenhead River separating into two groups during the common environment experiment (Fig. 11). If genetics was responsible for this differential reaction to the stress of the common environment, it may demonstrate the need to control for different genetic backgrounds in bioassessment studies.

4.4 Summary and Conclusions

In my study, crayfish metabolomes were sampled from seven streams divided amongst three distinct regions. Despite the environmental differences among the regions and streams, the region and stream scales predicted little of the variability among the metabolomes of crayfish residing in their native environments, as most of the variability among metabolomes was among individuals. As well, similarities among the crayfish metabolomes displayed on nMDS plots did not show groupings based on region or stream. These findings show that the crayfish metabolome was conserved across the environmental gradients included in this study and that the several crayfish populations could be used to create a single reference condition population. Thus, my study indicates that regionally specific reference metabolomes would be of no advantage to bioassessment studies employing the metabolome.

My study also performed a common environment experiment with crayfish collected from the same seven streams used in the field study. The amount of variability among the metabolomes of the common environment crayfish was similar to that seen in the field crayfish, indicating that genetics and/or epigenetics plays an important role in dictating differences among crayfish metabolomes. As well, the higher proportion of variability explained by the region and stream scales among the common environment crayfish indicated that the crayfish reacted differently to the common environment based on their region and stream of origin. This differential reaction may suggest that the crayfish were adapted to their native environments, and future studies should investigate if local adaptation is the mechanism allowing crayfish metabolomes to appear similar across environmental gradients.

My study indicated that regional landscape differences are not an impediment to creating a cross-regional reference crayfish metabolome. The next research step in investigating the

metabolome as a bioassessment tool will be to determine how clearly metabolomes from impacted streams can be differentiated from metabolomes from reference streams. Although further work needs to be done, the metabolome has thus far shown promise as a bioassessment tool.

References

Afaro, A. C. & Young, T. (2018). Showcasing metabolomic applications in aquaculture: a review. *Reviews in Aquaculture, 10,* 135-152. doi: 10.1111/raq.12152

Allan, J. D. (2004). Influence of land use and landscape setting on the ecological status of rivers. *Limnetica, 23*(3–4), 187–198. doi:10.1146/annurev.ecolsys.35.120202.110122

Aru, V., Engelsen, S. B., Savorani, F., Culurgioni, J., Sarais, G., Atzori, G., Cabiddu, S., & Marincola, F. C. (2017). The Effect of Season on the Metabolic Profile of the European Clam *Ruditapes decussatus* as Studied by ^{1}H-NMR Spectroscopy. *Metabolites, 7*(36), 1-13. doi:10.3390/metabo7030036

Bahamonde, P. A., Feswick, A., Isaacs, M. A., Munkittrick, K. R., & Martyniuk, C. J. (2016). Defining the role of omics in assessing ecosystem health: Perspectives from the Canadian environmental monitoring program. *Environmental Toxicology and Chemistry*, 35(1), 20-35. doi: 10.1002/etc.3218

Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2004). *Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Barbaresi, S., Santini, G., Tricarico, E., & Gherardi, F. (2004). Ranging behaviour of the invasive crayfish, *Procambarus clarkii* (Girard). *Journal of Natural History, 38,* 2821-2832. doi: 10.1080/00222930410001663308

Barnett, Z. C., Adams, S. B., Ochs, C. A., & Garrick, R. C. (2020). Crayfish populations genetically fragmented in streams impounded for 36–104 years. *Freshwater Biology, 00,* 1-18. doi: 10.1111/fwb.13466

Bearden, D. W. (2012). Environmental Metabolomics. *eMagRes 1*(2). doi: 10.1002/9780470034590.emrstm1256

Bergman, D. A. & Moore, P. A. (2003). Field Observations of Intraspecific Agonistic Behavior of Two Crayfish Species, *Orconectes rusticus* and *Orconectes virilis*, in Different Habitats. *The Biology Bulletin, 205,* 26-35.

Berrill, M., Hollett, L., Margosian, A., & Hudson, J. (1985). Variation in tolerance to low environmental pH by the crayfish *Orconectes rusticus, 0. propinquus,* and *Cambarus robustus*. *Canadian Journal of Zoology, 63,* 2586-2589.

Bollard, M. E., Stanley, E. G., Lindon, J. C., Nicholson, J. K., & Holmes, E. (2005). NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR in Biomedicine, 18,* 143-162. doi:10.1002/nbm.935

Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in Aquatic Insect Biomonitoring: A Comparative Analysis of Recent Approaches. *Annual Review of Entomology, 51*(1), 495–523. doi: 10.1146/annurev.ento.51.110104.151124

Bonvillain, C. P., Rutherford, D. A., Kelso, W. E., & Green, C. C. (2012). Physiological biomarkers of hypoxic stress in red swamp crayfish *Procambarus clarkii* from field and laboratory experiments. Comparative Biochemistry and Physiology, Part A, 163, 15-21. doi:10.1016/j.cbpa.2012.04.015

Bowman, M. F. & Somers, K. M. (2005). Considerations when Using the Reference Condition Approach for Bioassessment of Freshwater Ecosystems. *Water Quality Research Journal of Canada, 40*(3), 347-360.

Bubb, D. H., Lucas, M. C., & Thom, T. J. (2002). Winter movements and activity of signal crayfish *Pacifastacus leniusculus* in an upland river, determined by radio telemetry. *Hydrobiologia, 483*, 111-119.

Bundy, J. G., Davey, M. P., & Viant, M. R. (2009). Environmental metabolomics: A critical review and future perspectives. *Metabolomics, 5*, 3-21. doi: 10.1007/s11306-008-0152-0

Bunn, S. E. & Davies, P. M. (2000). Biological processes in running waters and their implications for the assessment of ecological integrity. *Hydrobiologia, 422/423*, 61-70.

