

UVM ScholarWorks

Cisco Science: Using Omics To Answer A Range Of Key Questions

| | |
|---------------|---|
| Item Type | thesis;article |
| Authors | Lachance, Hannah |
| Download date | 2026-06-11 09:57:23 |
| Link to Item | https://hdl.handle.net/20.500.14849/3045 |

CISCO SCIENCE: USING OMICS TO ANSWER A RANGE OF KEY QUESTIONS

A Thesis Presented

by

Hannah Lachance

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Natural Resources

August, 2019

Defense Date: June 27, 2019

Thesis Examination Committee:

Jason Stockwell, Ph.D., Advisor

Melissa Pespeni Ph.D., Chairperson

Ellen Marsden, Ph.D.

Wendylee Stott, Ph.D.

Cynthia J. Forehand, Ph.D., Dean of the Graduate College

ABSTRACT

Coregonines, including cisco (*Coregonus artedi*), kiyi (*Coregonus kiyi*), and bloater (*Coregonus hoyi*), are a focus for prey fish conservation and restoration efforts throughout the Laurentian Great Lakes. However, fundamental questions about coregonine ecology and genetics remain. For example, we know little about how the early life stages of coregonines respond to environmental change at either the genotypic or phenotypic level. We also have limited knowledge about how to identify different species at the larval stage and the genetic relationships among species, which makes the different species difficult to study at the larval stage. To increase the probability for success in restoration efforts, current and future research need to integrate traditional and novel approaches to better understand what leads to current and future coregonine successes. We used DNA and RNA omics tools, genomics and transcriptomics to boost our comprehension of current coregonine populations and to help understand how *C. artedi* may respond to environmental change. During the winter of 2017, we conducted a pilot experiment to evaluate how *C. artedi* eggs may respond to increased light exposure resulting from current and expected reductions in annual ice and snow cover due to global warming. We used transcriptomics to assess differences in gene expression between a continuous light and continuous dark treatment. Our results indicate that light is an environmental factor that could lead to earlier hatch dates, smaller yolk sacs, changes in mortality and differential gene expression in metabolic related and other functionally important genes. In 2018, we sampled larval coregonines in the Apostle Islands of Lake Superior each week from hatch in May until late July. We used genomic sequencing to genetically identify 197 larvae to species: *C. artedi*, *C. hoyi*, and *C. kiyi*. The larval demographic characteristics of each species was assessed and revealed that length ranges, growth rates, yolk sac condition, and effective population size varied among species. Larvae of all three species were found throughout the entirety of the Apostle Islands and the genetic diversity within each species appears high. The results from our pilot experiment and field observations help advance our understanding of the important early life stages of coregonines and how changes in light exposure or growth rates could affect their success or failure in a changing climate.

ACKNOWLEDGEMENTS

Graduate school has been a fun, exciting, exhausting, and exhilarating experience that I'm incredibly grateful to have had the chance to partake in. I would like to formally thank all those who have contributed intellectual and emotional support throughout the design, implementation and writing of my thesis. To start, I would like to thank my advisor, Jason Stockwell. Not many advisors would take a risk on a student with such minimal fish experience, yet you trusted me and let me grow and expand my knowledge and skill set both through my own research projects and by allowing me to help with other field research. I am forever indebted to you for your trust, support, and for introducing me to the best community of scientists and professionals I have ever met. My committee members, Dr. Melissa Pespeni, Dr. Ellen Marsden, and Dr. Wendylee Stott, have served as invaluable sources of knowledge and experience that I was able to draw on throughout my time as a master's student. Beyond my committee, I also received vital support from Andy Evans (from the Vermont Advanced Computing Core (VACC)), Dr. Peter Euclide and Dr. Brian O'Malley (former members of the "Rube Lab"), Dr. Trevor Krabbenhoft (my transcriptome guru), and Dr. Mark Vinson (the one who started Taylor Stewart and I on our cisco journeys). Andy, along with Google, helped ease me into the world of bioinformatics and the VACC was generous to fund my server needs throughout my master's. Peter and Brian are wonderful co-workers and early career scientists who helped me spitball ideas, connect with collaborators and

navigate the various graduate school milestones. Mark has been a wonderful source of Great Lakes data, information and ideas that have helped guide the direction of my research. I've also been blessed with the support of many wonderful past and present departments and labs including the faculty, staff, co-workers and grad students from the Department of Animal and Veterinary Sciences, the Rubenstein School of Environment and Natural Resources, the Rubenstein Ecosystem Science Laboratory, the Biology Department (especially the Pespeni Lab), and the Stockwell and Marsden labs. Thank you all for making me laugh, expanding my knowledge and helping with my research. Special thanks to Taylor Stewart for co-parenting the cisco babies and being a wonderful resource. Additionally, Verena Lucke, Maddi Sorrentino, and Kaitlyn Maines' help with fish rearing, collecting, counting, measuring and lab work was invaluable and much appreciated.

I'm blessed with an incredible group of friends and family, including but certainly not limited to Kelsey, Aly, Brianne, Celine, Brittney, Maura, Rosie, Allie, my Mom and my Dad. Through countless dinners, spontaneous adventures, phone calls, visits, and café work sessions they have provided me with the fuel I needed to keep chugging along. Perhaps most noteworthy is the fact that my friends and family have spent 2.5 years listening to my excited ramblings on fish and genetics and still chose to love me.

TABLE OF CONTENTS

| | |
|---|-----|
| ACKNOWLEDGEMENTS | ii |
| LIST OF TABLES | vi |
| LIST OF FIGURES | vii |
| CHAPTER 1: COREGONINES AND OMICS..... | 1 |
| Coregonines in the Great Lakes..... | 1 |
| Omics..... | 6 |
| Genomics | 7 |
| Transcriptomics | 8 |
| Using omics for coregonine restoration..... | 9 |
| CHAPTER 2: SHEDDING LIGHT ON AN UNDERSTUDIED ENVIRONMENTAL FACTOR: HOW DOES LIGHT IMPACT CISCO (COREGONUS ARTEDI) EGGS AND LARVAE?..... | 11 |
| 2.1. Introduction..... | 11 |
| 2.2. Methods and Materials | 15 |
| Egg collection and rearing | 15 |
| Life-history trait analysis | 16 |
| Transcriptome preparation and analysis | 17 |
| 2.3. Results | 19 |
| Life-history trait analysis | 19 |
| Transcriptome (all) | 22 |
| Transcriptome (Egg) | 23 |

| | |
|---|----|
| Transcriptome (Larvae) | 23 |
| 2.4 Discussion..... | 23 |
| 2.6 Supplemental Information | 27 |
| Section 1: Water quality methods..... | 27 |
| Section 2: RNA extraction protocol | 28 |
| | |
| CHAPTER 3: WHO’S THERE? HARNESSING POPULATION GENOMICS TO IDENTIFY CISCO (COREGONUS ARTEDI), KIYI (C. KIYI), AND BLOATER (C. HOYI) LARVAE FROM THE APOSTLE ISLANDS OF LAKE SUPERIOR. | 56 |
| 3.1. Introduction..... | 56 |
| 3.2. Methods and Materials | 59 |
| Sample collection..... | 59 |
| Demographic measurements..... | 60 |
| DNA extraction and barcoding..... | 63 |
| Library preparation, sequencing and genotyping | 64 |
| Population genomic analyses..... | 65 |
| 3.3. Results | 67 |
| DNA extraction, barcoding, library prep, sequencing and genotyping | 67 |
| Population genomic analyses..... | 67 |
| Demographic measurements..... | 70 |
| 3.4. Discussion..... | 80 |
| | |
| BIBLIOGRAPHY..... | 84 |

LIST OF TABLES

| Table | Page |
|---|------|
| Table 1: Current status of the major coregonines species in the Laurentian Great Lakes. Recreated from Eshenroder et al. (2016)..... | 3 |
| Table 2: Genes and proteins that were differentially expressed in both eggs and larvae (excluding uncharacterized proteins). Negative stat indicates up-regulated in light, down-regulated in dark and positive stat indicates up-regulated in dark and down-regulated in light. | 29 |
| Table 3: Significantly differentially expressed genes (padj <0.01) with GO terms found when comparing the light and dark treatment egg samples. (Excludes “uncharacterized proteins”). Negative stat indicates up-regulated in light, down-regulated in dark and positive stat indicates up-regulated in dark and down-regulated in light. | 31 |
| Table 4: Gene ontology (GO) terms and the frequency at which they were significantly DE (padj <0.01) when comparing the light and dark treatment egg samples..... | 34 |
| Table 5: Significantly differentially expressed gene (padj <0.01) with GO terms found when comparing the light and dark treatment larval samples. Negative Stat indicates up-regulated in light, down-regulated in dark and positive Stat indicates up-regulated in dark and down-regulated in light..... | 36 |
| Table 6: Gene ontology (GO) terms and the frequency at which they were significantly DE (padj <0.01) when comparing the light and dark treatment in the larval samples..... | 49 |
| Table 7: Species-specific data for the 197 larval samples that were sequenced from the spring/summer 2018 sampling in the Apostle Islands, Lake Superior. The number of larvae that assigned to each species along with the length range, week range N_e , H_o and H_e for each species is reported..... | 70 |

LIST OF FIGURES

| Figure | Page |
|---|------|
| <p>Figure 1: Ice coverage (%; blue line) and light ($\mu\text{mol}/\text{m}^2$; black bars) relationship based on light sensors set at 10-m depth off Sand Island, Lake Superior WI during the winter of 2017. Ice coverage data for the Apostle Islands, Lake Superior taken from the U.S. National Ice Center database.....</p> | 14 |
| <p>Figure 2: Life-history trait measurements from each light treatment including (a) hatch dates, (b) mean yolk sac volume, (c) mean length-at-hatch, and (d) growth of larvae over time.</p> | 21 |
| <p>Figure 3: PCA of the 82,588 transcripts identified by life stage (shape) and treatment (color).....</p> | 22 |
| <p>Figure 4: BVRA gene expression differences between individuals. Color represents treatment and life stage is represented by shape.....</p> | 26 |
| <p>Figure 5: Adults of species in the genus <i>Coregonus</i> currently found in the Great Lakes (Eshenroder et al. 2016).</p> | 57 |
| <p>Figure 6: Map of the larval catch totals per week at each of the 10 sites sampled between May 14 and July 25, 2018 in the Apostle Islands, Lake Superior, WI. Size of dots indicate density of coregonine larvae collected.</p> | 61 |
| <p>Figure 7: Representative photos of each coregonine species found in the Apostle Islands, Lake Superior during the spring/summer of 2018. Individuals pictured were collected during week 25 or week 29 and are within 0.02 mm of each other.....</p> | 62 |
| <p>Figure 8: <i>Structure</i> style stacked bar graph demonstrating AssignPop species assignments for the 197 coregonine larvae that were sequenced. Each bar is a separate individual and the y- axis is the percent that the larval genome matched to known adult samples for <i>C. artedi</i> (blue), <i>C. hoyi</i> (green), <i>C. kiyi</i> (orange), and an unknown <i>Coregonus</i> adult (purple) found in Lake Superior.</p> | 68 |
| <p>Figure 9: PCA showing the separation among species identified in 197 larval coregonines that were sequenced from the Apostle Islands, Lake Superior, during spring/summer 2018. Each dot is a separate individual.....</p> | 69 |
| <p>Figure 10: Map of the weeks and locations where the 197 sequenced larvae were collected in the Apostle Islands, Lake Superior during the spring/summer of 2018. Each panel is a different week and the colors indicate genetically identified</p> | |

species. Circle size indicates abundance and proportion of the circle indicates proportion of the catch that identified as each species. 72

Figure 11: Length-frequency histograms of the 197 sequenced larvae captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Each panel is a different week and the colors indicate species. 73

Figure 12: Length-frequency histograms of all 605 larvae (sequenced and not sequenced) captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Each panel is a different week and the colors indicate species. 74

Figure 13: Length frequency histogram of all 605 larvae (sequenced and not sequenced) captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Colors indicate species. 75

Figure 14: Length-frequency histogram of the 197 sequenced larvae captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Colors indicate species. 76

Figure 15: Yolk sac condition (YSC) of 197 sequenced coregonine larvae collected in the Apostle Islands of Lake Superior during the spring/summer of 2018. Symbol represents species and color represents yolk sac condition. 78

Figure 16: Length distribution of yolk sac condition as function of species identity for 197 sequenced coreognine larvae collected in the Apostle Islands of Lake Superior during the spring/summer of 2018. Color represents yolk sac condition. 79

CHAPTER 1: COREGONINES AND OMICS

Coregonines in the Great Lakes

Coregonines are key prey species found in many northern hemisphere lakes including the Laurentian Great Lakes. Coregonines are ecologically important as prey resources year-round. For example, cisco (*Coregonus artedi*) eggs are an excellent nearshore lipid source for benthic species such as lake whitefish (*C. clupeaformis*) during the winter when food resources are limited (Stockwell et al. 2014). Coregonine larvae are also key spring-time food items for fish such as rainbow smelt (*Osmerus mordax*), and juveniles and adults are prey for fish such as lake trout (*Salvelinus namaycush*) (Matuszek et al. 1990, Myers et al. 2014). For example, coregonines in the western U.S. waters of Lake Superior were consumed more than any other prey type (Stockwell et al. 2009).

In addition to coregonines' ecological importance, coregonines are economically important to the Great Lakes in the past and in the present. Historically, commercial fishing exerted heavy pressure on coregonines across the Great Lakes (Eshenroder et al. 2016). For example, *C. johannae* (deepwater cisco) and *C. nigripinnis* (blackfin cisco) were highly exploited in Lake Michigan during the mid-1800s to the early 1900s (Smith 1964). Currently, lake whitefish and cisco are very popular in local restaurants (especially with the locavore movement) (George 2016). Cisco and whitefish have also become an international economic commodity due to growing popularity of their caviar in northern Europe (George 2016).

The Great Lakes have historically been home to a variety of coregonine species, each seemingly adapted to a different ecological niche (Todd and Smith 1992, Eshenroder et al. 2016). For example, *C. reighardi* was the only spring spawning deepwater coregonine and was found in Lakes Michigan, Huron, and Ontario (Table 1) (Koelz 1929, Webb and Todd 1995, Scott and Crossman 1998, Eshenroder et al. 2016). *Coregonus kiyi* and *C. hoyi* are both deepwater species and are the smallest of the Great Lakes coregonines (Eshenroder et al. 2016). However, *C. hoyi* is considered a nearshore deepwater (<80 m) planktivore and *C. kiyi* is considered an offshore (>80 m) deepwater planktivore which consumes more *Mysis diluviana* than *C. hoyi* (Sierszen et al. 2014, Eshenroder et al. 2016). *Coregonus artedi* is the only pelagic coregonine in the Great Lakes and historically was the most widespread and abundant (Table 1).

Many coregonine populations, however, have been reduced and some species are no longer found in the Great Lakes due to overfishing, invasive species, and introgression (Table 1) (Todd and Smith 1992, Eshenroder et al. 2016). The need to restore prey species as a means to ensure a healthy ecosystem has been highlighted (Zimmerman and Krueger 2009), and coregonines are among the top prey species identified for restoration in the Great Lakes due to their important roles in the ecosystem (Bronte et al. 2017, George et al. 2018).

Table 1: Current status of the major coregonines species in the Laurentian Great Lakes. Recreated from Eshenroder et al. (2016).

| Species | Lake | | | | | |
|-----------------------|-----------|------------|--------------|------------|--------------|---------|
| | Superior | Michigan | Huron | Erie | Ontario | Nipigon |
| <i>C. alpenae</i> | - | Extinct | Introgressed | Extinct | - | - |
| <i>C. artedi</i> | Extant | Extant | Extant | Extirpated | Extant | Extant |
| <i>C. hoyi</i> | Extant | Extant | Introgressed | - | Reintroduced | Extant |
| <i>C. johanna</i> | - | Extinct | Extinct | - | - | - |
| <i>C. kiyi</i> | Extant | Extirpated | Introgressed | - | Extirpated | - |
| <i>C. nigripinnis</i> | Uncertain | Extinct | Extinct | - | - | Extant |
| <i>C. reghardi</i> | Uncertain | Extinct | Introgressed | - | Extinct | - |
| <i>C. zenithicus</i> | Extant | Extirpated | Introgressed | - | - | Extant |

Goals for coregonine restoration include 1) identifying any impediments to reestablishment and growth, 2) identifying gaps in knowledge of life history, genetics, and propagation, and 3) understanding roles of genetics and environment on the development of phenotype (Bronte et al. 2017).

To identify gaps in knowledge of coregonine life history, we must be able to identify the species. We need to determine if all coregonines are actually separate species or if they are simply plastic adaptations of one overarching species. Various theories have argued for or against coregonines as separate species, but a good test is to evaluate how genetically different each proposed species is from one another. Initial studies using seven microsatellite markers concluded that coregonines should be considered a single taxon (*C. artedi*) because of insufficient genetic diversity among the proposed species (Turgeon and Bernatchez 2003). However, others argued that the absence of newly adapted forms to fill the vacant ecological niches of the extirpated forms indicated coregonines are separate species instead of plastic adaptations of *C. artedi* (Eshenroder et al. 2016). To resolve this conundrum, analysis of more of the genome was required to identify potential differences that the seven microsatellite markers did not detect (Stott et al. 2018)

Whether coregonines are separate species or forms, the ecological importance of each “species” has been determined to be worth conserving, meaning we want to conserve each coregonine separately instead of treating all the species as one conservation unit. Therefore, being able to tell species apart is important. As adults, coregonines can be visually identified to species through the use of morphometric attributes, such as body and

head shape, and meristic measurements such as gill raker counts (Koelz 1929, Eshenroder et al. 2016). However, during early life stages no method exists that can successfully differentiate coregonine larvae species. Genetic barcoding of the mitochondrial cytochrome C oxidase I gene can differentiate *C. clupeaformis* from other coregonines (Schlei et al. 2008, George et al. 2017). To develop a method that can differentiate the remaining coregonines, we need to look at a larger portion of the genome beyond the small number of loci investigated with microsatellites or barcoding.

The inability to identify coregonine species at early life stages makes studying each species' life history during such a crucial part of their life cycle difficult. Knowledge of life history at each life stage is essential to obtain a full understanding of what leads to the success or decline of each species. For example, the egg and larval stages are the key to recruitment success or failure (Miller et al. 1988). If we cannot study egg and larval life history through assessing their demographic characteristics then we cannot unlock the key to understanding why coregonine experience variable recruitment, which is a foundational question that needs to be answered to aid conservation and restoration efforts (Miller et al. 1988, Bronte et al. 2017).

Additional goals to aid coregonine restoration efforts include understanding the roles of genetics and environment on the development of phenotype and identifying any impediments to reestablishment and growth (Bronte et al. 2017). Both of these goals are tied to understanding how environmental shifts, such as climate change, could impact coregonine success. A potential threat to coregonines resulting from climate change is reduced ice cover. Ice coverage on the Laurentian Great Lakes has declined over the last

four decades and is projected to continue to decline (Wang et al. 2012, Sharma et al. 2019). Such changes in ice cover could have significant consequences for fish species whose eggs incubate over the winter, including coregonine species such as *C. artedi* (Cahn 1927). Because *C. artedi* has already experienced significant population declines, reduction in ice coverage could pose an additional threat to their populations (Stockwell et al. 2009). To understand how reduced ice cover might impact *C. artedi*, we can identify the environmental factors influenced by ice cover (i.e. temperature and light) and then rear *C. artedi* eggs under simulated ice cover scenarios.

In addition to assessing demographic attributes, such as survival, growth, hatch dates and yolk sac volume, we can address the restoration goal of understanding the roles of genetics and environment on the development of phenotype (Bronte et al. 2017). Phenotype is the combined effect of genotype and environment, so if we manipulate the environment in a controlled laboratory setting we can then analyze the environmental effect on the genetic response. One way environment influences phenotype is by initiating the expression of different genes to instigate various functional responses to the environment (Andrews et al. 2016). Therefore, by studying any differences in gene expression we can understand how the environment influences phenotype.

Omics

While traditional sampling and analysis methods (i.e. acoustics, morphometrics, survival assessments) can be used to help address coregonine restoration goals, some questions need to be addressed with the use of advanced molecular techniques. Omics is

the “comprehensive, or global, assessment of a set of molecules” and includes genomics (DNA), transcriptomics (RNA), and proteomics (proteins) (Hasin et al. 2017). Harnessing omics for conservation and restoration purposes has been touted as a key to future conservation success (Allendorf et al. 2010, Andrews et al. 2016, Connon et al. 2018, Meek and Larson 2019). Already, areas have been identified in coregonine research that could benefit from the use of one or more omics techniques. For example, genomic sequencing will be needed to investigate the genetic differences among species to identify larvae to the species level. Transcriptomics can be used to assess the impact of climate change related environmental shifts on *C. artedi* (Connon et al. 2018). Additional uses of omics in coregonine research have recently been published or are currently underway, including transcriptomics and genomics used to study divergence history in lake whitefish, genomics used to identify species differences in Great Lakes coregonines, and transcriptomics used to compliment diet and meristic measurements of Great Lakes coregonines (Bernal pers. comm., Ackiss pers. comm., Jeukens et al. 2009 Rougeux et al. 2018). The addition of omics to our tool belt will help answer many key questions that we could not answer otherwise, enhance current datasets, and provide a deeper level of understanding of species demographics, genetic diversity, response to climate change and much more.

Genomics

The ability to efficiently and cost effectively use sequencing techniques that look at all, or the majority of, the genome is relatively new (Kohn et al. 2006, Allendorf et al. 2010, Meek and Larson 2019). Whole genome sequencing is currently the most

comprehensive method available (Hasin et al. 2017). Methods that capture large, but not all, portions of the genome include restriction site-associated DNA sequencing (RADseq) (Davey and Blaxter 2010, Hasin et al. 2017). These methods provide significantly more loci than microsatellites without the higher cost in time, money, and sequencing power needed to capture the whole genome. Generally, the more of the genome or transcriptome you can capture the better, but realistically choosing a sequencing approach is a tradeoff between costs and the questions. For conservation questions such as species identification and general population genomics questions, a technique such as RADseq can provide enough power to answer these questions at a reasonable cost (Andrews et al. 2016).

Transcriptomics

Transcriptomics uses techniques such as RNAseq to sequence all the mRNA present at the time of sample preservation (Wang et al. 2009). By focusing on mRNA instead of DNA, we can focus on only the genes that were being used, or expressed, at the time of sampling (Wang et al. 2009, Connon et al. 2018). The benefits of using transcriptomics to study fish include determining organismal responses to environmental stressors (e.g., heat wave), identifying functional response pathways to environmental cues, and understanding biological processes in fish, such as development, adaptive evolution, host immune response, and stress response (Qian et al. 2014, Connon et al. 2018). Transcriptomics has been used to test for differential gene expression between dwarf and normal lake whitefish species (*C. clupeaformis* spp.) and is currently used in the Great

Lakes to aid in coregonine restoration efforts (Jeukens et al. 2009, Krabbenhoft pers. comm.).

Using omics for coregonine restoration

The concurrent application of traditional methods with omics methods will complement and expand our understanding of coregonines and how they might respond to restoration efforts, management, and threats such as climate change. Throughout my master's research, I used traditional (e.g., measuring total length, yolk sac size, growth rates and mortality) and novel molecular (genomics and transcriptomics) methods to 1) determine what, if any, impacts light exposure resulting from reduced ice coverage has on *C. artedi* egg and subsequent larvae development and survival and 2) identify (for the first time) different species of wild coregonine larvae in Lake Superior.

We can employ traditional methods to provide a general understanding of development and survival differences in a light and no-light environment by rearing *C. artedi* eggs in a controlled laboratory. Adding molecular techniques, such as RNAseq, can provide insights into gene expression differences resulting from environmental factors such as light (Wang et al. 2009). The benefits from combining traditional methods with transcriptomics have been realized by fish researchers, policy makers and resource managers across the world (Connon et al. 2018). For example, while we can obtain a “surface level” understanding of any general responses to light (i.e. increased mortality or decreased yolk sac size) with traditional methods, we can also transcriptomics to ask why that response might have occurred (i.e. metabolic genes were up-regulated due to the light).

Therefore, we decided to use both traditional methods and transcriptomics to provide a more comprehensive understanding of any potential impacts changes in winter light might have on *C. artedi* eggs and larvae.

Microsatellites and visual identification have been unable to identify all of the different larval coregonine species in the Great Lakes. Consequently, we used genomic sequencing (RADseq) to capture a larger portion of the genome, compared to microsatellites, and increase our ability to identify genetic differences among species. We also assessed the genetic diversity of each population with the RADseq data. Once the species were identified, we used demographic information obtained through traditional methods (i.e., length, yolk sac condition, and growth rate calculations) to provide the first glimpse at the differing demographic characteristics of each species at the larval stage.

