5-2014

Inter-and Intra-Population Variability across the Transcriptome of Lake Baikal's Endemic Copepod with Ramifications for Adapting to Climate Change

Larry L. Bowman Jr
East Tennessee State University

Inter-and Intra-Population Variability across the Transcriptome of Lake Baikal's Endemic Copepod with Ramifications for Adapting to Climate Change

---

A thesis

presented to

the faculty of the Department of Biological Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Biology

---

by

Larry L Bowman, Jr.

May 2014

Lev Y. Yampolsky, Chair

Thomas C. Jones

Thomas F. Laughlin

Keywords: comparative genomics, population genetics, *Epischura baikalensis*, zooplankton, plasticity, local adaptation, differential gene expression, genetic differentiation, single nucleotide polymorphisms (SNPs), Next Generation Sequencing (*Illumina*, *454 Rosche*)

ABSTRACT

Inter-and Intra-Population Variability across the Transcriptome of Lake Baikal's Endemic Copepod with Ramifications for Adapting to Climate Change

by

Larry L Bowman, Jr.

The future of Lake Baikal's biodiversity is uncertain in response to climate change. Unlike its diverse benthos, Lake Baikal's zooplankton is species poor, with up to 96% of its biomass being composed of a single Calanoid copepod species, *Epischura baikalensis*. This study characterizes the genetic differentiation and differential gene expression of *E. baikalensis*. Using partial-transcriptome sequences obtained by *454 Rosche* and *Illumina* sequencing technologies, the genetic differentiation at inferred single nucleotide polymorphism (SNP) sites and differential gene expression in populations sampled from various parts of the lake were analyzed. The functional genomics of genes showed significant differential expression among the lake's regions with some genes being highly up-or down-regulated. High genetic differentiation among regions suggests isolated subpopulations. Moreover, significantly differentially expressed transcripts were significantly more genetically differentiated than transcripts exhibiting no differential expression. These results suggest high potential phenotypic plasticity and adaptability in response to climate change, e.g., temperature.

ACKNOWLEDGEMENTS

I would like to thank Dr. Lev Yampolsky for serving as a mentor and advisor for my graduate work at East Tennessee State University. I would also like to thank Drs. Thomas Jones and Thomas Laughlin for their support as my committee members, the Department of Biological Sciences, and the Biology graduate students at ETSU.

Many thanks go to the National Science Foundation Lake Baikal Dimensions of Biodiversity Grant: DEB-1136637; PIs Drs. Stephanie Hampton (UCSB/WSU), Lev Yampolsy (ETSU), Marianne Moore (Wellesley College), Ed Theriot (UT-Austin), Elena Litchman (MSU), and Chris Klausmeier (MSU) for financial support of my research. In addition to the PIs of the Lake Baikal Dimensions of Biodiversity Grant, I would like to thank our Russian collaborators Kirill Shchapov, Lena Pislegena, Sasha Pislegen, Lyubov Izmest'eva, Eugene Silow, and Devard Stom; for without their cooperation, our research would not be possible. I would also like to thank the other team members (and lifelong friends) of the Lake Baikal Dimensions of Biodiversity Grant who made my time in Siberia most enjoyable: Dr. Steve Katz, Dr. Ted Ozersky, Teofil Nakov, Kara Woo, and Katie Wright.

Additionally, many thanks to the National Science Foundation GK-12 "*Science First!*" Grant: DGE-0742364; PIs Drs. Gordon Anderson (ETSU), Aimee Govett (ETSU), Anant Godbole (ETSU), Sharon Pickering (NSES), and Mrs. Valerie Orfield (NSES) for the financial support of my graduate fellowship as a National Science Foundation Graduate Fellow in K-12 Education. Special thanks to Daniel Ressler, Colton Watts, Jennifer Price, Caitlyn King, and Allison Smith for keeping me sane along the way.

TABLE OF CONTENTS

Page

ABSTRACT 2

LIST OF TABLES 9

LIST OF FIGURES 10

Chapter

1. INTRODUCTION 11

Lake Baikal 11

Geography and Hydrography 14

Epischura baikalensis 16

2. RESEARCH OBJECTIVES 17

Zooplankton Distribution 17

Differential Gene Expression 17

Genetic Differentiation 18

3. MATERIALS, METHODS, AND RATIONALE 20

Zooplankton Distribution 20

Sample Collection 20

Depth Sampling in Chivrikuj Bay Summer 2013 20

Sample Storage 21

Sampling Rationale 21

Summer 2012 21

Summer 2013 21

Sample Processing 22

Copepoda 22

Cladocera 22

Chapter Page

Statistical Analysis 22

Whole Lake 23

Chivrikuj Bay 23

Reference Transcriptome 27

RNA Extraction 27

DNase Digestion 27

Library Construction and *454 Rosche* Sequencing 28

Differential Gene Expression 29

Sample Collection 29

Sample Storage 29

Sampling Rationale 30

Sample Processing 32

RNA Extraction 32

Library Construction and *Illumina* Paired-End Sequencing 32

Analysis 32

tBLASTn 33

Gene Ontology 33

Genetic Differentiation 34

Sample Collection 34

Sample Storage 34

Sample Processing 34

Analysis..... 34

4. ZOOPLANKTON DISTRIBUTION 36

Results 36

Whole Lake 36

Epischura baikalensis 36

Chapter Page

Daphnia longispina 36

Cyclops spp. 36

Diaptomus graciloides 37

Macrohectopus branickii 37

Chivrikuj Bay 44

Zooplankton Abundance 44

Abiotic Factors 44

Discussion 48

Whole Lake 48

Chivrikuj Bay 50

5. DIFFERENTIAL GENE EXPRESSION 52

Results 52

Correlation Analysis 52

Transcript Annotation and Gene Ontogeny 57

Discussion 63

Cluster Analysis 63

Cluster 1 Transcripts 63

Cluster 2 Transcripts 63

Cluster 3 Transcripts 64

Cluster 4 Transcripts 64

6. GENETIC DIFFERENTIATION 65

Results 65

Correlation Analysis 65

Discussion 68

7. CONCLUSIONS 70

Zooplankton Distribution 70

Chapter Page

Differential Gene Expression 70

Genetic Differentiation 71

Ramifications for Climate Change 71

REFERENCES 72

APPENDICES 75

Appendix A: Lake Baikal Fauna..... 75

Appendix B: Research Vessels 78

Appendix C: Bolshie Koty (Больши́е Ко́ты), Irkutsk Oblast, Russia 79

VITA 80

LIST OF TABLES

Table Page

1. Means Comparison *Epischura baikalensis* Abundance 38
2. Means Comparison *Daphnia longispina* Abundance 39
3. Means Comparison *Cyclops* spp. Abundance 40
4. Means Comparison *Diaptomus graciloides* Abundance 41
5. Means Comparison *Macrohectopus branickii* Abundance 42
6. Transcript Gene Ontology by Cluster 58
7. Transcript Gene Ontology by Database 59
8. Genetic Differentiation Analysis Summary 67

LIST OF FIGURES

Figure Page

1. Map of Lake Baikal Regions 15
2. Summer 2012 Sampling Sites 24
3. Summer 2013 Sampling Sites 25
4. Summer 2013 Chivrikuj Bay Sampling Sites 26
5. Summer 2012 Transcriptome Sampling Sites 31
6. *Epischura baikalensis* Abundance ANOVA 38
7. *Daphnia longispina* Abundance ANOVA 39
8. *Cyclops* spp. Abundance ANOVA 40
9. *Diaptomus graciloides* Abundance ANOVA 41
10. *Macrohectopus branickii* Abundance ANOVA 42
11. Zooplankton Abundance across Lake Baikal 43
12. Bivariate Analyses of Zooplankton Abundance in Chivrikuj Bay 45
13. Zooplankton Abundance across Chivrikuj Bay 46
14. Bivariate Analyses of Abiotic Factors in Chivrikuj Bay 47
15. Maloe More and Chivrikuj Bay Average Water Temperature ANOVA 49
16. Dendrogram and Heatmap of Differential Gene Expression in *E. baikalensis* 53
17. Parallel Plots by Transcript Cluster for Multivariate Cluster Analysis 54
18. Principal Components Analysis (PCA) on Correlations of Transcripts 56
19. Principal Components Analysis (PCA) on Correlations of Sample Sites 56
20. Color Map on Sample Correlations 57
21. Gene Ontologies of All Annotated Transcripts 61
22. Gene Ontologies of Transcripts by Cluster 62
23. Principal Components Analysis (PCA) on Correlations of SNP Frequency 65
24. Predicted and Actual F_{ST} Values 66
25. Mean F_{ST} Values by Cluster Membership 68

CHAPTER 1

INTRODUCTION

It has been shown that the more plasticity an organism displays, the better adapted it may become to climate change (Gomulkiewicz and Holt 1995). Further, a population can be rescued from a critically low level of individuals by natural selection, but rescue is largely dependent on stochastic events that may instead force extinction (Gomulkiewicz and Holt 1995). Changes in demography are critical to the success of populations in peril (Barton and Partridge 2000). Adaptive evolution can theoretically allow a population to occupy previously uninhabitable environments (Gomulkiewicz and Holt 1995). Adaptation to climate change requires genetic variation within populations; therefore, to get a well-informed picture of the future success of a declining population, it is crucial to understand genetic variation within and among populations (Barton and Partridge 2000).

Lake Baikal

Drastic changes are particularly dangerous for hotspots of biodiversity, where species are limited by range or endemism. One of these hotspots is the ancient rift lake, Lake Baikal, Siberia, Russian Federation. Lake Baikal boasts many unique characteristics: largest (by volume), deepest, and projected oldest (Moore et al. 2009). Although the lake is superlative in physical characteristics, Lake Baikal is also home to many endemic species from the benthos to the world's only freshwater seal (*Pusa siberica*). Species diversity has been thoroughly investigated by numerous researchers (Timoshkin 1995; Sherbakov 1999), with much focus on the benthos; however, there is little research on the plankton of Lake Baikal at present.

Lake Baikal shares many characteristics with other ancient lakes, specifically Lakes Tanganyika, Malawi, Biwa, and Ohrid. Understanding the cold, oligotrophic Lake Baikal system may inform responses to projected climatic changes in the other ancient lakes, additional hotspots of biodiversity. Increased warming has also occurred in all of the aforementioned lakes (Verburg, et al. 2003; Matzinger et al. 2006; Albrecht and Wilke 2009; Tsugeki, et al. 2009; Tierney et al. 2010), allowing for the opportunity to conduct comparative studies among ancient lake systems.

With Lake Baikal's many distinct characteristics, however, come many distinct challenges related to global climate change. The temperature of Lake Baikal has been increasing over the past century and has been found to be rising twice as quickly as the global average, with a 1.2°C over the past century (Shimaraev et al. 2002; Hampton et al. 2008). Currently, projected temperatures for the next century are estimated to be 4.5°C warmer, especially in shallow bays, which will present acute challenges to populations adapted to living in colder temperatures and those vulnerable to large increases in temperature (Moore et al. 2009).

A 1.2+°C temperature increase over the next century will increase the possibility of extinction for many of Lake Baikal's endemics, including the plankton (Moore et al. 2009). The plankton has a special place in the ecosystem and is largely regarded as the major energy source within the food web (Sherbakov 1999; Sideleva 2003; Bondarenko et al. 2006; Moore et al. 2009). Endemic diatoms (Afanasyeva 1998) and zooplankton (Yoshii et al. 1999) feed the higher trophic level organisms, including many commercially important fish species (Moore et al. 2009). Endemic plankton are also decreasing in Lake Biwa (Japan) because of warming (Tsugeki et al. 2010) and may suffer a similar fate in the Lake Baikal system. Changing the structure of the bottom trophic levels of this system will unequivocally affect higher trophic

levels, even the benthos species (Bondarenko et al. 2006): amphipods (Sherbakov 1999) and benthic fishes (Sideleva 2003). Considering that endemic diatoms are the endemic zooplankter *Epischura baikalensis*'s primary energy source, one can surmise the remarkable support the 2 provide for the ecosystem.

Epischura baikalensis, though an integral part of the ecosystem, is perhaps in the most fragile of positions. It has been shown that *E. baikalensis* is especially susceptible to increased temperatures with a host of problems occurring (Moore et al. 2009), such as increased mortality (Afanasyeva 1998) and fungal infections by *Saprolegnia* sp. (Kozhova and Beim 1993). Recent invasions by cosmopolitan species, especially in shallow bays, present an additional dimension to the stress of climate change. Cosmopolitan species, such as *Daphnia longispina* and *Cyclops* spp., are projected to compete better in warmer temperatures than in the colder open waters of Lake Baikal (Melnik et al. 1998; Moore et al. 2009; Richardson et al. 2000). Because the endemic plankton of Lake Baikal are smaller than the invading cosmopolitan species (Moore et al. 2009), cosmopolitan introductions will no doubt affect energy transfer from bottom trophic levels to top trophic levels. The biogeochemical cycle is also likely to be affected, as has been shown in other lakes (Sekino et al. 2007).

An increase in cladocerans (*Daphnia* spp.) and cosmopolitan copepods (*Cyclops* spp.) has been observed, while endemic copepods (*E. baikalensis*) have declined especially in warming parts of Lake Baikal such as Chivrikuj Bay (Hampton et al. 2008). Unless genetic variation proves to be high, thereby allowing the selection of beneficial mutations, the endemic copepod population may be in danger of extinction with increasing water temperatures. However, even with high genetic variation in the endemic population, cosmopolitan species may outcompete endemic species in warmer waters. *Epischura baikalensis* must not only be able to

survive a major climate change but must be more fit than invading species that may be better adapted to warmer waters.

As previously mentioned, the benthos of Lake Baikal has been thoroughly studied (Efremova et al. 2002; Macdonald III et al. 2005; Sherbakov 1999), but the plankton has not. Though many of the other functional groups show high species richness, the zooplankton group is remarkably species poor. Because at current levels *E. baikalensis* composes 90% of the zooplankton biomass of Lake Baikal (Afanasyeva 1998), it is imperative that we understand its current state of genetic variation to help predict its future. The elimination of this key species could be detrimental to the many other links in the lake and have far-reaching ramifications in the commercial and political realms. Finding high genetic diversity among the zooplankton could predict *E. baikalensis*'s success in the future Lake Baikal ecosystem.

Geography and Hydrography

Lake Baikal is a northern rift lake at approximately 53°40'N and 109°0'E. Lake Baikal is split into 3 distinct basins: Northern, Central, and Southern Basins (Figure 1). The Northern and Central basins are split by the underwater Academician Ridge, while the Central and Southern Basins are separated by narrowness and sedimentation caused by the Selenga River delta. There are 3 larger bays: Chivrikuj Bay, Barguzin Bay, and Proval Bay (Figure 1). An additional region of interest is the shallow, narrow inlet Maloe More on the western bank of the lake. Olkhon Island borders Maloe More on the eastern side. The Selenga, Upper Angara, and Barguzin Rivers empty into the lake, and the Angara River is the lake's only outflow.

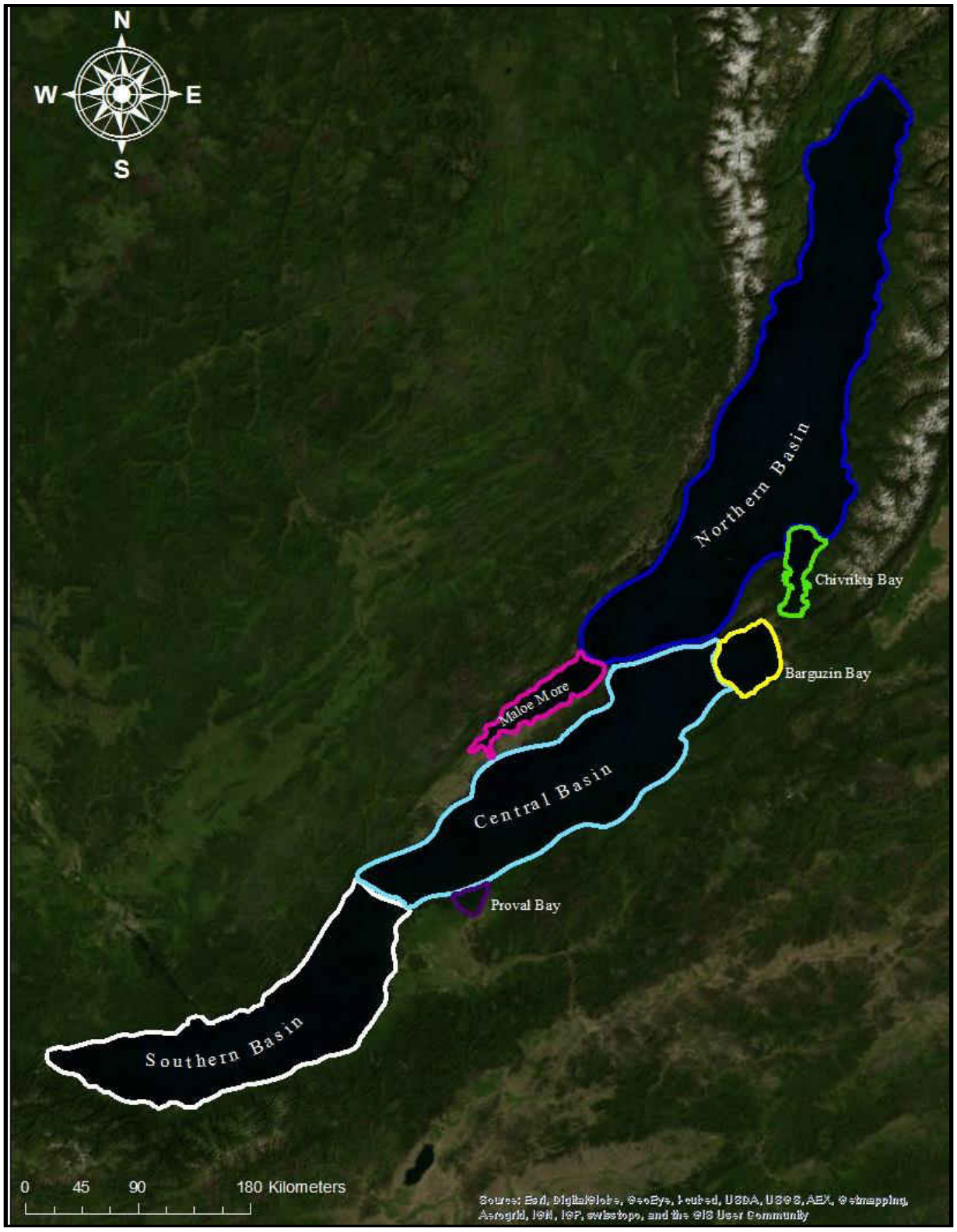


Figure 1 represents the major geographic regions of Lake Baikal. The outlines represent as follows: Northern Basin (dark blue), Central Basin (light blue), Southern Basin (white), Maloe More (pink), Chivrikuj Bay (green), Barguzin Bay (yellow), and Proval Bay (purple).

Epischura baikalensis

Epischura baikalensis (Crustacea: Copepoda) is the chosen study species of this project. This species makes up almost 90% of the zooplankton biomass and a large portion of total biomass in Lake Baikal (Afanasyeva 1998). As a cold-water stenotherm, this threatened endemic copepod species can be found throughout the water column and throughout all parts of Lake Baikal (Afanasyeva 1998). The organism has yet to be successfully maintained in the laboratory, so all samples were collected directly from Lake Baikal.

CHAPTER 2

RESEARCH OBJECTIVES

The main objective of this project is to characterize the zooplankton distribution, differential gene expression, and genetic differentiation within and among the zooplankton populations, specifically *Epischura baikalensis*, within the Lake Baikal system by location and temperature.

Zooplankton Distribution

It is important to track the spread of invasions in fragile systems such as Lake Baikal. Characterizing zooplankton assemblages across the lake's regions may suggest what parameters are necessary for invasions to be successful and what parameters are especially hard for the endemic *Epischura baikalensis* to overcome.

Understanding the zooplankton assemblages across the lake in response to certain abiotic factors may lead to predictions of how far the invasions may spread in the future and their immediate outlook with increasing temperatures in the short term. Collecting data on abiotic factors along with current assemblage data will inform the genetics portions of the research objective, quantifying the ecological reasons for up-or down-regulated differential gene expression or heightened or lowered genetic differentiation. Informing the genetic data with demographic data will allow for more informed predictions for the outlook of *E. baikalensis* in Lake Baikal's warming future.