Carpenter, S. R., Stanley, E. H., & Vander Zanden, M. J. (2011). State of the World's Freshwater Ecosystems: Physical, Chemical, and Biological Changes. Annual Review of Environment and Resources, 36, 75-99. doi: 10.1146/annurev-environ-021810-094524

Cappello, T, Maisano, M., Mauceri, A., & Fasulo, S. (2017). ^{1}H NMR-based metabolomics investigation on the effects of petrochemical contamination in posterior adductor muscles of caged mussel *Mytilus galloprovincialis*. *Ecotoxicology and Environmental Safety, 142*, 417-422. doi: 10.1016/j.ecoenv.2017.04.040

Chong, J., Wishart, D.S. and Xia, J. (2019). Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics, 68*(1). doi: 10.1002/cpbi.86

Clark, M. S., Thorne, M. A. S., King, M., Hipperson, H., Hoffman, J. I., & Pedk, L. S. (2018). Life in the intertidal: Cellular responses, methylation and epigenetics. *Functional Ecology, 32*, 1982-1994. doi: 10.1111/1365-2435.13077

Connon, R. E., Geist, J., & Werner, I. (2012). Effect-based tools for monitoring and predicting the ecotoxicological effects of chemicals in the aquatic environment. *Sensors, 12*, 12741-12771. doi: 10.3390/s120912741

Coombes, K. R. & Wang, M. (2019). PCDimension: Finding the Number of Significant Principal Components. R package version 1.1.11. https://CRAN.R-project.org/package=PCDimension

Cooper, C. A., Mayer, P. M., & Faulkner, B. R. (2014). Effects of road salts on groundwater and surface water dynamics of sodium and chloride in an urban restored stream. *Biogeochemistry, 121*, 149-166. doi: 10.1007/s10533-014-9968-z

Del Ramo, J., Diaz-Mayans, J., Torreblanca, A., & Nunez, A. (1987). Effects of Temperature on the Acute Toxicity of Heavy Metals (Cr, Cd, and Hg) to the Freshwater Crayfish, *Procambarus clarkfi* (Girard). *Bulletin of Environmental Contamination and Toxicology, 38,* 736-741.

Ecological Stratification Working Group. (1995). A National Ecological Framework for Canada. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, Ottawa/Hull.

Edwards, B. A., Jackson, D. A., & Somers, K. M. (2015). Evaluating the effect of lake calcium concentration on the acquisition of carapace calcium by freshwater crayfish. *Hydrobiologia, 744*, 91-100. doi:10.1007/s10750-014-2059-2

Ekman, D. R., Ankley, G. T., Blazer, V. S., Collette, T. W., Garcia-Reyero, N., Iwanowicz, L. R., Jorgenson, Z. G., Lee, K. E., Mazik, P. M., Miller, D. H., Perkins, E. J., Smith, E. T., Tietge, J. E., & Villeneuve, D. L. (2013). Biological Effects–Based Tools for Monitoring Impacted Surface Waters in the Great Lakes: A Multiagency Program in Support of the Great Lakes Restoration Initiative. *Environmental Practice, 15*(4), 409-426. doi: 10.10170S1466046613000458

Fanjul-Moles, M. L., Bosques-Tistler, T., Prieto-Sagredo, J., Castanon-Cervantes, O., & Fernandez-Rivera-Rio, L. (1998). Effect of Variation in Photoperiod and Light Intensity on Oxygen Consumption, Lactate Concentration and Behavior in Crayfish *Procambarus clarkii* and *Procambarus digueti*. *Comparative Biochemistry and Physiology, 119A*(1), 263-269.

Fernandez-Cisnal, R., Garcia-Sevillano, M. A., Garcia-Barrera, T., Gomez-Ariza, J. L., & Abril, N. (2018). Metabolomic alterations and oxidative stress are associated with environmental pollution in *Procambarus clarkii*. *Aquatic Toxicology*, 205, 76-88. doi: 10.1016/j.aquatox.2018.10.005

Fero, K., Simon, J. L., Jourdie, V., & Moore, P. A. (2007). Consequences of social dominance on crayfish resource use. *Behaviour, 144*, 61-82.

Fetzner Jr, J. W., Sheehan, R. J., & Seeb, L. W. (1997). Genetic implications of broodstock selection for crayfish aquaculture in the Midwestern United States. *Aquaculture, 154,* 39- 55.

France, R. L. & Graham, L. (1985). Increased microsporidian parasitism of the crayfish *orconectes virilis* in an experimentally acidified lake. *Water, Air, and Soil Pollution, 26,* 129-136.

Fujimori, T. & Abe, H. (2002). Physiological roles of free D- and L-alanine in the crayfish *Procambarus clarkii* with special reference to osmotic and anoxic stress responses. *Comparative Biochemistry and Physiology Part A, 131*, 893-900,

Gade, G. (1984). Effects of oxygen deprivation during anoxia and muscular work on the energy metabolism of crayfish *Orconectes limosus*. *Comparative Biochemistry and Physiology, 77A*(3), 495-502.

Gago-Tinoco, A., Gonzalez-Dominguez, R., Garcia-Barrera, T., Blasco-Moreno, J., Bebianno, M. J., & Gomez-Ariza, J. L. (2014). Metabolic signatures associated with environmental pollution by metals in Doñana National Park using *P. clarkii* as bioindicator. *Environmental Science and Pollution Research, 21*, 13315-13323. doi: 10.1007/s11356-014-2741-y

Gaspar, B. Bossdorf, O., & Durka, W. (2019). Structure, stability and ecological significance of natural epigenetic variation: a large-scale survey in *Plantago lanceolate. New Phytologist, 221*, 1585-1596. doi: 10.1111/nph.15487

Gessner, M. O. & Chauvet, E. (2002). A case for using litter breakdown to assess functional stream integrity. *Ecological Applications*, *12*(2), 498-510.

Gidman, E. A., Jones, M. L. M., Bussell, J. A., Malham, S. K., Reynolds, B., Seed, R., Causton, D. R., Johnson, H. E., & Gwynn-Jones, D. (2007). A methodology for screening haemolymph of intertidal mussels, *Mytilus edulis*, using FT-IR spectroscopy as a tool for environmental assessment. *Metabolomics, 3*(4), 465-473. doi: 10.1007/s11306-007-0060-8

Goessmann, C., Hemelrijk, C., & Huber, R. (2000). The formation and maintenance of crayfish hierarchies: behavioral and self-structuring properties. *Behavioral Ecology and Sociobiology, 48*, 418-428.