My two master's research projects demonstrated the benefits of pairing traditional and various omics methods to enhance the quality and quantity of information from each study. Future research should continue to combine methods to increase the efficiency and effectiveness of research to reach conservation and restoration goals faster.

**CHAPTER 2: SHEDDING LIGHT ON AN UNDERSTUDIED
ENVIRONMENTAL FACTOR: HOW DOES LIGHT IMPACT CISCO
(*COREGONUS ARTEDI*) EGGS AND LARVAE?**

2.1. Introduction

Ice cover plays an important role for fish species that spawn before winter and their eggs which incubate under ice (Magnuson et al. 1997). In rivers, marine systems and lakes ice cover provides ideal oxygen, flow, temperature and light for fish eggs such as Atlantic salmon (*Salmon salar*), Arctic cod (*Boreogadus saida*), and cisco (*Coregonus artedi*) (Welch et al. 1987, Drolet et al. 1991, Magnuson et al. 1997, Cunjak et al. 1998, Stockwell et al. 2009, Nguyen et al. 2017). However, changes in winter severity and ice cover could result in changes in the hatch dates, development and survival of overwintering eggs (Brown et al. 1993, Karjalainen et al. 2015, Collingsworth et al. 2017). For example, European coregonines (vendace *C. albula* and European whitefish *C. lavaretus*) reared under naturally long and projected short winter temperature conditions had expedited hatch dates with increased winter and spring temperatures (Karjalainen et al. 2015).

Ice coverage on the Laurentian Great Lakes has declined over the last four decades due to climate change (Wang et al. 2012, Sharma et al. 2019). Such changes in ice cover could have significant consequences for fish species whose eggs incubate over the winter, including cisco, lake whitefish (*C. clupeaformis*), and lake trout (*Salvelinus namaycush*) (Cahn 1927, Brown et al. 1993, Magnuson et al. 1997). Most, if not all, of these species have already experienced declines in their historical abundance due to exploitation, habitat destruction, and invasive species (Lawrie 1978, Schneeberger et al. 2005, Stockwell et al.

2009). Therefore, the reduction in ice coverage could pose an additional threat to their populations.

Recent conservation and restoration efforts have highlighted the need to restore prey species to ensure a healthy ecosystem. Such efforts depart from historical strategies to restore top predators (i.e. lake trout) (Zimmerman and Krueger 2009, Bronte et al. 2017, George et al. 2018). Coregonines are among the top prey species identified for restoration in the Great Lakes (Bronte et al. 2017, George et al. 2018). Goals for coregonine restoration include identifying any impediments to reestablishment and growth, identifying gaps in knowledge of coregonine life history and demographics, genetics, and propagation, and understanding what the roles of genetics and environment are on the development of phenotype (Bronte et al. 2017). Therefore, understanding how declining ice coverage could impact cisco early life history is an important part of the restoration and conservation efforts because this knowledge will 1) expand our knowledge of cisco life history and demographics, 2) help us understand how shifts in the environment affect cisco's phenotypic response, and 3) determine whether the response to reduced ice coverage could impede restoration efforts. While reduction in ice coverage might pose a threat to cisco populations, the direct impact of reduced ice coverage on cisco has not been specifically addressed.

Reduced ice cover may impact overwintering eggs through a variety of environmental factors including increased water flow (Nguyen et al. 2017), increased air and water temperatures (Assel and Robertson 1995, Mortsch and Quinn 1996, Magnuson et al. 1997, Hampton et al. 2016), and increased light (Greenbank 1945, Welch et al. 1987).

The most studied environmental factor linked to climate change is temperature. For example, increased temperatures result in increased development rates for cisco eggs (Colby and Brooke 1973). Therefore, we might expect reduced ice cover to increase the rate of egg development. However, other environmental factors might also be important for development and survival. For example, light is a well-known environmental cue in many molecular pathways such as circadian rhythm, stress response, DNA repair, and metabolism (Whitmore et al. 2000, Weger et al. 2011, Jordan and Lamia 2013). Additionally, light expedites egg development and early growth of rainbow trout (*Salmo gairdneri*) while lower light treatments had slower growth rates (Kwain 1975). Ice and snow cover strongly affect the light environment in lakes and can reduce light transmittance from 83% in open water to 62% under ice coverage, and to $\leq 10\%$ under snow and ice coverage (Bolsenga and Vanderploeg 1992). Light sensors deployed off Sand Island in Wisconsin waters of Lake Superior at 10-m depth measured $7.107\mu\text{mol m}^{-2} \text{s}^{-1}$ under 0% ice, $4.241\mu\text{mol m}^{-2} \text{s}^{-1}$ under 50% ice and $2.155\mu\text{mol m}^{-2} \text{s}^{-1}$ under 90% ice (Figure 1). Consequently, increased light during winter as a result of reduced ice and snow cover could lead to molecular changes that result in accelerated egg development rates, earlier hatch dates and/or increased mortality (Kwain 1975). While climate change could negatively impact or alter phenology of cisco egg development due to warmer temperatures and earlier spring, the additional impacts of increased light could also be important.

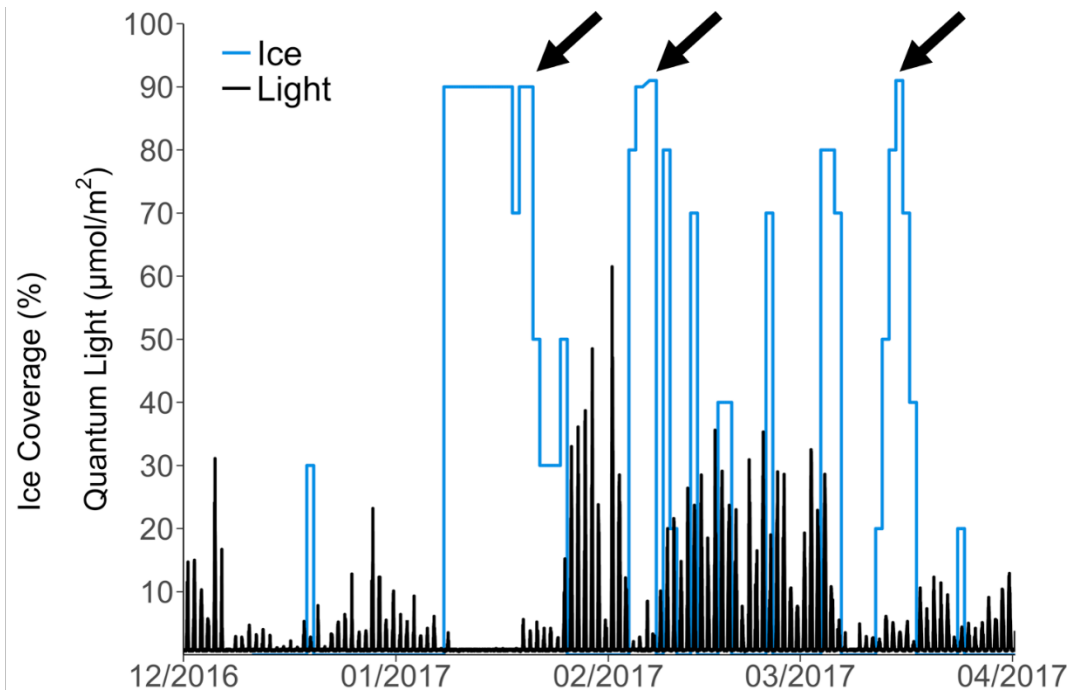


Figure 1: Ice coverage (%) (blue line) and light ($\mu\text{mol}/\text{m}^2$; black bars) relationship based on light sensors set at 10-m depth off Sand Island, Lake Superior WI during the winter of 2017. Ice coverage data for the Apostle Islands, Lake Superior taken from the U.S. National Ice Center database.

In this study, we used traditional (i.e. monitoring weekly development, measuring total length and yolk sac size, and mortality) and novel molecular (transcriptomics) methods to evaluate if light had any impact on cisco egg and subsequent larvae development and survival. Eggs were raised under three light treatments until hatch, and then transferred to rearing tanks.

2.2. Methods and Materials

Egg collection and rearing

Two female and three male cisco were collected with a trap-net deployed overnight in Chaumont Bay, Lake Ontario (44.06228 N, 76.15266 W) on November 29, 2016. The fish were transported to USGS Tunison Laboratory of Aquatic Science (Cortland, NY) and held live until fertilization on December 1, 2016. Adult cisco were anesthetized with tricaine methane-sulfonate (MS-222; dose to effect) prior to gamete removal. Fertilization was performed by pooling the eggs and milt using dry fertilization methods. Individuals were sorted by sex, total length (mm) and weight (g) was recorded, and otoliths were extracted for age estimation.

Eggs were transported within 12 hours of fertilization to the Rubenstein Ecosystem Science Laboratory (Burlington, VT). Upon arrival, eggs were homogenized to ensure the varying families were evenly distributed into one of three treatments; 24-hr light (herein “light” treatment), regular photoperiod, and 24-hr dark (herein “dark” treatment). Extreme light and dark treatments were chosen to help address if light has *any* effect on cisco early life history. Only one tank was used per treatment, due to resource constraints. Each treatment included an estimated 10,000 eggs which were reared in 113.5 L tanks. Each tank was isolated to eliminate exposure to nontreatment light. Daylight was simulated with full spectrum white LED lights (AquaShift® MLA-WH) and regulated by an AgriShift® Master Controller (Once® Inc, Plymouth, MN, USA). Peak day light intensity was 25.427 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for both the light and regular photoperiod treatments and 0.603 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for the

dark treatment. One hundred minute linear transitions were used to simulate sunrise and sunset in the regular photoperiod treatment.

Throughout egg and larvae development each tank was connected to a recirculating system where the flow, water quality, water temperature and all other environmental factors were consistent (see Supplemental section 1). Water was chilled to 3.8°C ($\pm 1.7^{\circ}\text{C}$), aerated, and sterilized with UV light. Dead eggs and larvae were counted and removed daily. Within 12 hours of hatch, larvae were moved to a 570 L tank that housed individuals from the same treatment. Larvae from all three treatments were then reared under the same photoperiod treatment using the overhead florescent lights in the lab. Larvae were fed *Artemia* spp. from the start of the hatching period until five days after the last hatching date. Feed was gradually transitioned to Otohime A (Reed Mariculture Inc., Campbell CA) and larvae were fed an estimated 5% body weight twice a day until May 31, 2017.

Life-history trait analysis

Egg survival was estimated as the percent of hatched eggs from the total initial number of eggs for each light level. Because we did not assess development stages, days since fertilization was our rate of embryonic development. Therefore, incubation period was calculated as a development rate in the number of days post fertilization. Total number of hatched larvae each day were recorded. A sample of up to 30 newly hatched larvae per day from each treatment were photographed under a dissecting scope and each larva was treated as a separate unit of observation (aka replicate) for downstream analyses. Total length and yolk-sac volume (YSV) were measured with Image-Pro[®] (Media Cybernetics,

Inc., Rockville, MD). YSV was calculated assuming the shape of an ellipse: $YSV = \frac{4}{3}\pi a b^2$ (with a, major radius and b, minor radius). Growth rates of larval cisco were measured to assess the influence of light intensity during incubation on the continued success/failure of larvae after they were removed from the varying light treatments. Total length was measured from 30 larvae from each incubation light treatment every 3-weeks. Larval survival was calculated as the percent of larvae that survived from hatch to the termination of the study.

The relationship between life-history traits (i.e., incubation period, length-at-hatch, and YSV) and incubation light treatments were analyzed in R (v3.5.2, Team, 2018). Length-at-hatch and YSV violated the ANOVA assumptions of normality and homogeneity of variance, even after transformations. Therefore, we used the nonparametric Kruskal-Wallis to test for differences in length-at-hatch and YSV among incubation light treatments and the Pairwise Wilcoxon rank sum tests to determine pairwise differences among treatments. Hatching periods were calculated by transforming the number of hatched larvae per day into cumulative proportions; where zero equals no individuals hatched and one equals all individuals hatched. Cumulative hatching proportions were fit using logistic regression models with binomial distributions and Tukey pairwise comparisons made with `mcp()` from the `multcomp` package (v1.4-8, Hothorn, 2008) to test for differences in hatching periods among treatments.

Transcriptome preparation and analysis

Six eggs and six freshly hatched larvae were collected from the light and dark treatments when each tank reached 50% hatch (March 8, 2017 in light; March 16, 2017 in

dark) for a total of 24 individuals. All samples were immediately stored in RNAlater (Invitrogen™, Life Technologies, Carlsbad, CA, USA) and then placed in a -80°C freezer. RNA was extracted from each individual using a protocol adapted from Krabbenhoft and Turner (2016) (see Supplemental Section 2). RNAseq library preparation was performed by Polar Genomics (Ithaca, NY) using the protocol described by Zhong et al. (2011). The prepared libraries were then sent to the Cornell University Institute of Biotechnology (Ithaca, NY) and run on a single lane of Illumina HiSeq2500 (Illumina, San Diego, CA, USA) to generate > 12 million 150 bp paired end reads per individual.

The raw reads were processed using the bioinformatics pipeline available on Github (<https://github.com/hmlrocks/Light-and-Cisco/blob/master/Transcriptome%20Bioinformatics.md#Trinity>). Briefly, the quality of the raw reads was assessed using FastQC (v.0.11.3; Andrews 2010), the adapters used in sequencing were trimmed off using Trimmomatic (v.0.36; Bolger et al. 2014) and the trimmed reads were reassessed with FastQC. Because cisco do not have a reference genome, a *de novo* transcriptome was assembled. Three *de novo* transcriptomes were assembled; one by pooling the reads from all egg and larval samples (n=24), a second by pooling only the reads from the egg samples (n=12) and the third by pooling only the reads from the larvae samples (n=12). The *de novo* transcriptomes were assembled using Trinity (v.2.4.0; Grabherr et al. 2011) and assessed using Trinity script ‘TrinityStats.pl’ and BUSCO (v.3.0.2; Waterhouse et al. 2017). Counts of transcripts, contig N50 stats based on all transcript contigs, and contig N50 stats based on only the longest isoform per “gene” were used from ‘TrinityStats.pl’ to assess the quality and completeness of the *de novo*

assemblies. BUSCO assessed transcriptome completeness for each assembly. After comparing the quality results from TrinityStats.pl and BUSCO for the three *de novo* assemblies, the *de novo* transcriptome assembled from all 24 individuals was chosen for all further analyses. Salmon (v.0.9.1; Patro et al. 2017), a "quasi-mapping" program, was used to compare individual reads to the *de novo* transcriptome.

DESeq2 (version 1.20.0; Love et al. 2014) was used in R (v. 3.5.1, Team 2018) to detect differentially expressed (DE) genes (p-value < 0.01). For the differential gene expression analyses, we analyzed each life stage (egg or larvae) separately and compared differential gene expression between treatments (dark vs light), within each stage. Our results were compared to the Uniref90 database, and the zebrafish (*Danio rerio*) and northern pike (*Esox lucius*) reference genomes using BLAST (version 2.7.1; Altschul et al. 1990). The Uniprot website (Consortium 2018) was employed to identify the genes, proteins and functions associated with DE genes. To identify any molecular pathways that are commonly enriched within a treatment (that is, enriched functions), two resources were used; GO_MWU.R (Wright et al. 2015), and Panther (Thomas et al. 2003).

2.3. Results

Life-history trait analysis

Egg mortality was similar across all three treatments (22.21% \pm 1.99%). Eggs in the light treatment hatched a full week earlier than the dark and photoperiod treatments (z = -3.061 and -3.179, p-value = 0.006 and 0.004, respectively; Figure 2a). Throughout the

larval stage mortality rates were 98.2% in the light treatment, 81.9% in the dark treatment and 80.3% in the regular photoperiod treatment, yet we are not confident that the mortalities seen are from the treatment and not from flow fluctuations that occurred in the later weeks of larval rearing. Mean yolk sac condition (YSV) was different among all three treatments with the smallest mean YSV observed in the light treatment and the highest in the dark treatment (light-photoperiod p-value = 0.012, light-dark p-value = <0.001, photoperiod-dark p-value = 0.028; Figure 2b). Length-at-hatch was smaller in the dark treatment compared to the light and photoperiod treatments (both p-values < 0.01; Figure 2c). Larval growth was slower in the dark treatment compared to the light and photoperiod treatments (p-value = 0.023 and 0.022, respectively; Figure 2d) however, again we cannot be sure that the growth rate differences were not from tank effects seen towards the end of the larval rearing.

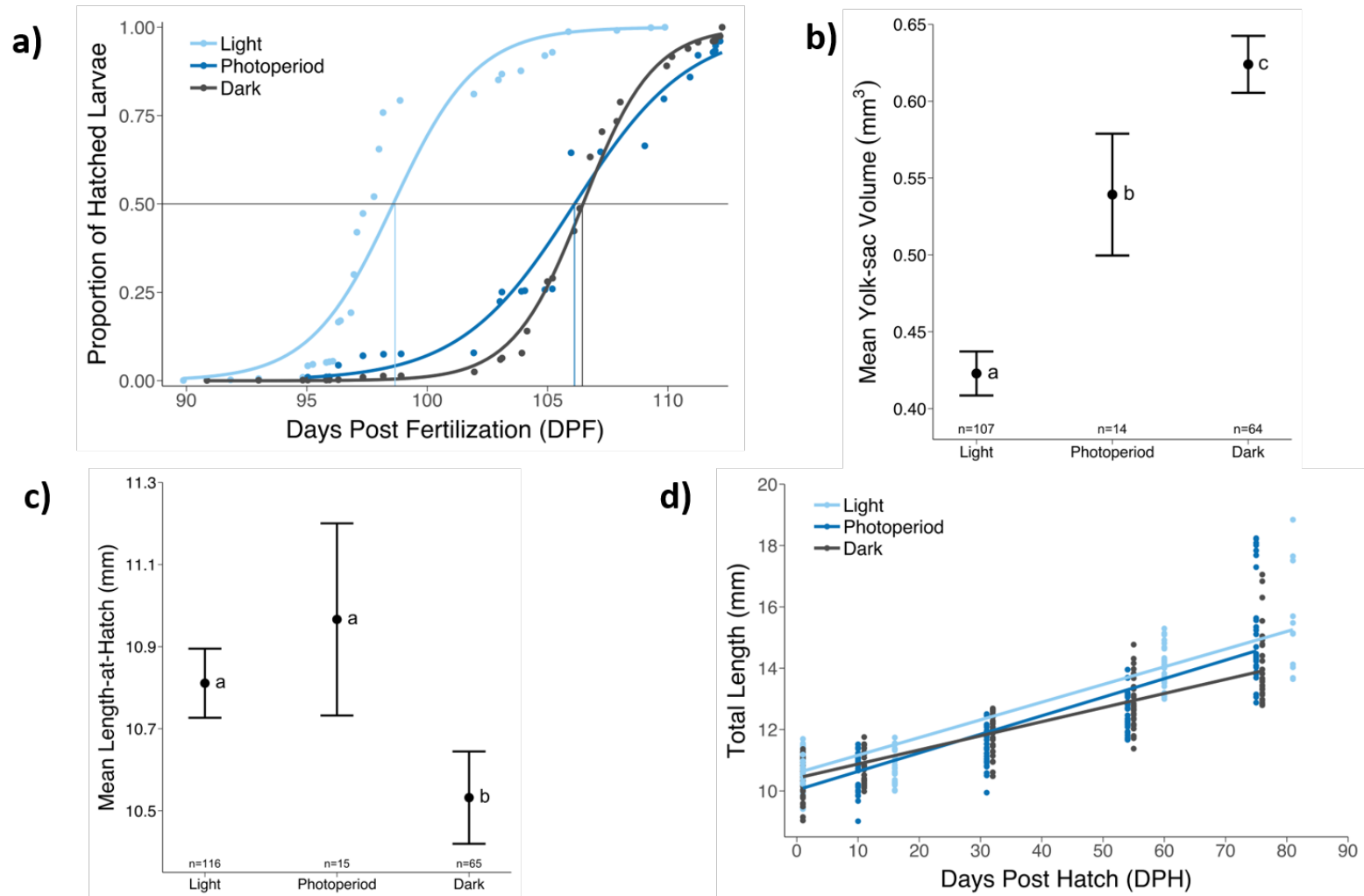


Figure 2: Life-history trait measurements from each light treatment including (a) hatch dates, (b) mean yolk sac volume, (c) mean length-at-hatch, and (d) growth of larvae over time.

Transcriptome (all)

RNAseq produced a total of 295,747,420 cleaned reads: 147,769,201 cleaned reads from the 12 egg samples and 147,968,219 cleaned reads from the 12 larvae samples. In the eggs, 171 genes were DE (adjusted p-value <0.01). Seventy-five genes were up-regulated in the dark treatment (down-regulated in light) and 96 genes were down-regulated in the dark treatment (up-regulated in light). In the larval samples, 991 genes were DE. Four-hundred genes were up-regulated in the dark treatment (down-regulated in light) and 591 genes were down-regulated in the dark treatment (up-regulated in light). Thirteen genes were differentially expressed in both eggs and larvae and had gene ontology (GO) terms (Table 2). Overall transcriptome-wide gene expression exhibited some grouping of individuals based on treatment (light vs dark) but no grouping was seen for life stage (egg vs larvae) (Figure 3).

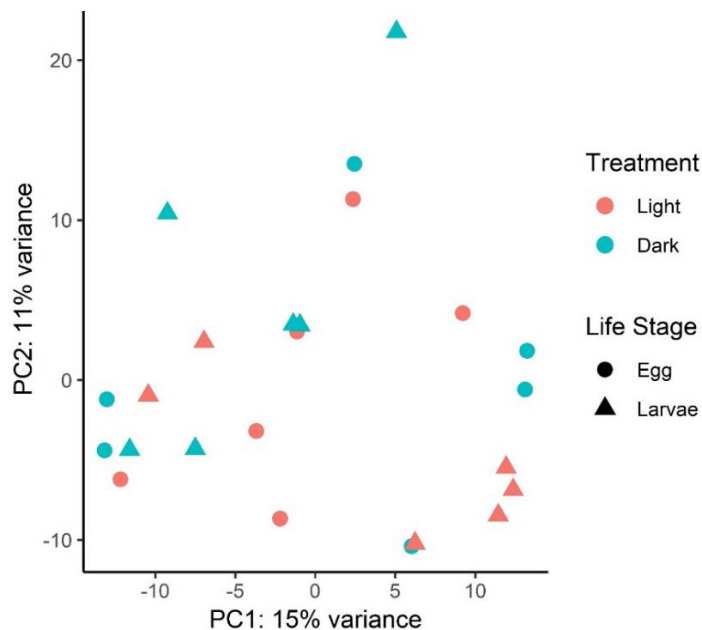


Figure 3: PCA of the 82,588 transcripts grouped by life stage (shape) and treatment (color).

Transcriptome (Egg)

In the egg samples, 64/171 of the DE genes were annotated and had Gene Ontology (GO) terms (Table 3). Broad functional groups for the DE genes (adjusted p-value <0.01) included integral component of membrane, ATP binding, metal ion binding, nucleus, and DNA binding (Table 4). Functional enrichment analysis for the significantly DE genes in the egg samples yielded no functionally enriched pathways with Panther.

Transcriptome (Larvae)

In the larvae samples, annotation of the significantly DE genes resulted in 368/991 genes that matched with GO terms (Table 5). The functional groups for the DE genes ($p_{adj}<0.01$) included ATP binding, metal ion binding, RNA binding, DNA binding, GTP binding, oxidoreductase activity, and zinc ion binding (Table 6). Functional enrichment analysis for the DE genes in the larval samples yielded no functional enrichment with Panther.

2.4 Discussion

Light is an environmental factor that merits further investigation into its potential effects on early life history traits of cisco. Differences in hatch dates, yolk sac size and gene expression indicate that light could be differentially driving growth and survival at both egg and larval stages. Reduced YSV and increased expression of energy-related functional genes in the light treatment suggests reduced ice cover could negatively

impact cisco eggs and subsequent larvae. Future investigations should include tank replicates to confirm that the differences we observed are from the light treatments and not any potential tank effects.

Earlier hatch dates in the light treatment could benefit or impede cisco survival depending on what food or predators are present at the time of hatch. Either cisco could hatch before other larvae and have increased resources available and less competition for those resources, or they could hatch before the food is available leading to starvation or expedited depletion of their already reduced yolk sacs (Hjort 1914). Additionally, if cisco hatch before other larvae they might be the first food resource for larval fish predators and therefore experience heavier predation than if they hatch at the same time as other larvae. Yet, if the same environmental factors (i.e. light, temperature, etc.) drive other larvae to hatch earlier as well, then expedited hatching would not be a benefit.

