Approval

Name: Scott Anthony Pavey
Degree: Doctor of Philosophy

Title of Thesis:
Functional ecology and evolution of sockeye salmon (Oncorhynchus nerka) life history in the dynamic environments of Aniakchak and Katmai

Examining Committee:
Chair: Dr. R. Ydenberg, Professor

Dr. F. Breden, Professor, Senior Supervisor
Department of Biological Sciences, S.F.U.

Dr. J. Reynolds, Professor
Department of Biological Sciences, S.F.U.

Dr. J. Nielsen, Research Scientist
USGS Alaska Science Center

Dr. T. Hamon, Chief of Natural Resource Management and Research
Katmai National Park and Preserve

Dr. M. Hart, Associate Professor
Department of Biological Sciences, S.F.U.
Public Examiner

Dr. D. Noakes, Professor and Senior Scientist
Oregon Hatchery Research Centre, Oregon State University
External Examiner

July 23 2010
Date Approved
Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website at: <http://ir.lib.sfu.ca/handle/1892/112>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada
STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Simon Fraser University
Burnaby, BC, Canada

Last update: Spring 2010
Abstract

Sockeye salmon exhibit great ecological diversity among populations. During the last glacial period, ice covered much of what today are freshwater habitats. As the glaciers retreated, sockeye colonized new freshwater habitats from relatively few glacial refugia. Colonizing populations adapted at a very fine spatial scale among river drainages, tributaries and lakes within rivers, and even divergent habitats within lakes. All of this occurred within the past 15,000 years since the last glacial maximum. This resulted in many thousands of locally adapted populations and a grand display of the process of evolution within a species. In this dissertation, I explore genetic and phenotypic diversity in the dynamic and changing environments of Aniakchak National Monument and Preserve as well as Katmai National Park in southwest Alaska. Recent eruptions at Aniakchak include events 500 and 79 years ago and the caldera presently contains sockeye salmon populations spawning in different habitats. Using genetic tools, I find that ecological divergence occurred in egg size and body depth in less than 500 years or 100 generations. Secondly, sockeye salmon exhibit a broad life history division by rearing habitat; some populations rearing in lakes (lake-type sockeye) and others rearing in rivers (riverine sockeye). I describe differences in juvenile body shape and relate these to differences in foraging strategy and predation. Finally, I apply gene expression technology to
understand the life history differences and the molecular trade-offs in sockeye salmon populations. I start with a review of recent technological advances that relate gene expression to ecology, evolution, and the formation and maintenance of new species. I then relate functional, expressed genes in muscle tissue to lake-type and riverine juvenile populations. This provides an ecological context to genes that are normally only described in artificial situations. Taken together, this work furthers the understanding of the interaction of ecology and evolution, from genes to populations to broad life history types.

Keywords: sockeye salmon, *Oncorhynchus nerka*, ecological genomics, ecological transcriptomics, lake-type, riverine, river-type, sea-type, ecological divergence, ecological speciation, divergence with gene flow, recent divergence, Aniakchak National Monument and Preserve, Katmai National Park and Preserve
Dedication

I dedicate this dissertation to my mother and father, Susan Pavey and Thomas Pavey, for their unyielding support of my goals.
Acknowledgements

I want to acknowledge the three primary mentors of my scientific career. Dr. Jennifer Nielsen, Dr. Troy Hamon, and Dr. Felix Breden. I also wish to thank Dr. Jocelyn Krebs, Dr. Frank von Hippel, Dr. Arne Mooers, Dr. Bernard Crespi, Dr. Michael Hart for valuable interactions during my academic training. I am thankful for the discussions in which I participated at FAB* Lab and the Vancouver Evolution Group (VEG). I thank Dr. Ben Koop and his entire lab at the University of Victoria for valuable training, lab space, and ongoing interaction. I thank Marlene Nyguyen for calmly leading me through the administration of graduate school. I thank the entire Breden Lab including Dr. Kristen Fay Gorman, Martin Brummell, Dr. Heather Alexander, Corey Watson and Ben Sandkam. I thank Dr. Frank von Hippel for teaching his graduate level evolution course at University of Alaska Anchorage. I thank Public Examiner: Dr. Michael Hart, External Examiner: Dr. David Noakes, Chair: Dr. Ronald Ydenberg and Committee Member: Dr. John Reynolds for helping to make my defense one to remember.
# Table of Contents

Approval........................................................................................................................................ ii
Abstract.......................................................................................................................................... iii
Dedication ........................................................................................................................................ v
Acknowledgements..................................................................................................................... vi
Table of Contents....................................................................................................................... vii
List of Figures............................................................................................................................. x
List of Tables................................................................................................................................. xi

## CHAPTER 1: GENERAL INTRODUCTION TO SOCKEYE SALMON
### EVOLUTION AND ECOLOGY ................................................................................. 1
1.1 Sockeye Salmon as a Model System ........................................................................... 1
1.2 Summary of Thesis Chapters ..................................................................................... 3
1.3 Literature Cited ........................................................................................................ 5

## CHAPTER 2: ECOLOGICAL DIVERGENCE DESPITE MIGRATION IN
### SOCKEYE SALMON (ONCORHYNCHUS NERKA)........................................................ 7
2.1 Abstract....................................................................................................................... 7
2.2 Introduction ................................................................................................................... 8
2.3 Materials and Methods ............................................................................................... 13
   2.3.1 Ecological Parameters .................................................................................... 13
   2.3.2 Morphological Traits .................................................................................. 13
   2.3.3 Time since Divergence and Migration ........................................................... 15
2.4 Results......................................................................................................................... 19
   2.4.1 Ecological Parameters .................................................................................... 19
   2.4.2 Morphological Traits .................................................................................. 19
   2.4.3 Time since Divergence and Migration ........................................................... 20
2.5 Discussion ..................................................................................................................... 22
   2.5.1 Morphology .................................................................................................... 23
   2.5.2 Divergence with Migration ........................................................................... 25
2.6 Acknowledgments ........................................................................................................ 27
2.7 Literature Cited ......................................................................................................... 28
2.8 Tables ......................................................................................................................... 33
2.9 Figures ......................................................................................................................... 38

## CHAPTER 3: CONTRASTING ECOLOGY SHAPES JUVENILE LAKE-TYPE
### AND RIVERINE SOCKEYE SALMON ..................................................................... 46
3.1 Abstract....................................................................................................................... 46
3.2 Introduction ................................................................................................................... 47
3.3 Methods ....................................................................................................................... 49
5.3.6 Array Normalization and Statistical Analysis ............................................. 144
5.3.7 Gene Ontology Analysis ............................................................................ 145

5.4 Results ........................................................................................................ 146
5.5 Discussion .................................................................................................. 147
5.6 Acknowledgments ....................................................................................... 153
5.7 Literature Cited .......................................................................................... 153
5.8 Tables .......................................................................................................... 158
5.9 Figures ......................................................................................................... 167

CHAPTER 6: GENERAL CONCLUSIONS AND FUTURE DIRECTIONS ............ 170

6.1 Conclusions ............................................................................................... 170
6.2 Future Directions for the Field ................................................................. 171
6.3 Literature Cited .......................................................................................... 173
List of Figures

Figure 2.1 Aniakchak National Monument and Preserve (ANMP) showing Aniakchak Caldera and the two study populations .................................................. 38
Figure 2.2 Size distribution of substrate for the two breeding locations ................. 40
Figure 2.3 Comparison of body depth in beach and outlet males............................ 41
Figure 2.4 Average and standard error of standardized egg weight for beach and outlet female sockeye salmon from Surprise Lake......................... 42
Figure 2.5 Posterior probability distributions for time since divergence in years ....... 43
Figure 2.6 Posterior probability of migration rates between populations .................. 44
Figure 2.7 Posterior probability of effective population sizes of beach, outlet, and ancestral populations ........................................................................ 45
Figure 3.1 Map of study area showing Katmai National Park and Preserve, Aniakchak National Monument and Preserve, the Bering Sea, the Pacific Ocean, and the six sampling locations ........................................ 73
Figure 3.2 The twelve landmarks used for morphometric analysis ........................... 75
Figure 3.3 Deformation grid based on the habitat variable of lake-type and riverine sockeye salmon both with and without Upper Q-Tip Lake (UQT) .............................................................. 76
Figure 3.4 Habitat canonical score plotted against centroid size ............................... 77
Figure 4.1 Conceptual diagram of the different ways the genetic and environmental components of gene expression might contribute to ecological speciation ........................................................................ 132
Figure 4.2 The effects of gene expression mediated phenotypic plasticity (GMPP: y-axis) on colonization of new environments, and subsequent population persistence ................................................................. 133
Figure 4.3 An example of the effects of gene expression in two genes (bone morphogenetic protein 4, Bmp4, and calmodulin, CaM) on phenotypic traits of likely importance for ecological speciation: in Geospiza Darwin’s finches ......................................................... 134
Figure 5.1 The two sampling locations: Surprise Lake and Albert Johnson Creek in southwest Alaska ................................................................................ 167
Figure 5.2 Heat map depicting all 1026 genes significantly differentially expressed (1.5 fold; P = 0.05; not corrected for multiple comparisons) between Albert Johnson Creek (AJC) and Surprise Lake (SL) ............... 169
List of Tables

Table 2.1 Male body shape and female egg mass data summary................................. 34
Table 2.2 ANCOVA table for male body depth comparison........................................ 34
Table 2.3 Estimated mutation rates based on mutation rate scalars.............................. 35
Table 2.4 Mutation rate estimates based on length of repeat method.......................... 35
Table 2.5 Comparison of the high point, average values, 95% credibility interval and 90% highest posterior density (HPD) of the posterior distribution of divergence onset dates in years before sample collection for all three runs of the IM program. ................................................................................... 35
Table 2.6 Comparison of the high point and average values of the posterior distribution of average migration dates................................................................. 36
Table 2.7 Effective migrants per generation for all three runs of the IM program......... 36
Table 2.8 Comparison of demographic parameters obtained with IM and MIGRATE programs ............................................................................................................. 37
Table 3.1 Sockeye salmon sampling locations ............................................................. 69
Table 3.2 Results of the nested MANCOVA for all six populations of sockeye salmon and also five populations excluding Upper Q-Tip Lake (UQT)........... 70
Table 3.3 Pearson’s correlation coefficients between superimposed landmark coordinates and the canonical axis ................................................................. 71
Table 3.4 Results of the discriminant function analysis with numbers of observed correct and incorrect classification into habitat of each population based on morphology ................................................................. 72
Table 4.1 Comparisons of gene expression methodologies ......................................... 125
Table 4.2. Examples of studies showing that gene expression is affected by ecological stress ............................................................................................................ 126
Table 4.3 Summary of what is known about gene expression and ecological speciation ............................................................................................................... 129
Table 5.1 Tecan HS 400 Pro hybridization protocol...................................................... 159
Table 5.2 Genes significantly over-expressed in Albert Johnson Creek sockeye salmon muscle compared to Surprise Lake sockeye salmon muscle .......... 160
Table 5.3 Genes significantly over-expressed in Surprise Lake sockeye salmon muscle compared to Albert Johnson Creek muscle ............................... 163
Table 5.4 Gene Ontology enrichment results............................................................... 166
CHAPTER 1: GENERAL INTRODUCTION TO
SOCKEYE SALMON EVOLUTION AND ECOLOGY

“...as the mane to the lion, the shoulder pad to the boar, and the
hooked jaw to the male salmon; for the shield may be as important
for victory, as the sword or spear.” (Darwin 1859)

1.1 Sockeye Salmon as a Model System

Does the study of ecology have value beyond applied fields of
conservation, management, and taxonomy? Or is the understanding of ecology
of value to other biological disciplines such as evolution, medicine, and
genomics? The application of ecology to other biological disciplines occurred in
steps, first with the relationship of ecology to taxonomy and evolution. More
recently, the concept of ecological speciation greatly enhanced the
understanding of the evolutionary process (Schluter 1998). Now with the
technological tools of rapid sequencing, bioinformatics, and transcriptome
screening, the role of ecology in the understanding of other aspects of biology
continues to increase (Gibson 2002; Thomas and Klaper 2004; Kammenga et al.
2007; Ouborg and Vriezen 2007).

Sockeye salmon (*Oncorhynchus nerka*) has been an important ecological
model for decades (Hendry and Stearns 2004). As with all models, there are
disadvantages with sockeye salmon, one of which is that their complex life
history and long life span complicates laboratory rearing and crossing. This
notwithstanding, there are many advantages of using sockeye salmon as a
model for ecology and evolution. First, all current populations are the result of postglacial colonization. This provides numerous replicated ecological situations with thousands of locally adapted and genetically distinct populations (Taylor 1991; Wood 1995). Second, there are many morphological and behavioral adaptations relating to ecology (Burgner 1991). Feeding and reproduction occur in discrete, non-overlapping life stages, separated by a profound metamorphosis upon returning to freshwater to spawn. This facilitates the compartmentalization in the study of these activities that are concurrent in most species. Third, since sockeye salmon are commercially important, long-term records of harvest and escapement are available (Stearns and Hendry 2004). The early development of many neutral genetic markers helped managers regulate harvest and define evolutionary significant units (Wood 1995). These same markers are tools to detect past and present demographic parameters such as gene flow, bottlenecks, and time of divergence. Fourth, genetic relationships over the entire species range are well defined (Beacham et al. 2006). Also, primarily due to their importance in aquaculture, closely related rainbow trout (Oncorhynchus mykiss) and Atlantic salmon (Salmo salar) bring a wealth of laboratory physiological and trait heritability experimentation that is applicable to sockeye salmon (Stearns and Hendry 2004). Finally, the phylogenetic relationship among salmonids are well characterized (Ramsden et al. 2003; Crespi and Fulton 2004), and the genomic relationships are being clarified (Phillips et al. 2009).