Gouin, N., Grandjean, F., & Souty-Grosset, C. (2006). Population genetic structure of the endangered crayfish *Austropotamobius pallipes* in France based on microsatellite variation: biogeographical inferences and conservation implications. *Freshwater Biology, 51*, 1369-1387. doi:10.1111/j.1365-2427.2006.01570.x

Groot, M. P., Wagemaker, N., Ouboug, N. J., Verhoeven, K. J. F., & Vergeer, P. (2018). Epigenetic population differentiation in field-and common garden-grown *Scabiosa columbaria* plants. *Ecology and Evolution, 8*, 3505-3517. doi: 10.1002/ece3.3931

Gross, R., Palm, S., Koiv, K., Prestegaard, T., Jussila, J., Paaver, T., Geist, J., Kokko, H., Karjalainen, A., & Edsman, L. (2013). Microsatellite markers reveal clear geographic structuring among threatened noble crayfish (*Astacus astacus*)

populations in Northern and Central Europe. *Conservation Genetics, 14*, 809-821. doi: 10.1007/s10592-013-0476-9

Guiasu, R. C. & Dunham, D. W. (1998). Inter-form agonistic contests in male crayfishes, *Cambarus robustus* (Decapoda, Cambaridae). *Invertebrate Biology, 117*(2), 144-154.

Hammer, K. M., Pedersen, S. A., & Storseth, T. R. (2012). Elevated seawater levels of CO2 change the metabolic fingerprint of tissues and hemolymph from the green shore crab *Carcinus maenas*. *Comparative Biochemistry and Physiology, Part D, 7*, 292-302. doi:10.1016/j.cbd.2012.06.001

Hawkins, C. P., Olson, J. R., & Hill, R. A. (2010). The reference condition: predicting benchmarks for ecological and water-quality assessments. *Journal of the North American Benthological Society, 19*(1), 312-343. doi: 10.1899/09-092.1

Hazlett, B., Rittschof, D., & Rubenstein, D. (1974). Behavioral Biology of the Crayfish *Orconectes virilis* I. Home Range. *The American Midland Naturalist, 92*(2), 301-319.

Herberholz, J., Issa, F. A., & Edwards, D. H. (2001). Patterns of Neural Circuit Activation and Behavior during Dominance Hierarchy Formation in Freely Behaving Crayfish. *The Journal of Neuroscience, 21*(8), 2759-2767.

Hines, A., Oladiran, G. S., Bignell, J. P., Stentiford, G. D., & Viant, M. R. (2007). Direct sampling of organisms from the field and knowledge of their phenotype: Key recommendations for environmental metabolomics. *Environmental Science and Technology, 41*(9), 3375–3381. doi: 10.1021/es062745w

Hird, F. J. R., Cianciosi, S. C., & McLean, R. M. (1986). Investigations on the origin and metabolism of the carbon skeleton of ornithine, arginine and proline in selected animals. *Comparative Biochemistry and Physiology, 83B*(1), 179-184.

Izral, N. M., Brua, R. B., Culp, J. M., & Yates, A. G. (2018). Developing metabolomics-based bioassessment: crayfish metabolome sensitivity to food and dissolved oxygen stress. Environmental Science and Pollution Research, 25, 6184–36193. doi: 10.1007/s11356-018-3518-5

Jegla, T. C., Ruland, C., Kegel, G., & Keller, R. (1983). The role of the Y-organ and cephalic gland in ecdysteroid production and the control of molting in the crayfish, *Orconectes limosus*. *Journal of Comparative Physiology B, 152*, 91-95.

Jeppe, K. J., Kouremenos, K. A., Townsend, K. R., MacMahon, D. F., Sharley, D., Tull, D. L., Hoffmann, A. A., Pettigrove, V., & Long, S. M. (2017). Metabolomic Profiles of a Midge (*Procladius villosimanus*, Kieffer) Are Associated with Sediment Contamination in Urban Wetlands. *Metabolites, 7*(64), 1-16. doi:10.3390/metabo7040064

Johnson, M. F., Rice, S. P., & Reid, I. (2010). Topographic disturbance of subaqueous gravel substrates by signal crayfish (*Pacifastacus leniusculus*). *Geomorphology, 123*, 269-278. doi: 10.1016/j.geomorph.2010.07.018

Johnson, R. K., Wiederholm, T., & Rosenberg, D. M. (1993). Freshwater Biomonitoring Using Individual Organisms, Populations, and Species Assemblages of Benthic Macroinvertebrates. In D. M. Rosenberg & V. H. Resh (Eds.), *Freshwater Biomonitoring and Benthic Macroinvertebrates* (pp. 75). New York, NY: Routledge, Chapman & Hall, Inc.

Jeremais, G., Barbosa, J., Marques, S. M., Asselman, J., Goncalves, F. J. M., & Pereira, J. L. (2018). Synthesizing the role of epigenetics in the response and adaptation of species to climate change in freshwater ecosystems. *Molecular Ecology, 27*, 2790-2806. doi: 10.1111/mec.14727

Kouba, A., Buric, M., & Kozak, P. (2010). Bioaccumulation and Effects of Heavy Metals in Crayfish: A Review. *Water, Air, & Soil Pollution, 211*, 5-16. doi: 10.1007/s11270-009-0273-8

Kozak, P., & Kuklina, I. (2016). Crayfish as Tools of Water Quality Monitoring. In T. Kawai, Z. Faulkes, & G. Scholtz (Eds.), *Freshwater Crayfish: A Global Overview* (pp. 275). Boca Raton, FL: CRC Press, Taylor and Francis Group.

Kunin, W. E., Vergeer, P., Kenta, T., Davey, M. P., Burke, T., Woodward, F. I., Quick, P., Mannarelli, M. E., Watson-Haigh, N. S., & Butlin, R. (2009). Variation at range margins across multiple spatial scales: environmental temperature, population genetics and metabolomic phenotype. *Proceedings of the Royal Society B, 276*, 1495-1506. doi:10.1098/rspb.2008.1767

Liu, P., Liu, Q., Wu, Y., & Jie, H. (2015). A pilot metabolic profiling study in hepatopancreas of *Litopenaeus vannamei* with white spot syndrome virus based on 1H NMR spectroscopy. *Journal of Invertebrate Pathology, 124*, 51-56. doi: 10.1016/j.jip.2014.09.008

Long, S. D., Rodrick, G. E., & Friedl, F. E. (1984). Succinate dehydrogenase in various tissues of *Anodonta Couperiana*, *Elliptio buckleyi* and *Mercenaria campechiensis* (Mollusca: Bivalvia). *Comparative Biochemistry and Physiology, 78B*(2), 467-472.

Longshaw, M. (2016). Parasites, Commensals, Pathogens and Diseases of Crayfish. In M. Longshaw & P. Stebbing (Eds.), *Biology and Ecology of Crayfish* (pp. 171) Boca Raton, FL: CRC Press, Taylor and Francis Group.