Yolk is a crucial resource both for the developing egg and the freshly hatched larvae (John and Hasler 1956). Cisco can use their yolk for weeks after hatch to 1) help bridge the learning curve to exogenous feeding, and 2) serve as a reserve if prey is limited. Because cisco in the light treatment hatched with smaller YSV, cisco could risk starvation if prey are unavailable or if they have a hard time transitioning to exogenous feeding.

Differences in mortality and growth rate at the larval stage should be viewed with caution. In the later weeks of larval rearing, flow rates, which were controlled through an automated system, became out of sync and the light larvae tank received lower flows and therefore slightly higher water temps. Because we did not have replicate tanks, we cannot

determine if the differences in larval mortality and growth we observed were due to the light treatment or due to the tank effects induced by the mechanical issues.

The large number of DE genes with annotations provides the opportunity to better understand what functional pathways could be up or down-regulated due to light. Functional enrichment analyses did not show any enrichment of full functional pathways. However, we do have annotations for the individual genes that provide an idea of what kinds of functions could be up or down-regulated. For example, we found a number of DE genes with ATP and GTP binding functions which are both related to energy consumption.

Additionally, DE genes that were up-regulated in both eggs and larvae reared in the light treatment can shed light on functions that are more likely linked to the light treatment. For example, biliverdin reductase A (BVRA) is a metabolic and catabolic gene that was up-regulated in both egg and larvae samples in the light treatment (Figure 4). BVRA has been linked to functions such as heme catabolism and insulin signaling; therefore, the up-regulation of BVRA in cisco raised in the light treatment could be linked to an increased breakdown of blood cells and/or an increased uptake of glucose and lipids (Lerner-Marmarosh et al. 2005, Maines 2005).

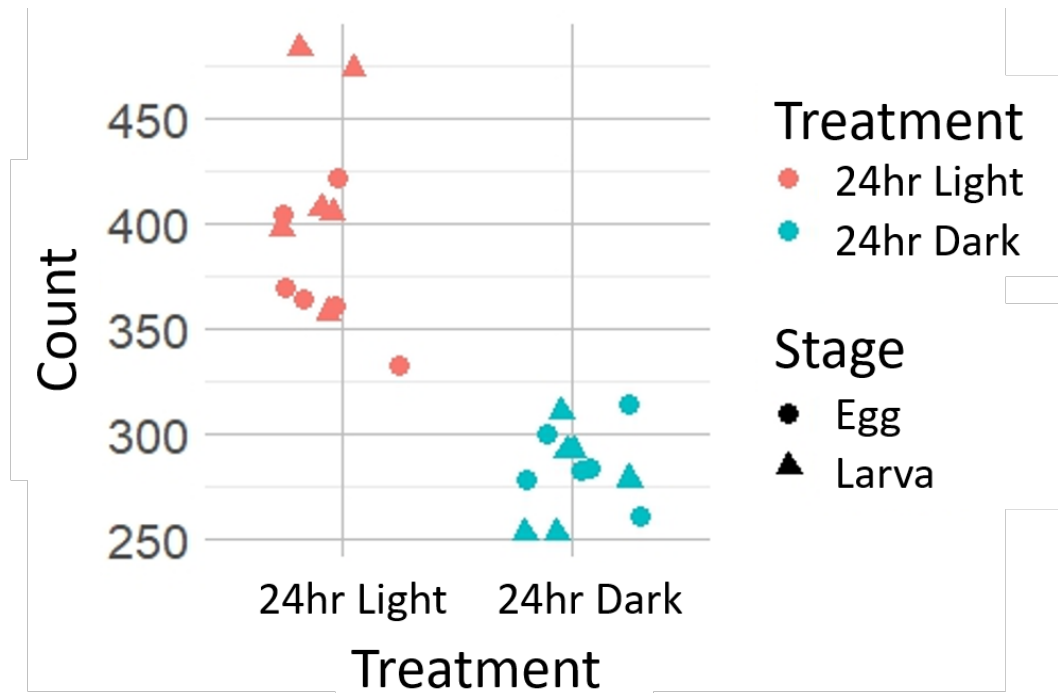


Figure 4: BVRA gene expression differences between individuals. Color represents treatment and life stage is represented by shape.

The up-regulation of BVRA and other energy consumptive genes could lead to the smaller YSA at hatch in the light treatment. Because BVRA is linked to increased lipid uptake and the yolk sac is predominantly lipids, the up-regulation of BVRA and other ATP and GTP related genes seems complementary to the life history YSV findings. Additionally, if the increased larval mortality was linked to differences in light (which is something we cannot be confident about due to lack of replicates), perhaps the mortality was a domino effect of the reduced lipid reserves due to increased expression of BVRA and other genes and therefore not enough energy reserves were available to help bridge the larvae to exogenous feeding.

The explanations provided are one of many potential reasons for the observed results. To improve this line of research, future studies should incorporate tank replicates

and multiple light intensities to better simulate climate change scenarios brought on by differing ice conditions. Additionally, gene annotations are biased towards “model organisms” and functions that are highly researched, such as metabolism. Increasing the breadth of organisms and genes that are annotated would help increase our confidence in linking the transcriptome findings to the life history analyses.

Despite some of the limitations of our pilot experiment, our results suggest 1) transcriptomics can enhance observational results and 2) the possible link from climate change to reduced ice cover to changes in light exposure to eggs is worth further investigation. Understanding how reduced ice cover could impact cisco early life history will help researchers and managers when making decisions on where to focus conservation and restoration efforts. For example, if light is crucial to early life history, stocking should occur in spawning areas that historically have high ice cover throughout the winter (i.e. Green Bay, Lake Michigan which experiences consistent ice coverage and has been deemed suitable habitat for cisco eggs) (Madenjian et al. 2011). Our work illustrates the potential of weaving molecular and traditional techniques together and foreshadows how powerful these methods could be in a fully replicated experiment.

2.6 Supplemental Information

Section 1: Water quality methods

Each week, water chemistry for egg incubation and larval rearing was monitored for pH, NH_4 , NO_2^- , NO_3^- , and tanks were visually inspected for high bacterial growth. An estimated 200 L of water was changed weekly. During egg rearing, tanks were formalin

treated three times a week, with a concentration that varied based on fungal growth from 1000-1600 ppm of formalin.

Section 2: RNA extraction protocol

Briefly, each individual egg or larva was placed in separate 2.0 ml screwcap tubes containing 500 μ l of TRIzol and two 5 mm stainless steel beads. The tubes were then run on a Mini-Beadbeater-16 (Biospec, Bartlesville, OK, USA) until the eggs/larvae were homogenized (15-30 seconds). Homogenized mixture was transferred to 1.5 ml tubes and 100 μ l of chloroform was added, shaken vigorously, incubated at room temperature for 2-3 min and centrifuged at 4°C for 15 min. After centrifugation, the upper (clear) layer was removed, placed in a new 1.5 ml tube and an equal volume of 70% ethanol was added to each sample and vortexed. The remainder of the extraction was performed using a Purelink RNA mini kit (Ambion[®], Life Technologies, Carlsbad, CA, USA) and followed the manufacturer's protocol with the addition of a 15 minute DNase I treatment (Ambion, Life Technologies, Carlsbad, CA, USA) after 350 μ l of Wash Buffer I was spun through the column and before the second 350 μ l of Wash Buffer I was added and spun. Extracted RNA was quantified using a Qubit[®] 3.0 Fluorometer (Invitrogen[™], Life Technologies, Carlsbad, CA, USA) and the quality was assessed through an RNA agarose gel.

Table 2: Genes and proteins that were differentially expressed in both eggs and larvae (excluding uncharacterized proteins). Negative stat indicates up-regulated in light, down-regulated in dark and positive stat indicates up-regulated in dark and down-regulated in light.

| Protein | Gene | Stat | Gene Ontology (GO) |
|--|--------|----------|--|
| Guanine nucleotide-binding protein subunit gamma | | 12.1196 | plasma membrane [GO:0005886]; signal transduction [GO:0007165] |
| Anterior gradient protein 3 homolog | AGR3 | 6.028271 | fin regeneration [GO:0031101] |
| Tubulin alpha chain | TUBA1A | 5.541001 | cytoplasm [GO:0005737]; microtubule [GO:0005874]; GTP binding [GO:0005525]; GTPase activity [GO:0003924] |
| Protein S100 (S100 calcium-binding protein) | | 5.090384 | metal ion binding [GO:0046872] |
| Ferritin | | 4.351921 | cell [GO:0005623]; metal ion binding [GO:0046872]; cellular iron ion homeostasis [GO:0006879] |
| Lon protease homolog, mitochondrial (| LONP1 | -4.25665 | mitochondrial matrix [GO:0005759]; ATP binding [GO:0005524]; ATP-dependent peptidase activity [GO:0004176]; sequence-specific DNA binding [GO:0043565]; serine-type endopeptidase activity [GO:0004252]; cellular response to oxidative stress [GO:0034599]; chaperone-mediated protein complex assembly [GO:0051131]; oxidation-dependent protein catabolic process [GO:0070407]; protein quality control for misfolded or incompletely synthesized proteins [GO:0006515] |
| Phosphate transporter | | -4.50514 | integral component of membrane [GO:0016021]; phosphate ion transport [GO:0006817] |
| Superoxide dismutase [Cu-Zn] | | -4.78632 | metal ion binding [GO:0046872]; superoxide dismutase activity [GO:0004784] |
| biliverdin reductase A | BVRA | -5.20545 | biliverdin reductase activity [GO:0004074]; zinc ion binding [GO:0008270]; heme catabolic process [GO:0042167] |
| Glycine cleavage system P protein | | -5.31711 | mitochondrion [GO:0005739]; glycine dehydrogenase (decarboxylating) activity [GO:0004375] |

| | | | |
|---|----------------|----------|---|
| Ribonucleoside-diphosphate reductase | | -6.13477 | ATP binding [GO:0005524]; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor [GO:0004748]; DNA replication [GO:0006260] |
| Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial | | -6.18279 | mitochondrial inner membrane [GO:0005743]; 2 iron, 2 sulfur cluster binding [GO:0051537]; 3 iron, 4 sulfur cluster binding [GO:0051538]; 4 iron, 4 sulfur cluster binding [GO:0051539]; metal ion binding [GO:0046872]; succinate dehydrogenase (ubiquinone) activity [GO:0008177]; tricarboxylic acid cycle [GO:0006099] |
| Succinate dehydrogenase assembly factor 2, mitochondrial (SDH assembly factor 2) (SDHAF2) | SDHAF2 PGL2 | -6.38261 | mitochondrial matrix [GO:0005759]; mitochondrial electron transport, succinate to ubiquinone [GO:0006121]; protein-FAD linkage [GO:0018293] |
| Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial | SDH5 SDHA | -6.94558 | mitochondrial inner membrane [GO:0005743]; succinate dehydrogenase (ubiquinone) activity [GO:0008177]; tricarboxylic acid cycle [GO:0006099] |

Table 3: Significantly differentially expressed genes ($p_{adj} < 0.01$) with GO terms found when comparing the light and dark treatment egg samples. (Excludes “uncharacterized proteins”). Negative stat indicates up-regulated in light, down-regulated in dark and positive stat indicates up-regulated in dark and down-regulated in light.

| Protein | Gene | Stat | GO terms |
|--|--------|----------|--|
| Guanine nucleotide-binding protein subunit gamma | | 12.1196 | plasma membrane [GO:0005886]; signal transduction [GO:0007165] |
| Poly [ADP-ribose] polymerase (PARP) | | 7.261808 | NAD+ ADP-ribosyltransferase activity [GO:0003950] |
| Histone-lysine N-methyltransferase | | 7.194263 | histone-lysine N-methyltransferase activity [GO:0018024]; metal ion binding [GO:0046872] |
| Anterior gradient protein 3 homolog | AGR3 | 6.028271 | fin regeneration [GO:0031101] |
| Galectin | | 5.70826 | carbohydrate binding [GO:0030246] |
| Tubulin alpha chain | TUBA1A | 5.541001 | cytoplasm [GO:0005737]; microtubule [GO:0005874]; GTP binding [GO:0005525]; GTPase activity [GO:0003924] |
| Calcium-transporting ATPase | | 5.350912 | integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; calcium ion transport [GO:0006816] |
| Ferritin | | 5.252838 | cell [GO:0005623]; metal ion binding [GO:0046872]; cellular iron ion homeostasis [GO:0006879] |
| Bloodthirsty-related gene family, member 12 | | 5.240118 | metal ion binding [GO:0046872] |
| Protein S100 (S100 calcium-binding protein) | | 5.090384 | metal ion binding [GO:0046872] |
| Phospholipid-transporting ATPase | | 5.078453 | integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; magnesium ion binding [GO:0000287]; phospholipid-translocating ATPase activity [GO:0004012] |
| Elongation factor 1-alpha | | 4.889833 | GTP binding [GO:0005525]; GTPase activity [GO:0003924]; translation elongation factor activity [GO:0003746] |

| | | | |
|--|-------|----------|--|
| RNA cytidine acetyltransferase (18S rRNA cytosine acetyltransferase) | NAT10 | -4.25351 | nucleolus [GO:0005730]; ATP binding [GO:0005524]; N-acetyltransferase activity [GO:0008080]; ribosomal small subunit biogenesis [GO:0042274]; rRNA modification [GO:0000154]; tRNA acetylation [GO:0051391] |
| Lon protease homolog, mitochondrial | LONP1 | -4.25665 | mitochondrial matrix [GO:0005759]; ATP binding [GO:0005524]; ATP-dependent peptidase activity [GO:0004176]; sequence-specific DNA binding [GO:0043565]; serine-type endopeptidase activity [GO:0004252]; cellular response to oxidative stress [GO:0034599]; chaperone-mediated protein complex assembly [GO:0051131]; oxidation-dependent protein catabolic process [GO:0070407]; protein quality control for misfolded or incompletely synthesized proteins [GO:0006515] |
| Glycylpeptide N- tetradecanoyltransferase | | -4.27663 | integral component of membrane [GO:0016021]; glycylpeptide N-tetradecanoyltransferase activity [GO:0004379] |
| MICOS complex subunit MIC60 (Mitofilin) | | -4.39756 | integral component of membrane [GO:0016021]; mitochondrial inner membrane [GO:0005743] |
| Phosphate transporter | | -4.50514 | integral component of membrane [GO:0016021]; phosphate ion transport [GO:0006817] |
| Serotransferrin | | -4.60502 | extracellular region [GO:0005576]; metal ion binding [GO:0046872]; ion transport [GO:0006811]; iron ion homeostasis [GO:0055072] |
| Superoxide dismutase [Cu-Zn] | | -4.78632 | metal ion binding [GO:0046872]; superoxide dismutase activity [GO:0004784] |
| Calmodulin | CALM | -4.79221 | calcium ion binding [GO:0005509]; calcium-mediated signaling [GO:0019722] |
| Histone H2A | | -4.8809 | nucleosome [GO:0000786]; nucleus [GO:0005634]; DNA binding [GO:0003677] |
| biliverdin reductase A | BVRA | -5.20545 | biliverdin reductase activity [GO:0004074]; zinc ion binding [GO:0008270]; heme catabolic process [GO:0042167] |
| Glycine cleavage system P protein | | -5.31711 | mitochondrion [GO:0005739]; glycine dehydrogenase (decarboxylating) activity [GO:0004375] |
| Suppressor of cytokine signaling 3a | | -5.93414 | protein ubiquitination [GO:0016567] |
| Ribonucleoside-diphosphate reductase | | -6.13477 | ATP binding [GO:0005524]; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor [GO:0004748]; DNA replication [GO:0006260] |
| Early growth response protein | | -6.15261 | nucleus [GO:0005634]; DNA binding [GO:0003677]; metal ion binding [GO:0046872] |

| | | | |
|---|-------------------------|----------|---|
| Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial | | -6.18279 | mitochondrial inner membrane [GO:0005743]; 2 iron, 2 sulfur cluster binding [GO:0051537]; 3 iron, 4 sulfur cluster binding [GO:0051538]; 4 iron, 4 sulfur cluster binding [GO:0051539]; metal ion binding [GO:0046872]; succinate dehydrogenase (ubiquinone) activity [GO:0008177]; tricarboxylic acid cycle [GO:0006099] |
| Succinate dehydrogenase assembly factor 2, mitochondrial (SDH assembly factor 2) (SDHAF2) | SDHAF 2 PGL2 SDH5 | -6.38261 | mitochondrial matrix [GO:0005759]; mitochondrial electron transport, succinate to ubiquinone [GO:0006121]; protein-FAD linkage [GO:0018293] |
| Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial | SDHA | -6.94558 | mitochondrial inner membrane [GO:0005743]; succinate dehydrogenase (ubiquinone) activity [GO:0008177]; tricarboxylic acid cycle [GO:0006099] |
| N-acetylglucosamine-6-sulfatase (Glucosamine-6-sulfatase) | | -7.4248 | lysosome [GO:0005764]; metal ion binding [GO:0046872]; N-acetylglucosamine-6-sulfatase activity [GO:0008449]; glycosaminoglycan metabolic process [GO:0030203] |

Table 4: Gene ontology (GO) terms and the frequency at which the GO terms appeared in the list of significantly DE ($p_{adj} < 0.01$) genes when comparing the light and dark treatment egg samples.

| GO ID | GO terms | GO # |
|------------|---|------|
| GO:0016021 | integral component of membrane | 20 |
| GO:0005524 | ATP binding | 8 |
| GO:0046872 | metal ion binding | 7 |
| GO:0005634 | nucleus | 7 |
| GO:0003677 | DNA binding | 6 |
| GO:0005743 | mitochondrial inner membrane | 4 |
| GO:0008177 | succinate dehydrogenase (ubiquinone) activity | 3 |
| GO:0006099 | tricarboxylic acid cycle | 3 |
| GO:0005623 | cell | 2 |
| GO:0006879 | cellular iron ion homeostasis | 2 |
| GO:0005737 | cytoplasm | 2 |
| GO:0005576 | extracellular region | 2 |
| GO:0005525 | GTP binding | 2 |
| GO:0003924 | GTPase activity | 2 |
| GO:0005759 | mitochondrial matrix | 2 |
| GO:0005886 | plasma membrane | 2 |
| GO:0016567 | protein ubiquitination | 2 |
| GO:0003723 | RNA binding | 2 |
| GO:0051537 | 2 iron, 2 sulfur cluster binding | 1 |
| GO:0051538 | 3 iron, 4 sulfur cluster binding | 1 |
| GO:0051539 | 4 iron, 4 sulfur cluster binding | 1 |
| GO:0003779 | actin binding | 1 |
| GO:0004176 | ATP-dependent peptidase activity | 1 |
| GO:0005923 | bicellular tight junction | 1 |
| GO:0004074 | biliverdin reductase activity | 1 |
| GO:0005509 | calcium ion binding | 1 |
| GO:0006816 | calcium ion transport | 1 |
| GO:0019722 | calcium-mediated signaling | 1 |
| GO:0030246 | carbohydrate binding | 1 |
| GO:0034599 | cellular response to oxidative stress | 1 |
| GO:0051131 | chaperone-mediated protein complex assembly | 1 |
| GO:0006260 | DNA replication | 1 |
| GO:0031101 | fin regeneration | 1 |
| GO:0004930 | G protein-coupled receptor activity | 1 |
| GO:0004375 | glycine dehydrogenase (decarboxylating) activity | 1 |
| GO:0030203 | glycosaminoglycan metabolic process | 1 |
| GO:0004379 | glycylpeptide N-tetradecanoyltransferase activity | 1 |
| GO:0042167 | heme catabolic process | 1 |
| GO:0018024 | histone-lysine N-methyltransferase activity | 1 |
| GO:0005882 | intermediate filament | 1 |

| | | |
|------------|--|---|
| GO:0006811 | ion transport | 1 |
| GO:0055072 | iron ion homeostasis | 1 |
| GO:0006869 | lipid transport | 1 |
| GO:0005764 | lysosome | 1 |
| GO:0000287 | magnesium ion binding | 1 |
| GO:0005874 | microtubule | 1 |
| GO:0006121 | mitochondrial electron transport, succinate to ubiquinone | 1 |
| GO:0005739 | mitochondrion | 1 |
| GO:0008449 | N-acetylglucosamine-6-sulfatase activity | 1 |
| GO:0008080 | N-acetyltransferase activity | 1 |
| GO:0003950 | NAD+ ADP-ribosyltransferase activity | 1 |
| GO:0005730 | nucleolus | 1 |
| GO:0000786 | nucleosome | 1 |
| GO:0070407 | oxidation-dependent protein catabolic process | 1 |
| GO:0006817 | phosphate ion transport | 1 |
| GO:0004012 | phospholipid-translocating ATPase activity | 1 |
| GO:0009881 | photoreceptor activity | 1 |
| GO:0007602 | phototransduction | 1 |
| | protein quality control for misfolded or incompletely synthesized proteins | 1 |
| GO:0006515 | | 1 |
| GO:0018298 | protein-chromophore linkage | 1 |
| GO:0018293 | protein-FAD linkage | 1 |
| | ribonucleoside-diphosphate reductase activity, thioredoxin | 1 |
| GO:0004748 | disulfide as acceptor | 1 |
| GO:0042274 | ribosomal small subunit biogenesis | 1 |
| GO:0000154 | rRNA modification | 1 |
| GO:0043565 | sequence-specific DNA binding | 1 |
| GO:0004252 | serine-type endopeptidase activity | 1 |
| GO:0007165 | signal transduction | 1 |
| GO:0005198 | structural molecule activity | 1 |
| GO:0004784 | superoxide dismutase activity | 1 |
| GO:0016740 | transferase activity | 1 |
| GO:0003746 | translation elongation factor activity | 1 |
| GO:0051391 | tRNA acetylation | 1 |
| GO:0008270 | zinc ion binding | 1 |

Table 5: Significantly differentially expressed gene ($p_{adj} < 0.01$) with GO terms found when comparing the light and dark treatment larval samples. Negative Stat indicates up-regulated in light, down-regulated in dark and positive Stat indicates up-regulated in dark and down-regulated in light.