Differential Gene Expression

Because of Lake Baikal's biogeographic isolation among its 3 basins (Northern, Central, and South Basins), each basin has distinct characteristics that may be reflected in the local

adaptations of the individuals/populations that live within them. Moreover, the drastic differences in climate seen in Lake Baikal's bays (Chivrikuj, Barguzin, and Proval Bays) and shallow inlet (Maloe More) could further be correlated with differential gene expression among their respective zooplankton populations.

Finding significant differential gene expression among zooplankton from open Baikal and zooplankton from warmer waters may suggest the ability of *E. baikalensis* to adjust to climate change. Quantifying differential gene expression may suggest what genes must be up-or down-regulated to compensate for hotter environments. Highly significant differential gene expression differences among the different demes from around the lake could suggest high phenotypic plasticity to changing environmental conditions that would afford *E. baikalensis* greater success as Lake Baikal continues to warm in the future.

Genetic Differentiation

Finding high genetic diversity among populations may suggest that the species has a greater chance of adapting to climate change; while exhibiting a relatively low level of genetic diversity may imply that the species lacks the diversity necessary to survive climate changes and to outcompete invading cosmopolitan species. Alternatively, the large zooplankton population may be fragmented due to seasonal, spatial, or depth gradations. Because Lake Baikal is split biogeographically into 3 basins, the separation of populations could be causing 3 smaller, more genetically diverse populations.

Because *Epischura baikalensis* inhabits the entire lake from its max depth to the surface, a depth gradient could be isolating populations. All of these isolating factors could be increasing genetic diversity and differentiation; but overall, the mixing and upwelling of the lake could be sufficient to keep individuals in a large, interbreeding population.

In this study, we plan to characterize the zooplankton abundance, differential gene expression, and genetic differentiation of 4 subpopulations from different geographic lake regions to better understand the inter-and intra-population variation of *Epischura baikalensis*. Samples from Maloe More that are naturally exposed to higher temperatures may serve as an example of *E. baikalensis*'s ability to respond and adapt to climate change.

CHAPTER 3

MATERIALS, METHODS, AND RATIONALE

Sampling included 2 field seasons on Lake Baikal: the summers of 2012 and 2013. Samples were collected from summer populations of zooplankton. Research efforts for both summers were locateded at the biostation at Bolshie Koty (Больши́е Ко́ты), Irkutsk Oblast, Russia, owned and operated by Irkutsk State University. The sampling plans described herein were different for each research component with similar sampling methods.

Zooplankton Distribution

Sample Collection

All zooplankton samples were collected with a 100 μm mesh-size plankton net with a diameter of 50 cm. Nets were extended overboard using ropes designated with distance intervals. Sampling was completed aboard 2 research vessels: *Professor Kozhov*, owned and operated by Irkutsk State University and *Professor Treskov*, owned and operated by the Baikal Museum. Data for temperature, dissolved oxygen, surface temperature, and chlorophyll concentrations were also collected using a YSI sonde. Samples were taken from 100 m to surface in all stations where depth was 100 m or greater. All other samples were taken at max depth, i.e., <100 m. Samples were filtered through a small diameter 100 μm mesh-size plankton sieve to condense into plastic specimen flasks.

Depth Sampling in Chivrikuj Bay Summer 2013. Sampling was completed in 3 traverses of Chivrikuj Bay following a transect along the 109th meridian East that spanned from the 125 m or greater depth of open Lake Baikal to the inner reaches of the inner bay. The 109th meridian East was used for ease of transect sampling because it runs the length of the bay with minutes

north of the 53rd parallel North being used to denote samples taken in a North to South direction. If a station was 125 m or greater in depth, a closing plankton net (100 µm mesh-size) was used to collect depth samples. Depth samples included all zooplankton in the water column from 125 m – 25 m, heretofore referred to as "deep" samples, and from 25 m – 0 m, heretofore referred to as "shallow" samples. The net was closed at 25 m to only capture zooplankton from 125 m – 25 m depths. The same was repeated for depths of 25 m – 0 m. If a station was <125 m in depth, max depth was sampled.

Sample Storage

Samples were stored in approximately 20 mL 70% ethanol for transportation back to East Tennessee State University, Johnson City, TN, where further processing and counts took place. Samples were refrigerated until counted.

Sampling Rationale

Summer 2012. Sampling in summer 2012 was intended for exploratory purposes; however, results were included from this sampling effort. A map of sampling sites from summer 2012 for zooplankton abundance is displayed in Figure 2. Sampling from this season included samples from all 3 of Lake Baikal's basins (Northern, Central, and Southern), Chivrikuj Bay, Barguzin Bay, and Maloe More.

Summer 2013. Sampling in summer 2013 was repeated for all 3 of Lake Baikal's basins (Northern, Central, and Southern), Chivrikuj Bay, Barguzin Bay, and Maloe More. However, extensive sampling was completed in Chivrikuj Bay to obtain higher resolution of zooplankton assemblages, temperature, dissolved oxygen, surface temperature, and chlorophyll concentrations. The only samples used for zooplankton distribution from summer 2013 were the Chivrikuj Bay samples; all others were for subsequent analyses.

From previous personal observations, a sharp gradient of exclusion between the endemic *Epischura baikalensis* and the invasive *Daphnia longispina* in Chivrikuj Bay existed. Extensive sampling was carried out to further investigate the parameters behind this sharp divide.

A map of all summer 2013 sampling sites of Lake Baikal is displayed in Figure 3, while a more detailed map of summer 2013 sampling sites of Chivrikuj Bay is displayed in Figure 4.

Sample Processing

Samples were diluted in ethanol and split into less dense subsamples. The counts were then corrected for their dilution factors. Zooplankton was identified to the subclass *Copepoda* or to the order *Cladocera*. All other zooplankton was classified to the subphylum *Crustacea*; specifically, the only zooplankton classified as "other *Crustacea*" was *Macrohectopus branickii*, Lake Baikal's endemic, planktonic amphipod (*Amphipoda*: *Gammaridea*).

Copepoda. Copepods were further classified into 3 categories: *Cyclops* spp., *Epischura baikalensis*, or *Diaptomus graciloides*. *Cyclops* spp. includes *Cyclops kolensis*, the subgenera *Mesocyclops*, *Acanthocyclops* sp., and *Cyclops vicinus*. *Cyclops* spp. were grouped as such due to difficulty in determining further species categories in organisms at various life stages.

Cladocera. Cladocerans were further classified into 4 categories: *Daphnia longispina*, *Chydorus* sp., *Bosmina* sp., and *Leptodora kindtii*.

Statistical Analysis

JMP 11.0 was used for statistical analysis of the zooplankton distribution samples (SAS Institute Inc. 2013). Raw counts were changed to percentages of full assemblage by species/taxonomic group in the water column for statistical analysis.

Whole Lake. One-way ANOVA were used to analyze the variation in percent abundance of different zooplankton groups between and among sampling sites across the lake in addition to

variation with response to abiotic factors (depth, average water temperature, dissolved oxygen, surface temperature, and chlorophyll concentrations). All pairs Tukey's HSD test with 95% confidence intervals were used to compare means in site-by-site comparisons.

Chivrikuj Bay. Bivariate analyses were used to analyze the variation in percent abundance of different zooplankton groups by distance along the 109th meridian East, the transect used for abundance sampling that runs the length of Chivrikuj Bay. Regressions were fit to the data and analyzed with ANOVA with 95% confidence intervals.

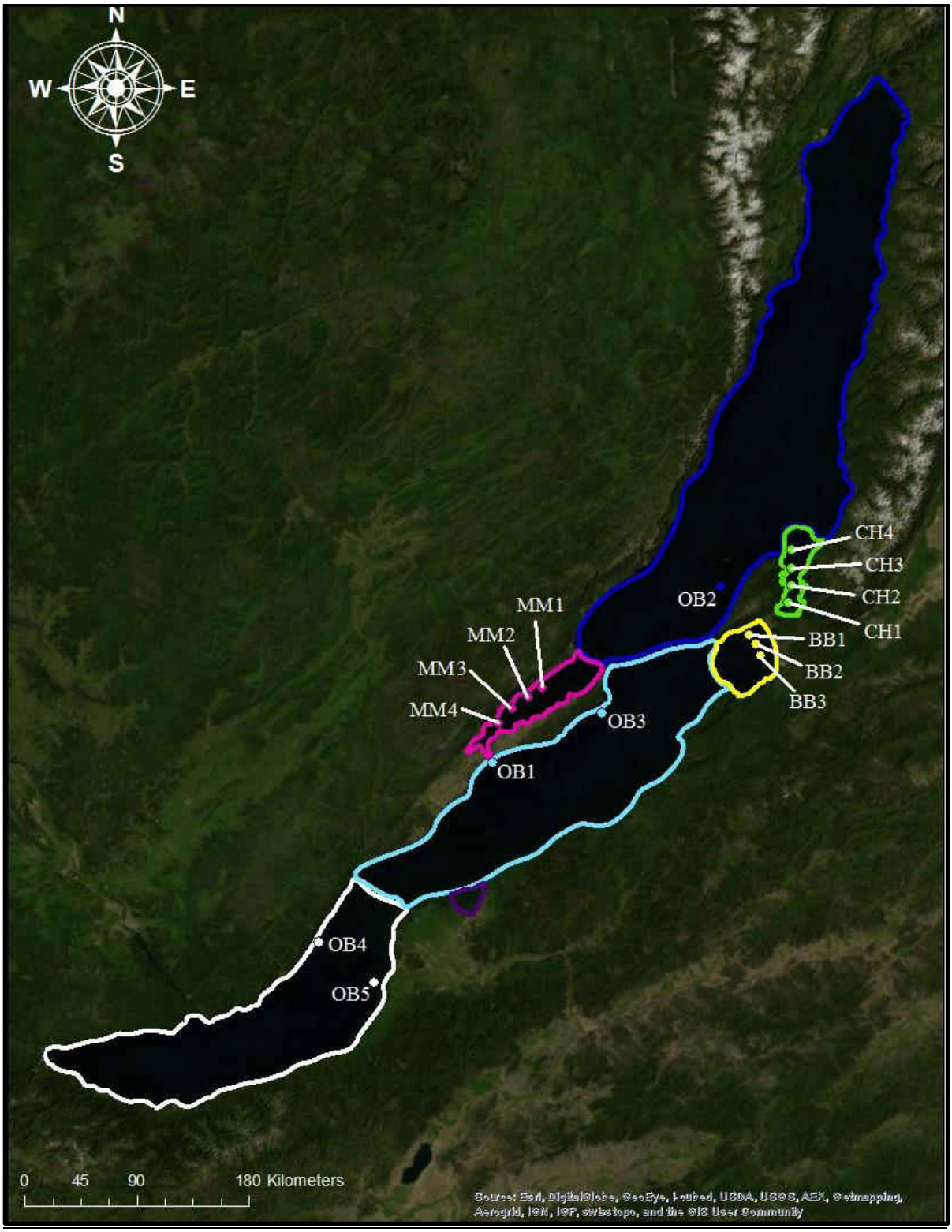


Figure 2 represents sampling sites from summer 2012 Lake Baikal sampling cruises aboard the *Professor Kozhov* and *Professor Treskov*. The sample codes are as follows: open Baikal (OB), Barguzin Bay (BB), Chivrikuj Bay (CH), and Maloe More (MM). Open Baikal samples include sampling sites in all 3 basins.

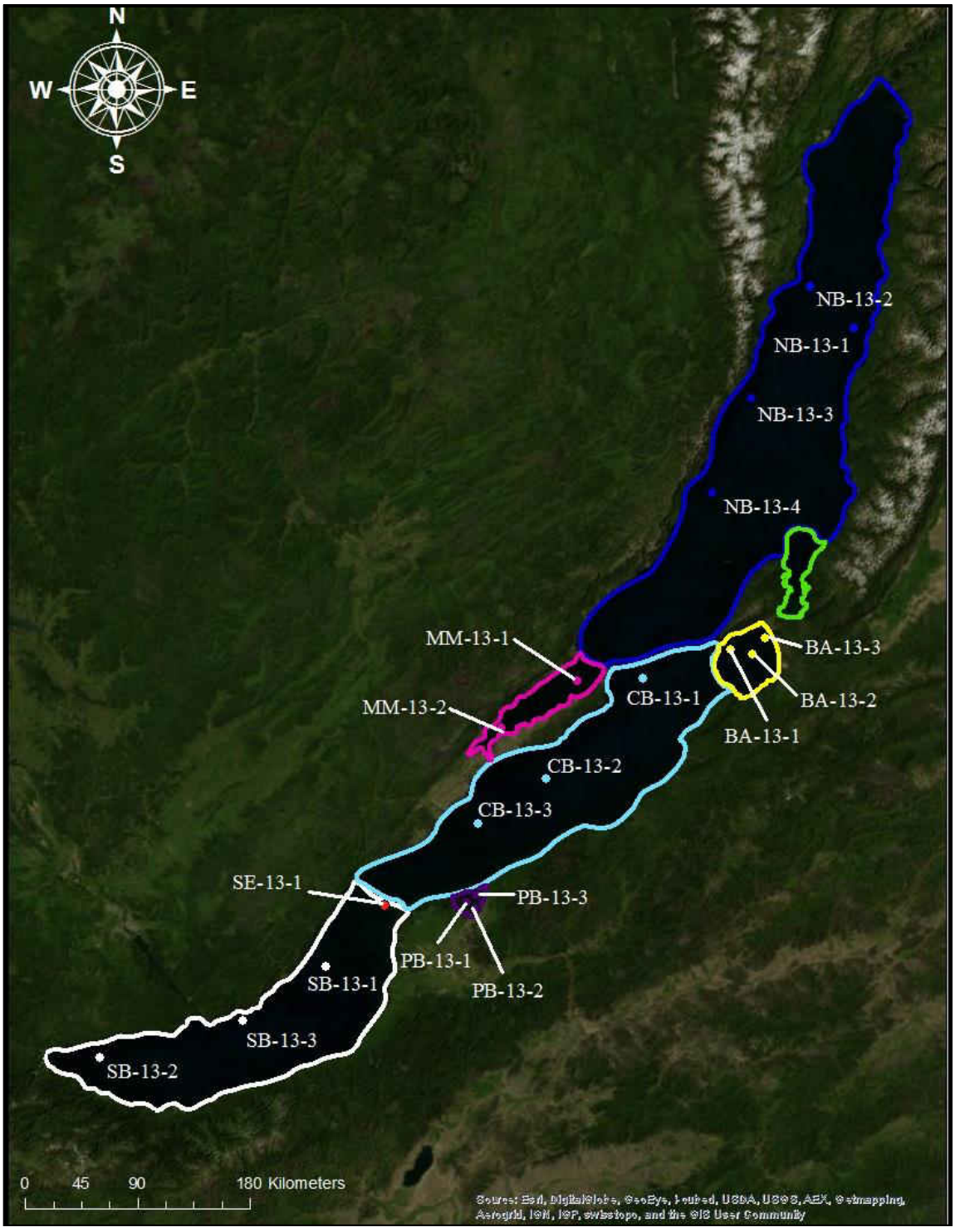


Figure 3 represents sampling sites from summer 2013 Lake Baikal sampling cruises aboard the *Professor Kozhov* and *Professor Treskov*. Sample codes are "Location-Year-Station" with "locations" as follows: Northern Basin (NB), Central Basin (CB), Southern Basin (SB), Selenga River Delta (SE), Proval Bay (PB), Barguzin Bay (BB), and Maloe More (MM). Chivrikuj Bay samples are not pictured due to density and can be found in Figure 4.

Figure 4 represents sampling sites from summer 2013 Chivrikuj Bay sampling cruise aboard the *Professor Kozhov*. Sample codes are "Location-Year-Traverse.Station" with "location" as follows: Chivrikuj Bay (CH). These samples represent 3 traverses of Chivrikuj Bay of a transect that runs North-South along the 109th meridian East.

Reference Transcriptome

A reference transcriptome was obtained from ~250 adult individuals collected from the Southern Basin during the winter sampling expedition in March 2013 (Figure 3, SB-13-3). This sample from the Southern Basin served as the reference against which all other sequenced transcriptomes are aligned. RNA was extracted from these samples per the following protocol.

RNA Extraction

Qiagen's RNeasy kit was used for RNA extraction as follows (Qiagen 2006). A sample of ~250 individuals was isolated and counted in RNAlater. The sample was dried with filter paper, flash frozen with liquid nitrogen, and macerated with a pestle in an RNase-free 2 mL microcentrifuge tube. Liquid nitrogen was allowed to evaporate without the sample thawing before whole-organism tissues were disrupted and homogenized with 600 µL lysis buffer (Buffer RLT) and mortar and pestle. The lysate was centrifuged for 3 min at 13,500 rpm. The supernatant (lysate) was removed and transferred to a new microcentrifuge tube.

Then, 600 µL 70% ethanol was added and mixed via pipette. From this, a 700 µL sample was taken (including precipitate) and pipetted into an RNeasy spin column within a 2 mL collector tube. The sample was centrifuged for 15 s at 13,500 rpm before an on-column DNase digestion was performed.

DNase Digestion. An on-column DNase digestion was performed on each sample by adding 350 µL membrane-washing buffer (Buffer RW1) and centrifuging for 15 s at 13,500 rpm. After the flow-through was discarded, 80 µL DNase I incubation mix (10 µL DNase I stock solution and 70 µL Buffer RDD) was added directly to the spin column membrane and incubated at room temperature for 15 min. After incubation, the column was washed with 350 µL Buffer

RW1 and centrifuged for 15 s at 13,500 rpm. The flow-through was discarded, and the column was used for the remainder of the RNA extraction.

After the DNase digestion, 500 µL of another membrane-washing buffer (Buffer RPE) was added to the column. The sample was centrifuged at 13,500 rpm, and the flow-through was discarded. A second wash using 500 µL Buffer RPE was completed. Two washes of Buffer RPE ensured that no ethanol that may hinder subsequent reactions continues in the protocol. The column was inserted into a new collection tube and centrifuged for 1 min at 13,500 rpm. In another new collection tube, the column's membrane was washed with 50 µL RNase-free water and centrifuged for 1 min at 13,500 rpm. The membrane was washed again with 50 µL RNase-free water to ensure elution of all RNA. The sample concentration (>10 ng/µL) was then confirmed with nanodrop and gel electrophoresis before *454 Rosche* sequencing libraries were constructed.

Library Construction and 454 Rosche Sequencing

Library construction and *454 Rosche* sequencing were performed by Beckman Coulter Genomics. Libraries (cDNA) for the *454 Rosche* sequencing were prepared according to the manufacturer's specifications (454 Life Sciences Corp. 2011). The resulting reference transcriptome served as the basis for alignment of the *Illumina* double-end reads used in subsequent analyses. *454 Rosche* sequencing was chosen to provide longer reads for better overlap with the goal of a more accurately aligned reference transcriptome.

The binary data file generated by the GS-FLX sequencing instrument software was converted to multiple-FASTA format and reads quality saved in multiple-QUAL format. Quality values express negative base-10 log of the basecall error probability. Transcriptome assembly

was conducted by Newbler (Fryslie 2012) and assembled contigs longer than 100 bases saved in a FASTA file.

Differential Gene Expression

Differential gene expressions samples were taken from the summer 2012 transcriptome samples (Figure 5). Samples included sites from the 3 major basins (Northern, Central, and Southern) and from the shallow inlet Maloe More.

Sample Collection

All zooplankton samples were collected with a 100 µm mesh-size plankton net. Nets were extended overboard using ropes designated with distance intervals. Sampling was completed aboard 2 research vessels: *Professor Kozhov*, owned and operated by Irkutsk State University, and *Professor Treskov*, owned and operated by the Baikal Museum. Data for temperature, dissolved oxygen, surface temperature, and chlorophyll concentrations were collected using a YSI sonde. Samples were taken from 150 m to surface in all stations where depth is 100 m or greater. Samples were filtered through a small diameter 100 µm mesh-size plankton sieve to condense into plastic 2 mL microcentrifuge tubes.