With the accelerating development of genomic tools, the utility of sockeye salmon is entering a new and exciting phase. The cGRASP (consortium for
Genomics Research on All Salmon) is describing the genome of salmon species (Davidson and Koop 2005; Koop et al. 2008). One of the products is a gene expression microarray that currently facilitates the simultaneous screening of 32,000 transcripts in a single assay (von Schalburg et al. 2008). Many of these genes have tentative gene function annotations. The goal of the consortium is identification and description of genes of all salmon. The application of genomic tools to sockeye systems is a powerful merger of disciplines with promise to uncover the ecological function of individual genes.

1.2 Summary of Thesis Chapters

The goal of this work is to synthesize divergent biological disciplines and recently developed analysis techniques to achieve a greater understanding of the evolution and ecology of sockeye salmon life history. The study systems are in Katmai National Park and Preserve and Aniakchak National Monument and Preserve in southwest Alaska. This area is home to great diversity in habitats and past disturbance, as well as an abundance of sockeye salmon populations.

In the second chapter, I describe a case of ecological divergence between two sockeye salmon populations that diverged more recently than others reported in the literature. The spawning and rearing habitat is within a volcanic lake with a history of recent eruption events. The geologic record of volcanic eruptions coupled with coalescence based genetic analyses suggests that two ecologically distinct spawning habitats within the volcano resulted in morphological adaptations in sockeye populations within 100 generations.
In the third chapter, I describe the morphological differences of juvenile sockeye that rear in ecologically distinct habitats. I employ an entire body approach and multivariate statistics yielding the first morphological comparison of riverine and lake-type juveniles. The hydrodynamic shape differences of riverine and lake-type sockeye may be due to contrasting foraging, predation, and habitat complexity in river and lake environments.

In the fourth chapter, I review the state-of-the-art of the application of gene expression in studies of ecological speciation. The plasticity of gene expression may be important for persistence in newly colonized environments. Colonization may play a large role in ecological speciation as it potentially exposes populations to divergent selection. In addition, changes in heritable levels of gene expression may result in evolving phenotypes more than changes in coding regions of individual genes (Gibson and Weir 2005). Regulatory regions of the genome may be the genetic elements evolving that have cascading effects on the gene expression and the ultimate phenotype.

In the fifth chapter, I use gene expression microarrays to screen the transcriptome of sockeye salmon juveniles in river and lake environments. I found individual genes as well as categories of gene function that associate with differing habitats and ecology. This molecular description of life history is a starting point for future study of the ecological function of genes in sockeye salmon.

In the sixth chapter, I conclude the dissertation. Taken together, this work combines multiple disciplines for the goal of a greater understanding of ecology
and evolution. By combining sockeye salmon as a long standing model for the understanding of ecology, evolution and behavior with recently developed analysis and genomic resources, I present a holistic approach to the study of life history and ecology.

1.3 Literature Cited


CHAPTER 2: ECOLOGICAL DIVERGENCE DESPITE MIGRATION IN SOCKEYE SALMON (ONCORHYNCHUS NERKA)

Published in the journal Evolution in 2010, volume 64, pages 1773-1783.

Author list: Scott A. Pavey, Jennifer L. Nielsen, and Troy R. Hamon.

Author contribution: Scott A. Pavey planned the study, conducted the fieldwork, conducted the lab work, ran the analysis, and was the primary writer of the paper.

2.1 Abstract

Ecological divergence may result when populations experience different selection regimes, but there is considerable discussion about the role of migration at the beginning stages of divergence before reproductive isolating mechanisms have evolved. However, detection of past migration is difficult in current populations and tools to differentiate genetic similarities due to migration versus recent common ancestry are only recently available. Using past volcanic eruption times as a framework, we combine morphological analyses of traits important to reproduction with a coalescent based genetic analysis of two proximate sockeye salmon (Oncorhynchus nerka) populations. We find that this is the most recent (~500 years, 100 generations) natural ecological divergence recorded in a fish species, and report that this divergence is occurring despite migration. While studies of fish divergence following the retreat of glaciers (10,000-15,000 years ago) have contributed extensively to our understanding of
speciation, the Aniakchak system of sockeye salmon provides a rare example of the initial stages of ecological divergence following natural colonization. Our results show that even in the face of continued migration, populations may diverge in the absence of a physical barrier.

**Keywords:** Colonization, divergence with migration, ecological speciation, isolation with migration (IM), rapid evolution, sympatric speciation.

### 2.2 Introduction

Populations subjected to different selection regimes can evolve reproductive isolation (Mayr 1947). This divergence ultimately may result in speciation arising from ecological differences. (Bush 1994; Schluter 1996a, 1996b, 2000; Feder et al. 2005; Rundle and Nosil 2005; Funk et al. 2006). In many cases of ecological divergence, a physical barrier to migration separates the populations in the initial stages (Schluter 2001; Schluter et al. 2001; Rundle and Nosil 2005). When the populations regain contact, isolating mechanisms (behavioral, morphological, etc.) have already evolved in the absence of migration. In populations that are currently sympatric, this may have occurred via a double colonization event; after one colonization occurs and a population locally adapts to a habitat or resource then a second colonization occurs, and adapts to an unoccupied niche (Schluter 2001; Schluter et al. 2001). Alternatively, colonization may have occurred from two different source populations that brought differences that evolved in allopatry (Schluter et al. 2001).
However, in some cases, divergence may occur with gene flow in early stages (Johnson et al. 1996; Filchak et al. 2000; Johannesson 2001; Barluenga et al. 2006; Hey 2006; Savolainen et al. 2006; Bolnick and Fitzpatrick 2007; Nosil 2008). Recently colonized populations may provide ideal systems for the study of ecological divergence, as initial reproductive isolation has a disproportionate effect on divergence that may not be apparent at later stages (Coyne and Orr 2004). Very recent cases (<1000 years) of colonization and ecological divergence demonstrate that this process can occur rather quickly (Hendry et al. 2007), but cases involving populations that were not the result of introductions or manipulations by humans are rare (Diamond et al. 1989; Carroll et al. 1997).

Here we present an example of ecological divergence following colonization that is both recent (~500 years, ~100 generations) and occurring with migration.

Sockeye salmon (*Oncorhynchus nerka*) reproduce in freshwater habitats throughout much of the North Pacific region (Burgner 1991). Several adult phenotypic traits are highly correlated with breeding environment and are believed to be the result of parallel evolution (Burgner 1991; Taylor 1991; Blair et al. 1993). Recent work has shown that adult body size and shape in sockeye are strongly related to depth and water velocity of their breeding habitat (Quinn et al. 2001). In general, sockeye males breeding along lake beaches have deeper bodies than those breeding in riverine habitats (Blair et al. 1993; Hendry et al. 2000). This appears to be a response to natural and sexual selection in the breeding environments (Quinn et al. 2001; Hamon and Foote 2005). Sockeye egg mass is correlated with breeding substrate size (Quinn et al. 1995). Since
these traits are adaptations to the ecology of the site of reproduction, they may
be traits that are directly responsible for reproductive isolation (Schluter 2001;
Rundle and Nosil 2005).

Aniakchak National Monument and Preserve (ANMP; Fig. 2.1) in southwest Alaska contains the most active volcano in the eastern Aleutian Arc, having erupted more than 40 times in the last 10,000 years (Neal et al. 2001). Several of these cataclysmic geologic events are well documented and provide a framework with which to evaluate the timing of divergence. A massive volcanic eruption 3,650 years ago formed a large caldera (Aniakchak Caldera) that filled with water forming a lake (McGimsey et al. 1994; Pearce et al. 2004). Sometime after this but before a more recent eruption that occurred 500 (standard error 369-565) years ago, the caldera wall collapsed resulting in a large flood and the formation of the Aniakchak River, which connects the caldera lake (Surprise Lake; elevation 321 m) with the Pacific Ocean through “The Gates”, a chasm breaching the caldera wall (see Fig. 2.1, McGimsey et al. 1994). Sometime after this connection, sockeye salmon colonized Surprise Lake. In addition to the well documented eruptions mentioned above, the volcano erupted again in 1931 (McGimsey et al. 1994). These eruptions probably affected breeding, rearing, and incubating conditions and may have impacted or eliminated any sockeye populations present in the caldera during that time. In fact, much of the inlet waters to the lake are presently devoid of dissolved oxygen as a result of volcanic activity (Cameron and Larson 1993) and a large portion of the associated beaches are unused by breeding sockeye. Current sockeye
populations in Aniakchak Caldera may have colonized after the original ocean access following the flood, after the substantial eruption 500 years ago, or following the most recent eruption (77 years ago).

Sockeye in Aniakchak Caldera use two breeding habitats, the outlet and the beaches, which are spatially separated by less than 1.5 km. These populations are genetically distinct ($F_{ST}=0.01; P<0.001$) and together form a clade that is distinct from other populations in the area (Pavey et al. 2007). There are several different scenarios that could have resulted in this current situation. First, divergence may have occurred prior to colonization of the caldera via two different source populations; i.e. a population adapted to breeding in outlets colonized the outlet and a population adapted to breeding at beaches colonized the beaches. Subsequent recent gene flow may result in convergence at neutral loci, while divergent ecology maintains adaptive differences. In this “two source” model, we would expect divergence time to be considerably earlier than caldera access, perhaps >10,000 years ago when most sockeye divergence occurred following glacial retreat. In this case, the ecology and proximity of the current habitats are not informative about population divergence.

Alternatively, these populations may represent a monophyletic group that colonized the caldera and subsequently diverged in response to selection, which would yield a more recent time for divergence in comparison to the timing of colonization. This colonization prior to divergence can be considered under two scenarios that describe relatively different roles of migration in divergence (Rundle and Nosil 2005; Xie et al. 2007). In the first of these two scenarios,
migration was greatly restricted through a double invasion of the habitat from a common source. In this scenario, colonization and local adaptation to one habitat occurred first, followed by a second invasion from a common source that colonized the unoccupied habitat (Schluter et al. 2001). Alternatively, in the second scenario, colonization may have occurred only once and populations diverged despite gene flow (Johannesson 2001).

In this study, we first measured ecological parameters of the breeding habitats (substrate size and rate of water flow) of these recently colonized populations. Next, we measured adult body depth and egg mass, morphological characters important to reproduction and shown to be correlated in relation to these ecological differences in many populations across the species range. Then, through applying coalescent techniques to a microsatellite database (Hey and Nielsen 2004; Won and Hey 2005), we estimated the time of onset of population divergence to see if the data suggest that the divergence took place after the availability of the habitat. Finally, we determined whether migration occurred after the onset of divergence as well as the relative timing of any migration events. If there is no detectible migration after divergence, the hypothesis of double colonization is supported. Migration after divergence is consistent with the hypothesis of ecological divergence despite gene flow.
2.3 Materials and Methods

2.3.1 Ecological Parameters

We determined average water velocity of the outlet from cross-sectional area and previously recorded discharge (Bennett 2004). Substrate composition was determined by Wolman pebble counts (Quinn et al. 1995).

2.3.2 Morphological Traits

2.3.2.1 Adult body shape

We captured 301 breeding adult males by net in 2001-2003. All measurements were to the nearest millimeter. Sampling consisted of measuring midorbital to hypural length (MOH; body length) and body depth at the anterior insertion of the dorsal fin. Body depth of breeding males was compared between habitats (outlet and beach) by Analysis of Covariance. The model included year, to account for variation in overall growth and size among the different sample years, and MOH as a covariate. In addition to the measurements, we assigned spawning condition to one of three categories for each individual. We recorded males as pre-spawning if the fish was bright red and in good physical condition but not expressing milt under gentle abdominal pressure. Males still in good physical condition but expressing milt were judged to be spawning, and males with extensive scarring, worn away skin, and showing a lack of slime production were categorized as senescent. We did not sample sockeye salmon showing silver coloration, as this indicates that they are still immature and their eventual spawning location and mature body shape are not finalized at that point.
2.3.2.2 Egg mass

Females were captured in August 2006 during spawning activity by net in the same manner as males were captured above, and MOH was measured in the same manner as for males. About twenty eggs were taken from each of fifty females at the beach habitats and thirty females at the outlet. We selected only females that were expressing eggs upon abdominal pressure. Eggs were preserved in 10% formalin. Once back in the lab, we blotted each batch of eggs with a Kimwipe® to remove external formalin solution. Then we measured each group of eggs to the nearest 0.1 mg. The source population of each sample was concealed during measurement. Of the 80 females for which egg samples were obtained, we eliminated seventeen samples from the beach collection and six samples from the outlet collection due to connective tissue attachment. We excluded eggs with adhesions or that did not freely separate from one another. As these samples may represent incomplete development, the eggs and the females that they came from were removed from the analysis. With each sample, we divided the total mass of all the eggs by the number of eggs collected to get an average mass. Egg mass was compared between habitats by Analysis of Covariance. The model included MOH as a covariate. Body length (MOH) accounts for some variation in egg mass, so the incorporation of body length in the analysis of covariance allowed us to perform residuals analysis to look at the effect of the habitat type on egg mass.
2.3.3 Time since Divergence and Migration

2.3.3.1 IM analysis

We performed an analysis using the program Isolation with Migration (Hey and Nielsen 2004) on a microsatellite DNA database from Pavey et al. (2007). We performed initial pilot runs of the program with large priors to make sure that the posterior probability area was contained within the priors. We then fine-tuned the priors in order to “zoom in” to show the detail of the posterior distributions while still encompassing the whole for each parameter in the model. After initial pilot runs of the program, we executed three long runs with 18 heated chains for ~10,000,000 steps. The command line for these runs was:

```
-q1 5 -m1 50 -m2 50 -t 1 -b 72.0 -I 24.0 -u 5 -p 4567 -n 18 -k 100 -fg -g1 0.6 -g2 0.95 -e 24.0.
```

The first four commands set the priors for all of the parameters. The “-b” command sets the program burn-in for 72 hours. The “-I” command tells the program to make an output file every 24 hours. The “-u” command sets the generation time of five years. The “-p” command sets the output options. The “-n” command sets the number of chains to 18. The “-k” command sets the number of swap attempts per step to 100. The “-fg” command sets the heating scheme to geometric. The “-g1” and “-g2” specify the degree of chain heating. The “-e” command creates a checkpoint file every 24 hours. We report high point and average posterior probability estimates from all loci for time since divergence onset, migration rate in each direction, and average date of migration events. Finally we report effective number of migrants for each population.
Since all parameters estimated in the IM model are in units of mutation, we need to estimate mutation rate in order to convert the parameter estimates into demographic units. Experimental work with other tetranucleotide microsatellites has demonstrated that mutation rate is often larger than the commonly used default mutation rate of $1 \times 10^{-4}$ (Weber and Wong 1993; Ellegren 1995; Leopoldino and Pena 2003). We estimated the mutation rate for each locus using two different methods. First, with the exception of One105, our markers are highly polymorphic, so we expect a larger than average mutation rate. We assigned the conservative mutation rate of $1 \times 10^{-4}$ to our one moderately polymorphic marker, One105, and used the mutation rate scalar estimates obtained from running the IM program to estimate the mutation rates of the other loci. Secondly, mutation rates in tetranucleotide microsatellites are shown to vary with length of repeat unit (Leopoldino and Pena 2003). We calculated a regression equation using the data from Leopoldino and Pena (2003) which was obtained from comparing observed mutation rate with geometric mean of number of repeats. We converted estimated parameters into demographic units with the method yielding the more conservative (slower) global mutation rate, which is the geometric mean of the mutation rates scaled from the model output for all loci.