| Protein | Gene | Stat | Gene ontology (GO) |
|---|--------|----------|---|
| Serotransferrin | | 11.51836 | extracellular region [GO:0005576]; metal ion binding [GO:0046872]; ion transport [GO:0006811]; iron ion homeostasis [GO:0055072] |
| 40S ribosomal protein S23 | RS23 | 8.093239 | small ribosomal subunit [GO:0015935]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412] |
| Guanine nucleotide-binding protein subunit gamma | | 6.721109 | plasma membrane [GO:0005886]; signal transduction [GO:0007165] |
| Caveolin | | 6.340389 | caveola [GO:0005901]; Golgi membrane [GO:0000139]; integral component of membrane [GO:0016021] |
| Gamma-aminobutyric acid receptor-associated protein | GBRAP | 6.187423 | cytoplasm [GO:0005737]; autophagy [GO:0006914]; head development [GO:0060322] |
| Tubulin alpha chain | TUBA1A | 5.934186 | cytoplasm [GO:0005737]; microtubule [GO:0005874]; GTP binding [GO:0005525]; GTPase activity [GO:0003924] |
| Anterior gradient protein 3 homolog | AGR3 | 5.660561 | fin regeneration [GO:0031101] |
| Tetraspanin | | 5.505337 | integral component of membrane [GO:0016021] |
| Si:ch211-212k18.7 | | 5.463345 | integral component of membrane [GO:0016021]; lysosomal membrane [GO:0005765] |
| Elongation factor 1-alpha | | 5.229225 | GTP binding [GO:0005525]; GTPase activity [GO:0003924]; translation elongation factor activity [GO:0003746] |
| Non-specific serine/threonine protein kinase | | 5.089999 | ATP binding [GO:0005524]; protein serine/threonine kinase activity [GO:0004674] |
| Sodium/potassium-transporting ATPase subunit alpha | | 4.969044 | integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; metal ion binding [GO:0046872]; potassium ion transport [GO:0006813] |

| | | | |
|--|--------|----------|---|
| Phosphoinositide phospholipase C | | 4.960886 | phosphatidylinositol phospholipase C activity [GO:0004435]; lipid catabolic process [GO:0016042]; signal transduction [GO:0007165] |
| Amine oxidase | | 4.903326 | metal ion binding [GO:0046872]; oxidoreductase activity [GO:0016491]; amine metabolic process [GO:0009308] |
| Elongation of very long chain fatty acids protein 6 (3-keto acyl-CoA synthase ELOVL6) (ELOVL fatty acid elongase 6) (ELOVL FA elongase 6) (Very long chain 3-ketoacyl-CoA synthase 6) (Very long chain 3-oxoacyl-CoA synthase 6) | ELOVL6 | 4.888346 | endoplasmic reticulum membrane [GO:0005789]; integral component of membrane [GO:0016021]; 3-oxo-arachidoyl-CoA synthase activity [GO:0102336]; 3-oxo-cerotoyl-CoA synthase activity [GO:0102337]; 3-oxo-lignoceroyl-CoA synthase activity [GO:0102338]; fatty acid elongase activity [GO:0009922]; very-long-chain 3-ketoacyl-CoA synthase activity [GO:0102756]; fatty acid elongation, monounsaturated fatty acid [GO:0034625]; fatty acid elongation, saturated fatty acid [GO:0019367]; long-chain fatty-acyl-CoA biosynthetic process [GO:0035338]; unsaturated fatty acid biosynthetic process [GO:0006636] |
| NADH-ubiquinone oxidoreductase chain 5 | ND5 | 4.869845 | integral component of membrane [GO:0016021]; mitochondrial inner membrane [GO:0005743]; respirasome [GO:0070469]; NADH dehydrogenase (ubiquinone) activity [GO:0008137]; ATP synthesis coupled electron transport [GO:0042773] |
| Deoxyribonuclease | | 4.840965 | endonuclease activity [GO:0004519] |
| Ferritin | | 4.821933 | cell [GO:0005623]; metal ion binding [GO:0046872]; cellular iron ion homeostasis [GO:0006879] |
| Protein S100 (S100 calcium-binding protein) | | 4.750881 | metal ion binding [GO:0046872] |
| Zgc:136908 | | 4.700657 | ATP binding [GO:0005524] |
| NADH dehydrogenase 1 alpha subcomplex subunit 4-like 2 | NUA4L | 4.274592 | integral component of membrane [GO:0016021] |
| Glyceraldehyde-3-phosphate dehydrogenase | | 4.259603 | glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity [GO:0004365]; NAD binding [GO:0051287]; glycolytic process [GO:0006096] |

| | | | |
|---|--------|----------|--|
| Polyadenylate-binding protein (PABP) | | 4.185736 | cytoplasm [GO:0005737]; RNA binding [GO:0003723] |
| Claudin | | 4.178295 | bicellular tight junction [GO:0005923]; integral component of membrane [GO:0016021]; plasma membrane [GO:0005886] |
| Draxin (Dorsal inhibitory axon guidance protein) (Dorsal repulsive axon guidance protein) | DRAXIN | 4.158637 | extracellular region [GO:0005576]; axon guidance [GO:0007411]; commissural neuron differentiation in spinal cord [GO:0021528]; dorsal spinal cord development [GO:0021516]; forebrain development [GO:0030900] |
| T-complex protein 1 subunit delta | | 4.13838 | cytoplasm [GO:0005737]; ATP binding [GO:0005524] |
| 40S ribosomal protein S4 | | 4.126159 | ribosome [GO:0005840]; rRNA binding [GO:0019843]; structural constituent of ribosome [GO:0003735] |
| RNA binding protein fox-1 homolog | | 4.122977 | nucleus [GO:0005634]; RNA binding [GO:0003723]; mRNA processing [GO:0006397]; RNA splicing [GO:0008380] |
| Calcium-transporting ATPase | | 4.053235 | integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; calcium ion transport [GO:0006816] |
| Cell cycle control protein | | 3.954891 | integral component of membrane [GO:0016021] |
| Integrin beta | | 3.954779 | integral component of membrane [GO:0016021]; cell adhesion [GO:0007155]; integrin-mediated signaling pathway [GO:0007229] |
| ATP-binding cassette, sub-family A (ABC1), member 12 | | 3.946484 | integral component of membrane [GO:0016021]; ATP binding [GO:0005524] |
| 39S ribosomal protein L30, mitochondrial | RM30 | 3.885298 | large ribosomal subunit [GO:0015934]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412] |
| 60S ribosomal export protein NMD3 | | 3.857834 | cytoplasm [GO:0005737]; nucleus [GO:0005634]; protein transport [GO:0015031] |
| O-acyltransferase | | 3.848701 | endoplasmic reticulum membrane [GO:0005789]; integral component of membrane [GO:0016021]; transferase activity, transferring acyl groups [GO:0016746] |
| Receptor expression-enhancing protein | REEP5 | 3.808704 | integral component of membrane [GO:0016021] |

| | | | |
|---|---------------------|----------|--|
| Spectrin beta chain | | 3.790497 | cytoplasm [GO:0005737]; cytoskeleton [GO:0005856]; actin binding [GO:0003779]; structural constituent of cytoskeleton [GO:0005200]; actin filament capping [GO:0051693] |
| Si:ch1073-184j22.2 | | 3.782672 | hydrolase activity [GO:0016787] |
| NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial | | 3.778128 | mitochondrial matrix [GO:0005759]; respirasome [GO:0070469]; oxidation-reduction process [GO:0055114] |
| 6-phosphogluconate dehydrogenase, decarboxylating | | -3.77832 | phosphogluconate dehydrogenase (decarboxylating) activity [GO:0004616]; D-gluconate metabolic process [GO:0019521]; pentose-phosphate shunt [GO:0006098] |
| Anthrax toxin receptor | | -3.78562 | integral component of membrane [GO:0016021]; metal ion binding [GO:0046872] |
| Dihydropyrimidine dehydrogenase [NADP(+)] (DHPDHase) (DPD) (Dihydrothymine dehydrogenase) (Dihydrouracil dehydrogenase) | | -3.79524 | 4 iron, 4 sulfur cluster binding [GO:0051539]; dihydropyrimidine dehydrogenase (NADP+) activity [GO:0017113]; beta-alanine biosynthetic process [GO:0019483] |
| Eukaryotic translation initiation factor 3 subunit L (eIF3L) | EIF3L | -3.81062 | eukaryotic 43S preinitiation complex [GO:0016282]; eukaryotic 48S preinitiation complex [GO:0033290]; eukaryotic translation initiation factor 3 complex [GO:0005852]; translation initiation factor activity [GO:0003743]; formation of cytoplasmic translation initiation complex [GO:0001732] |
| (Eukaryotic translation initiation factor 3 subunit 6-interacting protein) (Eukaryotic translation initiation factor 3 subunit E-interacting protein) | EIF3EIP EIF3S6IP | | |
| Fructose-bisphosphate aldolase | | -3.81062 | fructose-bisphosphate aldolase activity [GO:0004332]; glycolytic process [GO:0006096] |
| Aminopeptidase | | -3.82781 | integral component of membrane [GO:0016021]; aminopeptidase activity [GO:0004177]; metallopeptidase activity [GO:0008237]; zinc ion binding [GO:0008270] |

| | | | |
|---|--------------|------------|---|
| aminopeptidase N-like | LOC106560949 | -3.8278133 | integral component of membrane [GO:0016021]; aminopeptidase activity [GO:0004177]; metallopeptidase activity [GO:0008237]; zinc ion binding [GO:0008270] |
| Acid sphingomyelinase-like phosphodiesterase | | -3.83579 | extracellular space [GO:0005615]; hydrolase activity [GO:0016787]; metal ion binding [GO:0046872] |
| Sulfotransferase | | -3.83582 | transferase activity [GO:0016740] |
| Prolyl endopeptidase | PPCE | -3.8360244 | serine-type endopeptidase activity [GO:0004252]; serine-type exopeptidase activity [GO:0070008] |
| Probable cytosolic iron-sulfur protein assembly protein CIAO1 (WD repeat-containing protein 39) | CIAO1 | -3.85561 | iron-sulfur cluster assembly [GO:0016226] |
| Eukaryotic translation initiation factor 5A (eIF-5A) | | -3.85673 | ribosome binding [GO:0043022]; translation elongation factor activity [GO:0003746]; positive regulation of translational elongation [GO:0045901]; positive regulation of translational termination [GO:0045905]; translational frameshifting [GO:0006452] |
| Plasminogen | | -3.86239 | extracellular region [GO:0005576]; serine-type endopeptidase activity [GO:0004252]; blood coagulation [GO:0007596]; fibrinolysis [GO:0042730]; tissue remodeling [GO:0048771] |
| Protein disulfide-isomerase A4 | | -3.86444 | endoplasmic reticulum lumen [GO:0005788]; protein disulfide isomerase activity [GO:0003756] |
| Arginyl-tRNA--protein transferase 1 (Arginyltransferase 1) (R-transferase 1) (Arginine-tRNA--protein transferase 1) | ATE1 | -3.9075 | arginyltransferase activity [GO:0004057] |
| Protein transport protein SEC23 | | -3.94331 | cytosol [GO:0005829]; endoplasmic reticulum membrane [GO:0005789]; ER to Golgi transport vesicle membrane [GO:0012507]; metal ion binding [GO:0046872]; protein transport [GO:0015031]; vesicle-mediated transport [GO:0016192] |

| | | | |
|--|------------------|----------|--|
| Sodium/potassium-transporting ATPase subunit beta | | -3.949 | integral component of membrane [GO:0016021]; ion transport [GO:0006811] |
| Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) | | -3.96537 | endoplasmic reticulum membrane [GO:0005789]; integral component of membrane [GO:0016021] |
| NADPH--cytochrome P450 reductase (CPR) (P450R0) | POR | -3.9884 | endoplasmic reticulum membrane [GO:0005789]; integral component of membrane [GO:0016021]; flavin adenine dinucleotide binding [GO:0050660]; FMN binding [GO:0010181]; NADP binding [GO:0050661]; NADPH-hemoprotein reductase activity [GO:0003958] |
| Succinate dehydrogenase assembly factor 2, mitochondrial (SDH assembly factor 2) (SDHAF2) | SDHAF2 PGL2 SDH5 | -4.00408 | mitochondrial matrix [GO:0005759]; mitochondrial electron transport, succinate to ubiquinone [GO:0006121]; protein-FAD linkage [GO:0018293] |
| alpha-1,2-Mannosidase | | -4.07208 | hydrolase activity, acting on glycosyl bonds [GO:0016798]; metabolic process [GO:0008152] |
| Alkaline phosphatase | | -4.1367 | alkaline phosphatase activity [GO:0004035] |
| ELAV-like protein | | -4.13718 | RNA binding [GO:0003723] |
| Glutamine synthetase | | -4.14059 | ATP binding [GO:0005524]; glutamate-ammonia ligase activity [GO:0004356] |
| Protein phosphatase 1 regulatory subunit | | -4.1415 | glycogen metabolic process [GO:0005977] |
| Elongation of very long chain fatty acids protein (Very-long-chain 3-oxoacyl-CoA synthase) | | -4.14563 | integral component of membrane [GO:0016021]; 3-oxo-arachidoyl-CoA synthase activity [GO:0102336]; 3-oxo-cerotoyl-CoA synthase activity [GO:0102337]; 3-oxo-lignoceroyl-CoA synthase activity [GO:0102338]; very-long-chain 3-ketoacyl-CoA synthase activity [GO:0102756]; fatty acid biosynthetic process [GO:0006633] |

| | | | |
|---|--------------|----------|--|
| Meprin A subunit (Endopeptidase-2) | | -4.14851 | integral component of membrane [GO:0016021]; metalloendopeptidase activity [GO:0004222]; zinc ion binding [GO:0008270] |
| Cysteine dioxygenase | | -4.15224 | cysteine dioxygenase activity [GO:0017172]; iron ion binding [GO:0005506]; taurine biosynthetic process [GO:0042412] |
| Histone H2A | | -4.17369 | nucleosome [GO:0000786]; nucleus [GO:0005634]; DNA binding [GO:0003677] |
| Metalloendopeptidase | | -4.18284 | metal ion binding [GO:0046872]; metalloendopeptidase activity [GO:0004222] |
| Histone deacetylase | | -4.25629 | nucleus [GO:0005634]; NAD-dependent histone deacetylase activity (H3-K14 specific) [GO:0032041] |
| Peptidylprolyl isomerase | | -4.26385 | peptidyl-prolyl cis-trans isomerase activity [GO:0003755] |
| Glutamine-dependent NAD(+) synthetase (NAD(+) synthase [glutamine-hydrolyzing]) | | -4.30038 | ATP binding [GO:0005524]; NAD+ synthase (glutamine-hydrolyzing) activity [GO:0003952]; NAD biosynthetic process [GO:0009435] |
| Eukaryotic translation initiation factor 3 subunit C (eIF3c) (Eukaryotic translation initiation factor 3 subunit 8) | EIF3C EIF3S8 | -4.31052 | eukaryotic 43S preinitiation complex [GO:0016282]; eukaryotic 48S preinitiation complex [GO:0033290]; eukaryotic translation initiation factor 3 complex [GO:0005852]; translation initiation factor activity [GO:0003743]; formation of cytoplasmic translation initiation complex [GO:0001732] |
| Aldehyde dehydrogenase | | -4.33961 | integral component of membrane [GO:0016021]; oxidoreductase activity [GO:0016491] |
| Cyclin-dependent kinases regulatory subunit fibrillin-2-like | | -4.35841 | cell cycle [GO:0007049]; cell division [GO:0051301] |
| | LOC106573567 | -4.37523 | collagen-containing extracellular matrix [GO:0062023]; calcium ion binding [GO:0005509]; extracellular matrix structural constituent [GO:0005201] |
| Calponin | | -4.37548 | actin binding [GO:0003779]; calmodulin binding [GO:0005516] |
| NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial | | -4.38063 | mitochondrial inner membrane [GO:0005743]; respirasome [GO:0070469]; 4 iron, 4 sulfur cluster binding [GO:0051539]; |

| | | | |
|---|-----------------------|----------|--|
| Nuclear pore complex protein | | -4.40103 | metal ion binding [GO:0046872]; NAD binding [GO:0051287]; NADH dehydrogenase activity [GO:0003954] nuclear membrane [GO:0031965]; nuclear pore outer ring [GO:0031080]; structural constituent of nuclear pore [GO:0017056]; mRNA export from nucleus [GO:0006406]; posttranscriptional tethering of RNA polymerase II gene DNA at nuclear periphery [GO:0000973]; protein import into nucleus [GO:0006606] |
| Double-strand break repair protein | | -4.41025 | Mre11 complex [GO:0030870]; endonuclease activity [GO:0004519]; exonuclease activity [GO:0004527]; manganese ion binding [GO:0030145]; double-strand break repair [GO:0006302]; meiotic cell cycle [GO:0051321] |
| Transmembrane emp24 domain-containing protein 2 | TMED2 | -4.47107 | integral component of membrane [GO:0016021] |
| 5-demethoxyubiquinone hydroxylase, mitochondrial (DMQ hydroxylase (Timing protein clk-1 homolog) (Ubiquinone biosynthesis monooxygenase COQ7) | COQ7 | -4.4864 | extrinsic component of mitochondrial inner membrane [GO:0031314]; metal ion binding [GO:0046872]; oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen [GO:0016709]; ubiquinone biosynthetic process [GO:0006744] |
| FACT complex subunit SSRP1 | SSRP1 | -4.48701 | chromosome [GO:0005694]; nucleus [GO:0005634]; DNA binding [GO:0003677]; DNA repair [GO:0006281]; DNA replication [GO:0006260] |
| Dihydrolipoyl dehydrogenase | | -4.48893 | dihydrolipoyl dehydrogenase activity [GO:0004148] |
| Peptidyl-prolyl cis-trans isomerase (PPIase) | PPIB | -4.49668 | peptidyl-prolyl cis-trans isomerase activity [GO:0003755]; protein folding [GO:0006457] |
| 3-hydroxyisobutyryl-CoA hydrolase, mitochondrial (3-hydroxyisobutyryl-coenzyme A hydrolase) | GSONMT00 070980001 | -4.51556 | mitochondrion [GO:0005739]; 3-hydroxyisobutyryl-CoA hydrolase activity [GO:0003860]; valine catabolic process [GO:0006574] |

| | | | |
|--|--------|----------|--|
| Frizzled class receptor 1 | | -4.53338 | integral component of membrane [GO:0016021]; multicellular organism development [GO:0007275]; Wnt signaling pathway [GO:0016055] |
| Vacuolar protein sorting-associated protein 28 homolog | | -4.54354 | ESCRT I complex [GO:0000813]; protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway [GO:0043328] |
| Glycine cleavage system P protein | | -4.56325 | mitochondrion [GO:0005739]; glycine dehydrogenase (decarboxylating) activity [GO:0004375] |
| 10-formyltetrahydrofolate dehydrogenase | | -4.585 | formyltetrahydrofolate dehydrogenase activity [GO:0016155]; 10-formyltetrahydrofolate catabolic process [GO:0009258]; one-carbon metabolic process [GO:0006730] |
| Small nuclear ribonucleoprotein 35 (U11/U12) | | -4.64852 | RNA binding [GO:0003723] |
| Malate dehydrogenase | | -4.70634 | L-malate dehydrogenase activity [GO:0030060]; tricarboxylic acid cycle [GO:0006099] |
| Calreticulin 3b | | -4.71348 | endoplasmic reticulum [GO:0005783] |
| Superoxide dismutase [Cu-Zn] (EC 1.15.1.1) | | -4.74534 | metal ion binding [GO:0046872]; superoxide dismutase activity [GO:0004784] |
| Lon protease homolog, mitochondrial | LONP1 | -4.74725 | mitochondrial matrix [GO:0005759]; ATP binding [GO:0005524]; ATP-dependent peptidase activity [GO:0004176]; sequence-specific DNA binding [GO:0043565]; serine-type endopeptidase activity [GO:0004252]; cellular response to oxidative stress [GO:0034599]; chaperone-mediated protein complex assembly [GO:0051131]; oxidation-dependent protein catabolic process [GO:0070407]; protein quality control for misfolded or incompletely synthesized proteins [GO:0006515] |
| Carboxylic ester hydrolase | | -4.76444 | hydrolase activity [GO:0016787] |
| Sulfurtransferase | | -4.76792 | transferase activity [GO:0016740] |
| Serine/threonine-protein phosphatase | PPP1CC | -4.80403 | metal ion binding [GO:0046872]; phosphoprotein phosphatase activity [GO:0004721] |

| | | | |
|---|--------------|----------|--|
| Dimethylaniline monooxygenase [N-oxide-forming] | | -4.89847 | endoplasmic reticulum membrane [GO:0005789]; N,N-dimethylaniline monooxygenase activity [GO:0004499] |
| Solute carrier family 35 member F6 | | -4.94373 | integral component of membrane [GO:0016021]; mitochondrion [GO:0005739]; transmembrane transporter activity [GO:0022857]; negative regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway [GO:1901029]; positive regulation of cell population proliferation [GO:0008284] |
| DNA helicase | | -4.98461 | nucleus [GO:0005634]; ATP binding [GO:0005524]; DNA binding [GO:0003677]; helicase activity [GO:0004386]; DNA replication [GO:0006260] |
| Serine--pyruvate aminotransferase (Alanine--glyoxylate aminotransferase) | | -4.99392 | alanine-glyoxylate transaminase activity [GO:0008453]; serine-pyruvate transaminase activity [GO:0004760] |
| Importin subunit alpha | | -4.99415 | protein transport [GO:0015031] |
| SURF1-like protein | | -5.01703 | integral component of membrane [GO:0016021]; mitochondrial inner membrane [GO:0005743] |
| Staphylococcal nuclease domain-containing protein | | -5.02181 | cytoplasm [GO:0005737]; RISC complex [GO:0016442]; endoribonuclease activity [GO:0004521] |
| Eukaryotic translation initiation factor 3 subunit B (eIF3b) (Eukaryotic translation initiation factor 3 subunit 9) (eIF-3-eta) | EIF3B EIF3S9 | -5.06039 | eukaryotic 43S preinitiation complex [GO:0016282]; eukaryotic 48S preinitiation complex [GO:0033290]; eukaryotic translation initiation factor 3 complex [GO:0005852]; translation initiation factor activity [GO:0003743]; formation of cytoplasmic translation initiation complex [GO:0001732] |
| Rab GDP dissociation inhibitor | | -5.06937 | cytoplasm [GO:0005737]; GTPase activator activity [GO:0005096] |
| ER membrane protein complex subunit 3 | POB | -5.08046 | integral component of membrane [GO:0016021] |
| Annexin | | -5.12297 | calcium-dependent phospholipid binding [GO:0005544] |
| biliverdin reductase A | BVRA | -5.20545 | biliverdin reductase activity [GO:0004074]; zinc ion binding [GO:0008270]; heme catabolic process [GO:0042167] |

| | | | |
|---|--------------|------------|--|
| Kinesin-like protein | | -5.22629 | microtubule [GO:0005874]; ATP binding [GO:0005524]; motor activity [GO:0003774] |
| 5-aminolevulinate synthase (5-aminolevulinic acid synthase) (Delta-ALA synthase) (Delta-aminolevulinate synthase) | | -5.23444 | mitochondrial matrix [GO:0005759]; 5-aminolevulinate synthase activity [GO:0003870]; pyridoxal phosphate binding [GO:0030170]; protoporphyrinogen IX biosynthetic process [GO:0006782] |
| Ribonucleoside-diphosphate reductase | | -5.25457 | ATP binding [GO:0005524]; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor [GO:0004748]; DNA replication [GO:0006260] |
| Transmembrane 9 superfamily member | | -5.25849 | integral component of membrane [GO:0016021] |
| DNA mismatch repair protein | | -5.26433 | MutSalpha complex [GO:0032301]; DNA binding [GO:0003677]; nucleotide binding [GO:0000166]; mismatch repair [GO:0006298] |
| chromosomal protein D1-like | LOC106568700 | -5.2940568 | DNA binding [GO:0003677] |
| Transaldolase | | -5.3755 | sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate glyceronetransferase activity [GO:0004801]; pentose-phosphate shunt [GO:0006098] |
| Histone H3 | | -5.38475 | nucleosome [GO:0000786]; nucleus [GO:0005634]; DNA binding [GO:0003677] |
| Leukotriene A(4) hydrolase (LTA-4 hydrolase) (EC 3.3.2.6) | | -5.65261 | cytoplasm [GO:0005737]; leukotriene-A4 hydrolase activity [GO:0004463]; metal ion binding [GO:0046872]; metalloproteinase activity [GO:0008237]; leukotriene biosynthetic process [GO:0019370] |
| Adenosylhomocysteinase | | -5.65343 | adenosylhomocysteinase activity [GO:0004013]; one-carbon metabolic process [GO:0006730] |
| Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit (Oligosaccharyl transferase 48 kDa subunit) | | -5.73863 | endoplasmic reticulum membrane [GO:0005789]; integral component of membrane [GO:0016021]; protein N-linked glycosylation via asparagine [GO:0018279] |

| | | | |
|---|--------------|------------|---|
| Condensin complex subunit 2 | | -5.7956 | cell cycle [GO:0007049]; cell division [GO:0051301]; chromosome condensation [GO:0030261] |
| Phosphoglycerate mutase | | -5.86159 | bisphosphoglycerate mutase activity [GO:0004082]; glycolytic process [GO:0006096] |
| DNA topoisomerase 2 | | -6.0591 | ATP binding [GO:0005524]; DNA binding [GO:0003677]; DNA topoisomerase type II (ATP-hydrolyzing) activity [GO:0003918] |
| Protein disulfide-isomerase | | -6.16475 | protein disulfide isomerase activity [GO:0003756] |
| Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial | SDHA | -6.52711 | mitochondrial inner membrane [GO:0005743]; succinate dehydrogenase (ubiquinone) activity [GO:0008177]; tricarboxylic acid cycle [GO:0006099] |
| RBR-type E3 ubiquitin transferase | | -6.92083 | metal ion binding [GO:0046872]; transferase activity [GO:0016740] |
| Phosphate transporter | | -7.60177 | integral component of membrane [GO:0016021]; phosphate ion transport [GO:0006817] |
| 40S ribosomal protein S30 | UBIM | -7.85654 | ribosome [GO:0005840]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412] |
| Eukaryotic translation initiation factor 3 subunit I (eIF3i) (Eukaryotic translation initiation factor 3 subunit 2) (eIF-3-beta) (eIF3 p36) | EIF3I EIF3S2 | -7.95065 | eukaryotic 43S preinitiation complex [GO:0016282]; eukaryotic 48S preinitiation complex [GO:0033290]; eukaryotic translation initiation factor 3 complex [GO:0005852]; translation initiation factor activity [GO:0003743]; formation of cytoplasmic translation initiation complex [GO:0001732] |
| Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial | | -8.59054 | mitochondrial inner membrane [GO:0005743]; 2 iron, 2 sulfur cluster binding [GO:0051537]; 3 iron, 4 sulfur cluster binding [GO:0051538]; 4 iron, 4 sulfur cluster binding [GO:0051539]; metal ion binding [GO:0046872]; succinate dehydrogenase (ubiquinone) activity [GO:0008177]; tricarboxylic acid cycle [GO:0006099] |
| zinc finger CCHC domain-containing protein 7-like | LOC106595047 | -9.0250082 | nucleic acid binding [GO:0003676]; zinc ion binding [GO:0008270] |

| | | |
|--|--------------|---|
| arachidonate 15-lipoxygenase B-like | LOC106593018 | iron ion binding [GO:0005506]; oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen [GO:0016702] |
| Calmodulin | CALM | calcium ion binding [GO:0005509]; calcium-mediated signaling [GO:0019722] |
| Nattectin (galactose-specific lectin nattectin-like) | NATTE | carbohydrate binding [GO:0030246] |
| transmembrane and immunoglobulin domain-containing protein 1 | LOC106588017 | integral component of membrane [GO:0016021] |
| UDP-glucuronosyltransferase | tmigd1 | integral component of membrane [GO:0016021]; glucuronosyltransferase activity [GO:0015020] |