Sample Storage

Samples were dried with filter paper and stored in approximately 2 mL Qiagen RNAlater (RNA stabilization reagent) in 2 mL microcentrifuge tubes. Microcentrifuge tubes with samples were immediately flash frozen in liquid nitrogen and stored in a dry vapor shipper for transportation back to East Tennessee State University, Johnson City, TN, where further processing took place. Duplicates were flash frozen in liquid nitrogen alone. Samples were refrigerated in a -80°C unit until processed.

Sampling Rationale

For the most robust transcriptome analysis, as many individuals from as many environments as possible proved the most useful. However, because further analysis required whole-organism RNA extraction, care was taken to ensure only transcriptomes of the study organism were taken and not the transcriptomes of ciliates, algae, and other organisms that may have been present in the gut of *Epischura baikalensis*, i.e., the metatranscriptome. Because of this caveat and a preference for transcriptome study on *Epischura baikalensis* alone, samples were taken from the Northern, Central, and Southern Basins, and Maloe More only, where *E. baikalensis* appears in >90% of the zooplankton abundance. This sampling plan ensured RNA extraction from the most diverse habitats possible while minimizing contamination by other organisms' transcriptomes and the metatranscriptome of the sampling region. Three biological replicates were used from each transcriptome sample site (Figure 5).

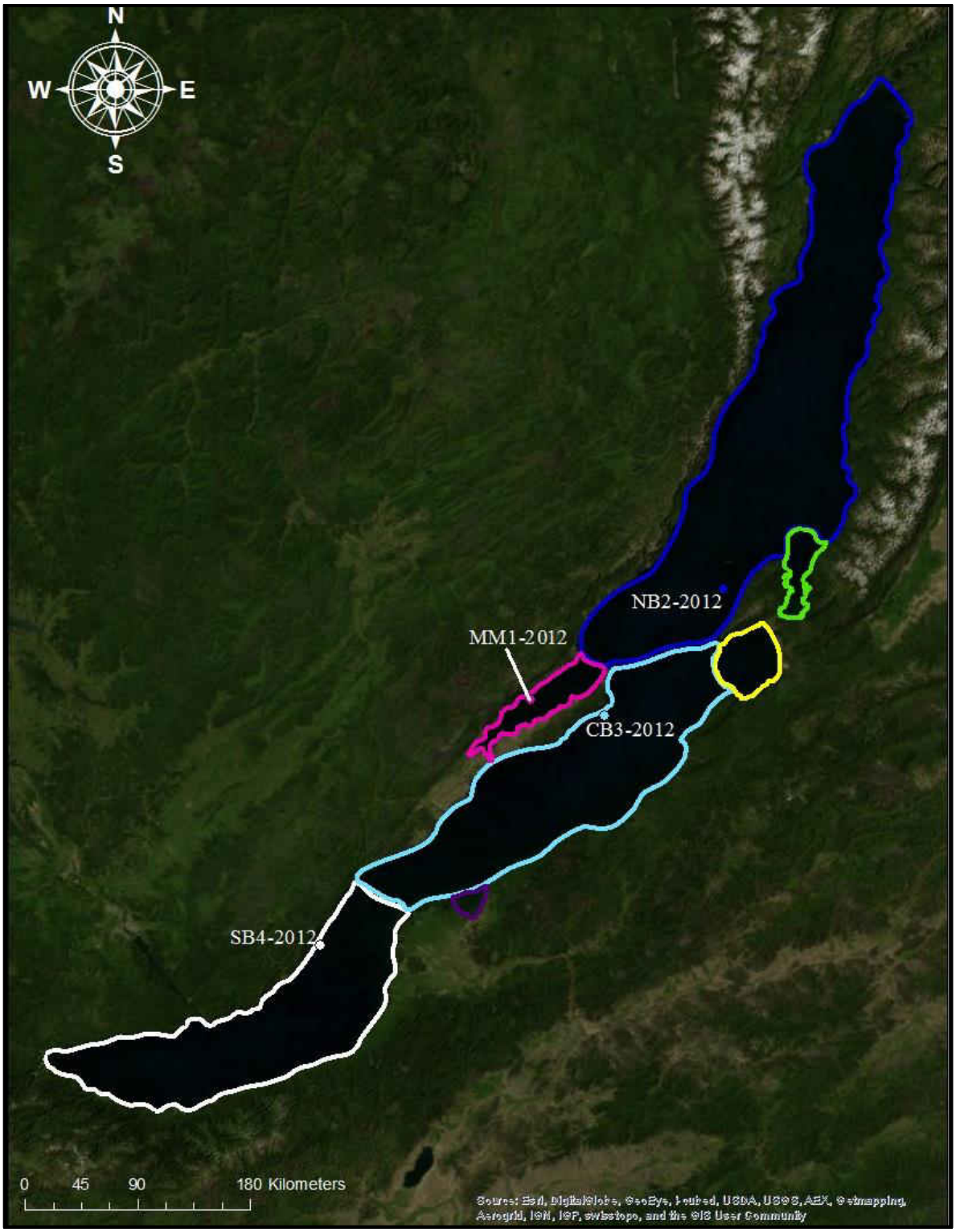


Figure 5 represents transcriptome sampling sites from summer 2012 Lake Baikal sampling cruise aboard the *Professor Kozhov* and *Professor Treskov*. Sample codes are "LocationStation-Year" with "location" as follows: Northern Basin (NB), Central Basin (CB), Southern Basin (SB), and Maloe More (MM).

Sample Processing

RNA Extraction. Qiagen's RNeasy kit was used for RNA extraction as previously explained in Reference Transcriptome: *RNA Extraction*. Samples of 120 adult individuals were isolated and counted in RNAlater from each of the 12 samples (3 from each of 3 basins, 3 from Maloe More, Figure 5). The sample concentrations (>10 ng/μL) were then confirmed with nanodrop and gel electrophoresis before cDNA libraries were constructed. All samples were placed in TE buffer in a foil-sealed 96-well PCR plate for library construction. The remaining volumes of samples were flash frozen and stored at -80°C.

Library Construction and Illumina Paired-End Sequencing. Library construction and *Illumina* paired-end sequencing were completed by Beckman Coulter Genomics. Libraries (cDNA) were constructed per the manufacturer's specifications (Illumina 2011). The binary data file generated by the GS-FLX sequencing instrument software was converted to multiple-FASTA format and reads quality saved in multiple-QUAL format. Quality values express negative base-10 log of the basecall error probability. Quality filtered *Illumina* data were saved in FASTQ containing both basecalls and ASCII encoded quality values.

Analysis

Illumina reads were mapped to the *454 Rosche* partial transcriptome using Trinity/RSM (Grabherr et al. 2011). Differential expression analysis was conducted by edgeR (Robinson et al. 2013) by Beckman Coulter Genomics on centered TMM-normalized FPKM (fragments per kilobyte of length per million reads). The 2-way cluster analysis was recreated by the author using JMP 11 (SAS Institute Inc. 2013). A principal components analysis (PCA) on correlations and a color map on sample correlations were completed.

tBLASTn. A subset of 863 transcripts with a very conservative false discovery rates of <0.001 (FDR=0.001) was made. These transcripts were used to search for similar sequences using the National Center for Biotechnology and Information's (NCBI) translated BLAST for nucleotide databases (tBLASTn) within the *Drosophila* database of reference RNA sequences (refseq_rna) (Geer et al. 2010). Additionally, these transcripts were used to search for similar sequences using NCBI's tBLASTn within the "Invertebrates" database of nucleotide collection (nr/nt) (Geer et al. 2010). The highest matched result of both searches for each transcript was annotated as the closest putative homologous transcript within *Epischura baikalensis*. If no sequence matched, the sequence was considered novel to *E. baikalensis*.

Gene Ontology. Gene ontology results as CG identification numbers from the *Drosophila* database tBLASTn were then probed against FlyBase, a database for *Drosophila* genes and genomes (St. Pierre et al. 2014). Putative molecular function, cellular component, and biological process were collected per transcripts that had a matching sequence from tBLASTn results. Output matches from the tBLASTn against "Invertebrates" that included gene ontology information were parsed for ontologies and merged with *Drosophila* putative ontologies to obtain the largest set of putative transcript functions.

The complete list of gene ontologies from both databases were placed into their respective categories of putative function, e.g. enzymes, structural proteins, chaperones. The remaining transcripts with annotations from gene ontologies were grouped into an "unknown/other" category. Functions for these transcripts were either unknown at the time of analysis, had poorly annotated gene ontologies, or did not fall into the other categories. Once the annotated transcripts had been classified into ontological groups, they were separated by the

cluster of transcripts from which they came. This allowed for a comparison of differential gene expression by cluster.

Genetic Differentiation

Genetic differentiation samples were taken from the summer 2012 transcriptome samples (Figure 5) and were the same biological samples as the Differential Gene Expression samples. Samples included sites from the 3 major basins (Northern, Central, and Southern) and from the shallow inlet Maloe More.

Sample Collection

Sample collection is the same as in Differential Gene Expression.

Sample Storage

Sample storage is the same as in Differential Gene Expression.

Sample Processing

Sample processing is the same as in Differential Gene Expression.

Analysis

Illumina reads were mapped to the *454 Rosche* partial transcriptome using Trinity/RSM (Grabherr et al. 2011). Allelic frequencies of variant alleles were calculated for each significantly differentiated ($p<0.05$) single nucleotide polymorphism (SNP) per transcript. Fixation indices for subpopulations to total population (F_{ST}) were found for individual SNPs using methods outlined by Weir and Cockerham (1984). Average F_{ST} values were found using mean numerator divided by mean denominator values as a more robust estimate of average F_{ST} for each SNP (Weir and Cockerham 1984) of transcripts that had >1 significantly differentiated SNP. All 131 347 average F_{ST} values were then bootstrapped to estimate the variance, i.e., standard error, with 400 replicate resamplings. One way ANOVA were used to determine

variance between subpopulations, and all pairs Tukey's HSD test with 95% confidence intervals were used to compare means in site-by-site comparisons.

A principal components analysis (PCA) based on minor allele frequency of 2000 randomly selected highly significantly differentiated SNPs was completed; expected F_{ST} values for randomly assigned alleles to subpopulations were found using a Poisson distribution and compared to actual F_{ST} values using JMP 11 (SAS Institute Inc. 2013). For population size estimates, a neutral mutation rate of 10^{-8} was assumed, and effective population size (N_e) and migration rate (N_m) were estimated from average F_{ST} values using methods and equations outlined by Weir and Cockerham (1984).

SNPs that were contained within transcripts that were enriched for FDR<0.001 or FDR<0.05 or that were in the top 5% of F_{ST} values were included in further analysis of gene ontologies. Gene ontologies collected from the method previously mentioned in *Differential Expression* were used for genetic differentiation analysis. Contingency tables were used to test for significant genetic differentiation in gene ontology groups with a Bonferroni correction for multiple comparisons.

CHAPTER 4

ZOOPLANKTON DISTRIBUTION

Results

Whole Lake

There were significant differences of zooplankton distribution among the lake's different regions (Figures 6 – 11, Tables 1 – 5).

Epischura baikalensis. *Epischura baikalensis* abundances were significantly higher in open Baikal samples (Northern, Central, and Southern Basins) and Maloe More samples than in either Barguzin or Chivrikuj Bays (Figure 6, Table 1). There were no significant differences between Maloe More and open Baikal or between Barguzin and Chivrikuj Bays.

Daphnia longispina. Meanwhile, *Daphnia longispina* abundances were significantly higher in Barguzin and Chivrikuj Bays than in open Baikal samples (Figure 7, Table 2). *Daphnia longispina* was significantly more abundant in Chivrikuj Bay than in Maloe More and nearly significantly more abundant in Barguzin Bay than in Maloe More. There were no significant differences between Maloe More and open Baikal or between Barguzin and Chivrikuj Bays.

Cyclops spp. *Cyclops* spp. followed the same trend as *Daphnia longispina* (Figure 8, Table 3). They were significantly more abundant in Barguzin and Chivrikuj Bays than in either open Baikal or Maloe More. However, there were significantly more *Cyclops* spp. in Chivrikuj Bay than in Barguzin Bay. There were no significant differences between Maloe More and open Baikal.

Diaptomus graciloides. *Diaptomus graciloides* were significantly more abundant in Chivrikuj Bays than in open Baikal or Maloe More (Figure 9, Table 4). There were no significant differences between Barguzin Bay, Maloe More and open Baikal abundances nor was there a significant difference between the abundances of Barguzin Bay and Chivrikuj Bay.

Macrohectopus branickii. No *Macrohectopus branickii* were found in the 2012 samples from Barguzin Bay, Chivrikuj Bay, or Maloe More. Several were found in open Baikal, but there was no significant difference in abundance throughout the lake (Figure 10, Table 5).

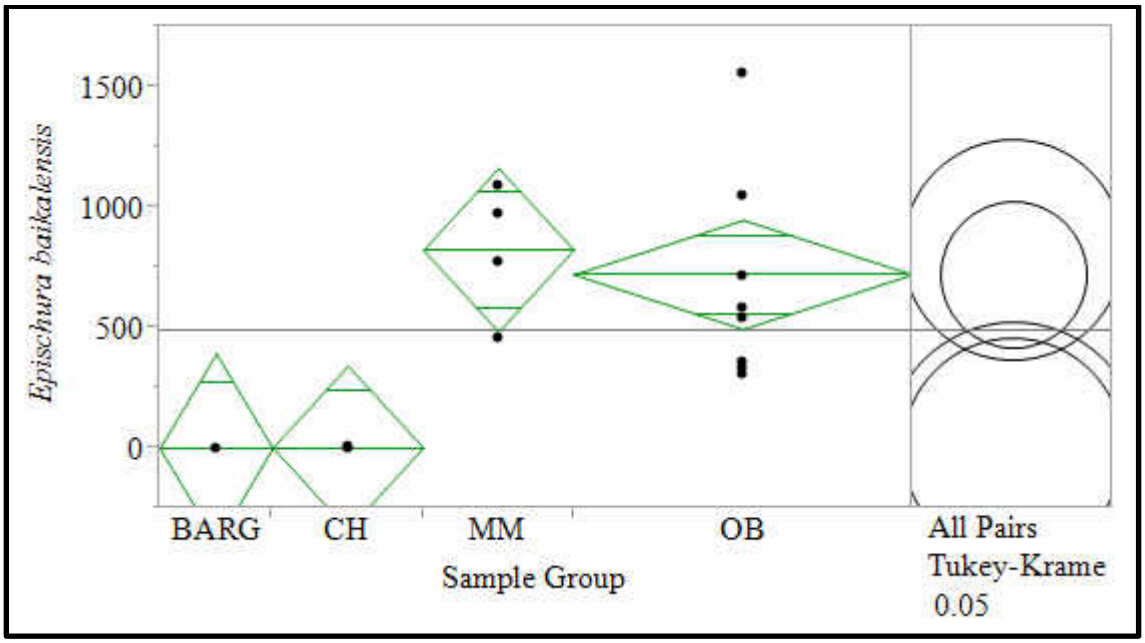


Figure 6 represents the differences in *Epischura baikalensis* raw abundance in samples from different regions of Lake Baikal, $R^2_{16} = 0.6125$, $p = 0.0014$. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins). Means comparisons by site from all pairs Tukey's HSD test with 95% confidence interval (right) can be found in Table 1.

Table 1 represents the means comparisons by sites from all pairs Tukey's HSD test with 95% confidence interval for *Epischura baikalensis* raw abundance in samples from different regions of Lake Baikal. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins).

Means Comparison Epischura baikalensis Abundance

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
MM	BARG	823.5000	244.8137	123.083	1523.917	0.0186*
MM	CH	821.5000	226.6534	173.040	1469.960	0.0110*
OB	BARG	721.2222	213.6909	109.848	1332.596	0.0182*
OB	CH	719.2222	192.6184	168.137	1270.307	0.0088*
MM	OB	102.2778	192.6184	-448.807	653.363	0.9503
CH	BARG	2.0000	244.8137	-698.417	702.417	1.0000

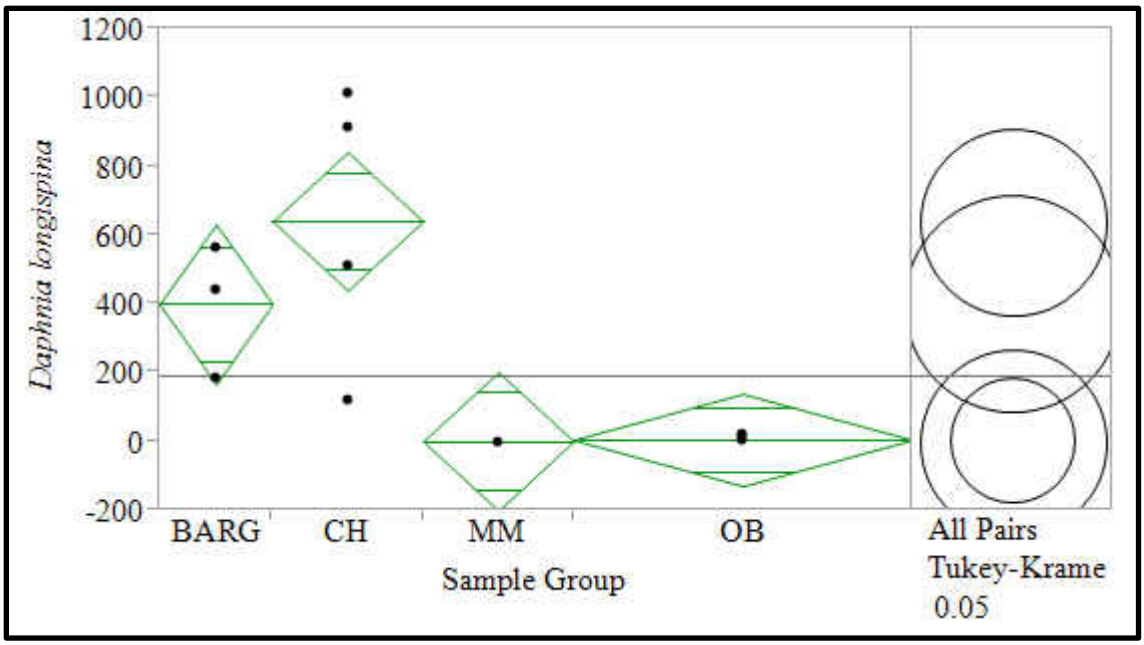


Figure 7 represents the differences in *Daphnia longispina* raw abundance in samples from different regions of Lake Baikal, $R^2_{16} = 0.7075$, $p = 0.0002$. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins). Means comparisons by site from all pairs Tukey's HSD test with 95% confidence interval (right) can be found in Table 2.

Table 2 represents the means comparisons by sites from all pairs Tukey's HSD test with 95% confidence interval for *Daphnia longispina* raw abundance in samples from different regions of Lake Baikal. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins).

Mean Comparisons Daphnia longispina Abundance

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
CH	MM	639.7500	134.2213	255.740	1023.760	0.0011*
CH	OB	635.5556	114.0662	309.210	961.901	0.0002*
BARG	MM	398.4167	144.9756	-16.361	813.195	0.0620
BARG	OB	394.2222	126.5451	32.174	756.270	0.0305*
CH	BARG	241.3333	144.9756	-173.445	656.111	0.3731
OB	MM	4.1944	114.0662	-322.151	330.540	1.0000

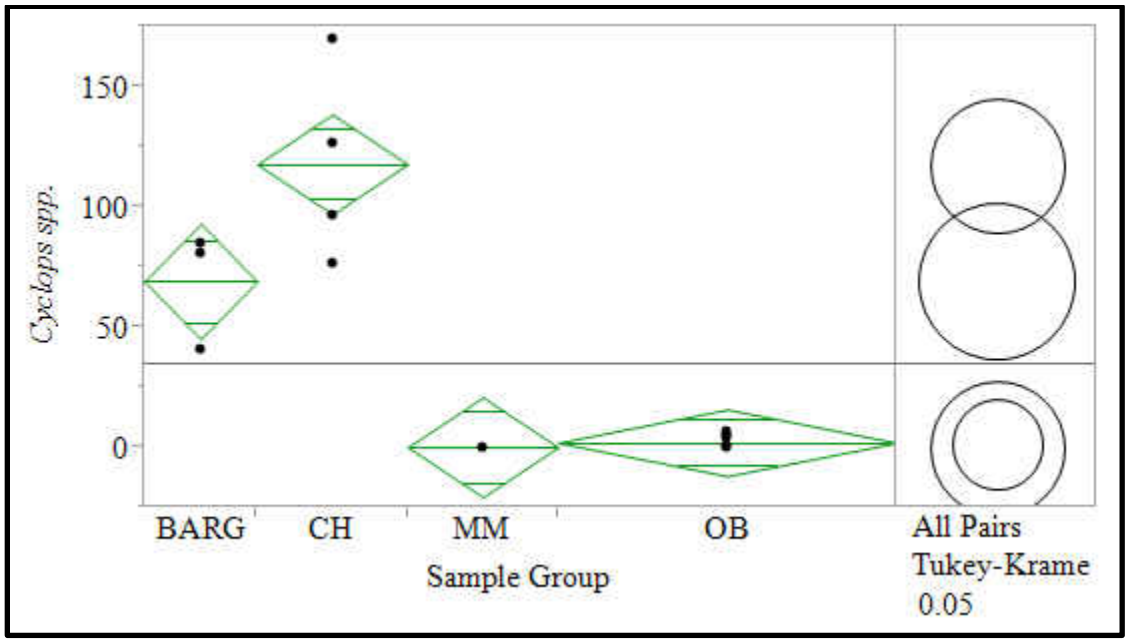


Figure 8 represents the differences in *Cyclops* spp. raw abundance in samples from different regions of Lake Baikal, $R^2_{16} = 0.8812$, $p = 0.0001$. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins). Means comparisons by site from all pairs Tukey's HSD test with 95% confidence interval (right) can be found in Table 3.