Lankadurai, B. P., Nagato, E. G., & Simpson, M. J. (2013). Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environmental Reviews, 21*, 180-205. doi: 10.1139/er-2013-0011

Loria, A., Cristescu, M. E., & Gonzalez, A. (2019). Mixed evidence for adaptation to environmental pollution. *Evolutionary Applications, 12*, 1259-1273. doi: 10.1111/eva.12782

Lower, W. R. & Kendall, R. J. (1990). Sentinel Species and Sentinel Bioassay. In J. F. McCarthy & L. R. Shugart (Eds.), *Biomarkers of Environmental Contamination* (pp. 309). Boca Raton, FL: CRC Press, Lewis Publishers.

Maity, S., Jannasch, A., Adamec, J., Gribskov, M., Nalepa, T., Hook, T. O., & Sepulveda, M. S. (2012). Metabolite profiles in starved Diporeia Spp. Using liquid chromatography-mass spectrometry (LS-MS) based metabolomics. *Journal of Crustacean Biology, 32*(2), 239-248. doi:10.1163/193724011X615578

Malev, O., Srut, M., Maguire, I., Stambuk, A., Ferrero, E. A., Lorenzon, S., & Klobucar, G. I. V. (2010). Genotoxic, physiological and immunological effects caused by temperature increase, air exposure or food deprivation in freshwater crayfish *Astacus leptodactylus*. Comparative Biochemistry and Physiology, Part C, 152, 433-443. doi: 10.1016/j.cbpc.2010.07.006

Martin III, A. L. & Moore, P. A. (2008). The Influence of Dominance on Shelter Preference and Eviction Rates in the Crayfish, *Orconectes rusticus*. *Ethology, 114*, 351-360. doi: 10.1111/j.1439-0310.2008.01473.x

Martyniuk, C. J. (2018). Are we closer to the vision? A proposed framework for incorporating omics into environmental assessments. *Environmental toxicology and pharmacology, 59*, 87-93. doi: 10.1016/j.etap.2018.03.005

Mathews, L. M., Adams, L., Anderson, E., Basile, M., Gottardi, E., & Buckholt, M. A. (2008). Genetic and morphological evidence for substantial hidden biodiversity in a freshwater crayfish species complex. *Molecular Phylogenetics and Evolution, 48*, 126-135. doi: 10.1016/j.ympev.2008.02.006

Mayer, F. L., Versteeg, D. J., McKee, M. J., Folmar, L. C., Graney, R. L., McCume, D. C., & Rattner, B. A. (1992). Physiological and Nonspecific Biomarkers. In R. J. Huggett, R. A. Klmerle, P. M. Mehrle, Jr., & H. L. Bergman (Eds.), *Biomarkers: Biochemical, Physiological, and Histological Markers of Anthropogenic Stress* (pp. 6). Boca Raton, FL: CRC Press, Taylor and Francis Group.

McLay, C. L. & van den Brink, A. M. (2016). Crayfish Growth and Reproduction. In M. Longshaw & P. Stebbing (Eds.), *Biology and Ecology of Crayfish* (pp. 62) Boca Raton, FL: CRC Press, Taylor and Francis Group.

Melvin, S. D., Lanctôt, C. M., Doriean, N. J. C., Carroll, A. R., & Bennett, W. W. (2018). Untargeted NMR-based metabolomics for field-scale monitoring: Temporal reproducibility and biomarker discovery in mosquitofish (*Gambusia holbrooki*) from a metal(loid)-contaminated wetland. *Environmental Pollution, 243*, 1096–1105. doi: 10.1016/j.envpol.2018.09.071

Meteorological Service of Canada (n. d.). *1971-2000 Climate Normals & Averages* [Raster Datasets]. Environment and Climate Change Canada [Distributor]. Retrieved from https://climate.weater.gc.ca/climate_normals/index_e.html

Miller, M. G. (2007). Environmental Metabolomics: A SWOT Analysis (Strengths, Weaknesses, Opportunities, and Threats). *Journal of Proteome Research, 6*, 2, 540-545.

Morgan, A. J., Kille, P., & Sturzenbaum, S. R. (2007). Microevolution and Ecotoxicology of Metals in Invertebrates. *Environmental Science and Technology, 41*, 1085-1096.

Norris, R. H. & Hawkins, C. P. (2000). Monitoring river health. *Hydrobiologia, 435*, 5-17.

Okuma, E., & Abe, H. (1994). Total D-amino and other free amino acids increase in the muscle of crayfish during seawater acclimation. *Comparative Biochemistry and Physiology, 109A*(1), 191-197.

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology, 22*, 5983-5999. doi: 10.1111/mec.12561

Palmer, M. A. & Febria, C. M. (2012). The Heartbeat of Ecosystems. *Science, 336*, 1393-1394. doi: 10.1126/science.1223250

Pomfret, S. M., Brua, R. B., Izral, N. M., & Yates, A. G. (2020). Metabolomics for biomonitoring: An evaluation of the metabolome as an indicator of aquatic ecosystem health. *Environmental Reviews*, *28*(1): 89-98. doi: 10.1139/er-2019-0003

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richards, E. J. (2006). Inherited epigenetic variation — revisiting soft inheritance. *Nature Reviews Genetics, 7*, 395-401.

Robinson, C. T., Tockner, K., & Ward, J. V. (2002). The fauna of dynamic riverine landscapes. *Freshwater Biology, 47*, 661-677.

Rodriguez, C. F., Becares, E., & Fernandez-Alaez, M. (2003). Shift from clear to turbid phase in Lake Chozas (NW Spain) due to the introduction of American red swamp crayfish (*Procambarus clarkii*). *Hydrobiologia, 506-509*, 421-426.

Roldan, B. M. & Shivers, R. R. (1987). The uptake and storage of iron and lead in cells of the crayfish (*Orconectes propinquus*) hepatopancreas and antennal gland. *Comparative Biochemistry and Physiology*, *86C*(1), 201-214.

Rosenberry, D. O., Bukaveckas, P. A., Buso, D. C., Likens, G. E., Shapiro, A. M., & Winter, T. C. (1999). *Water, Air, and Soil Pollution, 109*, 179-206.