Table 6: Gene ontology (GO) terms and the frequency at which they were significantly DE ($p_{adj} < 0.01$) when comparing the light and dark treatment in the larval samples

| GO Term | GO ID | GO # |
|---|--------------|-------------|
| integral component of membrane | GO:0016021 | 121 |
| ATP binding | GO:0005524 | 49 |
| metal ion binding | GO:0046872 | 42 |
| RNA binding | GO:0003723 | 23 |
| nucleus | GO:0005634 | 21 |
| DNA binding | GO:0003677 | 20 |
| cytoplasm | GO:0005737 | 14 |
| extracellular region | GO:0005576 | 13 |
| GTP binding | GO:0005525 | 9 |
| oxidoreductase activity | GO:0016491 | 9 |
| endoplasmic reticulum membrane | GO:0005789 | 8 |
| intermediate filament | GO:0005882 | 8 |
| motor activity | GO:0003774 | 8 |
| structural molecule activity | GO:0005198 | 8 |
| serine-type peptidase activity | GO:0008236 | 7 |
| transferase activity | GO:0016740 | 7 |
| metallopeptidase activity | GO:0008237 | 6 |
| mitochondrial inner membrane | GO:0005743 | 6 |
| zinc ion binding | GO:0008270 | 6 |
| actin binding | GO:0003779 | 5 |
| hydrolase activity | GO:0016787 | 5 |
| kinase activity | GO:0016301 | 5 |
| microtubule | GO:0005874 | 5 |
| plasma membrane | GO:0005886 | 5 |
| protein transport | GO:0015031 | 5 |
| cysteine-type peptidase activity | GO:0008234 | 4 |
| eukaryotic 43S preinitiation complex | GO:0016282 | 4 |
| eukaryotic 48S preinitiation complex | GO:0033290 | 4 |
| eukaryotic translation initiation factor 3 complex | GO:0005852 | 4 |
| formation of cytoplasmic translation initiation complex | GO:0001732 | 4 |
| lipid transport | GO:0006869 | 4 |
| mitochondrial matrix | GO:0005759 | 4 |
| mitochondrion | GO:0005739 | 4 |
| protein serine/threonine kinase activity | GO:0004674 | 4 |
| structural constituent of ribosome | GO:0003735 | 4 |
| translation | GO:0006412 | 4 |

| | | |
|---|------------|---|
| translation initiation factor activity | GO:0003743 | 4 |
| tricarboxylic acid cycle | GO:0006099 | 4 |
| 4 iron, 4 sulfur cluster binding | GO:0051539 | 3 |
| bicellular tight junction | GO:0005923 | 3 |
| DNA replication | GO:0006260 | 3 |
| endoplasmic reticulum | GO:0005783 | 3 |
| flavin adenine dinucleotide binding | GO:0050660 | 3 |
| glycolytic process | GO:0006096 | 3 |
| Golgi apparatus | GO:0005794 | 3 |
| GTPase activity | GO:0003924 | 3 |
| helicase activity | GO:0004386 | 3 |
| ion transport | GO:0006811 | 3 |
| monooxygenase activity | GO:0004497 | 3 |
| multicellular organism development | GO:0007275 | 3 |
| myosin complex | GO:0016459 | 3 |
| respirasome | GO:0070469 | 3 |
| serine-type endopeptidase activity | GO:0004252 | 3 |
| signal transduction | GO:0007165 | 3 |
| succinate dehydrogenase (ubiquinone) activity | GO:0008177 | 3 |
| vesicle-mediated transport | GO:0016192 | 3 |
| 3-hydroxyisobutyryl-CoA hydrolase activity | GO:0003860 | 2 |
| 3-oxo-arachidoyl-CoA synthase activity | GO:0102336 | 2 |
| 3-oxo-cerotoyl-CoA synthase activity | GO:0102337 | 2 |
| 3-oxo-lignoceronyl-CoA synthase activity | GO:0102338 | 2 |
| aminopeptidase activity | GO:0004177 | 2 |
| calcium ion binding | GO:0005509 | 2 |
| cell adhesion | GO:0007155 | 2 |
| cell cycle | GO:0007049 | 2 |
| cell division | GO:0051301 | 2 |
| endonuclease activity | GO:0004519 | 2 |
| glutamate-ammonia ligase activity | GO:0004356 | 2 |
| hydrolase activity | GO:0016787 | 2 |
| iron ion binding | GO:0005506 | 2 |
| metalloendopeptidase activity | GO:0004222 | 2 |
| NAD binding | GO:0051287 | 2 |
| nucleic acid binding | GO:0003676 | 2 |
| nucleosome | GO:0000786 | 2 |
| one-carbon metabolic process | GO:0006730 | 2 |
| pentose-phosphate shunt | GO:0006098 | 2 |
| peptidyl-prolyl cis-trans isomerase activity | GO:0003755 | 2 |

| | | |
|--|------------|---|
| protein disulfide isomerase activity | GO:0003756 | 2 |
| ribosome | GO:0005840 | 2 |
| structural constituent of cytoskeleton | GO:0005200 | 2 |
| transferase activity, transferring glycosyl groups | GO:0016757 | 2 |
| translation elongation factor activity | GO:0003746 | 2 |
| valine catabolic process | GO:0006574 | 2 |
| very-long-chain 3-ketoacyl-CoA synthase activity | GO:0102756 | 2 |
| Wnt signaling pathway | GO:0016055 | 2 |
| 10-formyltetrahydrofolate catabolic process | GO:0009258 | 1 |
| 2 iron, 2 sulfur cluster binding | GO:0051537 | 1 |
| 3 iron, 4 sulfur cluster binding | GO:0051538 | 1 |
| 5-aminolevulinate synthase activity | GO:0003870 | 1 |
| actin filament capping | GO:0051693 | 1 |
| adenosylhomocysteinase activity | GO:0004013 | 1 |
| alanine-glyoxylate transaminase activity | GO:0008453 | 1 |
| alkaline phosphatase activity | GO:0004035 | 1 |
| amine metabolic process | GO:0009308 | 1 |
| aminoacyl-tRNA ligase activity | GO:0004812 | 1 |
| arginyltransferase activity | GO:0004057 | 1 |
| aspartic-type endopeptidase activity | GO:0004190 | 1 |
| ATP synthesis coupled electron transport | GO:0042773 | 1 |
| ATP-dependent peptidase activity | GO:0004176 | 1 |
| autophagy | GO:0006914 | 1 |
| axon guidance | GO:0007411 | 1 |
| beta-alanine biosynthetic process | GO:0019483 | 1 |
| biliverdin reductase activity | GO:0004074 | 1 |
| bisphosphoglycerate mutase activity | GO:0004082 | 1 |
| blood coagulation | GO:0007596 | 1 |
| calcium ion transport | GO:0006816 | 1 |
| calcium-dependent phospholipid binding | GO:0005544 | 1 |
| calcium-mediated signaling | GO:0019722 | 1 |
| calmodulin binding | GO:0005516 | 1 |
| carbohydrate binding | GO:0030246 | 1 |
| carbohydrate metabolic process | GO:0005975 | 1 |
| caveola | GO:0005901 | 1 |
| cell | GO:0005623 | 1 |
| cellular iron ion homeostasis | GO:0006879 | 1 |
| cellular response to oxidative stress | GO:0034599 | 1 |
| chaperone-mediated protein complex assembly | GO:0051131 | 1 |
| chromosome | GO:0005694 | 1 |

| | | |
|--|------------|---|
| chromosome condensation | GO:0030261 | 1 |
| chromosome, telomeric region | GO:0000781 | 1 |
| collagen-containing extracellular matrix | GO:0062023 | 1 |
| commissural neuron differentiation in spinal cord | GO:0021528 | 1 |
| cysteine dioxygenase activity | GO:0017172 | 1 |
| cytoskeleton | GO:0005856 | 1 |
| cytosol | GO:0005829 | 1 |
| D-gluconate metabolic process | GO:0019521 | 1 |
| dihydrolipoyl dehydrogenase activity | GO:0004148 | 1 |
| dihydropyrimidine dehydrogenase (NADP+) activity | GO:0017113 | 1 |
| dioxygenase activity | GO:0051213 | 1 |
| dipeptidyl-peptidase activity | GO:0008239 | 1 |
| DNA repair | GO:0006281 | 1 |
| DNA topoisomerase type II (ATP-hydrolyzing) activity | GO:0003918 | 1 |
| dorsal spinal cord development | GO:0021516 | 1 |
| double-strand break repair | GO:0006302 | 1 |
| double-strand break repair via homologous recombination | GO:0000724 | 1 |
| endoplasmic reticulum lumen | GO:0005788 | 1 |
| endoribonuclease activity | GO:0004521 | 1 |
| ER to Golgi transport vesicle membrane | GO:0012507 | 1 |
| ESCRT I complex | GO:0000813 | 1 |
| exonuclease activity | GO:0004527 | 1 |
| extracellular matrix structural constituent | GO:0005201 | 1 |
| extracellular space | GO:0005615 | 1 |
| extrinsic component of mitochondrial inner membrane | GO:0031314 | 1 |
| fatty acid biosynthetic process | GO:0006633 | 1 |
| fatty acid elongase activity | GO:0009922 | 1 |
| fatty acid elongation, monounsaturated fatty acid | GO:0034625 | 1 |
| fatty acid elongation, saturated fatty acid | GO:0019367 | 1 |
| fibrinolysis | GO:0042730 | 1 |
| fin regeneration | GO:0031101 | 1 |
| FMN binding | GO:0010181 | 1 |
| forebrain development | GO:0030900 | 1 |
| formyltetrahydrofolate dehydrogenase activity | GO:0016155 | 1 |
| fructose-bisphosphate aldolase activity | GO:0004332 | 1 |
| glucuronosyltransferase activity | GO:0015020 | 1 |
| glutamine metabolic process | GO:0006541 | 1 |
| glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity | GO:0004365 | 1 |
| glycine dehydrogenase (decarboxylating) activity | GO:0004375 | 1 |
| glycogen metabolic process | GO:0005977 | 1 |

| | | |
|---|------------|---|
| Golgi membrane | GO:0000139 | 1 |
| GTPase activator activity | GO:0005096 | 1 |
| head development | GO:0060322 | 1 |
| heme catabolic process | GO:0042167 | 1 |
| hydrolase activity, acting on glycosyl bonds | GO:0016798 | 1 |
| integrin-mediated signaling pathway | GO:0007229 | 1 |
| iron ion homeostasis | GO:0055072 | 1 |
| iron-sulfur cluster assembly | GO:0016226 | 1 |
| isomerase activity | GO:0016853 | 1 |
| large ribosomal subunit | GO:0015934 | 1 |
| leukotriene biosynthetic process | GO:0019370 | 1 |
| leukotriene-A4 hydrolase activity | GO:0004463 | 1 |
| lipid catabolic process | GO:0016042 | 1 |
| L-malate dehydrogenase activity | GO:0030060 | 1 |
| long-chain fatty-acyl-CoA biosynthetic process | GO:0035338 | 1 |
| lyase activity | GO:0016829 | 1 |
| lysosomal membrane | GO:0005765 | 1 |
| manganese ion binding | GO:0030145 | 1 |
| meiotic cell cycle | GO:0051321 | 1 |
| membrane | GO:0016020 | 1 |
| metabolic process | GO:0008152 | 1 |
| methyltransferase activity | GO:0008168 | 1 |
| microtubule-based process | GO:0007017 | 1 |
| mismatch repair | GO:0006298 | 1 |
| mitochondrial electron transport, succinate to ubiquinone | GO:0006121 | 1 |
| Mre11 complex | GO:0030870 | 1 |
| mRNA export from nucleus | GO:0006406 | 1 |
| mRNA processing | GO:0006397 | 1 |
| MutSalpha complex | GO:0032301 | 1 |
| N,N-dimethylaniline monooxygenase activity | GO:0004499 | 1 |
| NAD biosynthetic process | GO:0009435 | 1 |
| NAD ⁺ synthase (glutamine-hydrolyzing) activity | GO:0003952 | 1 |
| NAD-dependent histone deacetylase activity (H3-K14 specific) | GO:0032041 | 1 |
| NADH dehydrogenase (ubiquinone) activity | GO:0008137 | 1 |
| NADH dehydrogenase activity | GO:0003954 | 1 |
| NADP binding | GO:0050661 | 1 |
| NADPH-hemoprotein reductase activity | GO:0003958 | 1 |
| negative regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway | GO:1901029 | 1 |
| nuclear membrane | GO:0031965 | 1 |
| nuclear pore outer ring | GO:0031080 | 1 |

| | | |
|--|------------|---|
| nucleotide binding | GO:0000166 | 1 |
| oxidation-dependent protein catabolic process | GO:0070407 | 1 |
| oxidation-reduction process | GO:0055114 | 1 |
| oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen | GO:0016709 | 1 |
| oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen | GO:0016702 | 1 |
| oxygen carrier activity | GO:0005344 | 1 |
| phosphate ion transport | GO:0006817 | 1 |
| phosphatidylinositol phospholipase C activity | GO:0004435 | 1 |
| phosphogluconate dehydrogenase (decarboxylating) activity | GO:0004616 | 1 |
| phosphoprotein phosphatase activity | GO:0004721 | 1 |
| positive regulation of cell population proliferation | GO:0008284 | 1 |
| positive regulation of translational elongation | GO:0045901 | 1 |
| positive regulation of translational termination | GO:0045905 | 1 |
| posttranscriptional tethering of RNA polymerase II gene DNA at nuclear periphery | GO:0000973 | 1 |
| potassium ion transport | GO:0006813 | 1 |
| protein disulfide isomerase activity | GO:0003756 | 1 |
| protein folding | GO:0006457 | 1 |
| protein import into nucleus | GO:0006606 | 1 |
| protein N-linked glycosylation via asparagine | GO:0018279 | 1 |
| protein quality control for misfolded or incompletely synthesized proteins | GO:0006515 | 1 |
| protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway | GO:0043328 | 1 |
| protein-FAD linkage | GO:0018293 | 1 |
| protoporphyrinogen IX biosynthetic process | GO:0006782 | 1 |
| pyridoxal phosphate binding | GO:0030170 | 1 |
| ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor | GO:0004748 | 1 |
| ribosome binding | GO:0043022 | 1 |
| RISC complex | GO:0016442 | 1 |
| RNA splicing | GO:0008380 | 1 |
| rRNA binding | GO:0019843 | 1 |
| sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphate glyceronetransferase activity | GO:0004801 | 1 |
| sequence-specific DNA binding | GO:0043565 | 1 |
| serine-pyruvate transaminase activity | GO:0004760 | 1 |
| serine-type exopeptidase activity | GO:0070008 | 1 |
| small ribosomal subunit | GO:0015935 | 1 |

| | | |
|---|------------|---|
| Smc5-Smc6 complex | GO:0030915 | 1 |
| structural constituent of nuclear pore | GO:0017056 | 1 |
| superoxide dismutase activity | GO:0004784 | 1 |
| taurine biosynthetic process | GO:0042412 | 1 |
| tissue remodeling | GO:0048771 | 1 |
| transferase activity, transferring acyl groups | GO:0016746 | 1 |
| translational frameshifting | GO:0006452 | 1 |
| transmembrane receptor protein tyrosine kinase activity | GO:0004714 | 1 |
| transmembrane transporter activity | GO:0022857 | 1 |
| ubiquinone biosynthetic process | GO:0006744 | 1 |
| unsaturated fatty acid biosynthetic process | GO:0006636 | 1 |

CHAPTER 3: WHO'S THERE? HARNESSING POPULATION GENOMICS TO IDENTIFY CISCO (*COREGONUS ARTEDI*), KIYI (*C. KIYI*), AND BLOATER (*C. HOYI*) LARVAE FROM THE APOSTLE ISLANDS OF LAKE SUPERIOR.

3.1. Introduction

The Great Lakes were historically home to at least eight coregonine species. Two have been extirpated leaving six extant species (Eshenroder et al. 2016); Figure 5). Each species serves different ecological roles in different parts of the lake. For example, cisco (*Coregonus artedi*) are generally considered to be pelagic planktivores that were found in all of the Great Lakes, yet now their populations have declined in all lakes and have been extirpated from Lake Erie (Table 1; Selgeby 1982, Stockwell et al. 2009, Ebener 2013, Eshenroder et al. 2016). Bloater (*C. hoyi*) are currently found in Lakes Superior and Michigan and are nearshore (<80 m) deepwater planktivores (Table 1, Figure 5; Anderson and Smith 1971, Crowder and Crawford 1984, Gorman and Todd 2007, Clemens and Crawford 2009, Eshenroder et al. 2016). Kiyi (*C. kiyi*) is an offshore (>80m) deepwater planktivore that consumes more *Mysis diluviana* than *C. hoyi* (Schmidt et al. 2011, Sierszen et al. 2014, Eshenroder et al. 2016). *C. kiyi* historically was found in all of the Great Lakes except Lake Erie, but now is found only in Lake Superior (Table 1; Miller et al. 1989, Eshenroder et al. 2016).

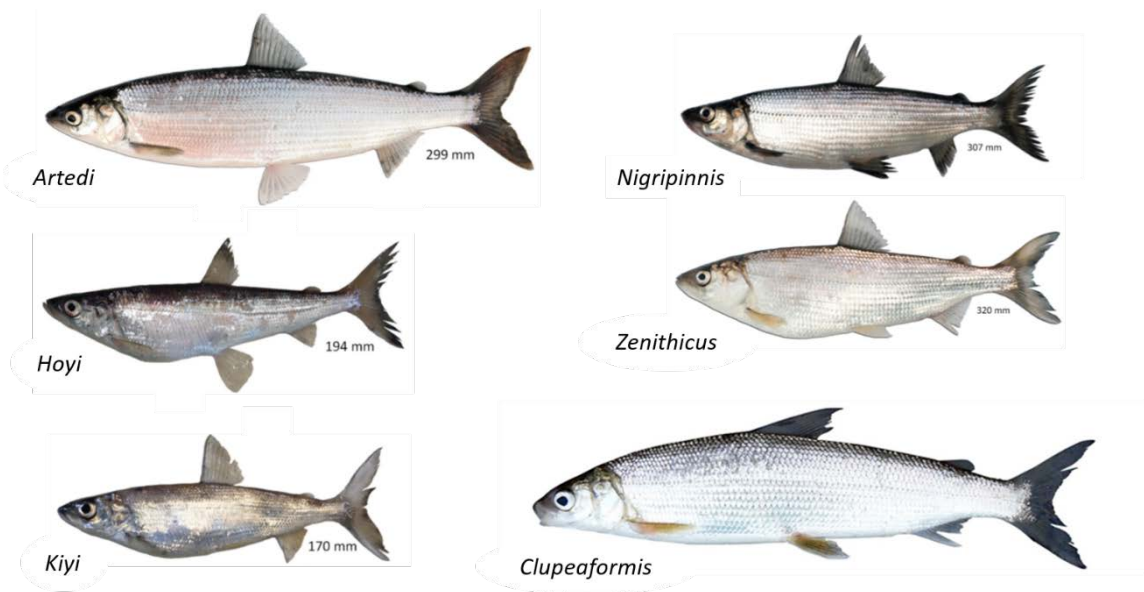


Figure 5: Adults of species in the genus *Coregonus* currently found in the Great Lakes (Eshenroder et al. 2016).

Coregonine species have adapted to different habitats and resources and serve as forage for a variety of predators, and thus fill various key ecological niches in the Great Lakes ecosystem (Eshenroder et al. 2016). Adaptation to various ecological roles and habitats has led to morphometric differences and similarities among and within species, and among and within lakes (Eshenroder et al. 2016). As adults, most coregonines can be visually identified using morphometric attributes, such as body shape and head shape, and meristic measurements including gill raker counts (Koelz 1929, Eshenroder et al. 2016). However, during early life stages, no method is available to differentiate coregonine larvae to the species level. Genetic barcoding of the mitochondrial cytochrome C oxidase I gene can be used to differentiate lake whitefish (*C. clupeaformis*) from other coregonines and ten microsatellite loci have been used to tell cisco from bloater (George et al. 2017), but no method has yet been developed to identify the remaining species from one another.

The inability to identify coregonine species at early life stages limits our ability to study the species' life histories during a crucial part of their life cycle. For example, if coregonine larvae cannot be identified, then we do not know exactly when each species hatches, where the larvae are found, what length ranges to expect or the growth rate. However, past research has drawn conclusions from wild-caught coregonine larvae, which results in misleading demographic information about a "species" when really, they are studying coregonines as a whole. For example, Oyadomari and Auer (2008) drew conclusions about *C. artedi* based on coregonine larvae collected off the Keweenaw peninsula (Michigan, USA), yet because they did not have a way to tell coregonine larvae apart, we cannot be confident that the growth rates reported are true representations of *C. artedi*. Because we cannot rely on demographic data collected from wild caught coregonines, we must rely on demographic data gathered from hatchery-reared larvae. For example, *C. artedi* and *C. hoyi* have successfully been reared in laboratory and hatchery settings and have length and growth rate estimates (McCormick et al. 1971, Rice et al. 1987). However, *C. kiyi* have not successfully been reared in captivity and therefore nothing is currently known about the demographics of early life. The ability to identify wild coregonine species would fill the knowledge gaps for *C. kiyi* and confirm and expand upon the existing demographic knowledge for *C. artedi* and *C. hoyi*.

Initial genetic analyses of coregonines indicated minimal genetic variation exists among species (Turgeon and Bernatchez 2003). However, Turgeon and Bernatchez (2003) used microsatellites which represent a small portion of the genome. Modern genomic sequencing techniques allow us to capture more of the genome, and therefore allow us to

better identify any differences between species. Recently, restriction site-associated DNA sequencing (RADseq) has been used to sequence known coregonine adults throughout the Great Lakes (Ackiss et al. 2019, in prep). The next step is to test the technique to identify unknown, wild-caught larvae by comparing larval sequences to known adult sequences.

In this study, we used RADseq to identify coregonine larvae collected in the Apostle Islands of Lake Superior. Larvae were identified by comparing RADseq data from known adult coregonines collected from the Apostle Islands and other parts of Lake Superior. Several demographic traits of each identified species were then assessed and compared as a preliminary evaluation of possible early life stage differences. Additional genetic analyses were conducted to provide genetic diversity information for each species, including effective population size (N_e), relatedness, observed heterozygosity (H_o) and expected heterozygosity (H_e).

3.2. Methods and Materials

Sample collection

Samples were collected weekly at 10 sites throughout the Apostle Islands, Lake Superior, between May 14 and July 25, 2018 (Figure 6). All samples were collected between 07:00 and 15:00. To keep sampling consistent and efficient, sites were divided into two equal groups which were always sampled together, and each group was sampled on a different day each week. The “Inner Islands” group consisted of Raspberry Point (46 56.136, 90 46.832), Inner Islands (47 00.003, 90 39.277), Outer Island (47 01.792, 90 28.778), Cat Island (46 59.651, 90 34.000), and Stockton Island (46 55.760, 90 39.330).

The “Outer Islands” group included Sand Island (46 58.739, 90 59.255), York Island (46 59.674, 90 52.322), Bear Island (47 01.344, 90 46.826), Devil’s Island (47 04.798, 90 44.084), and Rocky Island (47 04.553, 90 38.259). Site order varied according to weather constraints each day, such that each site was sampled at a variety of times of the day over the sampling period.

A bongo net equipped with 0.5-m diameter, 500- μ m mesh conical nets was used to collect larval fish. Each tow was conducted on a straight transect for 10 min with the net approximately 15 m behind the vessel, moving downwind. Tow speed was approximately 3.5 km/h to keep the top of the net frame visible at the surface. At the completion of each tow, larvae were immediately counted and preserved in 70% ethanol.

Demographic measurements

Larvae were imaged and assessed on a dissecting microscope equipped with a camera (Olympus SZX7, Waltham, MA) (Figure 7). Each larva was identified as *Coregonus* or non-*Coregonus* according to Auer (1982). *Coregonus* larvae were then visually sorted as whitefish or non-whitefish coregonines based on visual keys (Auer 1982) and downstream analyses were performed on the non-whitefish coregonines. Yolk sac condition (YSC) was assessed as either “yolk sac and oil globule present”, “oil globule only”, or “absorbed” (Taylor Stewart; pers. comm.).