Table 3 represents the means comparisons by sites from all pairs Tukey's HSD test with 95% confidence interval for *Cyclops* spp. raw abundance in samples from different regions of Lake Baikal. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins).

Means Comparisons Cyclops spp. Abundance

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
CH	MM	117.7500	13.86276	78.0884	157.4116	<.0001*
CH	OB	115.8611	11.78108	82.1552	149.5670	<.0001*
BARG	MM	69.0000	14.97350	26.1605	111.8395	0.0015*
BARG	OB	67.1111	13.06994	29.7178	104.5045	0.0005*
CH	BARG	48.7500	14.97350	5.9105	91.5895	0.0231*
OB	MM	1.8889	11.78108	-31.8170	35.5948	0.9985

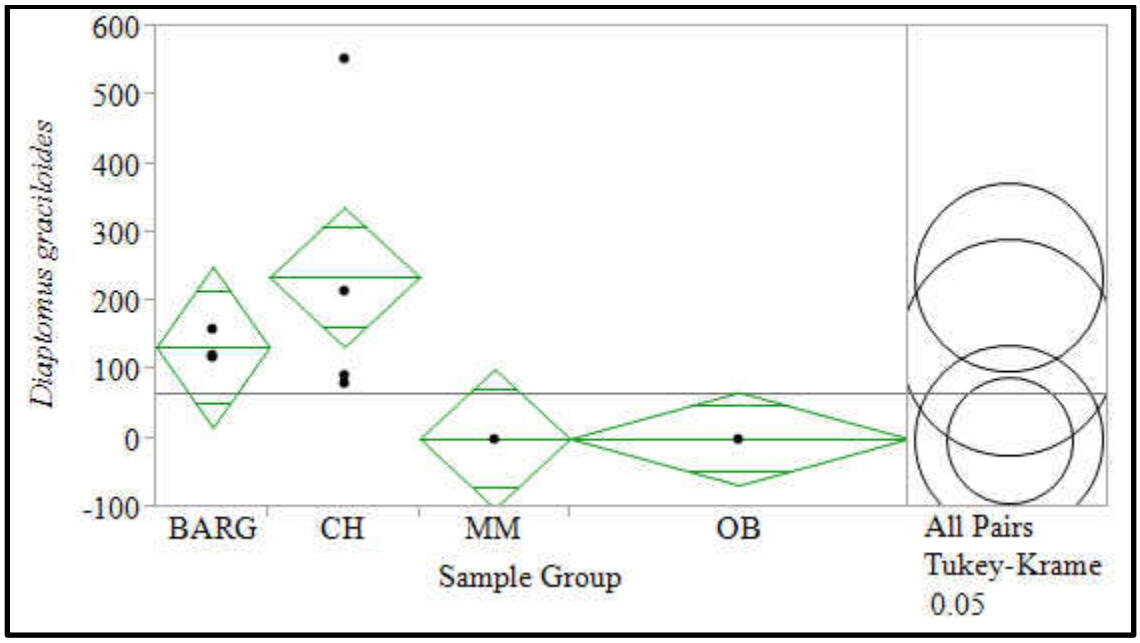


Figure 9 represents the differences in *Diaptomus graciloides* raw abundance in samples from different regions of Lake Baikal, $R^2_{16} = 0.5578$, $p = 0.0038$. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins). Mean comparisons by site from all pairs Tukey's HSD test with 95% confidence interval (right) can be found in Table 4.

Table 4 represents the means comparisons by sites from all pairs Tukey's HSD test with 95% confidence interval for *Diaptomus graciloides* raw abundance in samples from different regions of Lake Baikal. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins).

Means Comparisons Diaptomus graciloides Abundance

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
CH	MM	235.2500	67.66125	41.670	428.8302	0.0148*
CH	OB	235.2500	57.50100	70.738	399.7615	0.0043*
BARG	MM	133.3333	73.08250	-75.757	342.4238	0.2984
BARG	OB	133.3333	63.79163	-49.176	315.8425	0.1983
CH	BARG	101.9167	73.08250	-107.174	311.0072	0.5204
OB	MM	0.0000	57.50100	-164.512	164.5115	1.0000

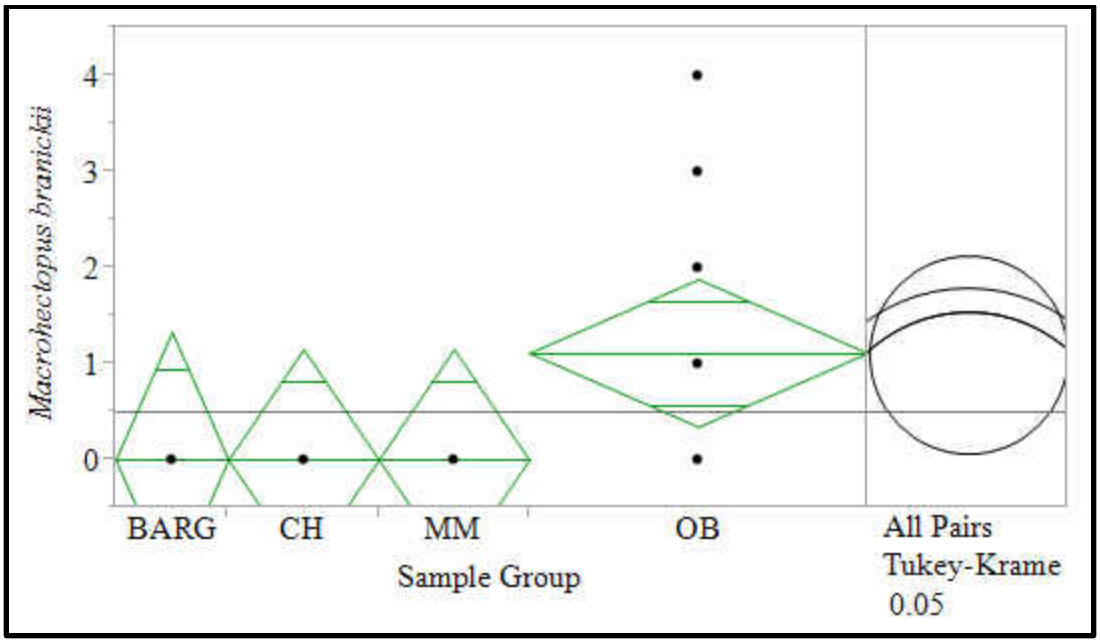


Figure 10 represents the differences in *Macrohectopus branickii* raw abundance in samples from different regions of Lake Baikal, $R^2_{16} = 0.2444$, $p = 0.2020$. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins). Mean comparisons by site from all pairs Tukey's HSD test with 95% confidence interval (right) can be found in Table 5.

Table 5 represents the means comparisons by sites from all pairs Tukey's HSD test with 95% confidence interval for *Macrohectopus branickii* raw abundance in samples from different regions of Lake Baikal. Sample groups are such as follows: Barguzin Bay (BARG), Chivrikuj Bay (CH), Maloe More (MM), Open Baikal (OB-includes Northern, Central, and Southern Basins).

Means Comparisons Macrohectopus branickii Abundance

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
OB	BARG	1.111111	0.7243558	-0.96129	3.183508	0.4416
OB	CH	1.111111	0.6529255	-0.75692	2.979144	0.3548
OB	MM	1.111111	0.6529255	-0.75692	2.979144	0.3548
CH	BARG	0.000000	0.8298538	-2.37423	2.374229	1.0000
MM	BARG	0.000000	0.8298538	-2.37423	2.374229	1.0000
MM	CH	0.000000	0.7682954	-2.19811	2.198109	1.0000

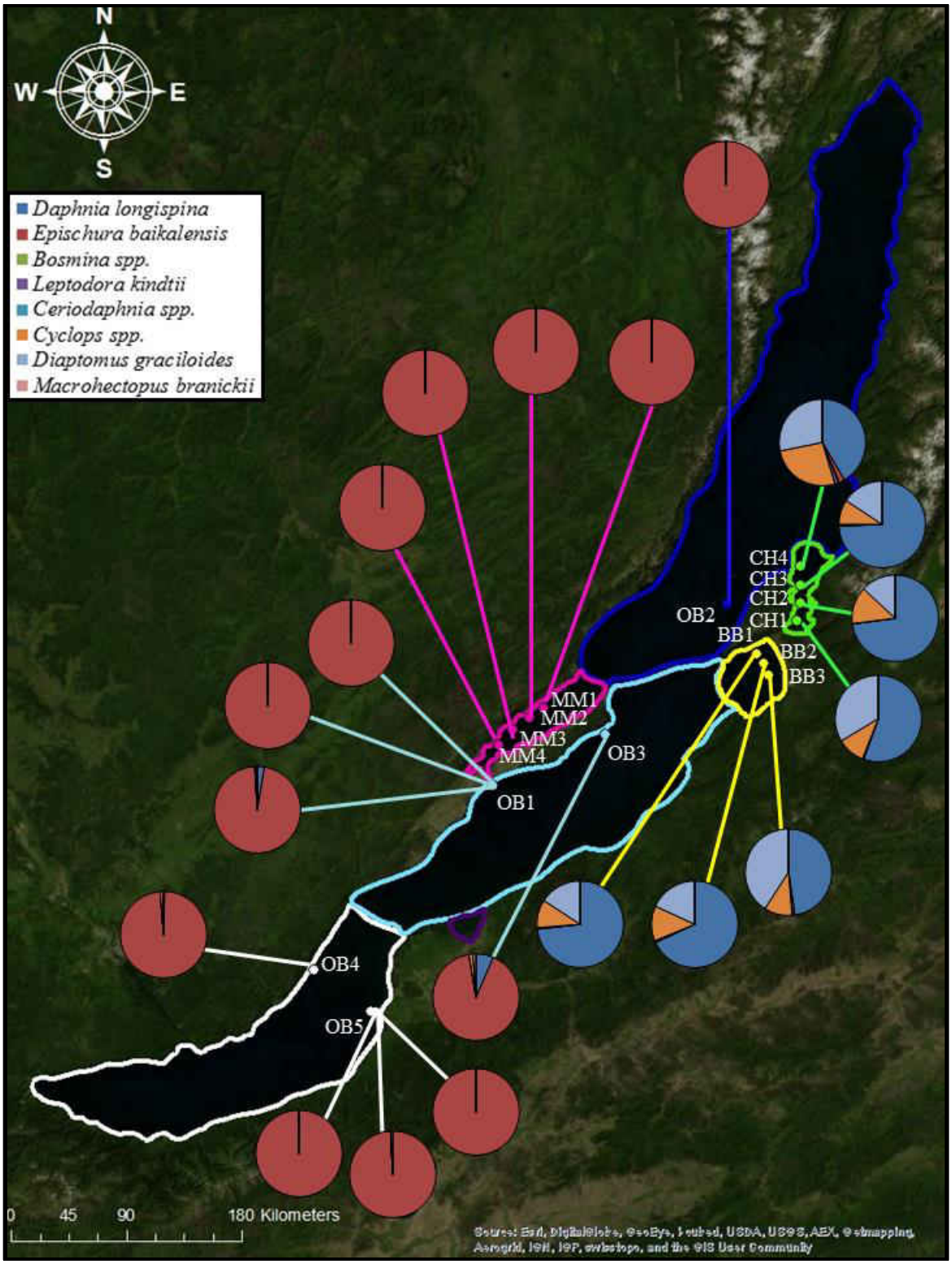


Figure 11 represents zooplankton raw abundance by sampling site from different regions of Lake Baikal. Sample code is as follows: "LocationStation" with "location" codes being open Baikal (OB), Chivrikuj Bay (CH), Barguzin Bay (BB), and Maloe More (MM). See Figure 1 for region color codes.

Chivrikuj Bay

Zooplankton Abundance. There were significant differences in zooplankton distribution within Chivrikuj Bay (Figures 12 – 13). Farther into the bay, i.e., more south, *Epischura baikalensis* percent abundances decreased significantly with distance into the bay, measured as minutes north of the 53rd parallel North (Figure 12). Conversely, *Daphnia longispina*, *Cyclops* spp., and *Diaptomus graciloides* percent abundances increased significantly with distance into the bay (Figure 12).

Abiotic Factors. There were significant differences in abiotic factors within Chivrikuj Bay (Figure 14). Farther into the bay, i.e., more south, water temperature, surface temperature, and percent dissolved oxygen increased significantly (Figure 14). Farther into the bay, depth and chlorophyll concentration decreased significantly (Figure 14).

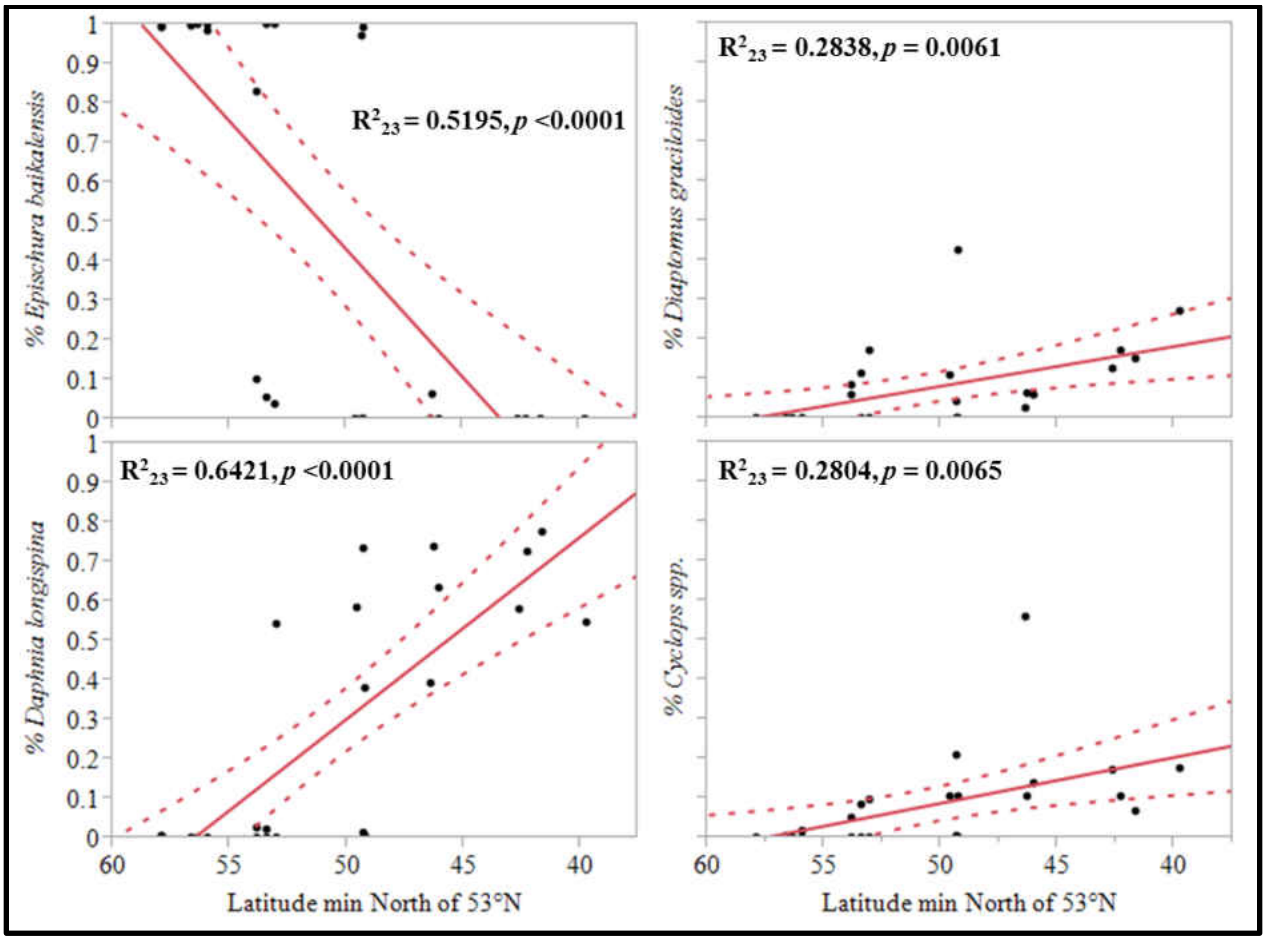


Figure 12 represents bivariate analyses of percent zooplankton abundances of Chivrikuj Bay, Lake Baikal. Minutes from the 53rd parallel North corresponds to a North – South direction, traveling farther south and farther into the bay, e.g., reading horizontal axis left to right corresponds to North (opening) to South (shore) direction. Solid lines represent regressions with 95% confidence intervals represented by dashed lines. Top left: *Epischura baikalensis*, R^2_{23} = 0.5195, p <0.0001. Above left: *Daphnia longispina*, R^2_{23} = 0.6421, p <0.0001. Top right: *Diaptomus graciloides*, R^2_{23} = 0.2838, p = 0.0061. Above right: *Cyclops* spp., R^2_{23} = 0.2804, p = 0.0065.