Schoenfuss, H. L. & Wang, R. L. (2015). In "The Challenge: Real-world application of 'omics endpoints." D. C. Bencic (Ed.) *Environmental Toxicology and Chemistry, 34*(4), 700-706. doi: 10.1002/etc.2839

Shi, W., Chen, X., Gao, L., Xu, C. Y., Ou, X., Bossdorf, O., Yang, J., & Geng, Y. (2019). Transient Stability of Epigenetic Population Differentiation in a Clonal Invader. *Frontiers in Plant Science, 9*, 1851. doi: 10.3389/fpls.2018.01851

Simmons, D. B. D., Benskin J. P., Cosgrove J. R., Duncker B. P., Ekman D. R., Martyniuk C. J., & Sherry J. P. (2015). Omics for aquatic ecotoxicology: Control of extraneous variability to enhance the analysis of environmental effects. *Environmental Toxicology and Chemistry*, *34*(8), 1693-1704. doi: 10.1002/etc.3002

Skelton, D. M., Ekman, D. R., Martinović-Weigelt, D., Ankley, G. T., Villeneuve, D. L., Teng, Q., & Collette, T. W. (2014). Metabolomics for in situ environmental monitoring of surface waters impacted by contaminants from both point and nonpoint sources. *Environmental Science and Technology, 48*, 2395–2403. doi: 10.1021/es404021f

Souty-Grosset, C. & Fetzner Jr, J. W. (2016). Taxonomy and Identification. In M. Longshaw & P. Stebbing (Eds.), *Biology and Ecology of Crayfish* (pp. 1). Boca Raton, FL: CRC Press, Taylor and Francis Group.

Statzner, B., Peltret, O., & Tomanova, S. (2003). Crayfish as geomorphic agents and ecosystem engineers: effect of a biomass gradient on baseflow and flood-induced transport of gravel and sand in experimental streams. *Freshwater Biology, 48*, 147–163.

Stoddard, J. L., Larsen, D. P., Hawkins, C. P., Johnson, R. K., & Norris, R. H. (2006). Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications, 16*(4), 1267-1276.

Styrishave, B., Rasmussen, A. D., & Depledge, M. H. (1995). The Influence of Bulk and Trace Metals on the Circadian Rhythm of Heart Rates in Freshwater Crayfish, *Astacus astacus*. *Marine Pollution Bulletin*, *31*(1-3), 87-92.

Tang, X., Zhang, Y., Zhou, Y., Liu, R., & Shen, Z. (2020). Quantitative proteomic analysis of ovaries from *Nosema bombycis*-infected silkworm (*Bombyx mori*). Journal of Invertebrate Pathology, 172, 107355. doi: 10.1016/j.jip.2020.107355

Taylor, C. A., Schuster, G. A., Cooper, J. E., DiStefano, R. J., Eversole, A. G., Hamr, P., Hobbs III, H. H., Robison, H. W., Skelton, C. E., Thoma, R. F. (2007). A Reassessment of the conservation Status of Crayfishes of the United States and Canada after 10+ Years of Increased Awareness. *Fisheries, 32*(8), 372-389.

Taylor, N. S., Gavin, A., Viant, M. R. (2018). Metabolomics Discovers Early-Response Metabolic Biomarkers that Can Predict Chronic Reproductive Fitness in Individual *Daphnia magna*. *Metabolites, 8*(42), 1-19. doi:10.3390/metabo8030042

Van Petegem, K. H. P., Renault, D., Stoks, R., & Bonte, D. (2016). Metabolic adaptations in a range-expanding arthropod. *Ecology and Evolution, 6* (18), 6556-6564. doi: 10.1002/ece3.2350

Vergeer, P. & Kunin, W. E. (2013). Adaptation at range margins: common garden trials and the performance of *Arabidopsis lyrata* across its northwestern European range. *New Phytologist, 197*, 989-1001. doi: 10.1111/nph.12060

Viant, M. R., Rosenblum, E. S., & Tjeerdema, R. S. (2001). Optimized method for the determination of phosphoarginine in abalone tissue by high-performance liquid chromatography. *Journal of Chromatography B, 765*, 107-111.

Viant, M. R. (2007). Metabolomics of aquatic organisms: The new "omics" on the block. *Marine Ecology Progress Series*, *332*, 301–306. doi: 10.3354/meps332301

Viant, M. R. (2008). Recent developments in environmental metabolomics. *Molecular BioSystems, 4*, 980-986. doi: 10.1039/b805354e

Vogt, G. (2012). Ageing and longevity in the Decapoda (Crustacea): A review. *Zoologischer Anzeiger, 251*, 1-25. doi: 10.1016/j.jcz.2011.05.003

von Schiller, D., Acuna, V., Aristi, I., Arroita, M., Basaguren, A., Bellin, A., Boyero, L., Butturini, A., Ginebreda, A., Kalogianni, E., Larranaga, A., Majone, B., Martinez, A., Monroy, S., Munoz, I., Paunovic, M., Pereda, O., Petrovic, M., Pozo, J., Rodriguez-Mozaz, S., Rivas, D., Sabater, S., Sabater, F., Skoulikidis, N., Solagaistua, L., Vardakas, L., & Elosegi, A. (2017). River ecosystem processes: A synthesis of approaches, criteria of use and sensitivity to environmental stressors. Science of the Total Environment, 596-597, 465-480. doi: 10.1016/j.scitotenv.2017.04.081

Vorburger, C., Rhyner, N., Hartikainen, H., & Jokela, J. (2014) A set of new and cross-amplifying microsatellite loci for conservation genetics of the endangered stone crayfish (*Austropotamobius torrentium*). *Conservation Genetics Resources, 6*, 629-631. doi: 10.1007/s12686-014-0157-0

Vorosmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Reidy Liermann, C., & Davies, P. M.

(2010). Global threats to human water security and river biodiversity. *Nature, 467*, 555-561. doi: 10:1038nature09440

Watanabe, M, Meyer, K. A., Jackson, T. M., Schock, T. B., Johnson, W. E., & Bearden, D. W. (2015). Application of NMR-based metabolomics for environmental assessment in the Great Lakes using zebra mussel (*Dreissena polymorpha*). *Metabolomics, 11*, 1302-1315. doi: 10.1007/s11306-015-0789-4

Wheatly, M. G. & Henry, R. P. (1992). Extracellular and Intracellular Acid-Base Regulation in Crustaceans. *The Journal of Experimental Zoology, 263*, 127-142.