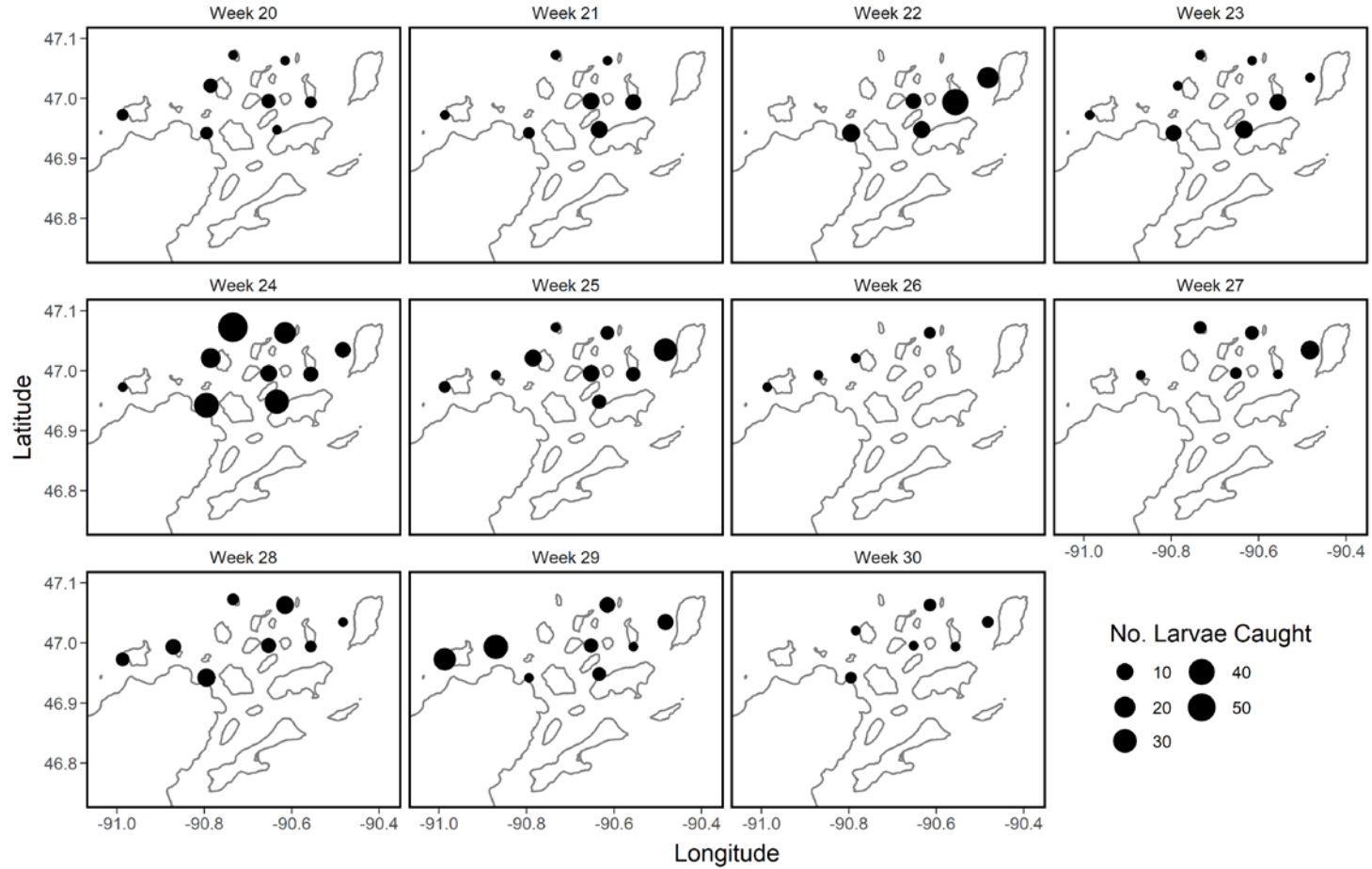


Figure 6: Map of the larval catch totals per week at each of the 10 sites sampled between May 14 and July 25, 2018 in the Apostle Islands, Lake Superior, WI. Size of dots indicate density of coregonine larvae collected.

To measure the length of each larvae to the nearest 0.01 mm on the microscope images, Leica The Application Suite was used (Leica Microsystems Inc, Buffalo Grve, IL) (Figure 7). The stomachs were then removed from each individual for a diet study and the rest of the body was used for genetic identification. During imaging and stomach removal the instruments and bench space were sterilized between each larva to prevent cross contamination or degradation of DNA needed for downstream genomics. Once the stomachs were removed, the remainder of the individual was preserved in 100% ethanol until DNA was extracted.

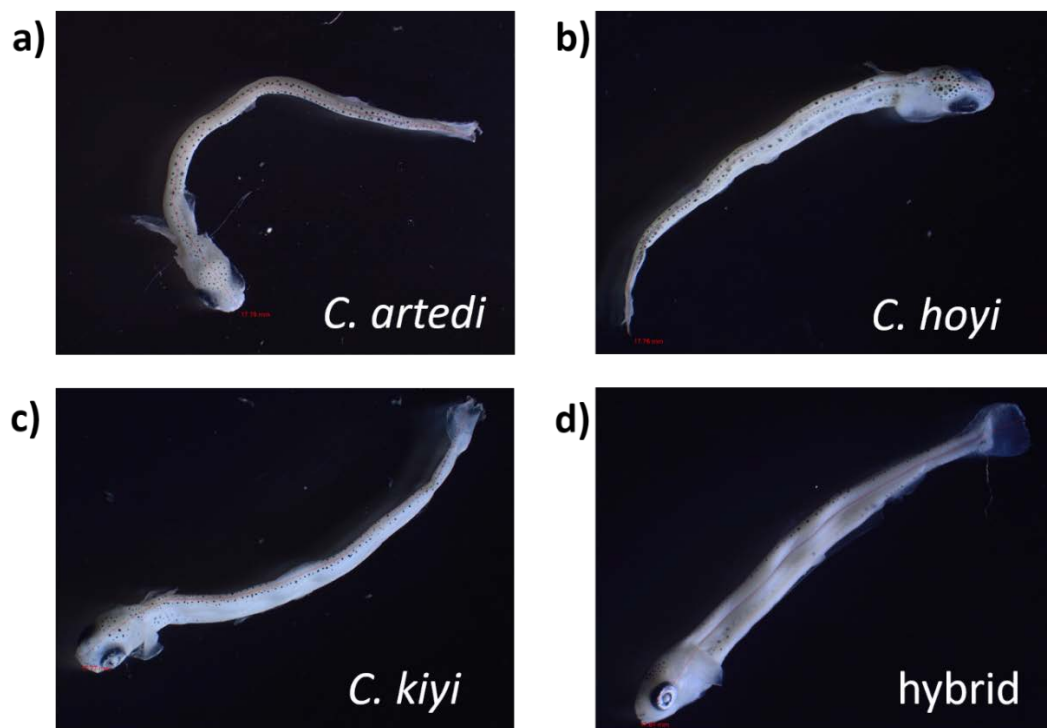


Figure 7: Representative photos of each coregonine species found in the Apostle Islands, Lake Superior during the spring/summer of 2018. Individuals pictured were collected during week 25 or week 29 and are within 0.02 mm of each other.

Once sequencing analyses were completed, we grouped and assessed the length data, YSC, location and date collected data for each species using R (v.3.6.0, Team 2018). Growth rates were estimated for each species using the slope of the linear regression of length as a function of time, using individual fish as the observation unit.

DNA extraction and barcoding

Larvae were subset first by time point and then by length to capture the main cohort present throughout the season. Two early season weeks, two mid-season weeks and four late season weeks were chosen to ensure we had relatively equal sample sizes throughout the beginning middle and end of the larval season. Within each week, length ranges from which to sample were set to limit outliers and follow the primary cohort of coregonine larvae throughout the larval stage.

DNA was extracted from 221 larvae using Qiagen DNeasy Blood and Tissue kits (Germantown, MD) after removal of stomachs. The 221 larvae were barcoded to identify any whitefish larvae using a genetic barcoding method (George et al. 2017). Briefly, PCR was used to amplify the mitochondrial cytochrome C oxidase I gene using the following PCR conditions: an initialization and denaturing step at 95°C for 2 min followed by 36 cycles of 95°C for 30 sec, 52°C for 40 sec, 68°C for 1 min, and finished with a 10 min final elongation at 95°C. Twenty microliters of the PCR product was combined with 1.0 µl of the restriction enzyme Eco0109I and 4 µl of the Eco0109I buffer (New England BioLabs, Ipswich, MA). The restriction enzyme solution was run for 2 hrs at 37°C. The digested

PCR product was then run on a 2% gel at 100V for at least 30 minutes so the species could be identified by the number and size of the DNA band(s) present.

Library preparation, sequencing and genotyping

RAD libraries were prepared using the BestRAD protocol (Ali et al. 2016) and followed the methods of Ackiss et al. (2019; in prep). Whole genomic DNA was normalized, digested with *SbfI*, and barcoded by individual. Libraries were then sonicated, and non-target sequences were removed before they were pooled into master libraries of 96 samples. Master library barcodes were ligated followed by size selection and enrichment. Library products were quantified and sent to Novogene (Sacramento, CA) for sequencing on two lanes of an Illumina® HiSeq 4000 (San Diego, CA).

Raw sequences were quality controlled with a Stacks (v.2.3d; Catchen et al. 2013, Catchen et al. 2011) pipeline at the same time as 96 adult coregonine sequences (Ackiss et al. 2019; in prep) to ensure the best comparison for downstream species identification. Briefly, reads were demultiplexed using barcodes and filtered by presence of cut site, quality controlled, and trimmed using *process_radtags* (parameter flags: `-e SbfI -c -q -r -t 140 --filter_illumina --bestrad`). Matching segments were stacked using *ustacks* (parameter flags: `--disable-gapped -m 3 -M 5 -H --max_locus_stacks 4 --model_type bounded --bound_high 0.05`) and a catalog of consensus loci was created using 126 individuals with the highest number of reads: 51 individuals from the *C. artedi* linkage map (Blumstein et al. 2019; in prep); 5 larvae and 5 adults of validated *C. hoyi*; 4 adults of validated *C. kiyi* from the Apostle Islands; and 5 adults each of validated *C. nigripinnis* and *C. zenithicus*

from various locations across Lake Superior (parameter flags: -n 3 -p 6 -disable_gapped). Individual reads were matched to the catalog and SNPs were called using *gstacks*. *Populations* was used to output the data as genepop and variant call format vcf files.

After Stacks, vcftools (v.0.1.15; Danecek et al. 2011) was used to filter out any loci that were 1) missing more than 70% of individuals, 2) individuals that were missing 50% of the loci, and 3) loci with a minor allele count (mac) of less than 3. To control for whole genome duplications within coregonines, HDplot (McKinney et al. 2017) identified any paralogous loci using a heterozygosity cutoff of > 0.55 and a read ratio deviation of > 5 or < -5 . Once paralogs were identified, sequences were moved back into vcftools to remove the paralogs. A perl script (Ackiss et al. 2019; in prep) was used to pull SNPs with highest minor allele frequency (maf) from each tag and only the highest maf was used for downstream analysis. PGDSpider (v2.1.1.5: Lischer and Laurent et al. 2011) was used to convert the vcf files to genepop files.

Population genomic analyses

A series of R (v 3.6.0; Team 2018) programs were used to analyze the population genomics data. First, AssignPop (v.1.1.4; Chen et al. 2018) was used to identify the species of each larvae by comparing the larval sequences to known adult coregonine sequences. A baseline was created with the adult species:

```
bse_data=read.Genepop("G80low50mac3.HDP.AB.ncode2.nolarvaenoNIGnoZE
N-Hoy3rem.FINAL.gen.txt",pop.names=c("Art","Hybrid","Hoyi","Kiyi"),
haploid = FALSE)
```

and the baseline was validated against itself using assign.kfold:

```
bse_data, k.fold=10,train.loci=1, multiprocessing=TRUE, dir="results1/"
```

Larvae were assigned to species using assign.X:

```
x1=bse_data, x2=mix_data, multiprocessing=TRUE, pac.method="mixed",  
pca.PCs = "kaiser-guttman", pca.loadings = T, mplot=TRUE, dir="results3/"
```

Assignments were considered accurate if the individual was $\geq 70\%$ similar to the known adult species. If the individual was $< 70\%$ similar to the known adults they were considered “hybrids”. Adegenet (v.2.1.1; Jombart, 2008) was used to create a PCA to visualize the larval genomic data.

Related (v.1.0; Pew et al. 2015) was used to calculate how related the larvae were to each other. *Coancestry* was run with error.rates = 0.001, ci95.num.bootstrap = 500, and ritland = 2. Larvae were considered half-siblings or greater level if the relatedness > 0.125 and unrelated if relatedness < 0.125 .

Two additional programs were used to calculate N_e , H_o , and H_e , to determine the genetic health and diversity of the coregonine populations. To calculate N_e , NeEstimator (v.2.1; Do et al. 2014) was used. “Linkage Disequilibrium – Random Mating” was selected, critical values were set to 0.05, and “LD locus paring” was run with a file of locus chromosome positions created from *C. artedi* data (Ackiss et al. 2019; in prep). Heterozygosity values (H_o and H_e) were calculated in GenoDive (v.2.0b23; Meirmans and Van Tienderen 2004) by running the genetic diversity analysis.

3.3. Results

DNA extraction, barcoding, library prep, sequencing and genotyping

Of the 221 barcoded *Coregonus* larvae, 5 were whitefish, 212 were not whitefish, and 4 had a novel banding pattern. From the 212 non-whitefish coregonines, 198 individuals were chosen for RAD sequencing based on the time and length sub-setting described above.

After Stacks, HDplot and vcftools were used, only 1 of the 198 larvae were removed due to the individual having a higher amount of missing data identified by filters used in Stacks, leaving 197 high-quality larval sequences for downstream analyses.

Population genomic analyses

Of the 197 individuals that passed the post-sequencing filtration steps, 104 were identified as *C. artedi*, 59 were identified as *C. kiyi*, 27 were identified as *C. hoyi*, and 7 were identified as hybrids (Figure 8, Figure 9). The seven hybrids appear to be made up of two F1s between *C. hoyi* and *C. kiyi*, one F1 between *C. hoyi* and *C. artedi*, and four backcrosses between *C. hoyi* and *C. artedi*, for which two appeared closer to *C. hoyi* and two closer to *C. artedi* (Figure 8, Figure 9).

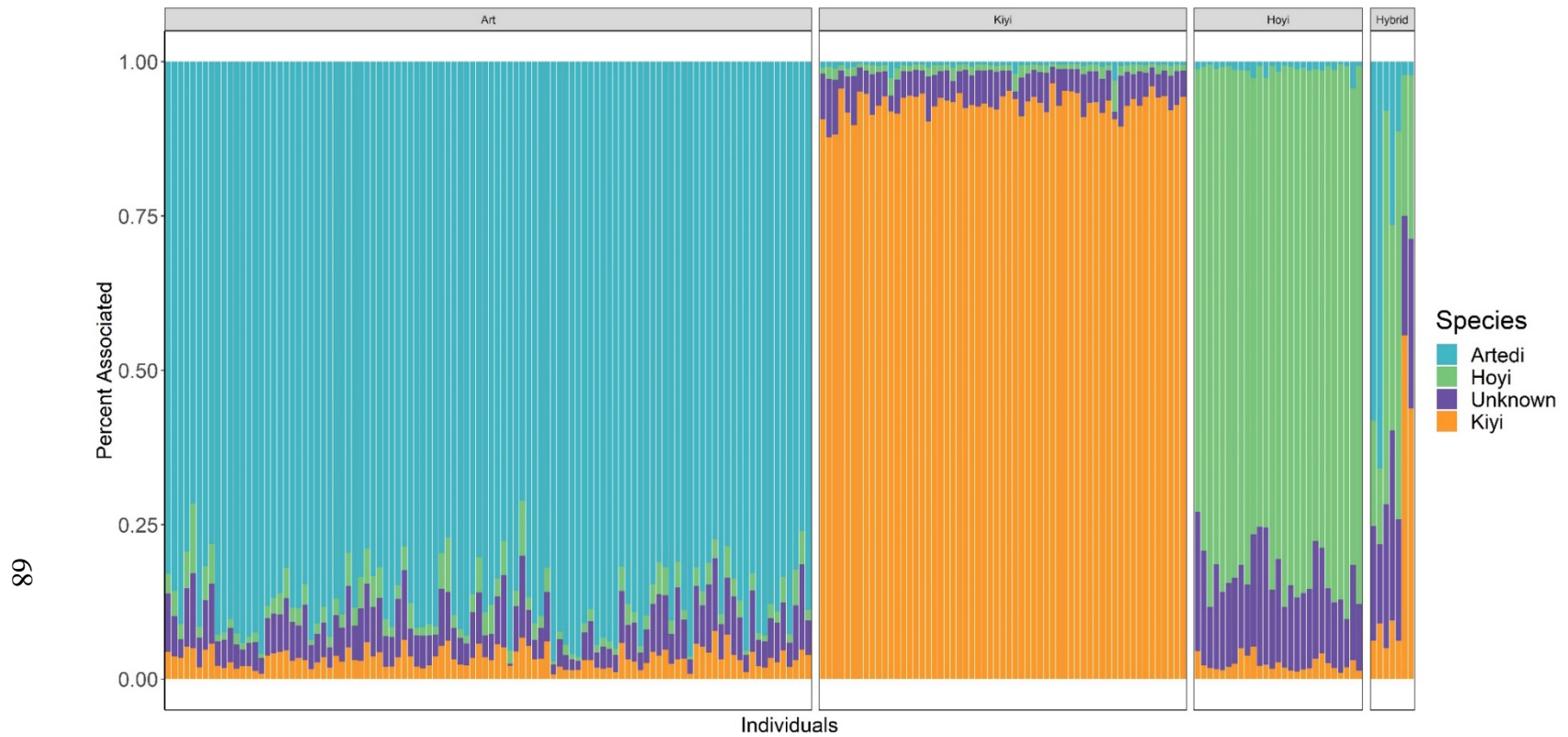


Figure 8: Structure style stacked bar graph demonstrating AssignPop species assignments for the 197 coregonine larvae that were sequenced. Each bar is a separate individual and the y- axis is the percent that the larval genome matched to known adult samples for *C. artedi* (blue), *C. hoyi* (green), *C. kiyi* (orange), and an unknown *Coregonus* adult (purple) found in Lake Superior.

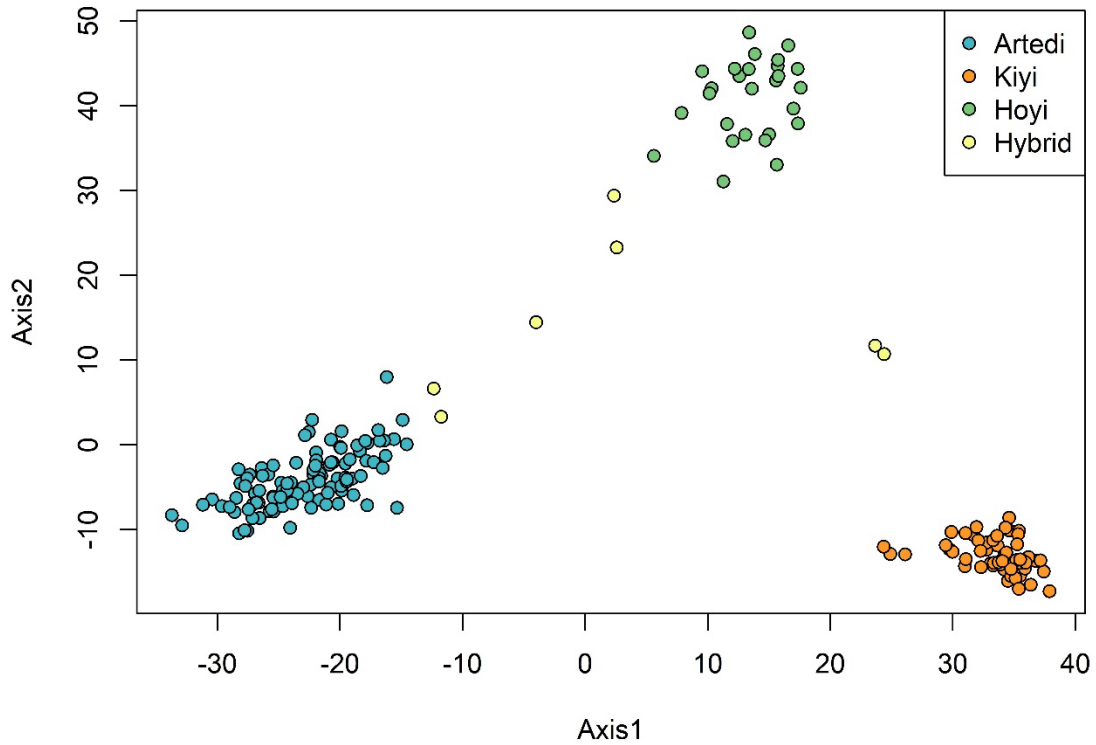


Figure 9: PCA showing the separation among species identified in 197 larval coregonines that were sequenced from the Apostle Islands, Lake Superior, during spring/summer 2018. Each dot is a separate individual.

The relatedness analysis revealed that 175 of the 197 individuals were distantly related to each other (cousins or more distant). The highest level of relatedness detected was half sibling (> 0.125). The 22 related pairs were all *C. hoyi* and their relatedness values ranged from 0.126 to 0.148.

The estimated N_e of *C. hoyi* and *C. artedi* was 4,006 (95% CI: 3,236 – 5,255) and 4,350 (95% CI: 4,071 – 4,670), respectively, but *C. kiyi* had the highest N_e of 13,459 (95% CI: 9,771 – 21,611) (Table 7). Heterozygosity values (H_o , H_e) and N_e indicate healthy

genetic variability and mixing and were all comparable to those found in the adult forms for each species (Table 7) (Ackiss et al. 2019; in prep).

Table 7: Species-specific data for the 197 larval samples that were sequenced from the spring/summer 2018 sampling in the Apostle Islands, Lake Superior. The number of larvae that assigned to each species along with the length range, week range N_e , H_o and H_e for each species is reported.

| Species | N | Length range (mm) | Date Range (Haas et al.) (Haas et al.) (Haas et al.) | N_e (95% CI) | H_o | H_e |
|------------------|-----|-------------------|--|-------------------------|-------|-------|
| <i>C. artedi</i> | 104 | 9.16-19.58 | 21-29 | 4,350 (4,071 – 4,670) | 0.208 | 0.215 |
| <i>C. kiyi</i> | 59 | 10.60-21.67 | 22-29 | 13,459 (9,771 – 21,611) | 0.203 | 0.211 |
| <i>C. hoyi</i> | 27 | 8.87-21.3 | 27-30 | 4,006 (3,236 – 5,255) | 0.233 | 0.245 |
| Hybrid | 7 | 11.05-19.58 | 22-29 | N/A | 0.224 | 0.236 |

Demographic measurements

Over the 10-week sampling period, 605 coregonines were collected (Figure 6). Each week's collection totals varied with the largest number of larvae collected during week 24 ($N = 189$) and the smallest number of larvae collected during week 30 ($N = 7$) (Figure 6). Overall, lengths ranged from 6.25 mm to 26.20 mm (Figure 12, Figure 13).

Coregonus artedi was the earliest species to be captured (week 21) followed by *C. kiyi* (week 22) and then *C. hoyi* (week 27) (Figure 10, Table 7). The first two hybrids appeared during week 22 and the other hybrids were found in weeks 25, 28, and 29 (Figure 10, Table 7). The length ranges and frequencies for each species match the

expected temporal pattern of emergence and growth; within each species the smallest larvae were found first, and individuals in each cohort were larger over time (Figure 11, Figure 12). *Coregonus hoyi* had the widest range of lengths (12.43 mm diff between largest and smallest larvae), while the range of lengths for *C. artedi* and *C. kiyi* were 10.12 mm and 11.07 mm respectfully (Figure 13, Figure 14).

The smallest *C. artedi* sequenced was 9.16 mm and the estimated growth rate was 0.147 mm/day (Table 7). The smallest *C. hoyi* sequenced was 8.87 mm and the estimated growth rate was calculated as 0.105 mm/day (Table 7). The smallest *C. kiyi* was 10.60 mm and the estimated growth rate was 0.1344 mm/day.