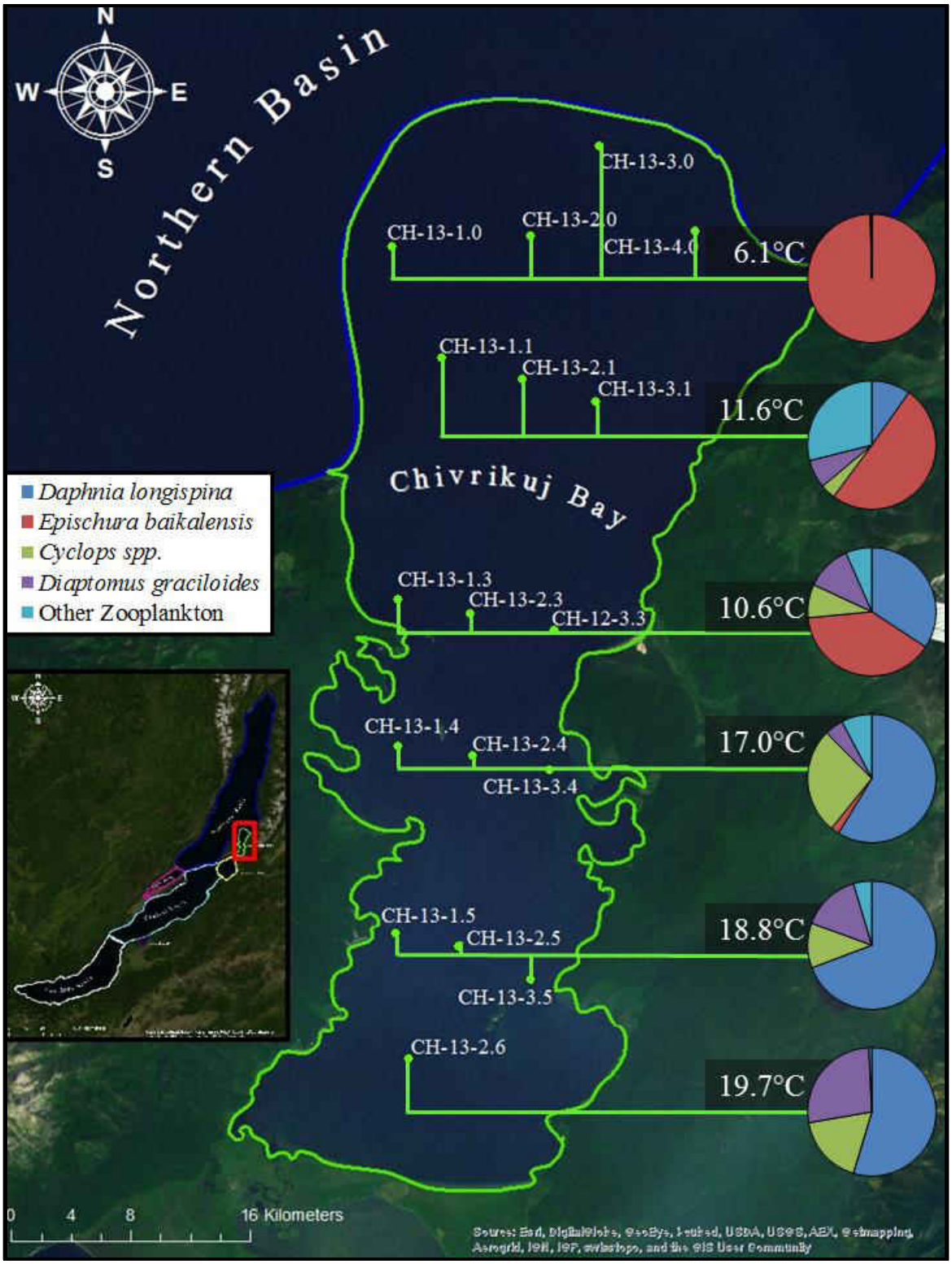


Figure 13 represents zooplankton raw abundance by sampling site from Chivrikuj Bay, Lake Baikal. Sample code is as follows: "Location-Year-Traverse.Station" with "location" codes being Chivrikuj Bay (CH). All samples from parallel stations of each respective traverse were combined as biological replicates.

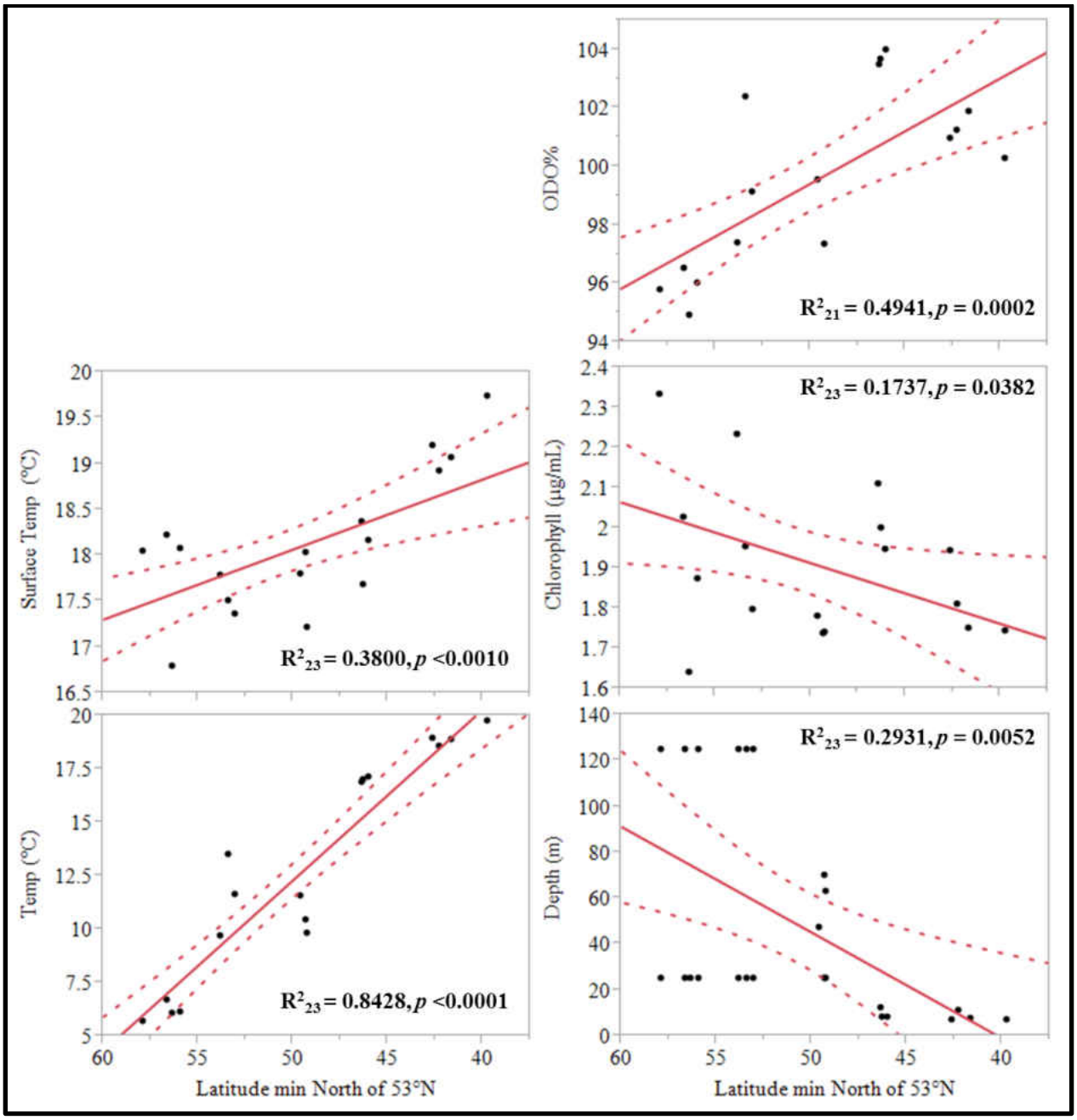


Figure 14 represents bivariate analyses of the abiotic factors of Chivrikuj Bay, Lake Baikal. Minutes from the 53rd parallel North corresponds to a North – South direction, traveling farther south and farther into the bay, e.g., reading horizontal axis left to right corresponds to North (opening) to South (shore) direction. Solid lines represent regressions with 95% confidence intervals represented by dashed lines. Top left: Surface Water Temperature, $R^2_{23} = 0.3800$, $p < 0.0010$. Above left: Water Temperature, $R^2_{23} = 0.8428$, $p < 0.0001$. Top right: Dissolved Oxygen, $R^2_{21} = 0.4941$, $p = 0.0002$. Center right: Chlorophyll Concentration, $R^2_{23} = 0.1737$, $p = 0.0382$; Bottom right: Depth, $R^2_{23} = 0.2931$, $p = 0.0052$.

Discussion

Whole Lake

The zooplankton population varies depending on the location within the lake. Evidence suggests that a major factor in having diverse assemblages, e.g. those that include invasive cosmopolitans, is temperature. However, Maloe More, with its temperature not being significantly different from that of Chivrikuj Bay (Figure 15) lacks this uniform assemblage of cosmopolitans. Maloe More's assemblage mimics that of open Baikal despite its different hydrologic parameters. Maloe More is significantly deeper than Chivrikuj Bay and is subject to stronger currents having an inflow and outflow. This could suggest that the thermocline of Maloe More may be providing *Epischura baikalensis* with a suitably low water temperature such that it remains competitive, e.g., *E. baikalensis* can migrate below the thermocline should surface water temperatures rise above its thermal tolerance. Whereas Chivrikuj Bay lacks the depth for a substantial thermocline, especially in its shallow reaches deep in the bay, *E. baikalensis* may have no suitable habitat from which to escape higher temperatures (>15°C).

The highly circulating currents of Maloe More may be preventing invasion by *Daphnia longispina* and the other cosmopolitans seen in Chivrikuj and Barguzin Bays. Any cosmopolitans that could survive in Maloe More may be getting quickly removed into the very cold water of open Baikal (4°C), where their survival is decreased. In addition to the lower temperatures, open Baikal lacks the chlorophyll concentrations usually required by indiscriminate-eating cosmopolitan species, such as *Daphnia longispina*.

It should be noted that there seems to be competitive exclusion between the endemics and invasive species. The samples were nearly, if not 100%, dominated by either cosmopolitan invasives or *Epischura baikalensis* with little overlap. It would be remiss to postulate that this

exclusion is entirely based on competitive exclusion for energy resources, however. Though energy resources may play a factor in the step-wise relationship between endemics and cosmopolitans, the relationship seems to be driven more by abiotic factors and niche specialization of the extremely stenothermic *E. baikalensis*, as was seen in the Chivrikuj Bay samples (Figures 12, 14).

The other zooplankton present throughout Lake Baikal should also be mentioned because *Cyclops* spp. and *Diaptomus graciloides* represent many of the same challenges as the invasive *Daphnia longispina*. Observing the highly significant abundances of these other generalist cosmopolitan invaders could lead to further pressures on the specialist *Epischura baikalensis*, should cosmopolitan colonies establish in open waters of Lake Baikal.

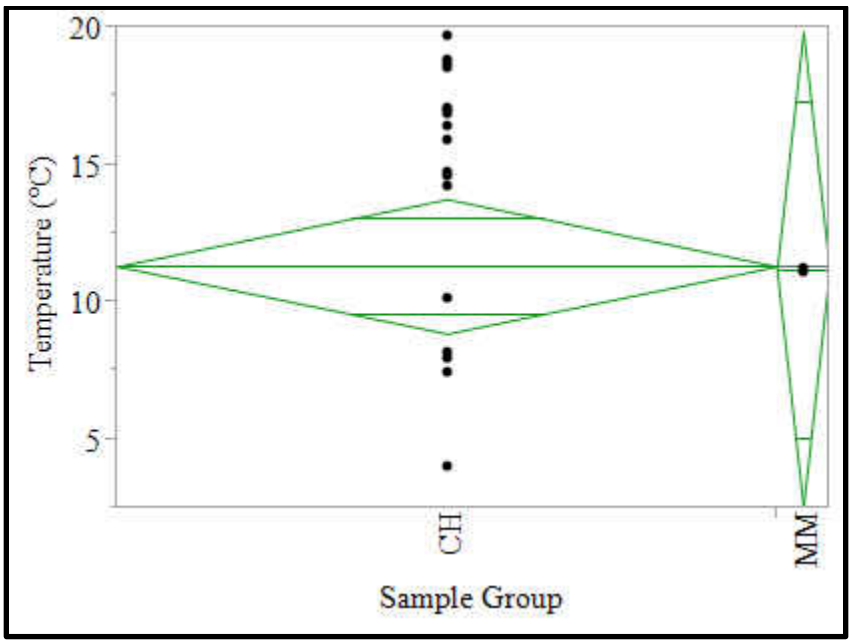


Figure 15 represents the comparison of water temperature variation between Maloe More (MM) and Chivrikuj Bay (CH), $R^2_{25} = 2.72e\text{-}5$, $p = 0.9794$.

Chivrikuj Bay

The zooplankton population within Chivrikuj Bay is quite dynamic from the bay's entrance to its shallower inner waters. The swift change from entirely *Epischura baikalensis* dominated assemblages to nearly entirely *Daphnia longispina* assemblages is striking. Both species' abundances mirror the temperature cline that exists along the 109th meridian East. Other abiotic trends mirror the steep decline in *E. baikalensis* abundance and the increase in the *D. longispina* abundance, but the most significant is temperature. Again, *E. baikalensis* seems to have a step-wise relationship with *D. longispina* where the only overlap zone is in station 3 of all 3 traverses (Figure 13). This zone also exhibits temperatures that both species can tolerate while maintaining a depth capable of giving ample thermocline protection to *E. baikalensis* should day time surface waters become too warm.

The combination of various abiotic factors is likely driving the trend in zooplankton abundance in Chivrikuj Bay. The same is likely true of Barguzin Bay; however, it is interesting that temperature may not be the only factor in the relationship. As seen in Maloe More, *E. baikalensis* has high abundance even with temperatures no different from those in Chivrikuj Bay. Hence, there is likely another factor in addition to water temperature or a factor that can mediate the major effects of higher water temperatures enabling *E. baikalensis* to survive in Maloe More but not in inner Chivrikuj Bay. The combination of shallow depth, possible lack of thermocline, and lack of strong currents is likely affecting *E. baikalensis* abundance. Meanwhile, higher dissolved oxygen in Chivrikuj Bay's shallow water, i.e., higher water-gas interface, may allow for a better habitat in which *D. longispina* and other cosmopolitan species can thrive.

A caveat of this study of zooplankton abundance is that the *Daphnia longispina* and other cosmopolitan species of Chivrikuj Bay may have not been able to colonize open Baikal and

Maloe More simply because of lack of contact with those areas. The circular current of Chivrikuj Bay could be excluding zooplankters from leaving the bay. Very few *Daphnia longispina* specimen have been found outside of the shallow bays, in either open Baikal or Maloe More, and their origin has yet to be determined. That is to say, however, should *Daphnia longispina* or other cosmopolitan species invade in substantial numbers in either Maloe More or open Baikal, colonization could be successful under the correct biotic and abiotic circumstances. Successful colonization in open Baikal's frigid waters (4°C) is unlikely for *D. longispina* or the other zooplankters studied here but could certainly happen should another stenothermic plankton species invade.

CHAPTER 5

DIFFERENTIAL GENE EXPRESSION

Results

Correlation Analysis

There was significant variation in gene expression in *Epischura baikalensis* from the lake's different regions (Figures 16, 17). From the FDR=0.001 transcripts, 4 distinct clusters were made from the differential expression data retrieved in this study. Cluster 1 transcripts had up-regulation in Maloe More samples and down-regulation in all other samples (Figures 16, 17). Cluster 2 transcripts had up-regulation in Central Basin and Southern Basin samples (Figures 16, 17). Cluster 3 transcripts were more notably down-regulated in Northern Basin samples and show mixed expression levels in Maloe More, Central Basin, and Southern Basin (Figures 16, 17). Finally, cluster 4 transcripts showed very high up-regulation in the Northern Basin and down-regulation in Maloe More; expression levels in Central and Southern Basin was mixed in cluster 4. There was a fifth cluster of genes, but despite significant differential gene expression in Lake Baikal samples, none of these transcripts mapped to either invertebrates or *Drosophila*. Hence, cluster 5 was excluded from further analysis due to lack of annotation for those transcripts. Cluster 5 transcripts showed down-regulation in Maloe More and Central Basin, while remaining up-regulated and mixed in Northern and Southern Basins, respectively.

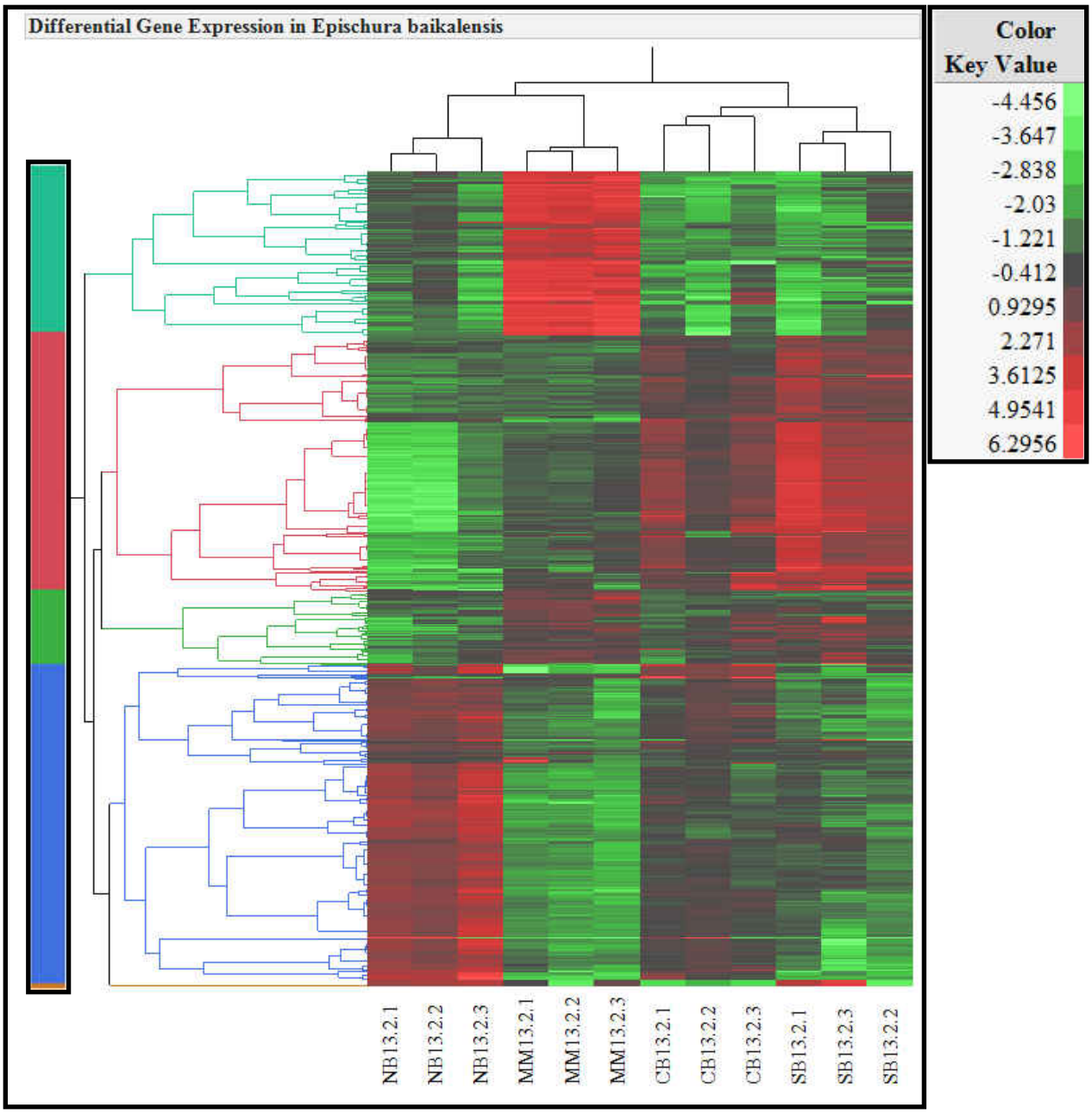


Figure 16 represents a dendrogram and heat map of different gene expression of 863 transcripts (FDR=0.0001). Down-regulation (green) and up-regulation (red) are clustered by transcript family (left) and by sample site (dendrogram: top; sample label: bottom). Sample name is "LocationYear.Station.Replicate" with "location" code as follows: Northern Basin (NB), Central Basin (CB), Southern Basin (SB), Maloe More (MM). Clusters, with their corresponding color and primary characteristic expression, are as follows: Cluster 1 (teal; MM up-regulation), Cluster 2 (red; NB down-regulation), Cluster 3 (green; NB and CB down-regulation), Cluster 4 (blue; NB up-regulation), Cluster 5 (orange; no primary characteristic pattern).