2.3.3.2 IM assumptions

The IM program has several assumptions about the input data. Perfect tetranucleotide microsatellites have been shown to exhibit mutations that are well described by the stepwise mutation model (SMM, Shiver et al. 1993; Leopoldino and Pena 2003), which is the model used in IM (Hey and Nielsen 2006). One
marker from this database, *One110*, did not meet the requirement of the IM program of a perfect repeat and was excluded. We included the five microsatellites that followed a perfect tetranucleotide repeat pattern from this database of 268 individuals. Another assumption about the input data is that the markers should not be physically linked. We tested for linkage disequilibrium and found no significant linkage disequilibrium in any of these markers (Pavey et al. 2007). A third assumption about the input data is that the markers are not under selection. This assumption is more difficult to explicitly test. One potential indication of selection is deviations from Hardy-Weinberg equilibrium (Conner and Hartl 2004) which is not present with these data (Pavey et al. 2007) or with similar data on these same markers in another study involving sockeye salmon in southwest Alaska (Olsen et al. 2004). Another potential indication of selection is outlier loci, or one or two loci being primarily responsible for the measured genetic differences. This was not indicated with these markers in another study (Olsen et al. 2004). In order to determine whether this was the case in our dataset, we used the program WHICHLOCI to assess the relative contribution of each marker to the measured genetic divergence. Because we ran IM without the *One110* locus (see above) we ran WHICHLOCI both including and excluding *One110*. Also, we sequentially dropped each locus to see if this substantially affected the $F_{ST}$ between these populations. We performed this analysis in GENEPOP (Raymond and Rousset 1995).

Another assumption of IM is that the two analyzed populations are more closely related to each other than they are to other populations. (Hey and Nielsen
The basic phylogenetic unit of lake-type sockeye salmon is the nursery lake (Wood 1995; Beacham et al. 2006). When glaciers recede and expose new lake habitats, sockeye colonize. Divergence also occurs among habitats within the nursery lake, but genetic differences are generally much smaller within lakes than between lakes. This is the situation at Aniakchak (Pavey et al. 2007). The $F_{ST}$ between the beach and outlet populations within Surprise Lake was smaller than the $F_{ST}$ between either of these populations and any other population outside of Surprise Lake. These relationships are further illustrated in our neighbor joining tree, where the bootstrap value for the Surprise Lake populations forming a clade received 96% support. Though the best information we have supports the hypothesis that the two Surprise Lake populations are genetically closer to each other than to other populations, genetic similarity is certainly not “proof” of common ancestry. Gene flow as well as common ancestry will result in close genetic relationships. This complication is precisely why we want to apply the IM model to this system, as it partitions these competing homogenizing processes.

As with all applications of the IM model between populations, we cannot rule out that there is some level of gene flow with other populations outside of Surprise Lake. However, due to the 300m elevation gain that may impose a substantial migratory barrier to outside populations, as well at the close proximity and the limitation of the study to the only populations sharing Surprise Lake as the nursery lake, we believe that “the history of a sample from two populations can
reasonably be described by an Isolation with Migration model” (Hey and Nielsen 2006).

2.3.3.3 MIGRATE analysis

We ran the program MIGRATE 3.0.3 (Beerli and Felsenstein 2001; Beerli 2006) on the same dataset in order to compare the output with the results from IM. This model estimates similar parameters to IM, except there is no “time since divergence” parameter. The model assumes that there has been sufficient time since divergence that migration and drift subsequent to divergence has a greater effect on current genetic relationships than shared ancestry prior to divergence. We used the same mutation rates from the method above. We used the Bayesian search strategy with slice sampling Markov chain Monte Carlo (MCMC) with four heated chains. We started with experimental runs with large priors and then performed long runs with uniform priors of 0-30 for both $\theta$ and $M$. We set the burn-in for 50,000 steps and collected data for 2,400,000 steps.

2.4 Results

2.4.1 Ecological Parameters

Surprise Lake outlet had larger substrate than Surprise Lake beaches (Fig. 2.2). Surprise Lake beaches have no measurable flow in the water column, but we calculated an average current of 0.455 m/sec in the outlet.

2.4.2 Morphological Traits

Male midorbital to hypural length averaged 497.9 mm (MOH; body length) and body depth averaged 186.6 mm. There was no difference in MOH between
the populations (P = 0.79). Analysis of covariance results indicated that male body depth was significantly correlated with length (P < 0.0001), and that both year (P < 0.01) and habitat (P < 0.001) were significant factors in determining body depth (Table 2.1, 2.2, Fig. 2.3).

Egg mass was significantly correlated with MOH of females (P < 0.01). In addition, habitat was a significant factor; the eggs of outlet sockeye were larger than those of beach sockeye (Table 2.1, Fig. 2.4; P < 0.02).

### 2.4.3 Time since Divergence and Migration

Our WHICHLOCI simulation indicated that individuals could be assigned to the correct population most of the time (86.0%, Standard error [SE] 0.11% including One110; 82.3%, SE 0.12% excluding One110). Also, the program indicated no outlier loci, as the relative assignment ability of each locus was evenly spread (minimum score 13.4% of 6 loci including One110, 16.5% of 5 loci excluding One110). Sequential dropping of loci did not substantially change the original FST of 0.0112 reported in Pavey et al. (2007). The range of sequential dropping was FST = 0.0110 – 0.0135.

Mutation rates calculated from the two methods were within the same order of magnitude. The geometric mean of mutation rates for all loci using the mutation rate scalar method was 7.91 x 10⁻⁴ per generation or 1.58 x 10⁻⁴ per year (Table 2.3). For the length of repeat method, the geometric mean of mutation rates of all loci was 3.40 x 10⁻³ per generation or 6.79 x 10⁻⁴ per year
We used the lesser rate obtained with the scalar method for all demographic conversions.

High point estimates of the posterior probability of the time since onset of divergence ranged from 47-123 years prior to sample collection (Table 2.5, Fig 2.5) however runs 2 and 3 exhibited 2 and 3 peaks respectively, and all peaks occurred between 47 and 400 years prior to sample collection. Mean distribution values of the entire posterior probability distributions for divergence times ranged between 389-503 years ago (Table 2.5). It is important to note that with all of the posterior probabilities, the y-axis scale is completely dependent on the number of bins in the x-axis (1000). The area under the curve is equal to one.

Migration occurred after the onset of population divergence. High point probability estimates of migration rate per generation (m) ranged between 0.00009875 and 0.002706 and mean probability estimates of migration rate ranged between 0.003377 and 0.006379 (Fig. 2.6). High point probability estimates of the average date of all migration events ranged between 42 years and 120 years ago (Table 2.6). Mean probability estimates of the average date of all migration events ranged between 177 and 305 years ago (Table 2.6). High point probability estimates of the effective number of migrants (parameters $\theta \times m/2$) ranged between 0.16 and 9.0 (Table 2.7). Mean probability estimates of the posterior probability of the effective number of migrants ranged between 8.9 and 18.4 (Table 2.7). High point estimates of $N_E$ ranged between 740-1655 for beach, 528-988 for outlet, and 5047-5588 for ancestral (Fig.2.7). Mean probability estimates of $N_E$ ranged between 1386-1709 for beach, 961-1394 for outlet, and
5561-5967 for ancestral (Fig. 2.7). Our results from our MIGRATE analysis were very similar to our results with IM in all common parameters, $N_E$ and migration rate (Table 2.8).

2.5 Discussion

We have described the most recent ecological divergence reported in a fish species following natural colonization. The divergence observed here is probably in a very early stage, but the morphological differences are consistent in direction with that documented for similar ecological differences in other sockeye populations. Our results indicate that migration occurred after divergence onset suggesting that this divergence is occurring despite migration.

There are many examples of ecological divergence following natural colonization in fish including threespine stickleback (*Gasterosteus aculeatus*, Lavin and McPhail 1985, 1993; Schluter 1996b; Rundle et al. 2000; Reusch et al. 2001), lake whitefish (*Coregonus clupeaformis*, Lu and Bernatchez 1999; Rogers et al. 2002; Derome et al. 2006) arctic char (*Salvelinus alpinus*, Gislason et al. 1999; Klemetsen et al. 2002; Johnston et al. 2004) and sockeye salmon (Blair et al. 1993; Wood and Foote 1996). However, all of these important examples are on a post glacial retreat timescale (~10,000-15,000 years). The present study demonstrates a very recent ecological divergence following natural colonization. This ecological divergence is extremely recent (~500 years, 100 generations), between populations of close geographic proximity (~1500 meters), and occurred despite migration.
It is possible that the actual mutation rates of the microsatellite markers used in this analysis were different than our estimates. This difference would proportionately change our converted demographic parameters: time since divergence onset, \( N_E \), average date of migration, as well as migration rate would all be affected by mutation rate. However, the relationship between divergence time and time of average migration would change similarly, making this relationship between them independent of mutation rate. The estimate for effective number of migrants is also independent of mutation rate, since mutation rate cancels itself out in the conversion process.

2.5.1 Morphology

Most sockeye salmon populations were established following glacial retreat on the order of 10,000 years ago (Wood 1995). Populations that breed in deep water along lake beaches consistently have greater average male body depth than populations breeding in flowing water environments (Blair et al. 1993). The differentiation of populations is variable, but for some populations the body depth as a function of body length is so different that there is little overlap between habitats (Hamon et al. 2000). This is particularly the case with access-limiting streams or inlets and high levels of predation (Quinn et al. 2001; Hamon and Foote 2005) which is not the situation with either of the habitats in this study. Also, gill net fisheries may impose selection on body depth (Hamon et al. 2000), but the only commercial fishery on sockeye in this study is a seine fishery. Our results indicate that the male sockeye in Aniakchak Caldera are deeper bodied along the lake beaches than in the riverine outlet breeding habitat, the predicted
nature of the difference being based on patterns of sockeye differentiation elsewhere.

Egg size of female sockeye salmon is also differentiated among other sockeye populations since the last glacia tion, with females that breed over larger substrate generally having larger eggs (Quinn et al. 1995). The substrate size along the breeding areas in Aniakchak Caldera is quite different between the beaches and the outlet river. The egg size of females in these locations has diverged in the manner that was expected; females breeding in larger substrate had larger eggs.

Both egg mass and body shape in salmon have genetic components (Gall and Huang 1988; Su et al. 1997; Kinnison et al. 2001; Kinnison et al. 2003; Gall and Neira 2004) but can also vary due to plastic responses. Outlet breeding salmon expend more energy after migration, which could lead to smaller eggs (Kinnison et al. 2001) and shallower bodies (Kinnison et al. 2003; Crossin et al. 2004). The difference in egg mass that we document here is in the opposite direction, while the differences in body size are consistent with energetic trade-offs. We also note that many of the environments experienced by these populations are similar. Incubating and spawning environments differ, but both populations have access to the same lake environments, migrate down the same river and the same distance to the ocean, have access to the same ocean environment, and again migrate up the same river for the same distance back to Surprise Lake. These populations have access to the same habitats, and the differences experienced are a consequence of an individual’s choice, with the
exception of incubation and emergence habitats, which are a consequence of an individual’s parent’s choice. Since we did not perform common garden rearing experiments, we cannot exclude the alternative hypothesis that phenotypic plasticity contributed to our measured morphological differences.

The observed pattern of divergence in body depth and egg size, taken together with a heritable basis for these traits established in closely related species, and the expected plastic response of egg size in the opposite direction of our measured difference, suggests some element of genetic divergence in these traits. However, the relatively small degree of divergence observed in these traits relative to other studies may have a number of causes. First, it may reflect relatively similar optimal phenotypes for the two environments in question. Second, it may result from the relatively recent divergence of these populations, and reflect that they have not yet reached the phenotypes that would be optimal for their breeding habitats. Third, it may be a plastic response not previously described that is in the opposite direction as found in Kinnison et al. (2001). Fourth, selection in this case for this trait may be relatively weak. Finally, limited divergence may result from migration and resulting gene flow between the habitats, constraining greater divergence.

2.5.2 Divergence with Migration

Our estimates obtained from the IM model indicated that divergence began recently (389-503 years; 78-100 generations ago) in a time frame that coincides with the 500 year old eruption event, and that migration has occurred since \( m = 0.003-0.006 \). Our IM migration and \( N_E \) estimates were similarly
estimated in MIGRATE. The MIGRATE 95% credibility intervals for all estimated parameters are within the bounds of the IM credibility intervals. The actual high point and mean parameter estimates are slightly higher than the IM estimates. Some differences are to be expected since MIGRATE does not have a time since divergence parameter in the model, but the overall convergence of the estimates suggest that demographic processes are more important in shaping the genetic structure than recent common ancestry.