Wheeler, J. O., Hoffman, P. F., Card, K. D., Davidson, A., Sanford, B. V., Okulitch, A. V., & Roest, W. R. (1996). *Geological map of Canada* [Shapefile]. Geological Survey of Canada [Distributor]. Retrieved from https://doi.org/10.4095/208175

Willig, A. & Keller, R. (1973). Molting Hormone Content, Cuticle Growth and Gastrolith Growth in the Molt Cycle of the Crayfish *Orconectes limosus*. *Journal of Comparative Physiology, 86*, 377-388.

Willis-Jones, E., Jackson, M. C., & Grey, J. (2016). Environmental Drivers for Population Success: Population Biology, Population and Community Dynamics. In M. Longshaw & P. Stebbing (Eds.), *Biology and Ecology of Crayfish* (pp. 251). Boca Raton, FL: CRC Press, Taylor and Francis Group.

Wu, H., Xu, L., Yu, D., & Ji, C. (2017). Differential metabolic responses in three life stages of mussels *Mytilus galloprovincialis* exposed to cadmium. *Ecotoxicology, 26*, 74-80. doi: 10.1007/s10646-016-1741-8

Appendix A

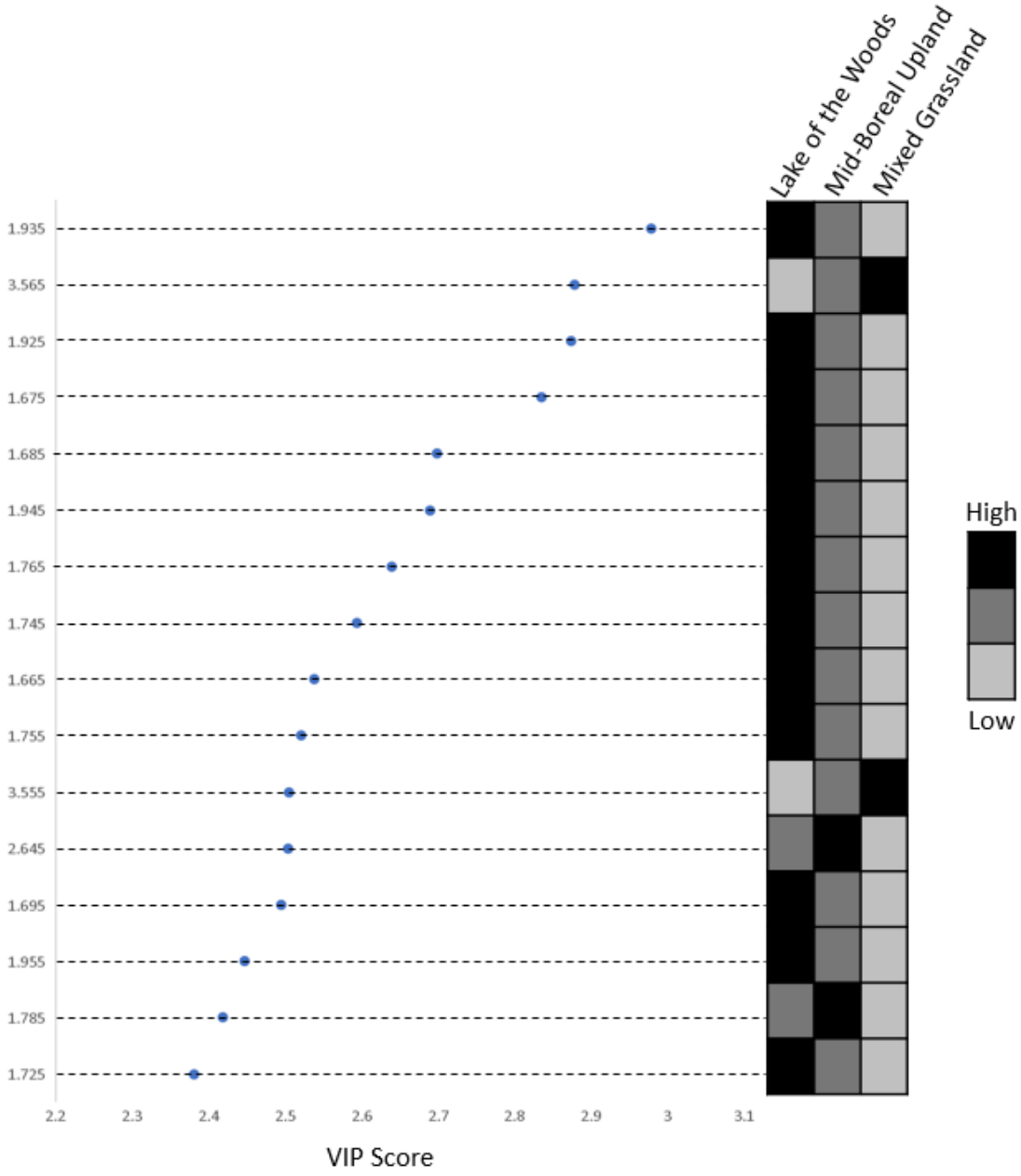
Appendix A.1. Analytical methods used to measure nutrients in the water collected from each stream and the detection limits associated with each nutrient (CFA = Continuous flow analysis, IR = Infrared spectroscopy, and FIA = Flow injection analysis).

Nutrient	Analytical Method	Detection Limit (µg/L)
Ammonia as Nitrogen (NH_3)	NH3-H Water Colour Salicylate CFA	5
Dissolved Organic Carbon (DOC)	DOC Water Non-Dispersive IR	100
Nitrate/Nitrite as Nitrogen	NO3-NO2 Water-Colour Cd Reduction CFA	10
Particulate Organic Nitrogen (PON)	POC-PON Water-Combustion	10
Soluble Reactive Phosphorus (SRP)	OrthoP/TP/DP Water-Colour Stannous Chloride CFA	2
Total Dissolved Nitrogen (TDN)	TN-TN Diss-Water-Alk Digest Colour Hydrazine FIA	15.0
Total Phosphorus (TP)	OrthoP/TP/DP Water-Colour Stannous Chloride CFA	2.0
Total Dissolved Phosphorus (TDP)	OrthoP/TP/DP Water-Colour Stannous Chloride CFA	2.0