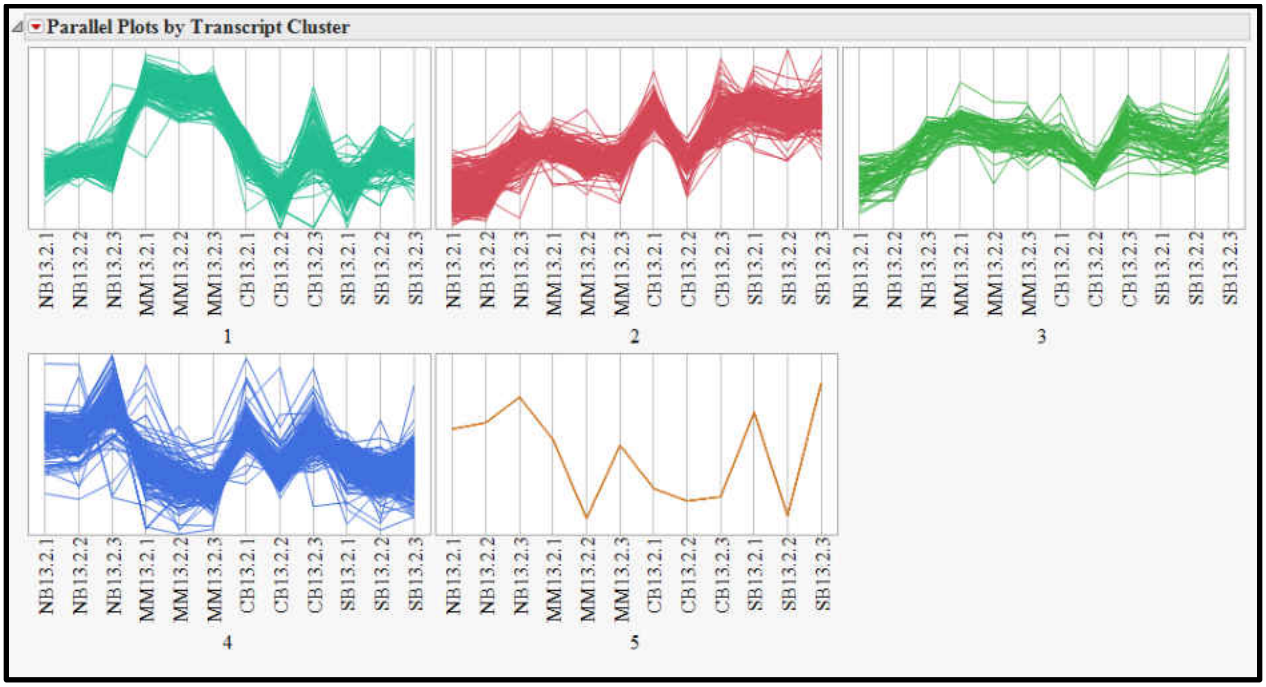


Figure 17 represents parallel plots by transcript cluster for the multivariate cluster analysis. Up and down trends correspond to up-and down-regulation of clustered transcripts. Cluster numbers and colors correspond to the color families from Figure 16. Sample name is "LocationYear.Station.Replicate" with "location" code as follows: Northern Basin (NB), Central Basin (CB), Southern Basin (SB), Maloe More (MM). Clusters, with their corresponding color and primary characteristic expression, are as follows: Cluster 1 (teal; MM up-regulation), Cluster 2 (red; NB down-regulation), Cluster 3 (green; NB and CB down-regulation), Cluster 4 (blue; NB up-regulation), Cluster 5 (orange; no primary characteristic pattern).

From the PCA on correlations, the transcripts clustered in tight clouds in their respective clusters within 2 dimensions (Figures 18, 19). Principal components 1 and 2 explained 85.9% of the variation within the correlations of transcripts (Figure 18). All sampling sites clustered heavily together by location. Nearly all transcripts variation was explained by 2 principal components expressing the strong correlations between transcript expression levels.

Two principal components explained 81.7% of the variation within the correlations of sample sites (Figure 19). Cluster 1 transcripts were dominated by the extreme down-regulation

of genes in the 3 basins (Northern, Central, and Southern). Cluster 2 transcripts were dominated by the up-regulation of genes in individuals from the Central and Southern Basins. Cluster 3 transcripts were dominated by the down-regulation in Maloe More, Central Basin, and Southern Basin. Finally, cluster 4 transcripts were strongly driven by the down-regulation of genes in Maloe More, Central Basin, and Southern Basin, coupled with the intense up-regulation of genes in the Northern Basin. From the correlations summary (Figure 19), samples from each area clustered near other samples from the same region without overlap. This analysis suggested that *Epischura baikalensis* from the 4 distinct areas sampled had very different gene expression levels largely with high correlation to location.

A sample by sample correlation color map helped visualize direct correlation between samples (Figure 20). The Central and Southern Basin expression levels had positive correlations to each other, while Maloe More and Northern Basin expression levels only correlated within their respective subsamples and, in fact, were highly negatively correlated with most other samples (Figure 20).

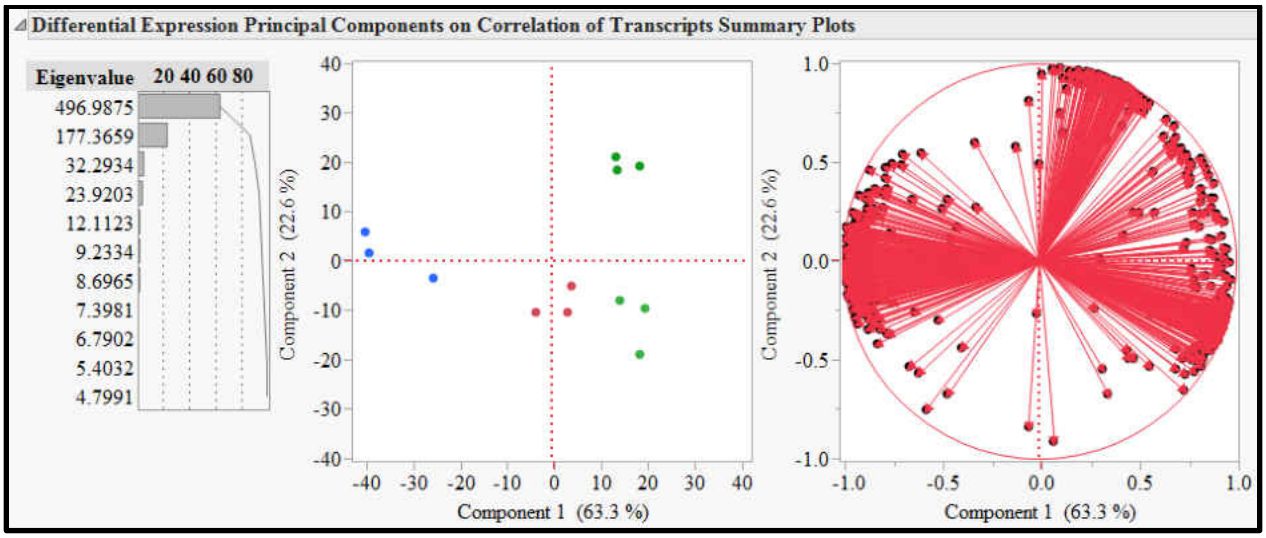


Figure 18 represents a Principal Components Analysis (PCA) on correlations of transcripts. Above left: eigenvalues for principal components. Above middle: principal components 1 and 2 explain 85.9% variation in correlations. Above right: summary map of transcript vectors. Marker color corresponds to sample site as follows: Northern Basin (blue), Central Basin (red), Southern Basin (green), Maloe More (teal).

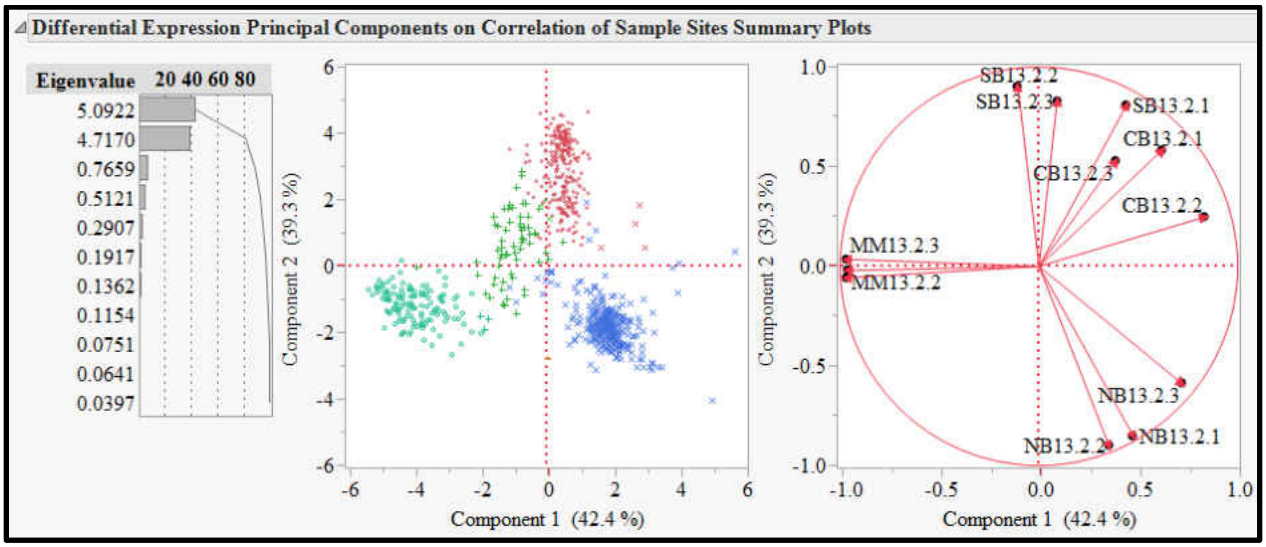


Figure 19 represents a Principal Components Analysis (PCA) on correlations of sample sites. Above left: eigenvalues for principal components. Above middle: principal components 1 and 2 explain 81.7% variation in correlations. Above right: summary map of sample vectors. Sample name is "LocationYear.Station.Replicate" with "location" code as follows: Northern Basin (NB), Central Basin (CB), Southern Basin (SB), Maloe More (MM).

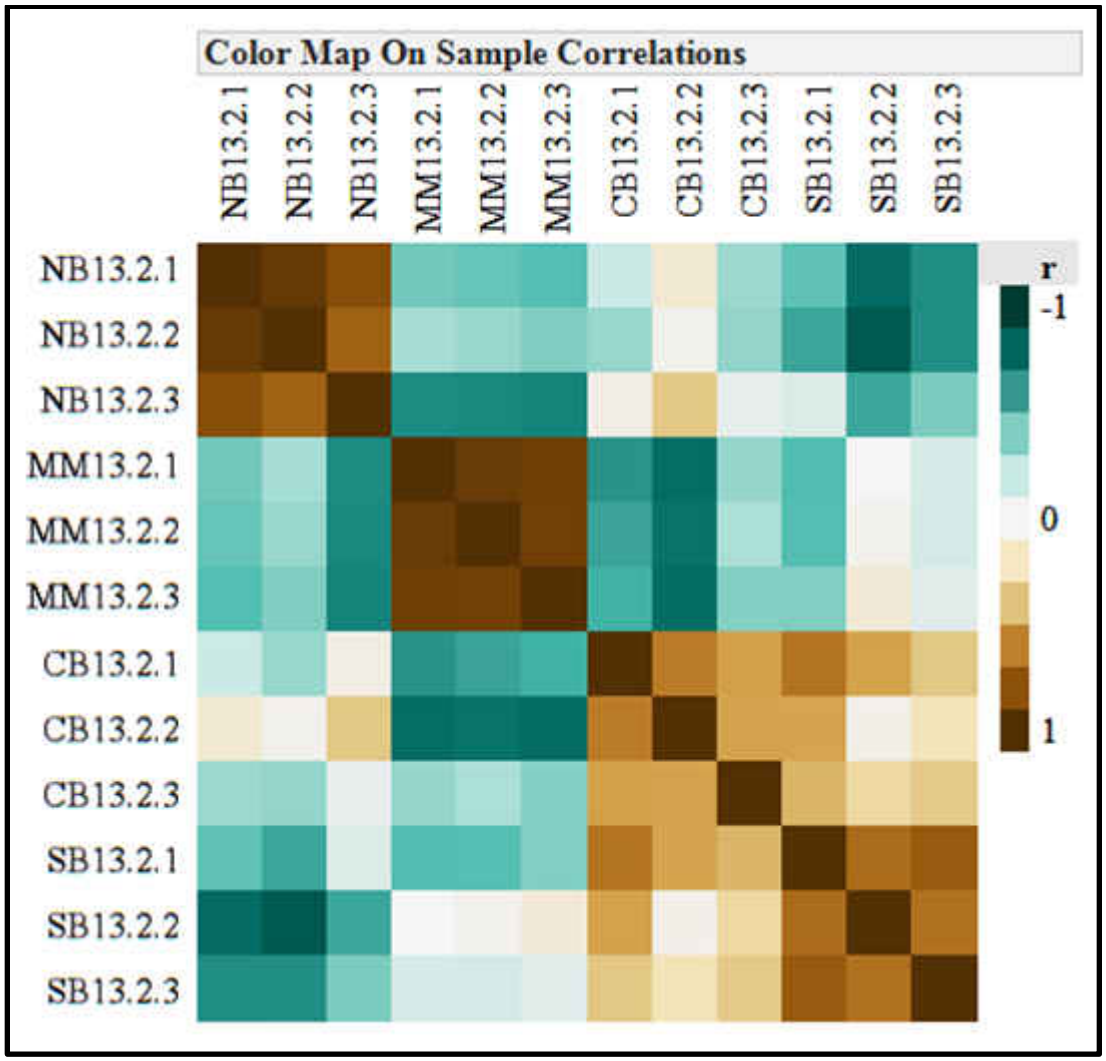


Figure 20 represents a color map of sample by sample correlations. Highly positive correlations (brown), no correlations (white), and highly negative correlations (teal) represent how similar replicates are within samples and among other samples. Sample name is "Location Year.Station.Replicate" with "location" code as follows: Northern Basin (NB), Central Basin (CB), Southern Basin (SB), Maloe More (MM).

Transcript Annotation and Gene Ontology

Many transcripts from the subset of 863 transcripts differentially expressed with FDR=0.001 were able to be mapped to either *Drosophila* or invertebrate sequences with known

ontologies. Ontologies grouped by cluster show variation in which groups of transcripts were either up-or down-regulated among the samples (Table 6).

Table 6 represents the transcripts belonging to different ontologies from all known putative functions of the 863 significantly differentially expressed (FDR=0.001) transcripts in *Epischura baikalensis* by cluster. Cluster numbers refer to transcript cluster families from Figure 16. Transcript functions are from alignments to *Drosophila* and invertebrate sequencing databases. Some transcripts matched both databases for the same putative function and were counted once toward that function, i.e. "total" may not equal total from 4 clusters combined. For a more detailed view of which annotations came from which database see Table 7.

Gene Ontology	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Actin Binding Proteins	1	1	8	8	18
ATP Binding Proteins	1	3	7	3	14
Chaperones (Protein Folding, HSP)	0	0	3	0	3
DNA Binding Proteins	1	0	7	3	11
Enzymes	28	21	36	28	113
Gene Expression (Splicing)	1	8	2	6	17
Glutathione Proteins	2	1	1	0	4
Histones	1	0	0	2	3
Nucleic Acid Binding Proteins	1	7	13	9	30
Oxidoreductases (Redox Enzymes)	1	0	8	5	14
Peptidases	17	14	5	8	44
Ribosomal Proteins	3	5	1	2	11
RNA Binding Proteins	0	6	1	3	10
Structural Proteins	0	5	1	12	18
Transcription Factors	0	0	3	1	4
Transcription Proteins	3	0	5	3	11
Translation Proteins	1	5	2	2	10
Transporter Proteins	13	2	8	3	26

Several transcripts aligned to both *Drosophila* and invertebrate databases with the same putative functions (Table 7). Transcripts were often annotated with several putative functions, and many fell into one or more categories, e.g. peptidases fall into both "peptidases" and "enzymes" categories (Tables 6, 7). Many transcripts aligned to no significant sequences in

either database, and many more aligned to genes/transcripts with unknown functions or were annotated poorly. These transcripts were placed into an "unknown/others" category not shown here due to difficulty in parsing known "unknown" functions from no significant alignments.

Table 7 represents the transcripts belonging to different ontologies from all known putative functions of the 863 significantly differentially expressed (FDR=0.001) transcripts in *Epischura baikalensis* by alignment database. Transcript functions are from alignments to *Drosophila* and invertebrate sequencing databases. Some transcripts matched both databases for the same putative function and were counted once toward that function, i.e. "total" may not equal total from 4 clusters combined.

Gene Ontology	***Drosophila***	**Invertebrates**	**Total Transcripts**
Actin Binding Proteins	12	6	18
ATP Binding Proteins	14	0	14
Chaperones (Protein Folding, HSP)	0	3	3
DNA Binding Proteins	11	0	11
Enzymes	110	3	113
Gene Expression (Splicing)	17	0	17
Glutathione Proteins	0	4	4
Histones	0	3	3
Nucleic Acid Binding Proteins	30	0	30
Oxidoreductases (Redox Enzymes)	14	0	14
Peptidases	40	5	44
Ribosomal Proteins	5	9	11
RNA Binding Proteins	10	0	10
Structural Proteins	18	0	18
Transcription Factors	3	1	4
Transcription Proteins	5	6	11
Translation Proteins	6	4	10
Transporter Proteins	5	21	26

All clusters had most function in the enzymes category (Figures 21, 22). The next highest variability expression transcripts were peptidases, nucleic acid binding proteins (included RNA and DNA binding proteins), and transporter proteins. Cluster 1 transcripts, primarily

characterized by up-regulation in Maloe More and down-regulation in the 3 basins, had high differential expression in peptidases and transporter proteins (Figure 22). Cluster 2 transcripts, primarily characterized by down-regulation in the Northern Basin, had high differential expression in peptidases and gene expression transcripts (splicing mediators) (Figure 22). Cluster 3 transcripts, primarily characterized by down-regulation in the Northern and Central Basins, had high differential expression in nucleic acid binding proteins (included RNA and DNA binding proteins), ATP binding proteins, actin binding proteins, oxidoreductases, and transporter proteins (Figure 22). Cluster 4 transcripts, primarily characterized by up-regulation in the Northern Basin, had high differential expression in peptidases, nucleic acid binding proteins (included RNA and DNA binding proteins), structural proteins, and actin binding proteins (Figure 22).

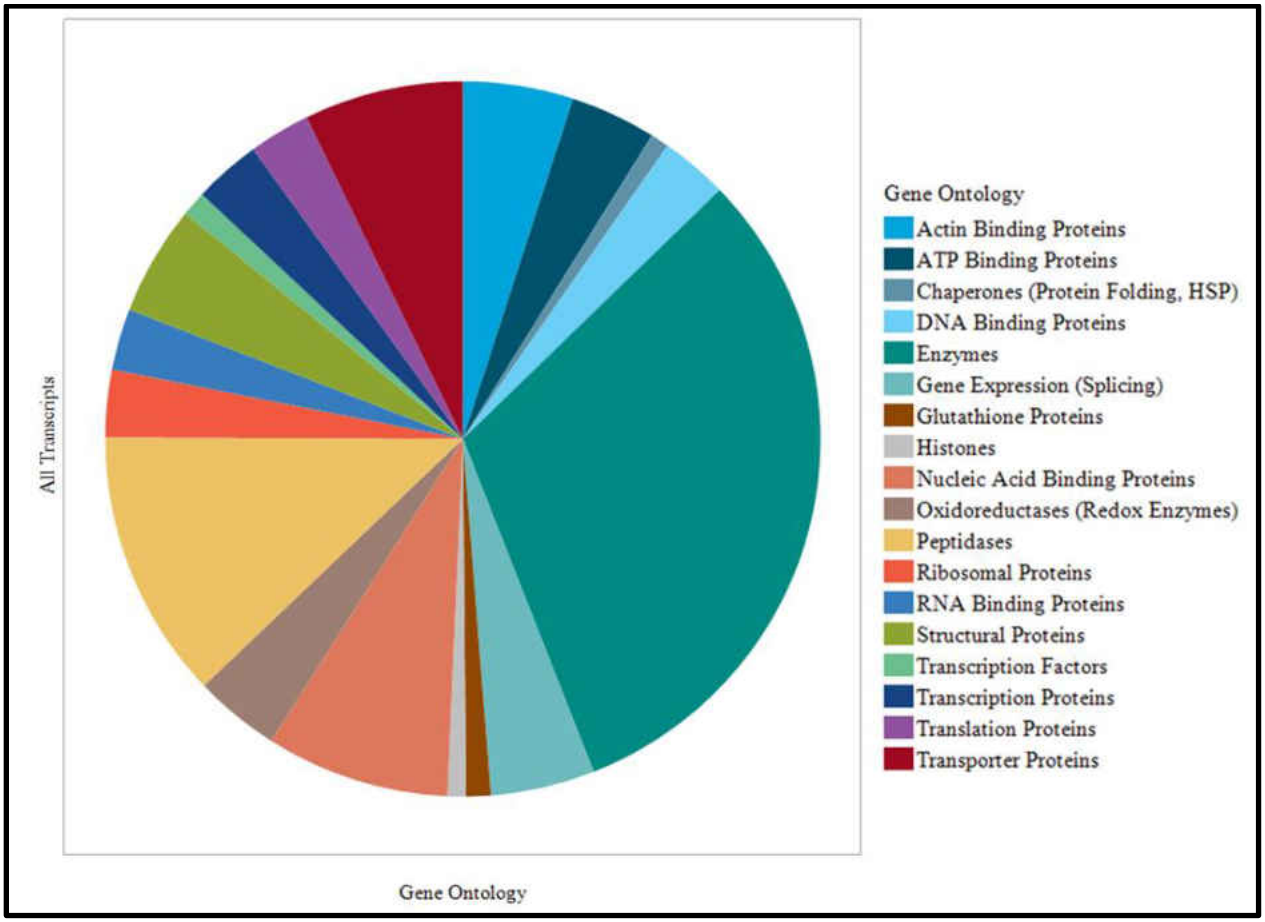


Figure 21 represents the transcripts belonging to different ontologies from all known putative functions of the 863 significantly differentially expressed (FDR=0.001) transcripts in *Epischura baikalensis*. Transcript functions are from alignments to *Drosophila* and invertebrate sequencing databases. Some transcripts matched both databases for the same putative function and were counted once toward that function. Transcripts may appear in multiple gene ontology categories.