These results allow us to reject our first scenario of two sources that diverged long ago with an unknown geographic relationship. Also, our measurement of migration since divergence suggests that migration is present in this ongoing divergence. In order to assess the relative importance of allopatry and sympatry, we compared the estimates of time of divergence onset with the average time of migration (Tables 2.4, 2.5). The time of average migration was estimated approximately midway between the estimated onset of divergence and time of sample collection in 2001-2003. This occurred in 2 of 3 runs in the comparison of high point posterior probabilities and 3 of 3 runs in the comparison of average posterior probabilities (Tables 2.4, 2.5). These results best support the scenario that migration was present between these populations for a substantial period after divergence. However because we do not estimate the distribution of actual migration events, only the average time of migration, we cannot compare the relative time periods of divergence with and without migration.
By applying genetic analysis techniques to a system with known temporal landmarks based on documented volcanic eruptions, we uncovered details of a case of very recent ecological divergence despite gene flow. This divergence began around 500 years or 100 generations before present. We measured migration that occurred since divergence. To our knowledge, this is the most recent ecological divergence ever reported in a fish species following natural colonization. In this case, it appears that this ecological divergence occurred despite migration.

2.6 Acknowledgments

We thank F. Breden, J. Hey, A. Mooers, C. Nice, P. Nosil, and three anonymous reviewers for comments on the manuscript. We thank the Evolution Discussion Group (FAB*-Lab) at Simon Fraser University and the Vancouver Evolution Group (VEG) for critical discussions of this work. We thank E. Wang, M. McBurney, J. E. Krebs, B. Girard, K. Bunney, T. Tingey, J. L. Miller, J. Marcus, R. Rumelhart and the Staff of Katmai National Park for field assistance. We thank R. Mackas for help with data entry and analysis. We thank W. D. Noon and J. L. Miller for help with graphics. We thank I. Williams, S. Graziano, and G. K. Sage for help in the lab. This work was funded by the National Park Service, the U.S. Geological Survey, Alaska EPSCoR, National Science Foundation, University of Alaska Anchorage, and Natural Sciences and Engineering Research Council operating grant to F. Breden, Simon Fraser University.
2.7 Literature Cited


2.8 Tables
Table 2.1 Male body shape and female egg mass data summary. We measured mid orbital to hypural length (MOH) along with body depth (BD) in males and egg mass in females.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Year</th>
<th>N</th>
<th>MOH (mm) (SE)</th>
<th>BD (mm) (SE)</th>
<th>N</th>
<th>MOH (mm) (SE)</th>
<th>BD (mm) (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>2001</td>
<td>51</td>
<td>514.31 (2.21)</td>
<td>194.25 (1.45)</td>
<td>50</td>
<td>507.08 (6.45)</td>
<td>185.90 (3.49)</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>53</td>
<td>500.60 (5.00)</td>
<td>191.49 (2.86)</td>
<td>50</td>
<td>495.00 (7.93)</td>
<td>186.00 (4.10)</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>48</td>
<td>479.50 (5.96)</td>
<td>180.46 (3.37)</td>
<td>49</td>
<td>489.61 (5.09)</td>
<td>180.43 (2.50)</td>
</tr>
<tr>
<td>All</td>
<td>152</td>
<td></td>
<td>498.54 (2.89)</td>
<td>188.93 (1.60)</td>
<td>149</td>
<td>497.28 (3.84)</td>
<td>184.13 (1.98)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>MOH (mm) (SE)</th>
<th>Egg mass (mg) (SE)</th>
<th>N</th>
<th>MOH (mm) (SE)</th>
<th>Egg mass (mg) (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>2006</td>
<td>30 468.83 (3.63)</td>
<td>99.16 (2.65)</td>
<td>22</td>
<td>475.36 (4.50)</td>
<td>107.38 (2.78)</td>
</tr>
</tbody>
</table>

Table 2.2 ANCOVA table for male body depth comparison. Year collected and habitat are categorical variables, mid orbital to hypural length (MOH) is included as a continuous covariate, and body depth is the dependent variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOH</td>
<td>1</td>
<td>953.5</td>
<td>0</td>
</tr>
<tr>
<td>Year</td>
<td>2</td>
<td>4.9</td>
<td>0.008</td>
</tr>
<tr>
<td>Habitat</td>
<td>1</td>
<td>11.7</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table 2.3 Estimated mutation rates based on mutation rate scalars.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Scalar Estimate Avg</th>
<th>SD</th>
<th>Calibrated Scalar per gen</th>
<th>Mutation Rates per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>One102</td>
<td>0.9092 (0.0097)</td>
<td></td>
<td>7.0046</td>
<td>0.0007</td>
</tr>
<tr>
<td>One105</td>
<td>0.1298 (0.0028)</td>
<td></td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>One108</td>
<td>4.406 (0.0812)</td>
<td></td>
<td>33.9448</td>
<td>0.00339</td>
</tr>
<tr>
<td>One109</td>
<td>0.7292 (0.0277)</td>
<td></td>
<td>5.6181</td>
<td>0.00056</td>
</tr>
<tr>
<td>One115</td>
<td>3.0129 (0.1403)</td>
<td></td>
<td>23.2116</td>
<td>0.00232</td>
</tr>
<tr>
<td>Geometric Mean</td>
<td></td>
<td></td>
<td></td>
<td>0.00079</td>
</tr>
</tbody>
</table>

Table 2.4 Mutation rate estimates based on length of repeat method. We include length of repeat and corresponding mutation rates for each marker.

<table>
<thead>
<tr>
<th>Locus</th>
<th>No. of Repeats</th>
<th>Mutation Rates per gen</th>
<th>per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>One102</td>
<td>15.22</td>
<td>0.00325</td>
<td>0.00065</td>
</tr>
<tr>
<td>One105</td>
<td>7.55</td>
<td>0.000085</td>
<td>0.000017</td>
</tr>
<tr>
<td>One108</td>
<td>18.44</td>
<td>0.00884</td>
<td>0.001768</td>
</tr>
<tr>
<td>One109</td>
<td>15.63</td>
<td>0.00374</td>
<td>0.000748</td>
</tr>
<tr>
<td>One115</td>
<td>25.69</td>
<td>0.0497</td>
<td>0.00994</td>
</tr>
<tr>
<td>Geometric means</td>
<td></td>
<td>0.003399</td>
<td>0.00068</td>
</tr>
</tbody>
</table>

Table 2.5 Comparison of the high point, average values, 95% credibility interval and 90% highest posterior density (HPD) of the posterior distribution of divergence onset dates in years before sample collection for all three runs of the IM program.

<table>
<thead>
<tr>
<th></th>
<th>High Point</th>
<th>Average</th>
<th>95% Cred. Int.</th>
<th>90% HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td>111</td>
<td>389</td>
<td>66-1725</td>
<td>35-1206</td>
</tr>
<tr>
<td>Run 2</td>
<td>47</td>
<td>503</td>
<td>54-4358</td>
<td>22-2528</td>
</tr>
<tr>
<td>Run 3</td>
<td>123</td>
<td>396</td>
<td>73-3022</td>
<td>28-1560</td>
</tr>
<tr>
<td>Mean</td>
<td>93.7</td>
<td>429.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>(40.9)</td>
<td>(63.9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.6 Comparison of the high point and average values of the posterior distribution of average migration dates in years before sample collection for all three runs of the IM program.

<table>
<thead>
<tr>
<th>Direction of migration</th>
<th>Run 1</th>
<th></th>
<th></th>
<th>Run 2</th>
<th></th>
<th></th>
<th>Run 3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>SD of five</td>
<td>Average</td>
<td>SD of five</td>
<td>Average</td>
<td>SD of five</td>
<td>Average</td>
<td>SD of five</td>
<td>Average</td>
</tr>
<tr>
<td>Outlet to beach</td>
<td>Point</td>
<td>loci</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 1</td>
<td>44</td>
<td>(3.4)</td>
<td>177</td>
<td>(1.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 2</td>
<td>120</td>
<td>(31.2)</td>
<td>276</td>
<td>(2.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 3</td>
<td>61</td>
<td>(2.8)</td>
<td>205</td>
<td>(1.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>74.9</td>
<td></td>
<td>219.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>(39.7)</td>
<td></td>
<td>(51.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beach to outlet</td>
<td>Run 1</td>
<td>42</td>
<td>(2.8)</td>
<td>185</td>
<td>(2.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 2</td>
<td>99</td>
<td>(28.1)</td>
<td>305</td>
<td>(6.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run 3</td>
<td>56</td>
<td>(3.5)</td>
<td>224</td>
<td>(3.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>66.0</td>
<td></td>
<td>238.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>(29.7)</td>
<td></td>
<td>(60.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.7 Effective migrants per generation for all three runs of the IM program. We include the high point and the average of the posterior probability distribution for both directions of migration.

<table>
<thead>
<tr>
<th>Run</th>
<th>Outlet to beach</th>
<th></th>
<th>Beach to outlet</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Point</td>
<td>Average</td>
<td>High Point</td>
<td>Average</td>
</tr>
<tr>
<td>1</td>
<td>0.16</td>
<td>16.15</td>
<td>1.23</td>
<td>12.25</td>
</tr>
<tr>
<td>2</td>
<td>9.00</td>
<td>16.67</td>
<td>4.41</td>
<td>12.71</td>
</tr>
<tr>
<td>3</td>
<td>3.83</td>
<td>18.41</td>
<td>1.01</td>
<td>8.86</td>
</tr>
</tbody>
</table>
Table 2.8 Comparison of demographic parameters obtained with IM and MIGRATE programs, including high point, mean and 95% credibility intervals for effective population size and migration rate for both populations.

<table>
<thead>
<tr>
<th>Program</th>
<th>Parameter</th>
<th>HiPoint</th>
<th>Mean</th>
<th>95Lo</th>
<th>95Hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>IM run 1</td>
<td>$N_E$ beach</td>
<td>794</td>
<td>1384</td>
<td>390</td>
<td>3889</td>
</tr>
<tr>
<td></td>
<td>$N_E$ outlet</td>
<td>528</td>
<td>961</td>
<td>311</td>
<td>3098</td>
</tr>
<tr>
<td></td>
<td>m outlet to beach</td>
<td>0.00010</td>
<td>0.00583</td>
<td>0.00018</td>
<td>0.03115</td>
</tr>
<tr>
<td></td>
<td>m beach to outlet</td>
<td>0.00024</td>
<td>0.00638</td>
<td>0.00045</td>
<td>0.03506</td>
</tr>
<tr>
<td>MIGRATE</td>
<td>$N_E$ beach</td>
<td>1979</td>
<td>2208</td>
<td>1417</td>
<td>2458</td>
</tr>
<tr>
<td></td>
<td>$N_E$ outlet</td>
<td>1062</td>
<td>2147</td>
<td>417</td>
<td>1875</td>
</tr>
<tr>
<td></td>
<td>m outlet to beach</td>
<td>0.00350</td>
<td>0.00335</td>
<td>0.00190</td>
<td>0.00439</td>
</tr>
<tr>
<td></td>
<td>m beach to outlet</td>
<td>0.00622</td>
<td>0.00625</td>
<td>0.00486</td>
<td>0.00747</td>
</tr>
</tbody>
</table>
2.9 Figures
Figure 2.1 Aniakchak National Monument and Preserve (ANMP) showing Aniakchak Caldera and the two study populations.
Figure 2.2 Size distribution of substrate for the two breeding locations.
Figure 2.3 Comparison of body depth in beach and outlet males. We standardized body depth by fitting regressions of body depth as a function of midorbital to hypural length for each population for each year. Then, we calculated the residual of body depth for each fish from the appropriate regression, and we used that residual to calculate a standardized body depth at the average size. Beach males show consistently deeper bodies each year, though the difference is smaller in 2002.
Figure 2.4 Average and standard error of standardized egg weight for beach and outlet female sockeye salmon from Surprise Lake. We took the residuals from regressions of egg weight against midorbital to hypural length.
Figure 2.5 Posterior probability distributions for time since divergence in years. All three runs of the IM program are illustrated; run 1 is black dashed, run 2 is black, run 3 is gray. The inset is an enlargement of years 1–1000.
Figure 2.6 Posterior probability of migration rates between populations. The composite figure of all three runs of the migration rate posterior probability distributions shows the beach to outlet migration as a gray line on a black field. The line is the average probability for the three runs, and the black field is the standard error field. The outlet to beach average probability is the thick white line, and the thin white lines depict the standard error space around the line.
Figure 2.7 Posterior probability of effective population sizes of beach, outlet, and ancestral populations. Solid lines are the average probability for all three runs of the IM program. The fields represent the standard error of the runs for each parameter. The inset shows detail.
CHAPTER 3: CONTRASTING ECOLOGY SHAPES JUVENILE LAKE-TYPE AND RIVERINE SOCKEYE SALMON

In press at the journal Transactions of the American Fisheries Society.

Author list: Scott A. Pavey, Jennifer L. Nielsen, Renae H. Mackas, Troy R. Hamon, and Felix Breden.

Author contribution: Scott A. Pavey planned the study, conducted the fieldwork, ran the analysis, and was the primary writer of the paper.

3.1 Abstract

Here we compare the body shape of juvenile sockeye salmon that rear in lakes (lake-type) with those that rear in rivers (riverine) and we relate rearing habitat to morphology and ecology. These habitats present different swimming challenges with respect to water flow, foraging strategy, habitat complexity, and predation level. We present morphological data from three riverine and three lake-type populations in southwest Alaska, and multivariate analyses conducted using geometric morphometrics to determine population and habitat specific body shape. As predicted, riverine sockeye salmon have a more robust body shape whereas lake-type sockeye salmon have a more streamlined body shape ($P < 0.001$). In particular, we found differences in caudal peduncle depth (riverine deeper), eye size (riverine larger) and overall body depth (riverine deeper). One lake-type population did not follow the predicted pattern and exhibited an overall
exaggerated riverine body shape. Differences between the habitats in predation, complexity, and foraging ecology are likely drivers of these differences. Allometry differed between life history types, suggesting habitat specific developmental differences.