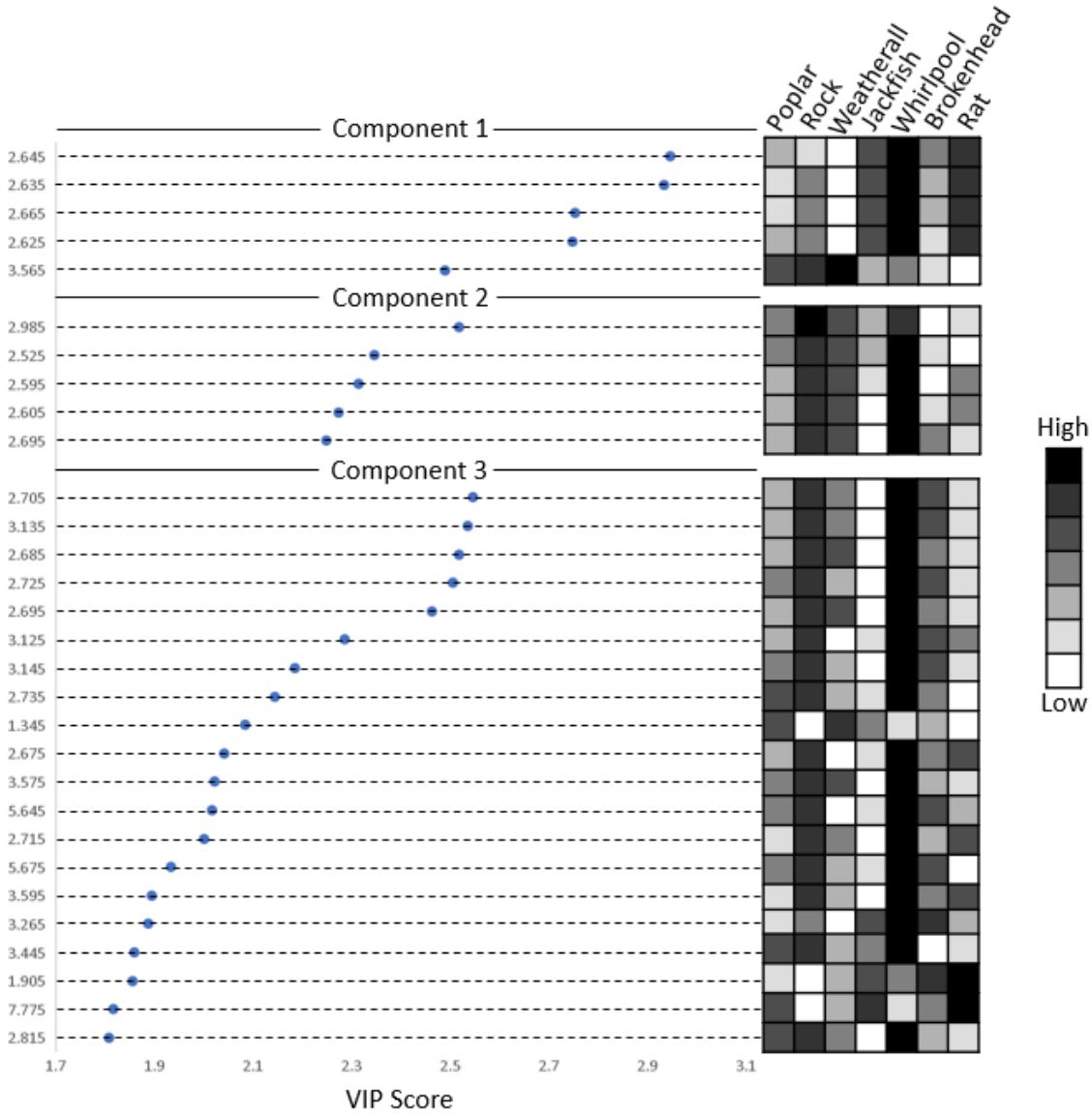
Appendix A.2. Analytical methods used to measure dissolved metals in the water collected from each stream and the detection limits associated with each dissolved metal.

Metal	Analytical Method	Detection Limit (µg/L)
Calcium	Inductively coupled plasma mass spectrometry	50
Magnesium	Inductively coupled plasma mass spectrometry	5
Sodium	Inductively coupled plasma mass spectrometry	5
Iron	Inductively coupled plasma mass spectrometry	0.5