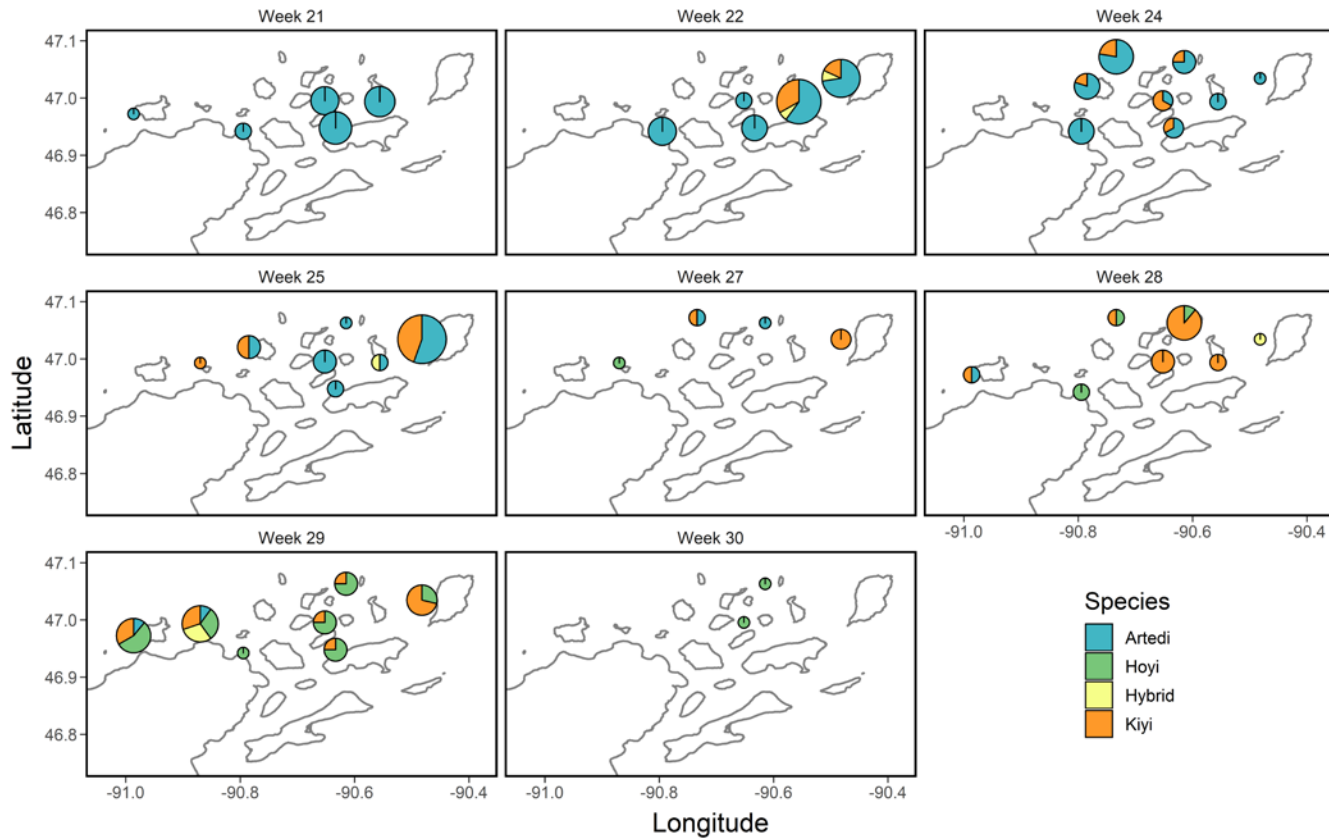


Figure 10: Map of the weeks and locations where the 197 sequenced larvae were collected in the Apostle Islands, Lake Superior during the spring/summer of 2018. Each panel is a different week and the colors indicate genetically identified species. Circle size indicates abundance and proportion of the circle indicates proportion of the catch that identified as each species.

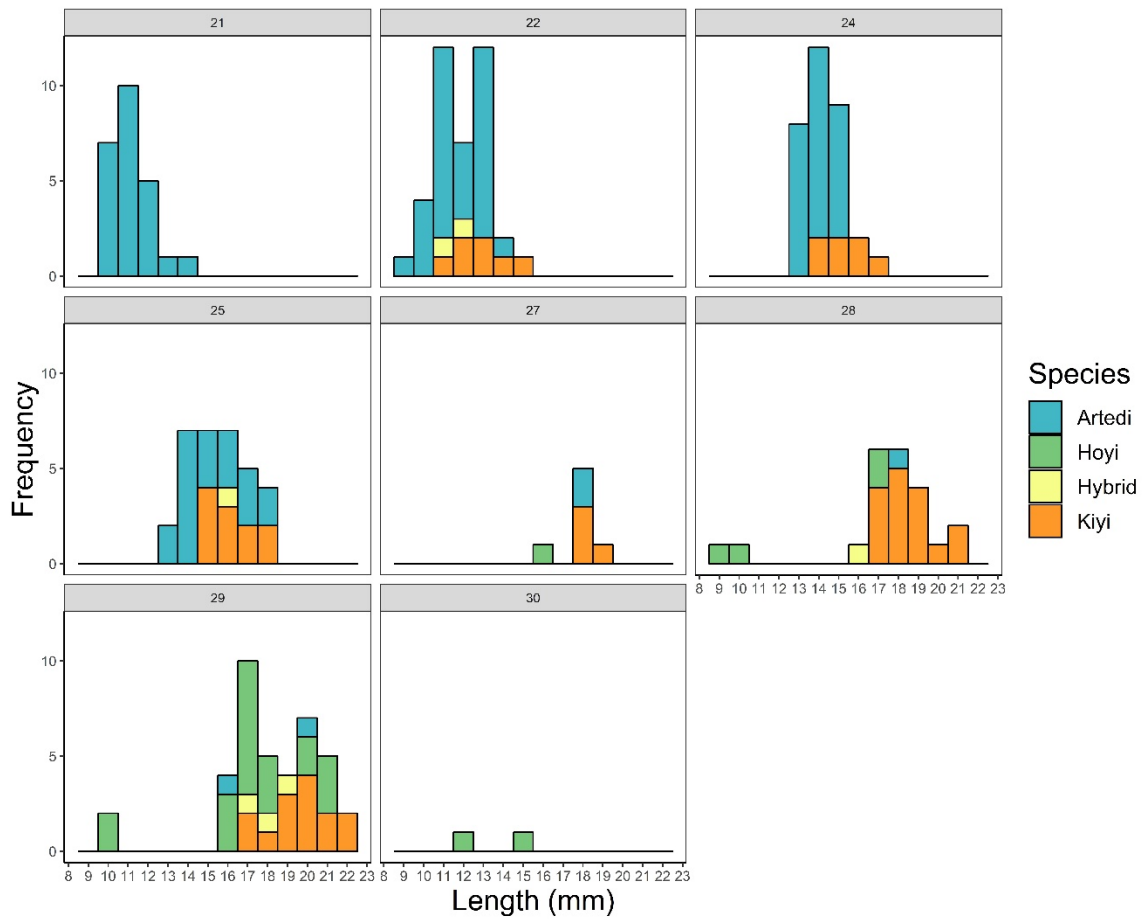


Figure 11: Length-frequency histograms of the 197 sequenced larvae captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Each panel is a different week and the colors indicate species.

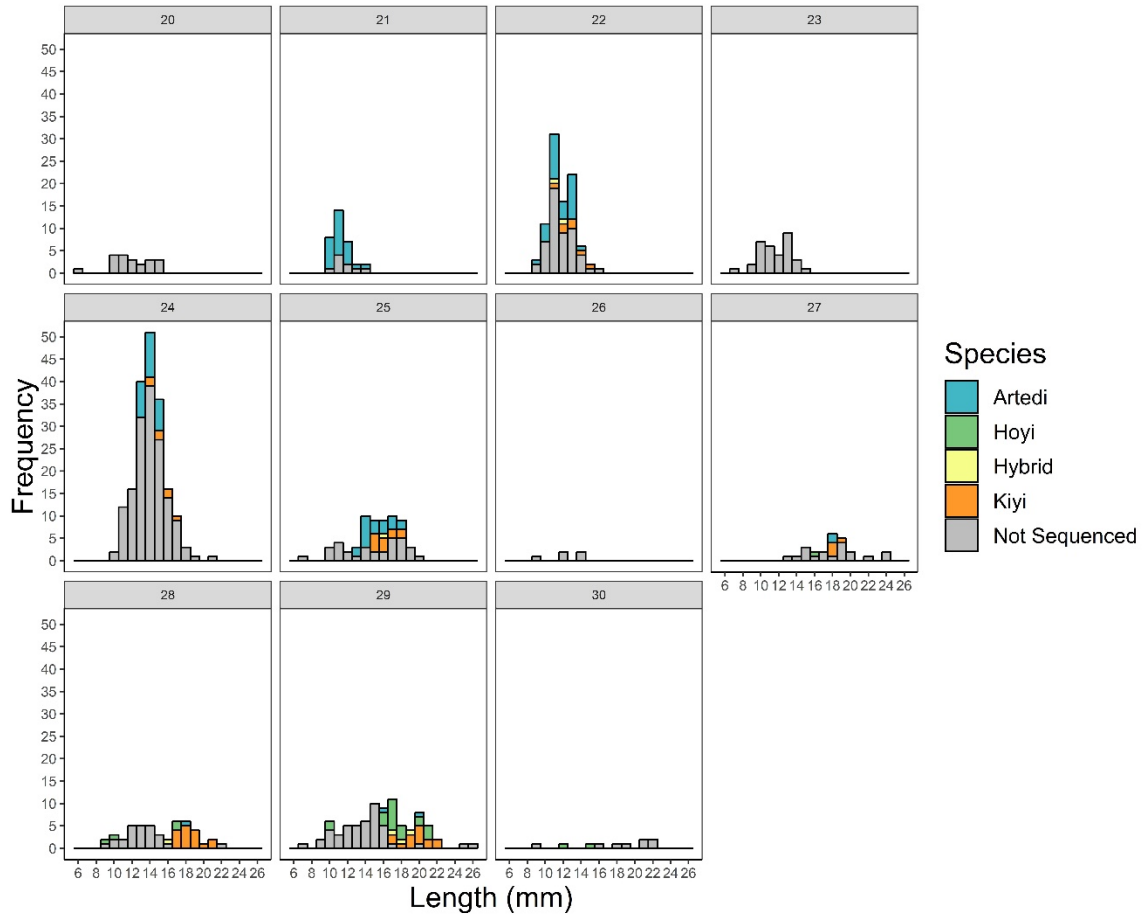


Figure 12: Length-frequency histograms of all 605 larvae (sequenced and not sequenced) captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Each panel is a different week and the colors indicate species.

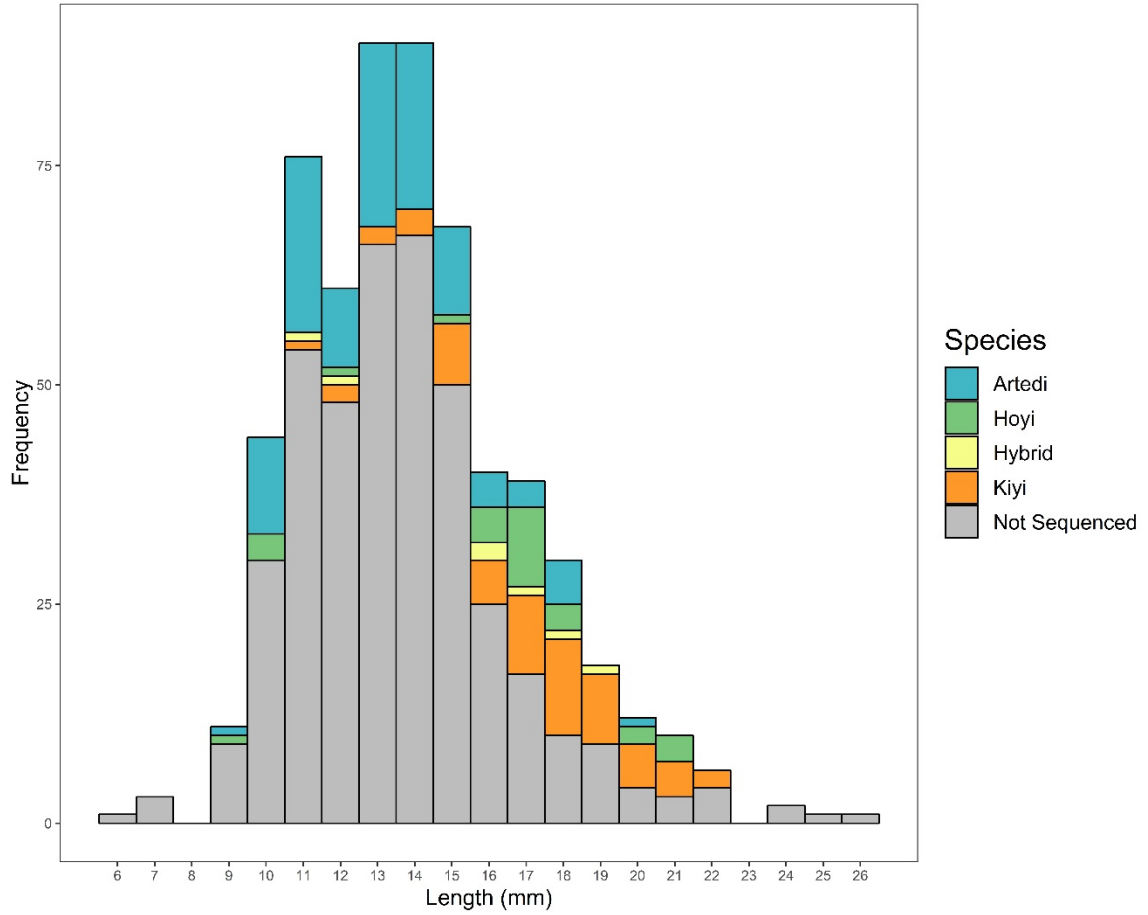


Figure 13: Length frequency histogram of all 605 larvae (sequenced and not sequenced) captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Colors indicate species.

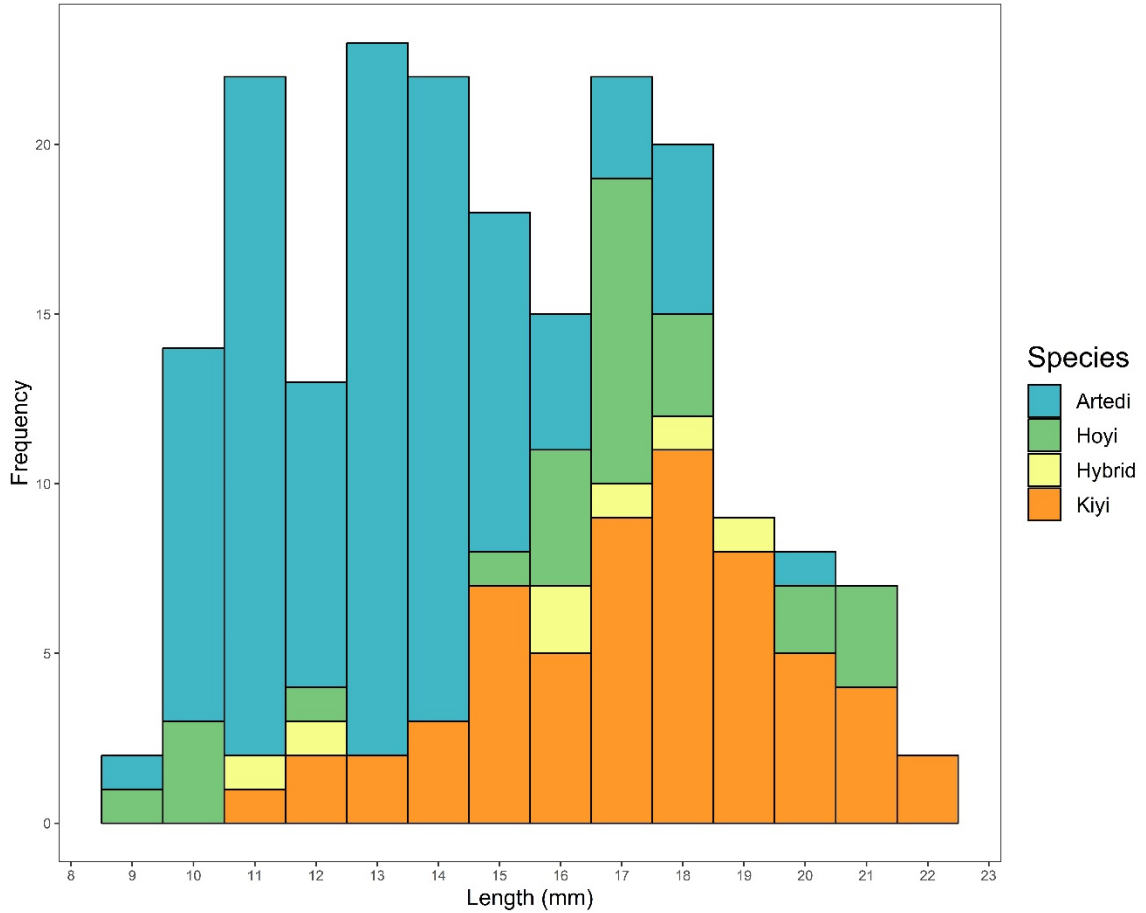


Figure 14: Length-frequency histogram of the 197 sequenced larvae captured in the Apostle Islands, Lake Superior during the spring/summer of 2018. Colors indicate species.

Coregonus artedi larvae with a yolk sac and oil globule were found in weeks 21-22 and ranged in length from 9.16 mm – 12.94 mm (Figure 15, Figure 16). *Coregonus artedi* with only the oil globule present were found in weeks 21-24 and ranged in size from 10.43 mm – 13.55 mm and *C. artedi* with the yolk sac fully absorbed were found in weeks 21-29 and ranged in size from 11.91mm – 19.58 mm (Figure 15, Figure 16). No *C. hoyi* with both yolk sac and oil globule were found. However, *C. hoyi* with just the oil globule were found in weeks 28 and 29 and ranged in length from 8.87mm – 10.33mm (Figure 15,

Figure 16). *Coregonus hoyi* with the yolk sac fully absorbed were found in weeks 27-30 and their lengths ranged from 11.70mm – 21.30mm (Figure 15, Figure 16). *Coregonus kiyi* were only found with either an oil globule (week 22; length range of 10.60mm – 14.60mm) or the yolk sac fully absorbed (weeks 22-29; length range of 12.30mm – 21.67mm) (Figure 15, Figure 16). Hybrids either had their yolk sac and oil globule (week 22; lengths 11.05mm and 12.06mm) or the yolk sac fully absorbed (weeks 25-29; length range of 15.87mm – 18.80mm) (Figure 15, Figure 16).

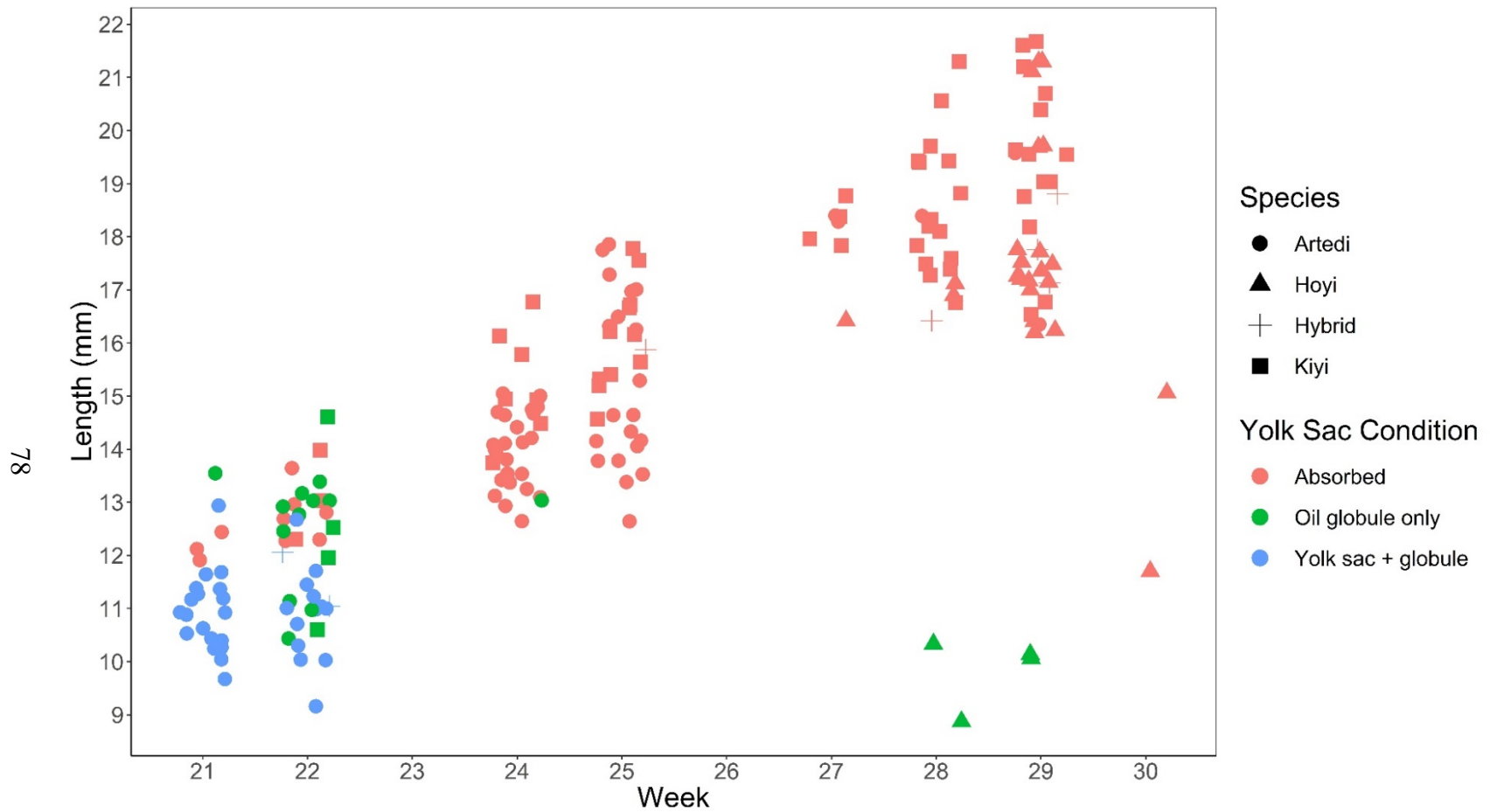


Figure 15: Yolk sac condition (YSC) of 197 sequenced coregonine larvae collected in the Apostle Islands of Lake Superior during the spring/summer of 2018. Symbol represents species and color represents yolk sac condition.

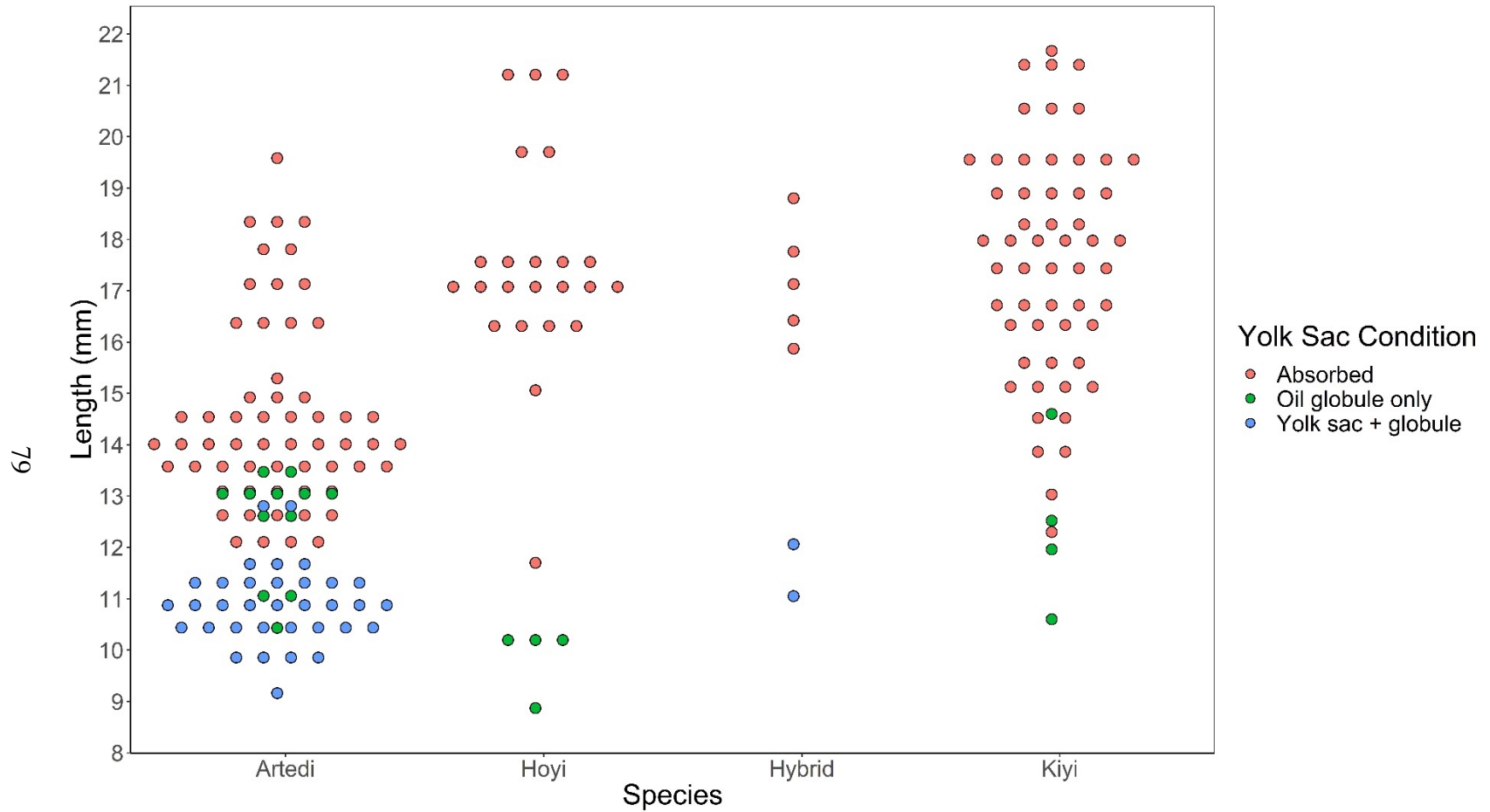


Figure 16: Length distribution of yolk sac condition as function of species identity for 197 sequenced coreognine larvae collected in the Apostle Islands of Lake Superior during the spring/summer of 2018. Color represents yolk sac condition.