**Keywords:** river-type, sea-type, sea/river-type, burst swimming, sustained swimming, unsteady swimming, steady swimming.

### 3.2 Introduction

Sockeye salmon (*Oncorhynchus nerka*) exhibit a high degree of morphological diversity and are locally adapted to their spawning habitats at a fine spatial scale (Blair et al. 1993; Hendry and Quinn 1997; Taylor et al. 1997; Pon et al. 2007). Flow conditions, predation, migration difficulty, incubation temperature and substrate size are some of the environmental factors that correlate with morphological or behavioral differences of sockeye salmon. Pon et al. (2007) found that juveniles emerging in lake outlets have body morphology and swimming ability that facilitates their migration against river current to arrive at their nursery lake. This is just one example demonstrating how morphology and behavior can vary at a fine spatial scale.

However, nearly all work documenting morphological divergence in sockeye salmon has focused on the adult stage and exclusively on a single life history type in which juveniles rear in lakes (lake-type). Sea-type (Semko 1954) and river-type sockeye salmon (Gilbert 1918) rear in riverine habitat and we refer to both as “riverine” (Beacham et al. 2004). In these life histories, migration to the
ocean occurs before the first winter (sea-type), or after one or more winters in freshwater (river-type). These populations rear in a variety of riverine habitats including slack water side channels, sloughs, or riffles (Murphy et al. 1989; Wood et al. 1994; Pavey et al. 2007; Wood et al. 2008). Though present riverine populations tend to be of lesser abundance than lake-type, during previous glaciations, they were likely widespread and abundant, and appear to have played an important role in recolonizing new watersheds as glaciers receded (Wood et al. 2008). Recent emphasis on population genetics of sockeye salmon with these alternative life histories has helped to clarify their relationship to lake-type sockeye salmon (Wood 1995; Beacham et al. 2004; Beacham et al. 2006a; Beacham et al. 2006b; Wood et al. 2008; McPhee et al. 2009). The recurrent and rapid evolution of sockeye salmon ecotypes in riverine and lake habitats, which differ in flow regime and available food items, presents an ideal opportunity to analyze correlations between body shape and ecology.

In this study, we compare juvenile body shape between lake-type and riverine sockeye salmon populations. We expect lake-type to have a streamlined body to facilitate continuous swimming and riverine to have a deep, robust body for burst swimming. We explore the ecological factors of water current, foraging, habitat complexity, and predator avoidance in shaping sockeye salmon morphology.
3.3 Methods

3.3.1 Study Area

Katmai National Park and Preserve and Aniakchak National Monument and Preserve (Fig. 3.1) in southwest Alaska provide a unique system to study the interactions of these sockeye salmon life history forms. There are sockeye salmon populations rearing in both lake and river environments, including in the Aniakchak River where both life history types coexist in the same system (Pavey et al. 2007). Our lake-type locations are Surprise Lake (SL), Upper Q-Tip Lake (UQT), and Lower Kaflia Lake (KAF). Our riverine locations are Albert Johnson Creek (AJC), Swikshak River (SWI) and Kamishak River (KMS). These riverine habitats appear to differ from those described in other studies (Wood et al. 1987; Murphy et al. 1989; Wood et al. 1994); our sampling locations are characterized by clear water, substantial current, and little or no glacial influence. All of our sample locations drain into the Gulf of Alaska in the northern Pacific Ocean, with the exception of UQT, which drains into Bristol Bay of the Bering Sea.

3.3.2 Methods

We sampled a total of 1000 juvenile sockeye salmon from the six locations in the summers of 2003, 2006, and 2007 (Table 3.1). We euthanized each fish with MS-222 and then photographed it in a standard position with a 35mm manual camera, a macro lens and a table stand. Slides were scanned with a Nikon Super Coolscan 5000 ED. We digitized 12 landmarks on each image using TpsDig (Fig. 3.2, Rohlf 2006). These landmarks represent homologous points that could be identified on every fish, and were selected to appropriately capture
the profile of each fish. Landmark 1 is the most anterior point on the dentary bone. Landmarks 2 and 3 represent the width of the orbit, and were taken at the anterior and posterior extremes of the orbit, respectively. Landmarks 4 and 5 together were included to capture the body depth in the region of the fish directly posterior to the head. Landmark 4 indicates the point along the body directly below the pectoral fin insertion point. Landmark 5 is the point on the dorsal side of the body directly above the most posterior point of the gill operculum. Landmark 6 is the anterior-most point of the dorsal fin along the body. Landmarks 7 and 8 indicate the point along the ventral side of the body directly below the pelvic and anal fin insertion points, respectively. Landmarks 9 and 10 together represent the caudal peduncle, and were defined as the anterior-most attachment points of the caudal fin onto the dorsal and ventral sides of the body. Landmark 11 lies at the hypural plate, and landmark 12 represents the center of the fork in the caudal fin.

Using the digitized landmarks, we analyzed shape variation using geometric morphometrics (Rohlf and Marcus 1993; Zelditch et al. 2004; Langerhans et al. 2007). We used TpsRelw (Rohlf 2007) to perform generalized Procrustes analysis, which translates, scales and rotates landmark configurations to remove information unrelated to shape and generates shape variables for each fish (partial warps and uniform components). Although this method does effectively remove all isometric effects of size on shape, allometric relationships remain. We calculated centroid size, which is the square root of the sum of the squared distances from each landmark to the arithmetic center. This is a more
complete proxy of size than fish length. We visualized shape differences between pairs of populations that had similar distributions of centroid size with TpsRegr, which accepts centroid size as a covariate so shape that differs only by habitat is visualized independently of size differences. We then performed a multivariate nested analysis of covariance (MANCOVA) to test for the significance of differences in profile of the body shape among populations and habitats (Langerhans et al. 2007). The partial warp scores, including the uniform components, are the dependent variables and centroid size, habitat (lake-type vs. riverine) and population nested in habitat are the independent variables. The inclusion of the centroid size variable captures allometric aspects of shape. The habitat variable captures shape differences due to habitat. And the inclusion of population nested within habitat captures shape differences among populations within habitats. We ran this analysis with and without the interaction term centroid size x habitat to determine whether allometry differs among habitats. We also ran the analysis with and without the inclusion of the UQT population, which was substantially different in body shape than the other lake-type populations. We then correlated each specimen’s habitat canonical score with the superimposed landmarks (Langerhans et al. 2007). This illustrates the extent to which each landmark differs between habitats. We performed the MANCOVA and canonical analysis in JMP 7.0.2 (SAS institute). We correlated the canonical score for each specimen with the superimposed landmark coordinates in SYSTAT 10.

We used a discriminant function analysis to examine whether the morphological information contained in the partial warp scores could be used to
distinguish riverine and lake-type populations. The analysis was conducted seven times, once with all populations included, and once after removal of the individuals from each population in turn. The percent of successful classifications of individuals to their originating habitat under jackknife procedures is reported for all populations as well as for the five populations remaining after removal of UQT. In addition, the percentage of classification of each population into the appropriate habitat types is reported without jackknifing. We also plotted the results of the discriminant function analysis against centroid size including all six populations as well as excluding UQT. The discriminant function analysis was performed in SYSTAT 10.

3.4 Results

We sampled a total of 1000 fish, ranging from 74 to 296 per population, at six locations (Table 3.1). In lakes, fish were captured in shallow beach habitat with a varied substrate. Riverine habitats sampled were clear, shallow, and flowing. We excluded two outliers and also all age 1 fish by creating fork length frequency distributions and eliminating individuals from the larger node if present. We used the resulting 798 age 0 individuals in subsequent analyses.

Two of the three lake-type compared with all three riverine populations followed the general pattern described from studies of other salmon species, with lake-type populations exhibiting a shallow body and riverine populations a deep body (Fig. 3.3). The striking exception was the lake-type UQT population, which was characterized by the deepest bodies of all populations. We found significant differences in body shape among populations and between habitats, both
including and excluding UQT, as indicated by the significant habitat term (P < 0.001; Table 3.2). Body shape was also different within habitats among populations, as indicated by the significant population(habitat) term (P < 0.001). The interaction term habitat × centroid size was significant (P < 0.001), so we retained this term in the models. However, retaining or excluding the interaction term yielded highly correlated canonical scores (P < 0.001, $r^2 = 0.97$). This was also true excluding UQT (P < 0.001, $r^2 = 0.99$). This means that shape did vary with size differently among habitats, but the overall effect was very small.

The correlations between canonical scores of the habitat effect and the superimposed landmarks changed substantially depending on the inclusion or exclusion of UQT (Table 3.3). These results depict which part of the fish shape changed based solely on habitat and how including and excluding UQT affected the life history type shape. Pearson’s correlation coefficients that relate to overall body depth, caudal peduncle depth, as well as orbital size were statistically significant (bold numbers in Table 3.3). The lake-type body shape is more fusiform than riverine in both cases, though this difference is more pronounced when UQT is excluded (Table 3.3; Fig. 3.3). The caudal peduncle also differs between habitats, shallow and long for lake-type and deep and short for riverine (Table 3.3; Fig. 3.3). Orbital size differed between life histories, with riverine sockeye salmon having a larger orbital region. We graphically compared lake-type UQT with the other populations that had individuals most closely matching in size, which included riverine KMS and AJC, as well as lake-type KAF (Table 3.1). Upper Q-Tip Lake sockeye salmon did not fit with the lake-type trend. This
population appears to have exaggerated riverine body depth and orbital size differences, but the caudal peduncle differences are more complex. Upper Q-Tip Lake sockeye salmon retain the caudal peduncle geometric shearing feature of the other lake-type populations but have similar caudal peduncle length and depth of riverine populations. Shearing is “translating landmarks along one axis by a distance proportional to their location along the other axis” (Zelditch et al. 2004). The shearing feature of the caudal peduncle is illustrated by comparing the lake-type and riverine deformation grids in Fig. 3.3. This shearing is more extreme with the removal of UQT (Fig. 3.3). In summary, riverine juveniles generally have deeper bodies, shorter and deeper caudal peduncles and larger orbits than lake-type sockeye salmon. However some of the characteristics of lake-type UQT appear to be exaggerated riverine.

In the discriminant function analysis, individuals were assigned to the proper habitat type with accuracy of 70% for lake type and 77% for riverine under jackknifing procedures (Table 3.4; Fig. 3.4). Removal of UQT resulted in jackknifed classifications of 80% for lake-type and 89% for riverine. Of all the analyses run with single populations removed, UQT was the only removal that improved assignment accuracy in all remaining populations as well as providing the best overall assignment accuracy. Assignment of individuals from the SL population was least accurately assigned in both datasets, but removal of UQT improved the classification accuracy of this population from 58% to 72% (Table 3.4; Fig. 3.4). The DFA makes use of the information that best allows separation of populations into habitats which makes for the similarities in the two sets of
 panels. Excluding UQT results in the addition of three of the body depth landmarks (and the loss of one) as significant contributors to the differentiation among habitats (Fig. 3.3). This change is largely responsible for the improved ability of the DFA to separate by habitat in the second set of panels with UQT excluded.

3.5 Discussion

We found that riverine juveniles were deeper bodied and had a deeper but shorter caudal peduncle and larger orbit, whereas lake-type juveniles were shallower bodied with a shallow, yet longer caudal peduncle and smaller orbit. This is an important initial step in understanding how morphology relates to the ecology of these life history types. These results are similar to other studies of juvenile salmon within and among species (Hoar 1958; Scott and Crossman 1973; Swain and Holtby 1989). However this pattern was not consistent between all population pairs; in particular UQT did not follow this pattern. For the purpose of this discussion, we assume that the analysis excluding UQT represents the general pattern, and UQT is atypical. We realize that the only way to verify this claim is a larger study with more populations from different places. To address our findings, we explore water current, foraging strategies, predator interactions, and habitat complexity as ecological factors influencing trends in body shape on three taxonomic levels: all fishes, among salmon species, and the sockeye populations reported here.
3.5.1 Ecology and Fish Shape: All Fishes

Fish body shape often represents a compromise between steady (continuous or sustained) and unsteady (burst and maneuverability) swimming. Langerhans (2008) developed a model relating morphology and swimming performance to flow regime and found that flowing water was correlated with a streamlined body shape in 42 of 58 intraspecific comparisons and 13 of 17 interspecific comparisons. The premise of the model is that fish in flowing water must swim continuously to maintain their position and this necessitates morphological and physiological adaptations for continuous swimming.

In addition to flow regime, predation and habitat complexity may also be important ecological aspects that affect morphology. Langerhans and Reznick (2010) found that in four divergent fish species, populations subject to high predation had larger caudal regions and superior burst swimming than populations with less predation. They also found variation in body shape across 32 species relating to habitat complexity, where species in open environments had higher endurance but lower ability for turning radius and acceleration than species in complex habitats (Domenici 2003; Langerhans and Reznick 2010).

3.5.2 Ecology and Fish Shape: Oncorhynchus sp.

In a simplistic assessment of the hydrodynamic differences and in the absence of other ecological and behavioral factors, we may expect riverine rearing of Pacific salmon species to have a more fusiform body shape than lake rearing salmon species (Langerhans 2008). However the general observation is the opposite morphological pattern in juveniles of Oncorhynchus (Hoar 1958;
Scott and Crossman 1973; Swain and Holtby 1989). Habitat specific foraging strategies may result in riverine species avoiding strong currents by holding positions behind rocks, close to the bottom or in side vegetation (Bisson et al. 1988). Foraging strategy, habitat complexity and predation regimes may influence body shape in juvenile salmon more than current in salmon (Swain and Holtby 1989).