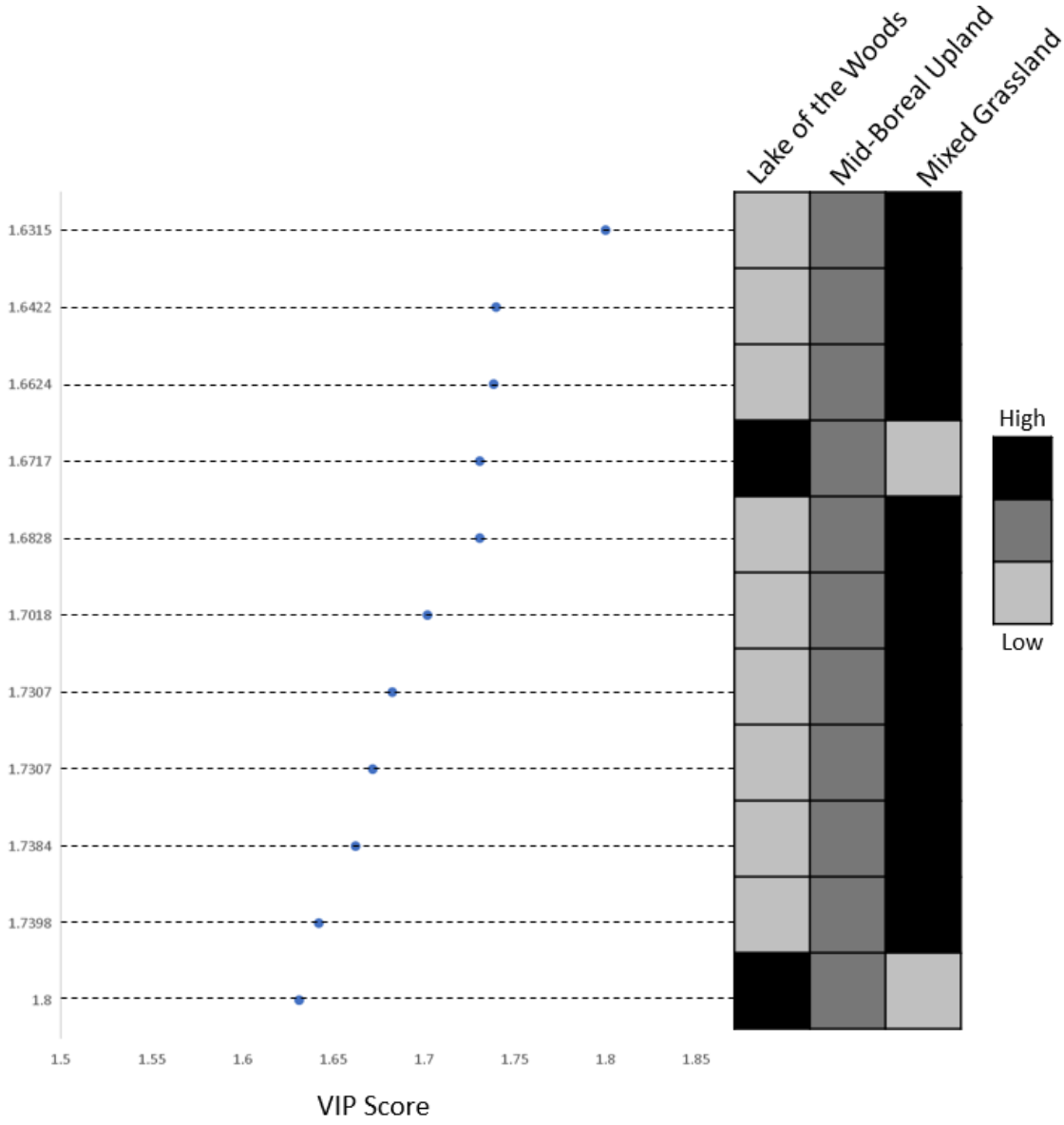
Appendix B



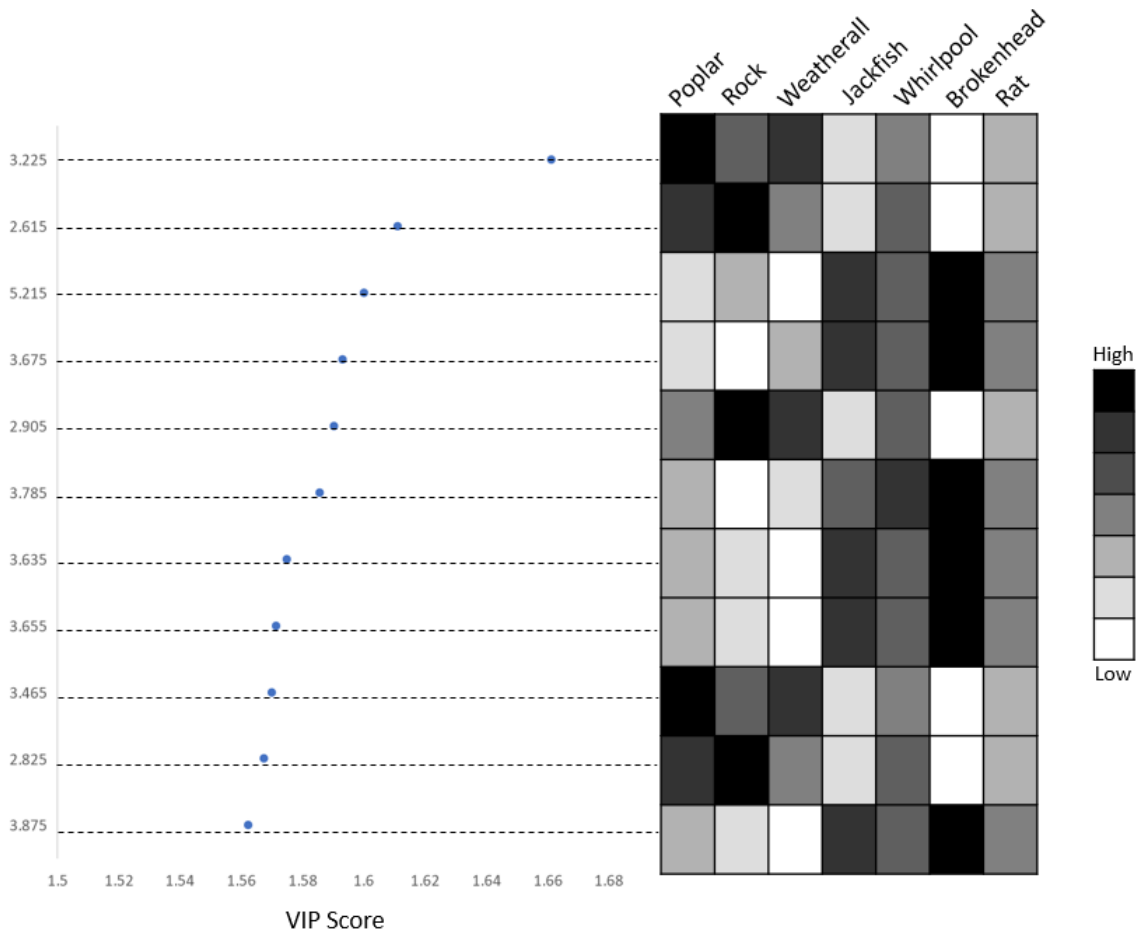
Appendix B.1. Bins with highest VIP scores distinguishing field crayfish by region. All bins are from the first component.



Appendix B.2. Bins with highest VIP scores distinguishing field crayfish by stream. Bins are from the first three components, all of which were significant to the PLS-DA model and which explained 15.8%, 15.6%, and 11.3% of the variability respectively.



Appendix B.3. Bins with highest VIP scores distinguishing common environment crayfish by region. All bins are from the first component.



Appendix B.4. Bins with highest VIP scores distinguishing common environment crayfish by stream. All bins are from the first component.

Curriculum Vitae

Name:	Cora Bilhorn
Post-secondary Education and Degrees:	Blackhawk Technical College Janesville, Wisconsin, USA 2004-2006, Associate Degree; Accounting The University of Wisconsin - Parkside Kenosha, Wisconsin, USA 2015-2018, B.Sc.; Molecular Biology The University of Western Ontario London, Ontario, Canada 2018-2020 M.Sc.; Geography
Honours and Awards:	Society of Freshwater Science Undergraduate Travel Award 2018 University of Wisconsin – Parkside Undergraduate Research Apprenticeship Award 2018 University of Wisconsin – Parkside Academic Achievement Award 2018
Related Work Experience	Research Assistant University of Wisconsin - Parkside 2018 Teaching Assistant University of Western Ontario 2018-2020