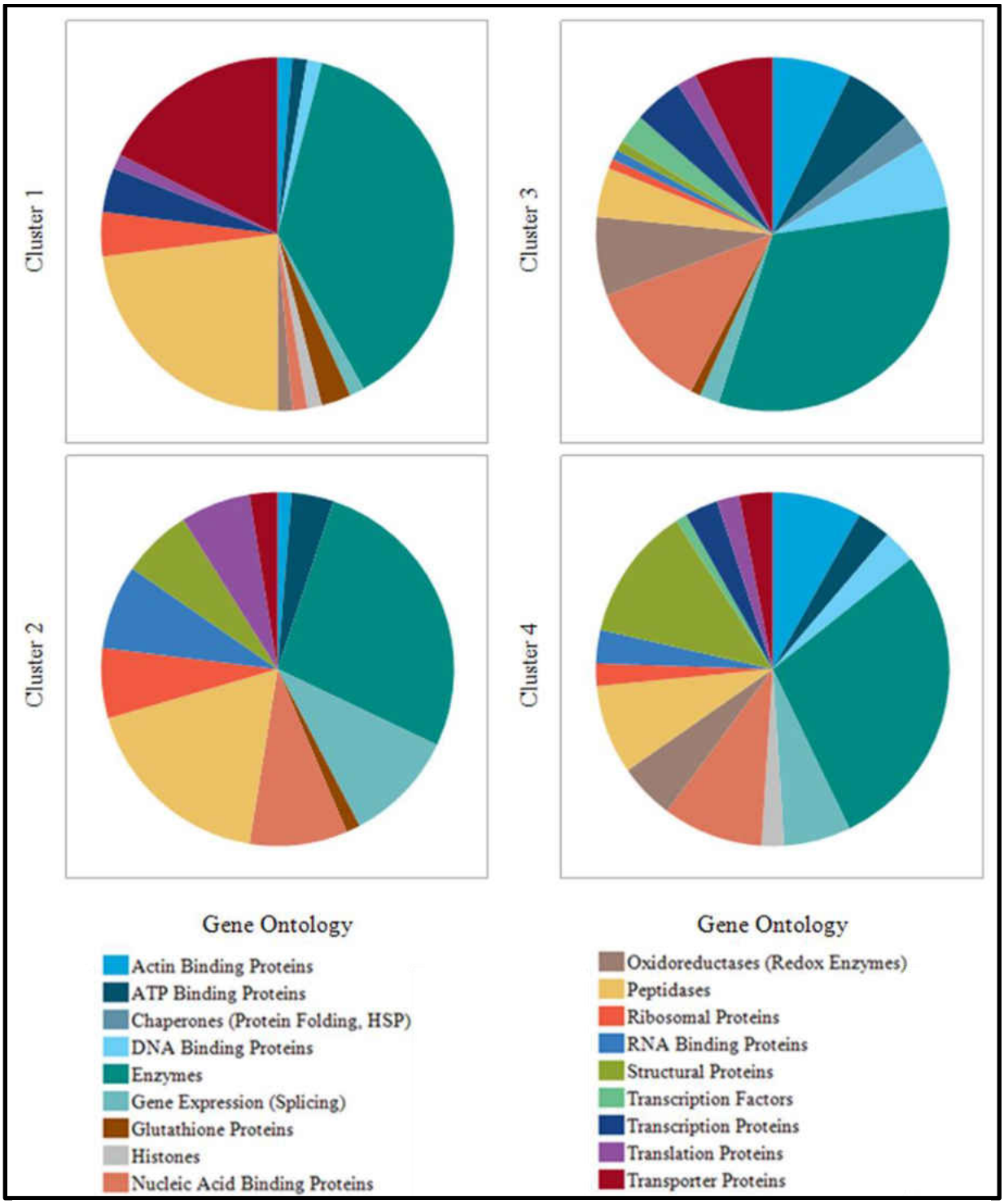


Figure 22 represents the transcripts by cluster (Figure 16) belonging to different ontologies from all known putative functions of the 863 significantly differentially expressed (FDR=0.001) transcripts in *Epischura baikalensis*. Transcript functions are from alignments to *Drosophila* and invertebrate sequencing databases. Some transcripts matched both databases for the same putative function and were counted once toward that function. Transcripts may appear in multiple gene ontology categories.

Discussion

The differential gene expression throughout Lake Baikal is very high. High variability in gene expression leads to greater plasticity. In the various climates of Lake Baikal, some differential expression was expected. Maloe More, especially, having a vastly different average water temperature and biota was expected to have very different gene expression than sites from the other basins within the lake. However, the differential expression between the 3 basins of the lake is somewhat surprising. With heavy mixing and strong currents, the water climate of Lake Baikal would presumably be static over all 3 basins with possible warming due to lower latitudes in the Southern Basin (Kozhova 1993). One would expect relatively uniform expression across the 3 basins, but there is significant differential gene expression among the 3 basins. This observation of high differential gene expression could suggest that even in a relatively static habitat *Epischura baikalensis* is exhibiting a high amount of phenotypic plasticity. Being able to induce the phenotypic changes in metabolism, growth, and reproduction could enable *E. baikalensis* to fare well as Lake Baikal climate continues to warm and change.

Cluster Analysis

Cluster 1 Transcripts. This transcript family had high differential expression in peptidases and structural proteins. With the higher temperatures *Epischura baikalensis* might be dealing with in Maloe More, there is likely increased metabolic rates, increased productivity, and increased growth rates. For these reasons, increases in peptidases are logical. Increased growth rate could also explain heightened up-regulation of structural proteins as well.

Cluster 2 Transcripts. This transcript family had high differential expression in peptidases and gene expression mediators. With the lower temperatures of Northern Baikal, *Epischura baikalensis* is likely exhibiting lower metabolic rates, lower productivity, and decreased growth

rates. For these reasons, decreases in peptidases are logical. Having decreased expression of splice variants and mediators could also be explained by decreased temperatures and lower metabolic rates if those gene expression mediators deal specifically with higher productivity functioning.

Cluster 3 Transcripts. This transcript family had high differential expression in several categories: nucleic acid binding proteins (included RNA and DNA binding proteins), ATP binding proteins, actin binding proteins, oxidoreductases, and transporter proteins. This cluster of transcripts was down regulated in the Central and Northern Basins, where temperatures are colder but not necessarily significantly colder than the Southern Basin. Because reducing energy consumption seems to be the overarching theme of these down-regulated transcripts, an energy expending activity must be driving this trend. These individuals could be preparing for a growth or reproduction event concurrently across the basins or the down-regulation could entirely be due to lowered temperatures.

Cluster 4 Transcripts. This transcripts family had high differential expression in peptidases, nucleic acid binding proteins (included RNA and DNA binding proteins), structural proteins, and actin binding proteins. The temperature in the Northern Basin is largely uniform year round but is slightly colder for a longer duration of the year. Enjoying colder temperatures could explain why many actin binding proteins, ATP binding proteins, and transporter proteins are up-regulated as the extremely stenothermic *Epischura baikalensis* thrives in its native habitat. Additionally, higher productivity at these lower temperatures could explain high expression of nucleic acid binding proteins as transcription and translation are maximized.

CHAPTER 6

GENETIC DIFFERENTIATION

Results

Correlation Analysis

From the PCA on correlations, SNP frequencies clustered in tight clouds in their respective geographic locations within 2 dimensions (Figure 23). Two principal components explained 35.4% of the variation within the correlations of transcripts (Figure 23). Much of the SNP frequency variation was explained by 2 principal components expressing the strong correlations between geographic locations.

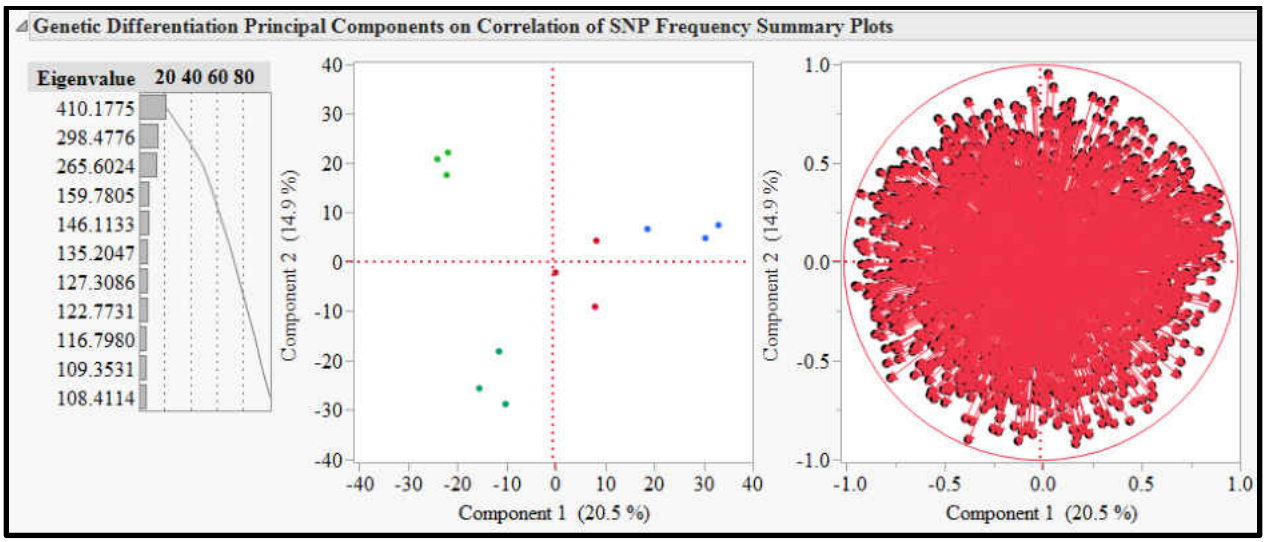


Figure 23 represents a Principal Components Analysis (PCA) on correlations of 2000 randomly sampled highly significantly differentiated SNP frequencies. Above left: eigenvalues for principal components. Above middle: principal components 1 and 2 explain 35.4% variation in correlations. Above right: summary map of SNP frequency vectors. Marker color corresponds to sample site as follows: Northern Basin (blue), Central Basin (red), Southern Basin (green), Maloe More (teal).

Randomly assigned F_{ST} values by a Poisson distribution predicted less genetic differentiation than actual F_{ST} values (Figure 24).

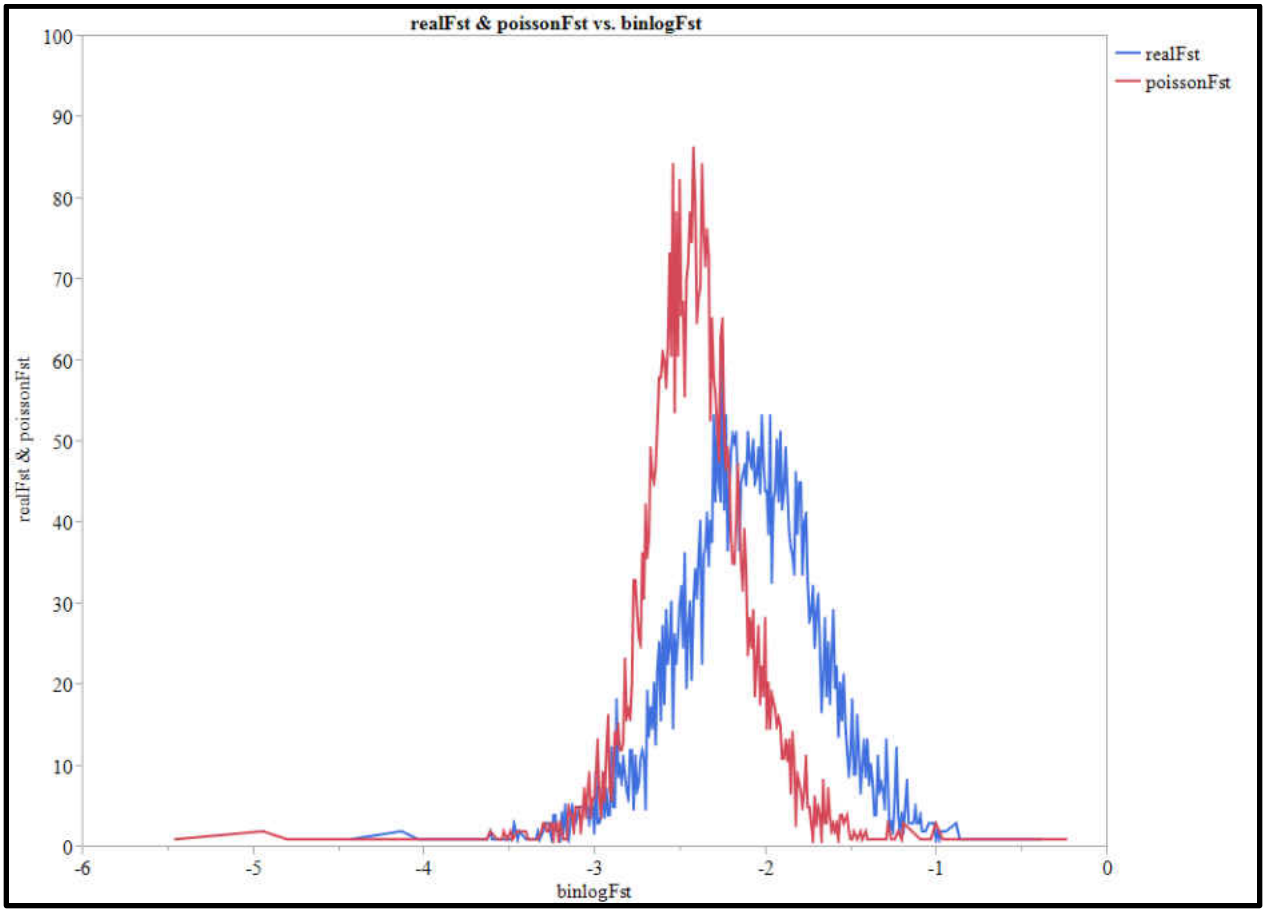


Figure 24 represents the actual (blue) and expected (red) F_{ST} values based on a random assignment of alleles via Poisson distribution to population distributions of F_{ST} values in the 4 sample groups shown in Figure 23.

One way ANOVA displayed significant differences in F_{ST} among subpopulations as grouped by differential expression clusters (Figure 25). There were no significant differences in gene ontology between differentially expressed and highly genetically differentiated transcripts except in nucleic acid binding proteins and gene expression related proteins ($p<0.05$). The effective population size (N_e) of Lake Baikal's population of *Epischura baikalensis* was estimated to be

306108 individuals, with a migration rate of 20.16 individuals per generation (6 months) (Table 8). The F_{ST} for the population of *E. baikalensis* was found to be 0.0122 ±3.428E-05 (Table 8).

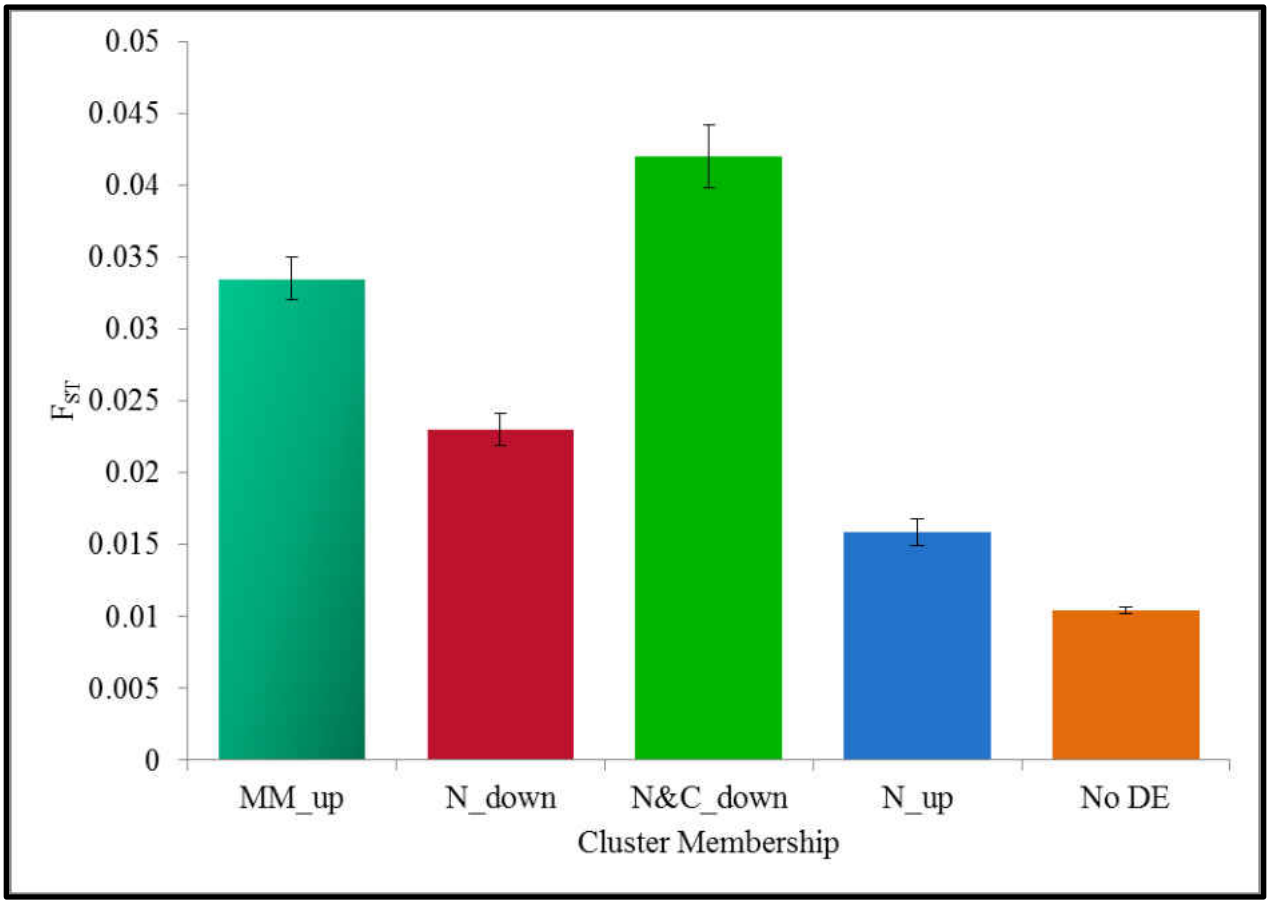


Figure 25 represents the subpopulation means comparisons of F_{ST} by cluster membership in differential expression. Colors correspond to clusters represented in Figure 16 with the *No DE* category representing SNP F_{ST}s that were not differentially expressed. Site codes are as follows: MM_up (teal; up-regulation in Maloe More), N_down (red; down-regulation in Northern Basins), N&C_down (green; down-regulation in Northern and Central Basins), N_up (blue; up-regulation in Northern Basin), and No DE (orange; no differential expression/no cluster membership). Standard error bars calculated using 95% CI of means. All means are significantly different by cluster membership.