In Pacific salmon juveniles, foraging behavior may be a more important factor in body shape than water velocity. Salmon species with extensive river residence tend to have a deep robust shape and exhibit agonistic behavior, whereas species inhabiting lakes tend to be more shallow bodied, more streamlined and less agonistic (Hoar 1951, 1954, 1958; Scott and Crossman 1973; Taylor and McPhail 1985a; Taylor and Larkin 1986). In salmon, a deep robust shape is thought to favor burst swimming (Pon et al. 2007; Langerhans 2008). Riverine populations hold and defend territories, and this robust shape may also increase the effectiveness of aggressive displays. A shallower, more streamlined body is better suited for lacustrine habitats, where the strategy is cruising in schools and feeding on zooplankton in the open water instead of territoriality with associated burst swimming. This dichotomy has also been observed intraspecifically among populations of coho salmon (O. kisutch) and Chinook salmon (O. tshawytscha) with different rearing environments (Taylor and McPhail 1985a; Taylor and Larkin 1986). Swain and Holtby (1989) found that riverine foraging coho salmon juveniles were more agonistic and had deeper, more robust bodies than those of an adjacent lake rearing population.
Other ecological factors such as predation and habitat complexity may also influence body shape in salmon (Swain and Holtby 1989; Langerhans and Reznick 2010). Rivers present greater structural complexity than the open water column of lakes and this is expected to favor a deeper and more maneuverable body shape (Langerhans and Reznick 2010). Common garden rearing experiments have confirmed that these morphological and behavioral differences have both genetic and environmental components (Taylor and McPhail 1985a, b; Rosenau and McPhail 1987; Pakkasmaa and Piironen 2001; Pon et al. 2007). In this study, we do not know the relative roles of genetics and the environment on body shape. Future research should focus on clarifying these roles through common garden and quantitative genetics experiments.

### 3.5.3 Ecology and Fish Shape: *Oncorhynchus nerka*

Like other lake rearing Pacific salmon, lake-type juvenile sockeye salmon feed on plankton in schools and are nonaggressive toward each other (Hoar 1954; Swain and Holtby 1989). This foraging strategy favors continuous swimming ability over burst swimming in salmon (Foerster 1968; Hartman and Burgner 1972; Eggers 1982; Burgner 1991). The general shape differences between lake-type and riverine life history forms of sockeye salmon found in this study are similar to the differences among species of *Oncorhynchus*. Superficially, riverine sockeye salmon rear in flowing water and lake-type sockeye salmon rear in still water. However, previous studies of riverine sockeye salmon habitat were found to have very low flow and turbid riverine habitats (Wood et al. 1987; Murphy et al. 1989). The riverine habitats where we captured...
sockeye in this study are neither low current (slack water) areas nor turbid, although we did not measure these parameters. If riverine sockeye salmon have a similar foraging strategy as other river rearing Pacific salmon, including aggressively defending territories near the substrate, this may have resulted in plastic or adaptive responses favoring burst swimming capability (Swain and Holtby 1989).

Predator interactions are likely an important element in sockeye salmon juvenile body shape. Sockeye salmon rearing in Alaskan lakes have a largely nonoverlapping distribution with potential predator species and therefore have low levels of predation (Roos 1959; Foerster 1968; Burgner 1991). Predation on lake-type sockeye salmon occurs mainly during the migration to the nursery lake and the seaward migration to lake outlets and in rivers (Burgner 1991). On the other hand, riverine sockeye salmon spend this rearing time in close proximity with predators including Dolly Varden (Salvelinus malma), lake trout (S. namaycush), and coho salmon. The presence of these predator species should favor burst swimming and a deep caudal region (Domenici et al. 2008; Langerhans and Reznick 2010). Gape-limited predation could also favor deeper body morphology (Nilsson and Bronmark 2000). In studies of riverine sockeye in British Columbia (Wood et al. 1987; Murphy et al. 1989), turbid water may limit the effects of predation, whereas the riverine study sites here were less glacially influenced and visually quite clear. In both foraging and predation, the riverine ecology favors burst swimming and a deep body shape.
Habitats of higher structural complexity like riverine rearing habitat are also expected to favor unsteady swimming and a deeper morphology due to maneuvering, breaking, and accelerating around structures (Langerhans and Reznick 2010). Since this attribute occurs also in the presence of higher predation and territorial foraging strategy, we cannot assess the relative contributions of these ecological aspects on sockeye salmon juvenile morphology.

Sockeye salmon from UQT were shaped differently than the other two lake-type populations. They have deeper bodies and larger orbits than any other population, including the riverine populations. Their caudal peduncle included shape aspects of both life histories. The basis for these differences is undetermined, and there are many nonexclusive potential explanations. It is possible that there are multiple ecotypes within this lake such as benthic/limnetic pairs found in other lake dwelling salmonids (Chouinard and Bernatchez 1998; Gislason et al. 1999). It is possible that sampling by beach seine preferentially sampled a deep bodied benthic ecotype at this location. This is the only population that was sampled in July; all others were sampled in the month of June. Although the individuals in UQT were of similar length as populations collected in June (Table 3.1), fish condition that changes as the summer progresses could explain some body depth differences.

This population is geographically separated from the others; UQT is the only sampling location that drains into the Bering Sea. There may be unique foraging strategies and/or predation regimes that may differ among lake-type
populations. Northern pike (*Esox lucius*) are present in lakes in the Bering Sea drainages but are not present in the north Pacific drainages. Dietary differences and alternative predatory regimes may force unique ecotypic response in these juvenile sockeye salmon. It is also possible that there is no general pattern in lake-type sockeye salmon, and that the variability within lake-type juveniles exceeds the variability between lake-type and riverine life histories. However, even with UQT included in the discriminate function analysis, the life history groups separate (Fig. 3.4). The separation is more distinct when UQT is removed (Fig 3.4). Also, removing UQT from the discriminate function analysis was the only removal that resulted in increased correct assignment for all other populations. Larger studies and more populations and geographic areas will clarify the generality of our findings.

In sockeye salmon, lake-type spawning adults have deeper bodies in still water habitats (lakes) compared to flowing habitats (lake outlets and inlets; (Quinn et al. 2001b; Pavey et al. 2010). These studies only involve lake-type life history and there are currently no studies comparing riverine spawning adults to lake-type. Spawning adults are large compared with rearing juveniles and probably cannot hide as effectively from water current, so the hydrodynamic advantage of a shallow body may be more important in adults (Pavey et al. 2010). Breeding adult salmon stop feeding upon re-entry into freshwater, so foraging energetics should not be a factor in adult body shape. However, selection in breeding populations has been shown to include effects of access limitation (Quinn and Buck 2001), sexual selection (Quinn and Foote 1994;
Hamon and Foote 2005) and predation (Quinn and Foote 1994; Quinn and Buck 2001; Quinn et al. 2001a; Quinn et al. 2001b; Hamon and Foote 2005). The difference in absolute size of individuals in relation to the depth of the water environment, as well as the changed focus from feeding to reproductive activities, appears to favor different patterns of phenotypic differentiation in adult and juvenile sockeye salmon.

The interaction term (csize x habitat) in the MANCOVA was significant and this indicates habitat specific allometric differences (Table 3.2, Fig. 3.4). Though the results of the other model covariates including and excluding this term were highly correlated, this suggests possible developmental differences among habitats and populations. The significant population nested within habitat term suggests population specific body shape characteristics (Table 3.2; Fig. 3.4). In fact, with UQT excluded, the population nested within habitat term has an equivalent effect on morphology as the habitat term (Table 3.2). Differing development within habitats as well as population specific ecological factors may contribute to fine tuning body shape for each population.

Our results indicate that juvenile sockeye salmon differ in body shape among populations rearing in different habitats. The specific differences may be affected by predation, plasticity, or other factors, but in our samples these differences seem to favor deep bodies in moving water rearing environments, likely as a result of the foraging environment. The generality of this pattern, or the frequency of “aberrant” populations like UQT, will only be known with the examination of more populations in different areas and varying ecological
contexts. We do not know to what extent genetic and environmental factors are influencing body shape in sockeye salmon juveniles. In many cases, phenotypic plasticity itself is adaptive; however this could also result in apparent phenotypic mismatches in individual populations. In addition, we cannot exclude nonadaptive genetic processes, such as drift or mutation resulting in morphological differences (Lande 1976). The difference in phenotypic pattern between adult and juvenile sockeye using flowing and still water environments likely relates to the shift in emphasis from foraging in juveniles to breeding activity in adults, as well as the difference in size of the animal in relation to its immediate environment.

3.6 Acknowledgements

We thank Drs. R. Beamish, J. Reynolds, B. Langerhans, F. Utter, C. Wood, and four anonymous reviewers for comments on the manuscript. We thank the Evolution Discussion Group (FAB*-Lab) at Simon Fraser University and the Vancouver Evolution Group (VEG) for critical discussions of this work. We thank E. Wang, M. McBurney, J.E. Krebs, B. Girard, K. Bunney, T. Tingey, J.L. Miller, J. Marcus, R. Rumelhart and the Staff of Katmai National Park for field assistance. We thank B. Langerhans and H. Alexander for help with analysis. We thank W.D. Noon and R. Wood for help with graphics. This work was funded by the National Park Service, the U.S. Geological Survey, Alaska EPSCoR, National Science Foundation, University of Alaska Anchorage, and Natural Sciences and Engineering Research Council operating grant to F. Breden, Simon Fraser
University. Use of any trade names or products is for descriptive purposes only and does not imply endorsement of the U. S. Government.

3.7 Literature Cited


3.8 Tables
Table 3.1 Sockeye salmon sampling locations, dates sampled, sample size (N), fork length population mean and standard deviation (SD), rearing habitat, and drainage ocean.

<table>
<thead>
<tr>
<th>Population</th>
<th>Date sampled</th>
<th>N</th>
<th>Fork length (mm) SD</th>
<th>Habitat</th>
<th>Drainage ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert Johnson Creek (AJC)</td>
<td>7 June 2003</td>
<td>296</td>
<td>36.1 4.9</td>
<td>River</td>
<td>Pacific</td>
</tr>
<tr>
<td>Kamishak River (KMS)</td>
<td>1-6 June 2007</td>
<td>133</td>
<td>37.0 2.8</td>
<td>River</td>
<td>Pacific</td>
</tr>
<tr>
<td>Swikshak River (SWI)</td>
<td>16-20 June 2006</td>
<td>91</td>
<td>30.0 3.1</td>
<td>River</td>
<td>Pacific</td>
</tr>
<tr>
<td>Lower Kaflia Lake (KAF)</td>
<td>6 June 2006</td>
<td>97</td>
<td>39.0 4.4</td>
<td>Lake</td>
<td>Pacific</td>
</tr>
<tr>
<td>Surprise Lake (SL)</td>
<td>3-4 June 2003</td>
<td>74</td>
<td>28.2 1.5</td>
<td>Lake</td>
<td>Pacific</td>
</tr>
<tr>
<td>Upper Q-Tip (UQT)</td>
<td>8-11 July 2006</td>
<td>108</td>
<td>36.3 4.9</td>
<td>Lake</td>
<td>Bering</td>
</tr>
</tbody>
</table>
Table 3.2 Results of the nested MANCOVA for all six populations of sockeye salmon and also five populations excluding Upper Q-Tip Lake (UQT). The dependent variables were the multivariate partial warp scores. We included all covariates including the interaction term, as well as habitats nested in populations. Variables in the model are centroid size (csize), habitat, populations nested within habitat (pop (habitat)), and the interaction term centroid size (csize) * habitat. Body shape changed in relation to all covariates and the significant interaction term means shape changed with size differently between habitats. Factor degrees of freedom are given before the comma and error degrees of freedom follow the comma.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Including UQT</th>
<th>Excluding UQT</th>
</tr>
</thead>
<tbody>
<tr>
<td>csize</td>
<td>df 19,773</td>
<td>19,666</td>
</tr>
<tr>
<td></td>
<td>F 166.61</td>
<td>100.49</td>
</tr>
<tr>
<td></td>
<td>P &lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>habitat</td>
<td>df 19,773</td>
<td>19,666</td>
</tr>
<tr>
<td></td>
<td>F 15.80</td>
<td>28.96</td>
</tr>
<tr>
<td></td>
<td>P &lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>pop (habitat)</td>
<td>df 76,3047.3</td>
<td>57,1986.6</td>
</tr>
<tr>
<td></td>
<td>F 40.78</td>
<td>33.11</td>
</tr>
<tr>
<td></td>
<td>P &lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>csize * habitat</td>
<td>df 19,773</td>
<td>19,666</td>
</tr>
<tr>
<td></td>
<td>F 9.27</td>
<td>6.70</td>
</tr>
<tr>
<td></td>
<td>P &lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Table 3.3 Pearson’s correlation coefficients between superimposed landmark coordinates and the canonical axis; larger numbers indicate important landmark-axes (see Fig. 3.2) in the morphological comparison of riverine and lake-type sockeye. Bold values are significant after Bonferroni correction. Results including and excluding Upper Q-Tip Lake (UQT) are presented.