3.4. Discussion

Our results provide the ability to identify *C. artedi*, *C. hoyi* and *C. kiyi* and provides many opportunities to investigate the early life demographics of these ecologically and economically important species. Our initial findings provide key insights into when each species was found, where they were found in the Apostle Islands of Lake Superior, and their length ranges, YSC, and growth rate. Additionally, we were able to use the RADseq data to estimate N_e , H_o , H_e , and relatedness which was used to determine that all three species have a healthy number of parents contributing to the larval population in the Apostle Islands and a healthy amount of genetic variability is present, which will help *C. artedi*, *C. hoyi* and *C. kiyi* adapt to changes in the future.

We were able to access and compare more of the genome and identify differences among species by using RADseq versus traditional microsatellites. *Coregonus artedi*, *C. hoyi* and *C. kiyi* were previously thought to be different forms of the same taxa (*C. artedi*). However, based on the differential clustering of each species and the ease by which we were able to identify species by comparing to known adult coregonines, we have shown that these three coregonines are more genetically distinct than previously thought.

The ability to identify *C. artedi*, *C. hoyi*, and *C. kiyi* allows us to study wild-caught larvae and assess their natural demographic characteristics at the species level. Not only can we compare differences in location, timing, lengths, YSC, and growth among species but also across conspecifics. Our study has provided an initial demographic baseline for *C. artedi*, *C. hoyi* and *C. kiyi* in the Apostle Islands of Lake Superior, but future use of our

identification technique can allow coregonines across the Great Lakes to be studied and compared.

One of the more interesting demographic findings was the temporal difference in catch among *C. artedi*, *C. hoyi* and *C. kiyi*. *Coregonus artedi* were caught earliest in the season and were mostly absent from our collections by the time *C. hoyi* appeared. *Coregonus kiyi* overlapped with *C. artedi* and *C. hoyi*. The timing of the catches indicates that each species likely hatches at different times. Because we were able to find *C. artedi* and *C. hoyi* that match the length-at-hatch sizes seen in hatcheries we are fairly confident that we captured each species just after hatch (McCormick et al. 1971, Rice et al. 1987). However, the absence of *C. hoyi* and *C. kiyi* with yolk sacs in our collections could indicate 1) that we missed the earliest post-hatch larvae due to *C. hoyi* and *C. kiyi* remaining in the hypolimnion near their hatch sites before reaching the surface where we sampled, 2) *C. hoyi* and *C. kiyi* hatch further offshore and were brought into the Apostle Islands after hatch via currents, 3) *C. hoyi* and *C. kiyi* hatch with minimal yolk sacs, or 4) we did not sequence enough small *C. hoyi* and *C. kiyi* (due to sub-sampling by length) to capture *C. hoyi* or *C. kiyi* with yolk sacs. While the lack of yolk sacs could be due to our sampling design, *C. hoyi* larvae remain in the hypolimnion for a short period before reaching the surface (Rice et al. 1987) and coregonine larvae can be advected by currents to new locations (Oyadomari and Auer 2008) which supports the hypothesis *C. hoyi* and *C. kiyi* may have hatched further offshore and were initially missed by sampling nearshore in surface waters.

Spatially, we found all three species widely distributed throughout the Apostle Islands. The overlapping spatial distribution could indicate that *C. artedi*, *C. hoyi* and *C.*

kiyi are spawning in similar locations; however, the mixing could be a result of currents carrying the larvae throughout the Apostle Islands (Oyadomari and Auer 2008).

Coregonus artedi and *C. kiyi* have similar growth rates but the growth rate of *C. hoyi*'s was lower. The lower growth rate we estimated could be a result of the low sample size and limited temporal span that was used for our growth rate calculations. Better growth rate estimates could be achieved through increasing sample size and capturing the entire *C. hoyi* larval stage.

Next steps include streamlining and standardizing the genetic identification process into a widely useable and cost-effective tool. A GTseq panel of ~200 SNPs is currently being developed at the University of Wisconsin Stevens Point (W. Larson, pers. comm.). Future studies could assess larval morphometric and meristic differences between species, but now that we can genetically identify larvae to species, the need for morphometric and meristic analyses *for identification purposes* is moot. Genetics identification is more accurate than visual identification and our results support the movement towards genetically identify the larvae instead of attempting visual identification (George et al. 2017). Other future directions include 1) identifying and assessing larval populations throughout the Great Lakes, 2) increasing sample sizes and sampling durations to improve growth rate, yolk sac and length range estimate, and 3) comparing larval CPUE at the species level to age-1 year class strength (YCS) which would further help us identify when YCS is set.

The ability to assess larvae caught in the field allows us to begin to tie information back to stock recruitment dynamics, which is key to understanding what leads to success

and failure of each species. Understanding key life stages for each species will help inform conservation and restoration efforts.

BIBLIOGRAPHY

- Ali, O. A., S. M. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller. 2016. RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics* **202**:389-400.
- Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**:697.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of molecular biology* **215**:403-410.
- Anderson, E. D., and L. L. Smith. 1971. A synoptic study of food habits of 30 fish species from western Lake Superior.
- Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**:81.
- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.
- Assel, R. A., and D. M. Robertson. 1995. Changes in winter air temperatures near Lake Michigan, 1851-1993, as determined from regional lake-ice records. *Limnology and Oceanography* **40**:165-176.
- Auer, N. A. 1982. Identification of larval fishes of the Great Lakes basin with emphasis on the Lake Michigan drainage.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114-2120.
- Bolsenga, S., and H. Vanderploeg. 1992. Estimating photosynthetically available radiation into open and ice-covered freshwater lakes from surface characteristics; a high transmittance case study. *Hydrobiologia* **243**:95-104.
- Bronte, C. R., D. B. Bunnell, S. R. David, R. Gordon, D. Gorsky, M. J. Millard, J. Read, R. A. Stein, and L. Vaccaro. 2017. Report from the Workshop on Coregonine Restoration Science. 2331-1258, US Geological Survey.
- Brown, R. W., W. W. Taylor, and R. A. Assel. 1993. Factors affecting the recruitment of lake whitefish in two areas of northern Lake Michigan. *Journal of Great Lakes Research* **19**:418-428.
- Cahn, A. R. 1927. An ecological study of southern Wisconsin fishes; the brook silversides (*Labidesthes sicculus*) and the cisco (*Leucichthys artedi*) in their relations to the region, 11. Illinois biological monographs; v. 11, no. 1.
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**:3124-3140.
- Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics* **1**:171-182.
- Chen, K. Y., E. A. Marschall, M. G. Sovic, A. C. Fries, H. L. Gibbs, and S. A. Ludsin. 2018. assignPOP: An r package for population assignment using genetic, non-

- genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution* **9**:439-446.
- Clemens, B. J., and S. S. Crawford. 2009. The ecology of body size and depth use by bloater (*Coregonus hoyi* Gill) in the Laurentian Great Lakes: patterns and hypotheses. *Reviews in Fisheries Science* **17**:174-186.
- Colby, P. J., and L. Brooke. 1973. Effects of temperature on embryonic development of lake herring (*Coregonus artedii*). *Journal of the Fisheries Board of Canada* **30**:799-810.
- Collingsworth, P. D., D. B. Bunnell, M. W. Murray, Y.-C. Kao, Z. S. Feiner, R. M. Claramunt, B. M. Lofgren, T. O. Höök, and S. A. Ludsin. 2017. Climate change as a long-term stressor for the fisheries of the Laurentian Great Lakes of North America. *Reviews in Fish Biology and Fisheries* **27**:363-391.
- Connon, R. E., K. M. Jeffries, L. M. Komoroske, A. E. Todgham, and N. A. Fangué. 2018. The utility of transcriptomics in fish conservation. *Journal of Experimental Biology* **221**:jeb148833.
- Consortium, U. 2018. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**:D506-D515.
- Crowder, L. B., and H. L. Crawford. 1984. Ecological shifts in resource use by bloaters in Lake Michigan. *Transactions of the American Fisheries Society* **113**:694-700.
- Cunjak, R., T. Prowse, and D. Parrish. 1998. Atlantic salmon (*Salmo salar*) in winter: "the season of parr discontent"? *Canadian Journal of Fisheries and Aquatic Sciences* **55**:161-180.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, and S. T. Sherry. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156-2158.
- Davey, J. W., and M. L. Blaxter. 2010. RADSeq: next-generation population genetics. *Briefings in functional genomics* **9**:416-423.
- Do, C., R. S. Waples, D. Peel, G. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources* **14**:209-214.
- Drolet, R., L. Fortier, D. Ponton, and M. Gilbert. 1991. Production of fish larvae and their prey in subarctic southeastern Hudson Bay. *Marine ecology progress series*. *Oldendorf* **77**:105-118.
- Ebener, M. P. 2013. Status of whitefish and ciscoes. Great Lakes Fishery Commission. *Special Publication* **13**:29.
- Eshenroder, R. L., P. Vecsei, O. T. Gorman, D. Yule, T. C. Pratt, N. E. Mandrak, D. B. Bunnell, and A. M. Muir. 2016. Ciscoes (*Coregonus*, subgenus *Leucichthys*) of the Laurentian Great Lakes and Lake Nipigon. Great Lakes Fishery Commission.
- George, E., D. Crabtree, M. Hare, J. Lepak, and L. Rudstam. 2018. Identifying Research Priorities for Cisco in Lake Ontario: A Workshop Summary Report.
- George, E. M. 2016. Cisco Spawning in Chaumont Bay, Lake Ontario. Cornell University.

- George, E. M., M. P. Hare, D. L. Crabtree, B. F. Lantry, and L. G. Rudstam. 2017. Comparison of genetic and visual identification of cisco and lake whitefish larvae from Chaumont Bay, Lake Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*.
- Gorman, O., and T. Todd. 2007. History of the shortjaw cisco (*Coregonus zenithicus*) in Lake Superior, 1895-2003. US Geological Survey.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, and I. Amit. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech.* **29**.
- Greenbank, J. 1945. Limnological Conditions in Ice-Covered Lakes, Especially as Related to Winter-Kill of Fish. *Ecological Monographs* **15**:343-392.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature protocols* **8**:10.1038/nprot.2013.1084.
- Hampton, S. E., A. W. Galloway, S. M. Powers, T. Ozersky, K. H. Woo, R. D. Batt, S. G. Labou, C. M. O'Reilly, S. Sharma, and N. R. Lottig. 2016. Ecology under lake ice. *Ecology Letters* **20**:98-111.
- Hasin, Y., M. Seldin, and A. Lusis. 2017. Multi-omics approaches to disease. *Genome Biology* **18**:83.
- Hjort, J. 1914. Fluctuations in the great fisheries of northern Europe viewed in the light of biological research. ICES.
- Hothorn, T., F. Bretz, P. Westfall, R. M. Heiberger, A. Schuetzenmeister, S. Scheibe, and M. T. Hothorn. 2016. Package 'multcomp'. Simultaneous inference in general parametric models. Project for Statistical Computing, Vienna, Austria.
- Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G.-J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assuncao, Y. Zhou, Y. Gu, J. Yen, J.-H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliott, G. Threadgold, G. Harden, D. Ware, B. Mortimer, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S.

- Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C. M. Dooley, Z. Ersan-Urun, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberlander, S. Rudolph-Geiger, M. Teucke, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, N. P. Carter, J. Harrow, Z. Ning, J. Herrero, S. M. J. Searle, A. Enright, R. Geisler, R. H. A. Plasterk, C. Lee, M. Westerfield, P. J. de Jong, L. I. Zon, J. H. Postlethwait, C. Nusslein-Volhard, T. J. P. Hubbard, H. R. Crollius, J. Rogers, and D. L. Stemple. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**:498-503.
- Jeukens, J., D. Bittner, R. Knudsen, and L. Bernatchez. 2009. Candidate genes and adaptive radiation: insights from transcriptional adaptation to the limnetic niche among coregonine fishes (*Coregonus* spp., Salmonidae). *Molecular Biology and Evolution* **26**:155-166.
- John, K. R., and A. D. Hasler. 1956. Observations on some factors affecting the hatching of eggs and the survival of young shallow-water Cisco, *Leucichthys artedi* LeSueur, in Lake Mendota, Wisconsin. *Limnology and Oceanography* **1**:176-194.
- Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403-1405.
- Jordan, S. D., and K. A. Lamia. 2013. AMPK at the crossroads of circadian clocks and metabolism. *Molecular and cellular endocrinology* **366**:163-169.
- Karjalainen, J., T. Keskinen, M. Pulkkanen, and T. Marjomaki. 2015. Climate change alters the egg development dynamics in cold-water adapted coregonids. *Environmental Biology of Fishes* **98**:979-991.
- Koelz, W. N. 1929. Coregonid fishes of the Great Lakes. *Bulletin of the United States Bureau of Fisheries* **43**:297-643.
- Kohn, M. H., W. J. Murphy, E. A. Ostrander, and R. K. Wayne. 2006. Genomics and conservation genetics. *Trends in ecology & evolution* **21**:629-637.
- Kwain, W.-h. 1975. Embryonic development, early growth, and meristic variation in rainbow trout (*Salmo gairdneri*) exposed to combinations of light intensity and temperature. *Journal of the Fisheries Board of Canada* **32**:397-402.
- Lawrie, A. 1978. The fish community of Lake Superior. *Journal of Great Lakes Research* **4**:513-549.
- Leggatt, R. A., and G. K. Iwama. 2003. Occurrence of polyploidy in the fishes. *Reviews in Fish Biology and Fisheries* **13**:237-246.
- Lerner-Marmarosh, N., J. Shen, M. D. Torno, A. Kravets, Z. Hu, and M. D. Maines. 2005. Human biliverdin reductase: a member of the insulin receptor substrate family with serine/threonine/tyrosine kinase activity. *Proceedings of the National Academy of Sciences* **102**:7109-7114.
- Lischer, H. E., and L. Excoffier. 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**:298-299.

- Madenjian, C. P., E. S. Rutherford, M. A. Blouin, B. J. Sederberg, and J. R. Elliott. 2011. Spawning habitat unsuitability: an impediment to cisco rehabilitation in Lake Michigan? *North American Journal of Fisheries Management* **31**:905-913.
- Magnuson, J. J., K. E. Webster, R. A. Assel, C. J. Bowser, P. J. Dillon, J. G. Eaton, H. E. Evans, E. J. Fee, R. I. Hall, L. R. Mortsch, D. W. Schindler, and F. H. Quinn. 1997. Potential effects of climate changes on aquatic systems: Laurentian Great Lakes and Precambrian shield Region. *Hydrological Processes* **11**:825-871.
- Maines, M. D. 2005. New insights into biliverdin reductase functions: linking heme metabolism to cell signaling. *Physiology* **20**:382-389.
- Matuszek, J. E., B. J. Shuter, and J. M. Casselman. 1990. Changes in lake trout growth and abundance after introduction of cisco into Lake Opeongo, Ontario. *Transactions of the American Fisheries Society* **119**:718-729.
- McCormick, J. H., B. R. Jones, and R. F. Syrett. 1971. Temperature requirements for growth and survival of larval ciscos (*Coregonus artedii*). *Journal of the Fisheries Board of Canada* **28**:924-927.
- McKinney, G. J., R. K. Waples, L. W. Seeb, and J. E. Seeb. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources* **17**:656-669.
- Meek, M. H., and W. A. Larson. 2019. The future is now: amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular Ecology Resources*.
- Meirmans, P. G., and P. H. Van Tienderen. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**:792-794.
- Miller, R. R., J. D. Williams, and J. E. Williams. 1989. Extinctions of North American fishes during the past century. *Fisheries* **14**:22-38.
- Miller, T. J., L. B. Crowder, J. A. Rice, and E. A. Marschall. 1988. Larval size and recruitment mechanisms in fishes: toward a conceptual framework. *Canadian Journal of Fisheries and Aquatic Sciences* **45**:1657-1670.
- Mortsch, L. D., and F. H. Quinn. 1996. Climate change scenarios for Great Lakes Basin ecosystem studies. *Limnology and Oceanography* **41**:903-911.
- Myers, J. T., D. L. Yule, M. L. Jones, H. R. Quinlan, and E. K. Berglund. 2014. Foraging and predation risk for larval cisco (*Coregonus artedi*) in Lake Superior: A modelling synthesis of empirical survey data. *Ecological Modelling* **294**:71-83.
- Nguyen, T. D., N. Hawley, and M. S. Phanikumar. 2017. Ice cover, winter circulation, and exchange in Saginaw Bay and Lake Huron. *Limnology and Oceanography* **62**:376-393.
- Oyadomari, J. K., and N. A. Auer. 2008. Transport and growth of larval cisco (*Coregonus artedi*) in the Keweenaw Current region of Lake Superior. *Canadian Journal of Fisheries and Aquatic Sciences* **65**:1447-1458.
- Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**:417.

- Pew, J., P. H. Muir, J. Wang, and T. R. Frasier. 2015. related: an R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources* **15**:557-561.
- Qian, X., Y. Ba, Q. Zhuang, and G. Zhong. 2014. RNA-Seq technology and its application in fish transcriptomics. *Omics: a journal of integrative biology* **18**:98-110.
- Rice, J. A., L. B. Crowder, and F. P. Binkowski. 1987. Evaluating potential sources of mortality for larval bloater (*Coregonus hoyi*): starvation and vulnerability to predation. *Canadian Journal of Fisheries and Aquatic Sciences* **44**:467-472.
- Rougeux, C., P.-A. Gagnaire, K. Praebel, O. Seehausen, and L. Bernatchez. 2018. Convergent transcriptomic landscapes under polygenic selection accompany inter-continental parallel evolution within a Nearctic *Coregonus* (*Salmonidae*) sister-species complex. *bioRxiv*.
- Schlei, O. L., A. CRÊTE-LAFRENIÈRE, A. R. Whiteley, R. J. Brown, J. B. Olsen, L. Bernatchez, and J. K. Wenburg. 2008. DNA barcoding of eight North American coregonine species. *Molecular Ecology Resources* **8**:1212-1218.
- Schmidt, S. N., C. J. Harvey, and M. J. Vander Zanden. 2011. Historical and contemporary trophic niche partitioning among Laurentian Great Lakes coregonines. *Ecological Applications* **21**:888-896.
- Schneeberger, P. J., M. P. Ebener, M. Toneys, and P. J. Peeters. 2005. Status of lake whitefish (*Coregonus clupeaformis*) in Lake Michigan. *in* Proceedings of a workshop on the dynamics of lake whitefish.
- Scott, W. B., and E. J. Crossman. 1998. Freshwater fishes of Canada. Oakville, Ont.: Galt House Pub.
- Selgeby, J. H. 1982. Decline of lake herring (*Coregonus artedii*) in Lake Superior: an analysis of the Wisconsin herring fishery, 1936–78. *Canadian Journal of Fisheries and Aquatic Sciences* **39**:554-563.
- Sharma, S., K. Blaggrave, J. J. Magnuson, C. M. O'Reilly, S. Oliver, R. D. Batt, M. R. Magee, D. Straile, G. A. Weyhenmeyer, and L. Winslow. 2019. Widespread loss of lake ice around the Northern Hemisphere in a warming world. *Nature Climate Change*:1.
- Sierszen, M. E., T. R. Hrabik, J. D. Stockwell, A. M. Cotter, J. C. Hoffman, and D. L. Yule. 2014. Depth gradients in food-web processes linking habitats in large lakes: Lake Superior as an exemplar ecosystem. *Freshwater biology* **59**:2122-2136.
- Smith, S. H. 1964. Status of the deepwater cisco population of Lake Michigan. *Transactions of the American Fisheries Society* **93**:155-163.
- Stockwell, J. D., M. P. Ebener, J. A. Black, O. T. Gorman, T. R. Hrabik, R. E. Kinnunen, W. P. Mattes, J. K. Oyadomari, S. T. Schram, and D. R. Schreiner. 2009. A synthesis of cisco recovery in Lake Superior: implications for native fish rehabilitation in the Laurentian Great Lakes. *North American Journal of Fisheries Management* **29**:626-652.
- Stockwell, J. D., D. L. Yule, T. R. Hrabik, M. E. Sierszen, and E. J. Isaac. 2014. Habitat coupling in a large lake system: delivery of an energy subsidy by an offshore

- planktivore to the nearshore zone of Lake Superior. *Freshwater biology* **59**:1197-1212.
- Stott, W., R. S. Cornman, C. Morrison, E. Roseman, K. Donner, and 2018. Development and Testing of Genomic Markers for *Coregonus* Species Identification and Population Assessment. Great Lakes Fishery Commission.
- Team, R. C. 2018. R: A Language and Environment for Statistical Computing.
- Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, and A. Narechania. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Research* **13**:2129-2141.
- Todd, T. N., and G. R. Smith. 1992. A review of differentiation in Great Lakes ciscoes. *Polskie Archiwum Hydrobiologii* **39**:261-267.
- Turgeon, J., and L. Bernatchez. 2003. Reticulate evolution and phenotypic diversity in North American ciscoes, *Coregonus* ssp. (Teleostei: Salmonidae): implications for the conservation of an evolutionary legacy. *Conservation Genetics* **4**:67-81.
- Wang, J., X. Bai, H. Hu, A. Clites, M. Colton, and B. Lofgren. 2012. Temporal and spatial variability of Great Lakes ice cover, 1973–2010. *Journal of Climate* **25**:1318-1329.
- Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**:57.
- Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, and E. M. Zdobnov. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**:543-548.
- Webb, S. A., and T. N. Todd. 1995. Biology and status of the shortnose cisco *Coregonus reighardi* Koelz in the Laurentian Great Lakes.
- Weger, B. D., M. Sahinbas, G. W. Otto, P. Mracek, O. Armant, D. Dolle, K. Lahiri, D. Vallone, L. Ettwiller, R. Geisler, N. S. Foulkes, and T. Dickmeis. 2011. The Light Responsive Transcriptome of the Zebrafish: Function and Regulation. *PLOS ONE* **6**:e17080.
- Welch, H. E., J. A. Legault, and M. A. Bergmann. 1987. Effects of Snow and Ice on the Annual Cycles of Heat and Light in Saqvaquac Lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **44**:1451-1461.
- Whitmore, D., N. S. Foulkes, and P. Sassone-Corsi. 2000. Light acts directly on organs and cells in culture to set the vertebrate circadian clock. *Nature* **404**:87-91.
- Wright, R. M., G. V. Aglyamova, E. Meyer, and M. V. Matz. 2015. Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC genomics* **16**:371.
- Zhong, S., J.-G. Joung, Y. Zheng, Y.-r. Chen, B. Liu, Y. Shao, J. Z. Xiang, Z. Fei, and J. J. Giovannoni. 2011. High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harbor Protocols* **2011**:pdb. prot5652.
- Zimmerman, M. S., and C. C. Krueger. 2009. An ecosystem perspective on re-establishing native deepwater fishes in the Laurentian Great Lakes. *North American Journal of Fisheries Management* **29**:1352-1371.