Table 8 represents summary statistics for the genetic differentiation analysis. F_{ST}, effective population size (N_e), migration rate, and genome size are included.

F_{ST}	**Eff. Pop. Size (N_e)**	**Migration Rate (N_m)**	**Est. Genome Size**
0.0122 ±3.428E-05	306108 individuals	20.16 ind./gen.	0.86 pg

Discussion

The populations of *Epischura baikalensis* exhibit high genetic differentiation as corroborated by the high F_{ST} value (Table 8). The genetic differentiation is significantly more than predicted by a random assignment of variant alleles (Figure 24), and most of the variation between and among populations can be described by geographic location (Figure 23). The effective population size of *E. baikalensis* is significantly smaller than actual population size but could have been much smaller given the possible geographic and other isolating mechanisms in the system. The migration rate suggests that about 20 individuals migrate into and out of their respective subpopulations per generation, which is estimated at 6 months, given the winter and subsequent summer populations. This small migration rate, given actual population size suggests that little mixing and interbreeding of these subpopulations occurs.

There was no significant difference in the transcripts that had the highest differential expression and those that had high F_{ST} values except in nucleic acid binding proteins and gene expression associated proteins. These 2 groups especially would likely be highly conserved across populations regardless of local selection pressures due to their "housekeeping" nature. The SNPs that were highly significantly differentiated were easily separated by geographic location further cementing the maintenance of local intermixing populations.

However, transcripts with high F_{ST} values were also highly differentially expressed (Figure 25). This suggests that genes that are undergoing differential expression have higher probabilities of becoming fixed in the local populations, suggesting that there is high selection pressure on transcripts that are being differentially expressed. Genes without differential expression were least differentiated, and the genes that would matter most for local adaptation, i.e., ones being highly differentially expressed, were the most genetically differentiated, further

corroborating the high adaptability potential of *E. baikalensis* in its varying biota with respect to projected climate change.

CHAPTER 7

CONCLUSIONS

Zooplankton Distribution

The fate of *Epischura baikalensis* may not ever be certain. From the zooplankton distribution analysis, invaders may easily invade and colonize open Baikal given higher temperatures. However, as long as there is a thermocline to retreat to, *E. baikalensis* may survive regardless due to its sheer abundance in open Baikal. The depth of Lake Baikal will likely always maintain a temperature within which *E. baikalensis* can thrive, but the true story will be told with how much *E. baikalensis* is competitively excluded by invading cosmopolitans. Future work would investigate whether cosmopolitans could survive if they were washed into open Baikal. Modeling the movement of cosmopolitans within the bays could predict whether transfer of many individuals is occurring and colonization is unsuccessful or whether cosmopolitans have simply not had the chance to colonize yet.

Differential Gene Expression

Epischura baikalensis exhibited a surprising amount of differential gene expression among the different lake regions. For the 3 basins, despite being relatively similar habitat, *E. baikalensis* is showing a vast array of different expression levels for many different transcripts with many different putative functions. With high ability to express plastic traits in different environments, the chances of survival in dramatic climate change events increases. Future work in this area would include temperature acclimation and tolerance experiments followed by subsequent sequencing of those individuals for temperature-specific differential expression. Finding the upper limits to *E. baikalensis*'s expression level in high temperature environments is

crucial to understanding how it might react to its changing natural habitat. Acclimation studies might forecast that if waters warm slowly, survivorship in *E. baikalensis* may remain unchanged.

Genetic Differentiation

Epischura baikalensis exhibited significant genetic differentiation among its subpopulations in the different lake regions. Even the 3 major basins have significantly different genetic differentiation among their individuals. This high genetic variation suggests that *E. baikalensis* may be well prepared to adapt to climate change. Across the entire population there is significant variation, and between its subpopulations, one of which is warming rapidly, there is selection on differentially expressed genes that have a high number of significant variant alleles. With high genetic differentiation in different biota, *E. baikalensis* possesses the ability to adapt rapidly to climate change. Future work in this area would include characterizing the SNPs as synonymous and nonsynonymous to further elucidate their importance in altering *E. baikalensis*'s adaptability.

Ramifications for Climate Change

Because *Epischura baikalensis* showed significant differential expression of many transcripts and those transcripts were highly genetically differentiated, this study predicts that *E. baikalensis* will have the ability to exhibit not only phenotypic plasticity but locally adapted traits that are heritable. As Lake Baikal warms, *E. baikalensis* populations may have the ability to offset the effects of climate change by inducing plastic traits and selecting for heritable mutations that are advantageous in warming waters. Future studies include investigating the limits of *E. baikalensis*'s phenotypic plasticity and analyzing the single nucleotide morphisms for synonymous and nonsynonymous changes with further study in gene ontologies and annotation.

REFERENCES

454 Life Sciences Corp. 2011. Rapid Library Preparation Method Manual. Branford, CT: 454 Life Sciences Corp.

Afanasyeva EL. 1998. Life cycle of *Epischura baicalensis sars* (Copepoda, Calanoida) in Lake Baikal. J Mar Syst 15(1–4):351-7.

Albrecht C and Wilke T. 2009. Ancient Lake Ohrid: Biodiversity and evolution. In: Patterns and processes of speciation in ancient lakes. Springer. 103 p.

Barton N and Partridge L. 2000. Limits to natural selection. Bioessays 22(12):1075-84.

Bondarenko N, Timoshkin O, Röpstorf P, Melnik N. 2006. The under-ice and bottom periods in the life cycle of *Aulacoseira baicalensis* (K. meyer) *simonsen*, a principal Lake Baikal alga. Hydrobiologia 568(1):107-9.

Efremova SM, Itskovich VB, Parfenova V, Drucker VV, Müller WE, Schroder HC. 2002. Lake Baikal: A unique place to study evolution of sponges and their stress response in an environment nearly unimpaired by anthropogenic perturbation. Cell Mol Biol 48(4):359-71.

Fryslie B. 2012. GS *De Novo* Assembler (Newbler) [computer program]. University of Arizona: Rosche Diagnostics Corporation. Newbler.

Geer LY, Marchler-Bauer A, Geer RC, Han L, He S, Liu C, Shi W, Bryant SH. 2010. The NCBI BioSystems database. Nucleic Acid Res (38):D492-6.

Gomulkiewicz R and Holt RD. 1995. When does evolution by natural selection prevent extinction? Evolution 49(1):201-7.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29(7):644.

Hampton SE, Izmetst'eva LR, Moore MV, Katz SL, Dennis B, Silow EA. 2008. Sixty years of environmental change in the world's largest freshwater lake – Lake Baikal, Siberia. Global Change Biol 14(8):1947-58.

Illumina I. 2011. Paired-End Sample Preparation Guide. San Diego, CA: Illumina, Inc.

Kozhova O and Beim A. 1993. Ecological monitoring of Lake Baikal. Ecology Publish, Moscow 352.

Macdonald III KS, Yampolsky L, Duffy JE. 2005. Molecular and morphological evolution of the amphipod radiation of Lake Baikal. Mol Phylogenet Evol 35(2):323-43.

Matzinger A, Spirkovski Z, Patceva S, Wüest A. 2006. Sensitivity of ancient Lake Ohrid to local anthropogenic impacts and global warming. J Great Lakes Res 32(1):158-79.

Melnik NG, Sheveleva NG, Pomazkova GI. 1998. Distribution of planktonic copepods of Lake Baikal. J Mar Syst 15(1–4):149-53.

Moore MV, Hampton SE, Izmest'eva LR, Silow EA, Peshkova EV, Pavlov BK. 2009. Climate change and the world's "sacred sea"--Lake Baikal, Siberia. (cover story). Bioscience 59(5):405-17.

Qiagen. 2006. RNeasy mini handbook.

Richardson TL, Gibson CE, Heaney SI. 2000. Temperature, growth and seasonal succession of phytoplankton in Lake Baikal, Siberia. Freshwat Biol 44(3):431-40.

Robinson M, McCarthy D, Chen Y, Smyth GK. 2013. edgeR: Empirical Analysis of Digital Gene Expression Data in R [computer program]. Version 3.4.2. Bioconductor.

SAS Institute Inc. 2013. JMP 11 [computer program]. Cary, NC: SAS Institute Inc.

Sekino T, Genkai-Kato M, Kawabata Z, Melnik NG, Logacheva NP, Belykh OI, Obolkina LA, Bondarenko NA, Khodzher TV, Gorbunova LA. 2007. Role of phytoplankton size distribution in lake ecosystems revealed by a comparison of whole plankton community structure between Lake Baikal and Lake Biwa. Limnology 8(3):227-32.

Sherbakov DY. 1999. Molecular phylogenetic studies on the origin of biodiversity in Lake Baikal. Trends in Ecology & Evolution 14(3):92-5.

Shimaraev MN, Kuimova LN, Sinyukovich VN and Tsekhanovskii VV. 2002. Manifestation of global climatic changes in Lake Baikal during the 20th century. Dokladt earth sciences C/C of doklady-akademiia naukInterperiodica Publishing. 288 p.

Sideleva VG. 2003. The endemic fishes of Lake Baikal. Backhuys.

St. Pierre SE, Ponting L, Stefancsik R, McQuilton P, and the FlyBase Consortium. 2014. FlyBase 102 - advanced approached to interrogating FlyBase. Nucleic Acids Res 42(D1):D780-8.

Tierney JE, Mayes MT, Meyer N, Johnson C, Swarzenski PW, Cohen AS, Russell JM. 2010. Late-twentieth-century warming in Lake Tanganyika unprecedented since AD 500. Nature Geoscience 3(6):422-5.

Timoshkin O. 1995. Biodiversity of Lake Baikal: Review of current state of knowledge and perspectives of studies. Guide and Key to Pelagic Animals of Lake Baikal with Ecological Notes [in Russian]. Nauka:25-51.

Tsugeki NK, Ishida S, Urabe J. 2009. Sedimentary records of reduction in resting egg production of *Daphnia galeata* in Lake Biwa during the 20th century: A possible effect of winter warming. J Paleolimnol 42(2):155-65.

Tsugeki NK, Urabe J, Hayami Y, Kuwae M, Nakanishi M. 2010. Phytoplankton dynamics in Lake Biwa during the 20th century: Complex responses to climate variation and changes in nutrient status. J Paleolimnol 44(1):69-83.

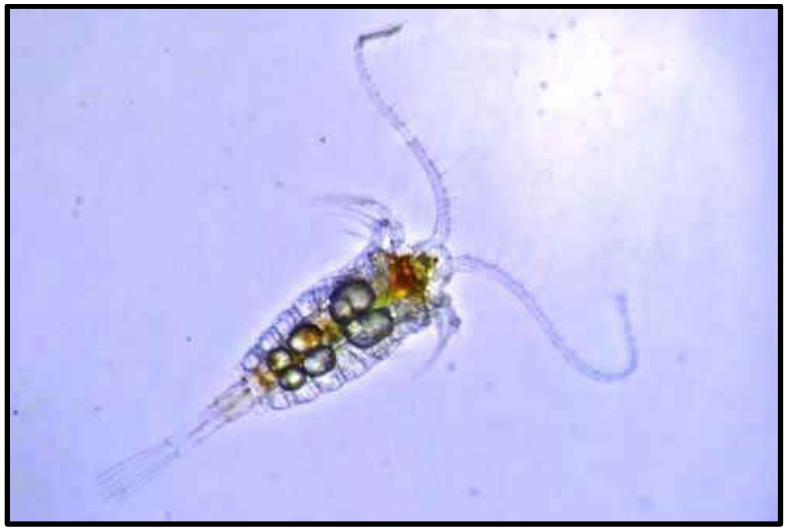
Verburg P, Hecky RE, Kling H. 2003. Ecological consequences of a century of warming in Lake Tanganyika. Science 301(5632):505-7.

Weir BS and Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358-70.

Yoshii K, Melnik NG, Timoshkin OA, Bondarenko NA, Anoshko PN, Yoshioka T, Wada E. 1999. Stable isotope analyses of the pelagic food web in Lake Baikal. Limnol Oceanogr 44(3, Part 1):502-11.

APPENDICES

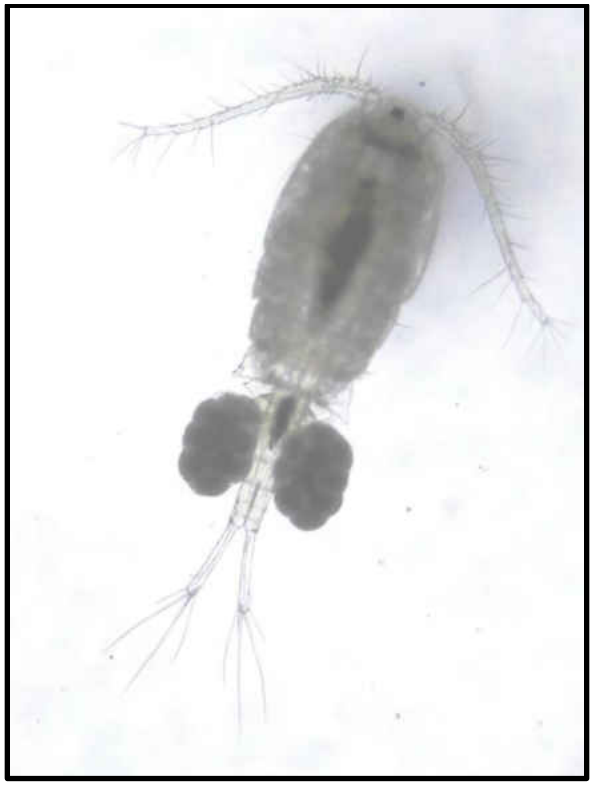
APPENDIX A: Lake Baikal Fauna



Epischura baikalensis

Daphnia longispina

Left: *Epischura baikalensis*; Right: *Daphnia longispina*



Cylcops kolensis

Macrohectopus branickii

Pusa sibirica

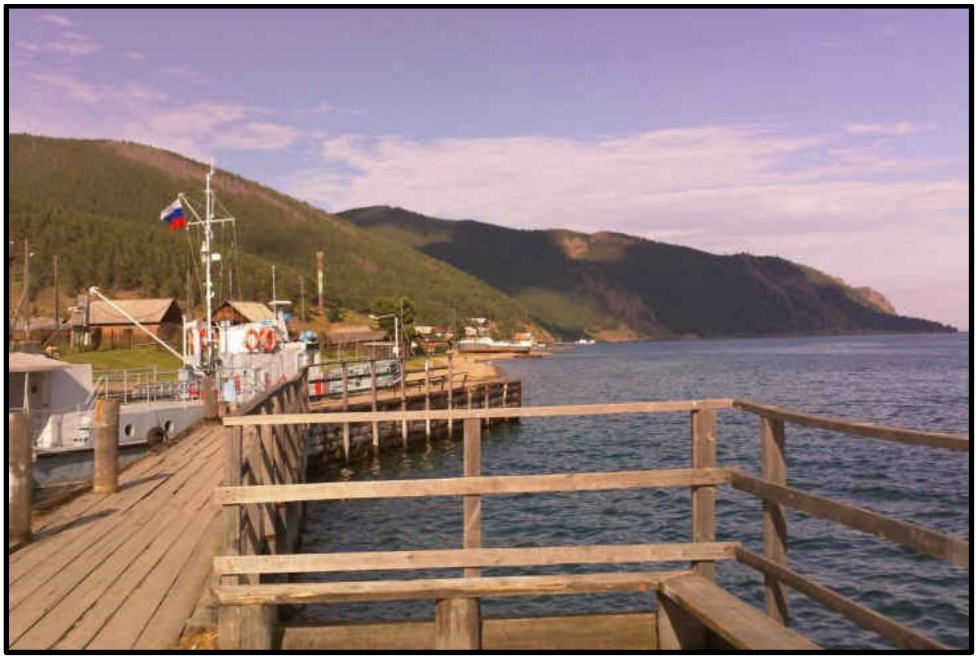
APPENDIX B: Research Vessels

Professor Kozhov

From Far Left to Right: *Professor Treskov* and *Professor Kozhov*

APPENDIX C: Bolshie Koty, Irkutsk Oblast, Siberia

View of biostation from main dock



View of Bolshie Koty from main dock

VITA

LARRY L BOWMAN, JR.

Education:	M.S. Biology, East Tennessee State University, Johnson City, Tennessee, 2014 Teach For America Summer Institute, Delta State University, Cleveland, Mississippi, 2011 A.B. Biological Sciences (Ecology and Evolutionary Biology), Minor: English Literature (20C American), Dartmouth College, Hanover, New Hampshire, 2011 Biology Foreign Study Program: Tropical Ecology in Costa Rica and the Cayman Islands, Dartmouth College, 2011 Williams-Mystic Maritime Studies Program, Maritime Studies and Marine Ecology, Williams College, Williamstown, Massachusetts & Mystic, Connecticut, 2009 Marion Senior High School, Marion, Virginia, 2007 Governor's School for Life Sciences and Medicine, Virginia Commonwealth University, Richmond, Virginia, 2006 Appalachian Regional Governor's School, Wytheville Community College, Wytheville, Virginia, 2002-2006
Professional Experience:	National Science Foundation Graduate Teaching Fellow in K-12 Education, East Tennessee State University, 2012 – 2014 Northeast Tennessee STEM Innovation Hub Advisory Council, 2014 – present Tennessee Science Teachers Association, Member, 2013 – present Association for the Sciences of Limnology and Oceanography, Member, 2012 – present Teach For America, Secondary Math faculty, Greene County High School, Eutaw, AL, 2011 SSI SCUBA Certification, Open Water Diver, 2011 Teaching Assistant, Intensive English Tutor and Admissions Intern, American University of Kuwait; Kuwait City, Kuwait, College of Arts and Sciences, 2010 Aquaria Technician, James T. Carlton Marine Science Center, Mystic, Connecticut, 2009 Admissions Assistant, Williams-Mystic Maritime Studies Program, 2009 U.S. Sailing, Motorboating and Sailing License, CT & VA, 2009
Academic & Research Experience:	Southeastern Ecology and Evolution Conference, Genetic Differentiation and Differential Gene Expression in Lake Baikal's Endemic Copepod, *Epischura baikalensis*, 2014 Innovation Academy, Kingsport, TN*, Professional Development

Presenter: Using NGSS Guidemaps, 2014
Northeast Tennessee STEM Innovation Hub Advisory Council Meeting*, NGSS-TNSS Correlation Guidemaps, 2014
Symposium on Effective Implementation of CCSS and NGSS*, Breakout Session Leader: NGSS Correlations to TN State Science Standards, 2013
Tennessee Science Teachers Association Professional Development Conference, Science First!: Adapting Painlessly to NGSS, 2013
International Teacher Scientist Partnership/AAAS Conference, Why Partnerships? Learn about Different Partnership Models and Their Benefits, 2013
Dartmouth College Field Methods in Ecology Summer Symposium, Estimating Population Growth in *Linaria vulgaris*, and Invasive Plant Species, 2010
Williams-Mystic Marine Ecology Symposium, Extrinsic Factors Effecting Vegetation Cover in Pannes on Barn Island Marsh, 2009
18th Annual Karen E. Wetterhahn Science Symposium, Identification of an Extracytoplasmic Function Sigma Factor involved in the *Bradyrhizobium*/Soybean Symbiosis, 2009
(* Denotes Invited Presentation)

Honors and Awards:

East Tennessee State University School of Graduate Studies and East Tennessee State University Graduate Council Service Projects that Enhance the Public Good Award, 2014
National Science Foundation GK-12 Fellowship, 2012 – 2014
Golden Key International Honour Society, 2013
The Honor Society of Phi Kappa Phi, 2013
International Teacher Scientist Partnership Travel Grant, $1000, 2013
Southeastern Population Ecology and Evolutionary Genetics Travel Award, $150, 2012
Academic Citation in Advanced Studies in Tropical Ecology, Dartmouth College, 2011
Milham Grant Recipient, Dartmouth College, Tuition Grant, 2011
Crowley Maritime Corporation Tuition Scholarship, Williams-Mystic, 2009
Dartmouth College Fund Scholar, Dartmouth College, 2007 – 2011
Howard Hughes Medical Institute Fellow, 2008 – 2009
Rotary International Scholarship, $2100, 2007
Valedictorian, Marion Senior High School, 2007