<table>
<thead>
<tr>
<th>Landmark</th>
<th>Including UQT</th>
<th>Excluding UQT</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>0.2918</td>
<td>0.1296</td>
</tr>
<tr>
<td>Y1</td>
<td>0.2468</td>
<td>0.5813</td>
</tr>
<tr>
<td>X2</td>
<td>0.0053</td>
<td>0.1746</td>
</tr>
<tr>
<td>Y2</td>
<td>-0.4094</td>
<td>-0.0958</td>
</tr>
<tr>
<td>X3</td>
<td>-0.0559</td>
<td>-0.0349</td>
</tr>
<tr>
<td>Y3</td>
<td>-0.4290</td>
<td>-0.2059</td>
</tr>
<tr>
<td>X4</td>
<td>-0.5177</td>
<td>-0.5111</td>
</tr>
<tr>
<td>Y4</td>
<td>0.2293</td>
<td>-0.4198</td>
</tr>
<tr>
<td>X5</td>
<td>-0.1883</td>
<td>-0.0217</td>
</tr>
<tr>
<td>Y5</td>
<td>0.0686</td>
<td>0.5135</td>
</tr>
<tr>
<td>X6</td>
<td>0.0671</td>
<td>0.1272</td>
</tr>
<tr>
<td>Y6</td>
<td>0.0133</td>
<td>0.5085</td>
</tr>
<tr>
<td>X7</td>
<td>0.1285</td>
<td>0.2104</td>
</tr>
<tr>
<td>Y7</td>
<td>0.2089</td>
<td>-0.3364</td>
</tr>
<tr>
<td>X8</td>
<td>0.2385</td>
<td>0.3374</td>
</tr>
<tr>
<td>Y8</td>
<td>0.2300</td>
<td>-0.2336</td>
</tr>
<tr>
<td>X9</td>
<td>-0.0154</td>
<td>-0.0951</td>
</tr>
<tr>
<td>Y9</td>
<td>0.0252</td>
<td>0.4362</td>
</tr>
<tr>
<td>X10</td>
<td>-0.1778</td>
<td>-0.0693</td>
</tr>
<tr>
<td>Y10</td>
<td>-0.2101</td>
<td>-0.1150</td>
</tr>
<tr>
<td>X11</td>
<td>0.1095</td>
<td>-0.0677</td>
</tr>
<tr>
<td>Y11</td>
<td>-0.1460</td>
<td>0.1894</td>
</tr>
<tr>
<td>X12</td>
<td>-0.0214</td>
<td>-0.3132</td>
</tr>
<tr>
<td>Y12</td>
<td>-0.1636</td>
<td>0.0048</td>
</tr>
</tbody>
</table>
Table 3.4 Results of the discriminant function analysis with numbers of observed correct and incorrect classification into habitat of each population based on morphology. We present results both including and excluding Upper Q-Tip Lake (UQT), as well as the percentage difference in successful assignment. Bold numbers are the correctly assigned fish of each population to the habitat type.

<table>
<thead>
<tr>
<th></th>
<th>Riverine</th>
<th></th>
<th>Lake-type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AJC</td>
<td>KMS</td>
<td>SWI</td>
<td>SL</td>
</tr>
<tr>
<td><strong>Including UQT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lake</td>
<td>53</td>
<td>24</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>riverine</td>
<td>243</td>
<td>109</td>
<td>59</td>
<td>31</td>
</tr>
<tr>
<td>% correct</td>
<td>82%</td>
<td>82%</td>
<td>65%</td>
<td>58%</td>
</tr>
<tr>
<td><strong>Excluding UQT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lake</td>
<td>25</td>
<td>19</td>
<td>7</td>
<td>53</td>
</tr>
<tr>
<td>riverine</td>
<td>271</td>
<td>114</td>
<td>84</td>
<td>21</td>
</tr>
<tr>
<td>% correct</td>
<td>92%</td>
<td>86%</td>
<td>92%</td>
<td>72%</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9%</td>
<td>4%</td>
<td>27%</td>
<td>14%</td>
</tr>
</tbody>
</table>
3.9 Figures
Figure 3.1 Map of study area showing Katmai National Park and Preserve, Aniakchak National Monument and Preserve, the Bering Sea, the Pacific Ocean, and the six sampling locations of Albert Johnson Creek (AJC), Kamishak River (KMS), Swikshak River (SWI), Lower Kaflia Lake (KAF), Surprise Lake (SL), and Upper Q-Tip Lake (UQT). In the first three locations, we captured riverine sockeye salmon and in the latter three locations we captured lake-type sockeye salmon.
Figure 3.2 The twelve landmarks used for morphometrics analysis. These landmarks were placed on the digital image of each juvenile sockeye salmon.
Figure 3.3 Deformation grid based on the habitat variable of lake-type and riverine sockeye salmon both with and without Upper Q-Tip Lake (UQT). Results were generated with tspRegr and are exaggerated 3 times the range of the sample to allow visual comparison of how the body shape is different between lake-type and riverine sockeye salmon. Riverine sockeye have a deeper body, deeper caudal peduncle, and larger orbit. The differences are more pronounced when the UQT population is removed because lake-type UQT had riverine morphological features, including a deep body.
Figure 3.4 Habitat canonical score plotted against centroid size. The left column of panels includes all six populations of sockeye salmon; three of which are riverine: Swikshak (SWI), Kamishak (KMS) and Albert Johnson Creek (AJC), and three of which are lake-type: Kafila Lake (KAF), Surprise Lake (SL) and Upper Q-Tip Lake (UQT). The right column is the same analysis excluding UQT. The top panels depict 67\% sample ovals; clear ovals are riverine populations and gray ovals are lake-type populations. The lower panels show the regression line and centroid circle for each population. The light and dark gray ovals in the lower panels encompass the riverine and lake-type populations respectively. Habitat separation is greater when the UQT population is removed because lake-type UQT had riverine morphological features, including a deep body.
CHAPTER 4: THE ROLE OF GENE EXPRESSION IN ECOLOGICAL SPECIATION


Author list: Scott A. Pavey, Hélène Collin, Patrik Nosil, and Sean Rogers.

Author contribution: Scott A. Pavey, along with the co-authors, was an active participant in the writing process, including drafting the original manuscript and revising the manuscript after comments.

4.1 Abstract

Ecological speciation is the process by which barriers to gene flow between populations evolve due to adaptive divergence via natural selection. A relatively unexplored area in ecological speciation is the role of gene expression. Gene expression may be associated with ecologically important phenotypes not evident from morphology and play a role during colonization of new environments. Here we review two potential roles of gene expression in ecological speciation: (1) its indirect role in facilitating population persistence, and (2) its direct role in contributing to genetically based reproductive isolation. We find indirect evidence that gene expression facilitates population persistence, but direct tests are lacking. We also find clear examples of gene expression having effects on phenotypic traits and adaptive genetic divergence, but links to the evolution of reproductive isolation itself remain indirect. Gene expression during adaptive divergence often seems to involve complex genetic architectures.
controlled by gene networks, regulatory regions, and ‘eQTL hotspots’. Nonetheless, we review how approaches for isolating the functional mutations contributing to adaptive divergence are proving to be successful. The study of gene expression has promise for increasing our understanding of ecological speciation, particularly when integrative approaches are applied.

**Keywords:** adaptive divergence, eQTL, microarray, phenotypic plasticity, population persistence, qPCR, reproductive isolation, speciation, *cis*-regulatory mutation, gene expression

### 4.2 Ecological Speciation

Natural selection is a central mechanism of evolutionary change within species. But to what extent is selection also responsible for the formation of new species (i.e. speciation)? Recent years have seen renewed efforts to address this question. Under one scenario, populations living in different ecological environments undergo adaptive genetic differentiation via divergent natural selection, and these same adaptive changes also result in the populations ceasing to exchange genes. Consistent with past work, we define this process of ‘ecological speciation’ as one in which barriers to genetic exchange evolve between populations as a result of ecologically based divergent natural selection (Funk 1998; Schluter 2000, 2001; Rundle and Nosil 2005; Schluter and Conte 2009). Ecological speciation generally occurs because phenotypic traits under divergent selection, or those genetically correlated with them, incidentally affect reproductive isolation (Muller 1942; Mayr 1963). Thus, ecological speciation can involve any type of reproductive barrier, and can occur under any geographic
arrangement of populations (allopatry, parapatry, sympatry, Mayr 1947; Funk 1998; Lu and Bernatchez 1999; Schluter 2000, 2001; Ogden and Thorpe 2002; Rundle and Nosil 2005; Funk et al. 2006; Vines and Schluter 2006). Ecological speciation is distinguished from models of speciation which do not involve ecologically based divergent selection, such as speciation via genetic drift or the fixation of different incompatible mutations in populations experiencing similar selection (Schluter 2009; Schluter and Conte 2009).

The process of ecological speciation makes explicit predictions. For example, it predicts that ecologically divergent pairs of populations will exhibit greater reproductive isolation than ecologically similar pairs of populations of similar age (Funk et al. 2002; Funk et al. 2006). Another prediction is that phenotypic traits involved in divergent adaptation will also cause reproductive isolation (Jiggins et al. 2001). For example, adaptive traits might directly reduce the fitness of immigrants and hybrids, due to a mismatch between immigrant and hybrid phenotypes and the ecological environment, generating ‘immigrant inviability’ and extrinsic postmating isolation, respectively (Rundle and Whitlock 2001; Nosil et al. 2005). Finally, ecological speciation predicts that neutral gene flow between populations will decrease as adaptive divergence increases (Ogden and Thorpe 2002; Nosil 2008). These predictions have now been supported numerous times using experiments with molecular data on levels of neutral gene flow (see Schluter 2001; Funk et al. 2002; Rundle and Nosil 2005; Schluter and Conte 2009). Additionally, some progress has been made in understanding the genetic basis of ecological speciation, but this stems primarily
from QTL and candidate gene studies (Rogers and Bernatchez 2007; Via and West 2008; Schluter and Conte 2009). Here we focus on a largely unexplored issue: the role of gene expression in ecological speciation.

The role of gene expression warrants consideration because two events need to occur during the process of ecological speciation (following Schluter 1998), and gene expression might strongly affect each of them. First, a key mechanism by which ecological divergence between populations occurs is via the colonization of new environments. In these cases, ecological speciation requires that newly founded populations persist in the colonized environments. Ernst Mayr (1947, 1963) especially espoused this ‘persistence view’ of the role of ecology in speciation (for review see Schluter 1998; Levin 2004). Second, populations in different environments need to evolve genetically based reproductive isolation. Gene expression might therefore promote ecological speciation in two ways: (1) indirectly by promoting population persistence or (2) more directly by affecting adaptive genetic divergence in traits causing reproductive isolation (Fig. 4.1, Price et al. 2003).

Here we review both putative roles for gene expression in ecological speciation. Because the study of gene expression and ecological speciation is in its infancy, our goals here are not only to review the existing literature and highlight what is already known, but also to provide a conceptual framework for thinking about the topic and to point to especially promising avenues for further research. We begin by providing more detail on why studying gene expression might be fruitful for understanding speciation. Next, we discuss how to measure
gene expression. We then review the roles of gene expression in population persistence and in affecting adaptive genetic divergence.

4.3 What Can the Study of Gene Expression Tell Us About Ecological Speciation?

Gene expression is shaped by both genetic and environmental components, and can therefore be considered as a “molecular phenotype” (Ranz and Machado 2006). For example, the transcription rate of a gene can vary among genotypes such that it is a heritable phenotype (Schadt et al. 2003; Gibson and Weir 2005; Whitehead and Crawford 2006b; Roelofs et al. 2009). Gene expression might provide novel insights into speciation because gene expression profiles have the ability to uncover phenotypes which would not readily be visible via traditional approaches. Our understanding of evolution has often been limited by our ability to define relevant phenotypes (Nevins and Potti 2007). For example, initial progress in understanding ecological speciation has necessarily focused on easily measured morphological, and to some extent, behavioral traits. In essence, gene expression might allow us to circumvent these limits by uncovering hidden phenotypes potentially of ecological relevance; phenotypes which are perhaps difficult or counterintuitive to measure. This could be especially critical given that genome annotations to date currently stem mostly from model genetic organisms, and consequently lack ecological relevance (Aubin-Horth and Renn 2009). Identifying ecologically relevant expressed genes will thus likely increase the efficacy of genomics to address questions related to
ecological speciation (Carroll and Potts 2006; Landry and Aubin-Horth 2007; Pena-Castillo and Hughes 2007).

For example, physiology has been grossly underrepresented in ecological speciation studies, presumably because of the difficulty associated with measuring these phenotypes (Schluter 2000; Rundle and Nosil 2005). With current gene expression technologies, we can now examine many metabolic and mechanistic processes that were previously difficult to measure (Dalziel et al. 2009; Scott et al. 2009). This may be important because evolutionary changes in expression of physiological genes might sometimes precede morphological changes (Mayr 1963). Overall, the sensitivity achievable from modern gene expression technology (large numbers of genes in one assay, low transcript genes, and subtle gene expression differences) has allowed the study of specific organs and tissues and revealed that hidden phenotypes may stem from genes expressed in all of these tissues. For all of these reasons, gene expression studies have the potential for testing numerous hypotheses (i.e., numerous ‘traits’ or genes), many of which an investigator would not necessarily think to test from previous research (Noor and Feder 2006). The power of gene expression profiles as surrogate phenotypes is well established in fields such as genetic studies of disease research in humans (Liotta and Petricoin 2000; Nevins and Potti 2007), but needs to be further implemented into ecological speciation studies.

Theoretically, this implementation seems possible. Johnson and Porter (2000) demonstrated that parallel directional selection on geographically isolated populations might lead to misregulation of gene expression that in turn may be
associated with hybrid incompatibility. By modeling the evolution of a regulated pathway wherein hybrid incompatibility can arise as a consequence of misregulated gene expression, Johnson and Porter (2000) showed that parallel selection is expected to yield reproductive isolation regardless of the underlying mechanisms relating genotype to phenotype. In their analyses, population pairs experienced identical selection conditions and thus did not experience divergent selection. Nonetheless, these results suggest that the detection of gene misregulation may be a very feasible starting point towards understanding the role of gene expression in ecological speciation, where the objective is measuring the level of hybrid incompatibility due to gene expression before finding the ultimate mutation responsible (Renaut et al. 2009).

In summary, gene expression studies may reveal the genes underlying adaptations that are difficult or impossible to measure in other ways, and these phenotypes may be of importance for initiating ecological divergence during speciation (Feder and Mitchell-Olds 2003). Consequently, patterns of gene expression should be integrated into studies of ecological speciation, with a need for clearer predictions about how gene expression affects ecological speciation. Gene expression patterns may also provide insight into the underlying genetic architecture of ecological speciation and importantly, if it differs from other types of speciation.

4.4 How to Study Gene Expression

Studies of gene expression measure the expression level of single genes, multiple genes, or the entire transcriptome (the latter defined as all the genes
expressed in a cell, tissue or organism). The measure of expression is the abundance of transcribed messenger RNA (mRNA) molecules and is specific to the tissue, developmental stage, point in time, and taxon in which it is measured (Thomas and Klaper 2004; Ranz and Machado 2006; Goetz and MacKenzie 2008). Protein and mRNA abundances are highly correlated, which is why mRNA levels can be used as a proxy for differences in protein products (Gracey 2007). A variety of methods are now available to quantify gene expression and can be subdivided into two broad categories: (1) those for which (candidate) genes must be known in advance of quantification of expression, and (2) those that quantify abundance for multiple genes and thus simultaneously identify genes of interest (Table 4.1). We treat each category in turn briefly here, and refer readers to previous reviews for greater detail (Yuan et al. 2006; Karlen et al. 2007).

We first outline methods that require candidate genes prior to analysis. These are cases where known genes are of a priori interest, for example because of their function or association with ecological variables. The original gene expression technique is Northern blotting where RNA is extracted from a specific tissue, and subjected to electrophoresis on a gel (Berger and Kimmel 1987). The gel is then transferred to a nylon membrane that is washed with a labeled probe specific to the candidate gene of interest. If the gene was expressed and the transcript is present, the probe will hybridize and anneal to the membrane. Other samples may then be compared for the expression of this same gene. Northern blots allow the detection and only semiquantification of mRNA target sequences (the darker the band, the greater the expression
Schlamp et al. 2008). More sensitive techniques have since been developed. One such technique is retro-transcriptase quantitative polymerase chain reaction (RT-qPCR or qPCR, Rasmussen et al. 1998; Van Straalen and Roelofs 2006). With qPCR, one converts mRNA to cDNA and then uses fluorescent probes specific to the cDNA in PCR to monitor the quantity of cDNA template. The PCR cycle associated with exponential growth of product is tightly associated with the quantity of the initial cDNA template, providing an estimate for the level of mRNA expression in the tissue. Depending on experimental design, qPCR can assess relative or absolute abundance of RNA. Since qPCR does not have the same technical problems as microarrays (see below), qPCR has emerged as a method to quantify and verify expression levels of candidate genes identified with large scale transcriptomic studies (Rajeevan et al. 2001; Etienne et al. 2004).

A second set of techniques do not require that candidate genes be chosen prior to analysis. Within this set, two techniques are no longer in common use or have a limited history of use in ecological studies. The first are differential display techniques, which with real-time PCR (DDRT-PCR), can describe differences in gene expression between species (Liang and Pardee 1992). This strategy is based on the amplification of partial cDNA sequences from a pool of mRNA (of unknown genes being expressed), and is only useful when genes are abundantly expressed. Another method, suppression subtractive hybridization (SSH) employs PCR to differentially amplify cDNA (Diatchenko et al. 1996). SSH has the advantage of identifying all the differentially expressed genes even at low abundance between nonmodel species (Berger and Kimmel 1987). Few studies
have used this method to compare gene expression profiles in divergent ecological conditions, presumably because of the availability of more sensitive and precise techniques (Shaw and Danley 2003; Jones et al. 2006).

Currently, the most common technique for extensively assessing global gene expression profiles is microarrays, which are generally akin to a reverse Northern Blot (Schena et al. 1995). Microarray experiments are performed by hybridizing “target” cDNA in solution from an experimental group or groups to the spots or “probes” that are fixed to the glass slide, often representing in the order of thousands of genes. Gene expression among groups for each spot are then compared according to their fluorescence intensities to detect up or down regulated genes. Treatment sample is either competitively hybridized and compared to a common reference or another treatment (two-color experiment) or just the absolute intensity of a single treatment sample is measured (one-color experiment, Patterson et al. 2006). The last decade has experienced an explosion of microarray studies in ecology and evolution, and we refer readers to past reviews for a more thorough treatment of the methodology (reviewed in Bowtell and Sambrook 2003; Childs et al. 2003; Shimkets 2004; Van Straalen and Roelofs 2006; Goetz and MacKenzie 2008; Nielsen and Pavey 2010).

Here, we focus on two critical points: susceptibility of type I error and repeatability of the results. First, microarrays are a powerful tool, but can be prone to type I errors stemming from the large number of comparisons involved and variation in experimental conditions (e.g. use of different tissues, treatments, ecological types, species, Thomas and Klaper 2004; Yang and Churchill 2007).
Several methods exist for comparing samples on the arrays (Oleksiak et al. 2002; Ball et al. 2003; Naidoo et al. 2005), along with many different data analysis programs, techniques for normalization, quality control, quantifying spot intensity, and correcting for multiple tests. Second, the repeatability of published microarray studies is arguably limited, especially when the sum of the expression data is unavailable (Allison et al. 2006; Ioannidis et al. 2009). These discrepancies appear to be primarily due to incomplete data annotation or specification of data processing and analysis, rather than technical limitations. Standardized analytical procedures do not exist which can lead to contentious interpretations of the data. Although many journals now require that the data be submitted to acceptable public repositories upon conditional acceptance (Ball et al. 2004), more strict publication rules enforcing public data availability and explicit description of data processing and analysis will be needed to ensure repeatability.

Along these lines, it is important to note that microarrays are intended to be used with the species for which the chip has been developed (Bar-Or et al. 2007), but some studies have demonstrated that microarrays can also be used in closely related species (Van Straalen and Roelofs 2006; Kammenga et al. 2007; Whiteley et al. 2008). Oligo microarrays have shorter fragments of cDNA spotted on the chip, so they are less ideal for cross species work, since few numbers of polymorphism may affect hybridization greatly (von Schalburg et al. 2008). However, cDNA arrays have longer DNA fragments, increasing their potential usefulness to nontarget species. For example, the Genomics Research on
Atlantic Salmon Project developed a salmonid microarray consisting of expressed sequence tags (ESTs) developed from both rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). This microarray has been successfully applied in many closely related salmonids, including other salmon (*Salmoninae*), whitefish (*Coregoninae*) and rainbow smelt (*Osmerus mordax*, von Schalburg et al. 2005; Koop et al. 2008). Thus, cDNA microarrays can bridge differences between systems with plentiful genomic resources that are poorly understood ecologically, and systems with a well known ecology but poorly developed genomic resources (Thomas and Klaper 2004; McKay and Stinchcombe 2008).

An additional technique that assembles many short sequence tags (9-10 bp) excised from cDNA and inserts them in a 1kbp vector for sequencing is serial analyses of gene expression (SAGE, Velculescu et al. 1995). Each tag in SAGE can be traced back to a single gene and the relative amount of that tag in the vector corresponds to the mRNA levels in the tissue collected. Thus, SAGE may be used to both identify expressed genes and quantify the relative amounts without any *a priori* ESTs or genomic resources. The latest iteration of this technique, SUPERSAGE assembles longer sequence tags (26 bp) from which primers for qPCR or potentially even oligo microarrays could be created (Matsumura et al. 2006).

With the advent of next generation sequencing techniques, it is now possible to routinely sequence the entire transcriptome of each sample (Wall et al. 2009), even for nonmodel organisms (Vera et al. 2008). Like SAGE, this
technique generates sequence data and transcript abundance (Goetz et al. 2010; Wolf et al. 2010). However, longer sequence reads have the power to discern alternative splice variants (Ferguson et al. 2010), alternative alleles (Fontanillas et al. 2010), and Single Nucleotide Polymorphisms (SNPs) within coding regions (Renaut et al. 2010). This method will therefore be more precise in identifying and quantifying the transcription of closely related genes. As costs decrease and read lengths increase, this may ultimately replace all other transcriptomic methods.

In conclusion, a quickly expanding and improving suite of methods exist for the study of differential gene expression. The best method will depend on the research question, study organism, budget, and whether reasonable candidate genes are known a priori. One important gap in the literature is the lack of a comprehensive study comparing the accuracy and precision of these techniques. We now turn to more conceptually oriented questions of how gene expression might affect ecological speciation.

4.5 Gene Expression and Population Persistence

“The creatures which can stand “the storm and the stress” of the physical influences of the environment [...] will live; while the others which cannot, will not.” (Baldwin 1896).

The first manner in which gene expression might affect speciation is via promoting population persistence. As exemplified by Baldwin’s quotation above, once a population colonizes a new environment, it must persist if it is to speciate. Population establishment and persistence in a new environment may be
facilitated by phenotypic plasticity (Via and Lande 1985; Robinson and Dukas 1999; Yeh and Price 2004; Reusch and Wood 2007; Svanback et al. 2009). Modulation of behavioral, morphological or physiological traits via phenotypic plasticity could therefore occur before any adaptive genetic evolution occurs (Price et al. 2003). Gene expression mediated phenotypic plasticity may be described as reaction norms in gene expression with the molecular phenotype of gene expression facilitating population persistence following colonization (Aubin-Horth and Renn 2009). Direct tests of this idea are lacking, but two lines of indirect evidence exist: (1) studies of plasticity in traits (morphology and behavior mostly) related to fitness and population persistence, and (2) studies of gene expression responses during ecological shifts, particularly those resulting in exposure to ecological stress.

First, studies of plasticity in phenotypic traits related to fitness provide evidence for a role for plasticity in population persistence. However, such studies routinely lack evidence on how (or if) gene expression itself was involved. For example, Yeh and Price (2004) studied two populations of dark-eyed Junco birds, a native population in the mountains and a newly established population on the University of California San Diego (UCSD) campus. The UCSD population persisted for years despite significant environmental differences compared to the native habitat. Studies on the length of the breeding season, a classic trait dependent on temperature, revealed that the breeding season of the UCSD populations was twice as long as that of the ancestral populations, presumably due to more favorable climate (e.g. lack of snow) in the newly established
population. Importantly, UCSD females displayed higher offspring production without a corresponding increase in mortality, suggesting that plasticity in breeding time was promoting colonization, population establishment, and persistence in the new environment. However, future studies are needed to examine if gene expression might be associated with the shifts in these life history traits.

Second, studies of gene expression response during ecological shifts support a role for gene expression in facilitating responses to ecological change (i.e., stress). By ecological stress we mean simply a shift in ecology that affects the fitness of a population. In the last decade, the frequency of studies reporting evidence for a role of gene expression during ecological stress has increased (Table 4.2). Although these studies are critical towards understanding physiological stress response, detailed analysis of the visible phenotypes, and their explicit effects on population persistence, are needed. For instance, McCairns and Bernatchez (2010) examined adaptive divergence between freshwater and marine sticklebacks in a common garden experiment. Specifically, they measured fitness and survival to explore the role of gene expression at four candidate genes in response to osmoregulation. They found a significant correlation between gene expression and fitness and their results thus supported the hypothesis that ancestral plasticity for osmoregulation promoted adaptive divergence via heritable osmoregulation expression (sodium-potassium ATPase). These results are consistent with the hypothesis that gene expression
modulation can promote adaptive divergence by allowing populations to persist in a changing environment, whereby fitness is maintained by plasticity.

Despite these advances, definitive tests demonstrating that gene expression facilitates population persistence are lacking. Such tests could be carried out using experimental evolution in the lab. For example, the genomics of *Drosophila* is increasingly well characterized, with some mutant lineages able or unable to cope with different stressors (Sorensen et al. 2009). Under controlled stress conditions, one could measure which genes are most strongly differentially expressed, while controlling for variation in ecologically relevant alternative alleles in different environments (Dalziel et al. 2009). The differentially expressed genes could then be knocked out in one ‘expression mutant’ treatment (e.g. by the use of RNAi or destroying or inhibiting the promoter regions). Both mutant and control treatments would then be exposed to stress, simulating colonization of a new environment, and the population persistence of each compared. The prediction is that population persistence would be weaker for the mutant treatment. In principle, this experiment could even be conducted in the field (Knight et al. 2006). Likewise, gene expression studies of natural populations colonizing new environments may identify genes and pathways whose plasticity is essential to persistence in new environments (Juenger et al. 2006).

We note two additional points about the importance of gene expression for population persistence. First, populations and genes with much pre-standing genetic variation may exhibit rapid evolution following the colonization of new environments (Barrett and Schluter 2008) and thus not require differential gene
expression as strongly for population persistence. Second, purely environmentally induced gene expression (non-heritable molecular phenotypes) can still play an indirect role in speciation by facilitating population persistence and “buying the population time” for divergence in other, less plastic, evolvable traits (Fig. 1 in Price et al. 2003).

4.6 Gene Expression and Adaptive Genetic Divergence

The second manner in which gene expression might affect ecological speciation is by being associated with adaptive genetic divergence and reproductive isolation (Fig. 4.1). This forces a consideration of the link between divergent selection, adaptive genetic divergence, and reproductive isolation: loci under divergent selection and loci causing reproductive isolation are similar in exhibiting reduced introgression (and thus greater divergence) between populations relative to other loci (Barton and Hewitt 1989; Mallet 1995; Wu 2001; Mallet 2005; Nosil et al. 2005). Indeed, an allele ‘a’ that confers a poor fit of the phenotype to the environment can be selected against and contribute to speciation, whether the afflicted allele resides in one of the parental species (immigrant homozygote ‘aa’) or in a hybrid individual (heterozygote ‘Aa’). Recognizing that the adaptive genetic divergence, which results in selection against immigrants and hybrids, represents reproductive isolation itself helps clarify the relatedness of the two processes. Additionally, we stress that adaptive genetic divergence might incidentally cause the evolution of any form of reproductive isolation, including ‘non-ecological’ forms such as sexual isolation.
and intrinsic genetic incompatibilities in hybrids (Rundle and Nosil 2005; Rogers and Bernatchez 2006).

Understanding the heritable component of gene expression will be fundamental towards understanding the genetics of ecological speciation. This is conceptually possible because the expression level for any given transcript is a phenotype that is influenced by both genetics and the environment. The genetic basis of gene transcription itself may exist prior to colonization of new environments or may actually evolve via genetic assimilation (Gibson 2008). We consider here two fundamental questions: (1) how substantial is the genetic component of gene expression, and (2) can we elucidate whether or not this genetic component of gene expression is associated with adaptive divergence and reproductive isolation? Each question is addressed in a separate section. The main findings, as well as explicit directions for future research, are summarized in Table 4.3.

4.7 Genetic Architecture of Gene Expression: Heritability and eQTL Mapping

4.7.1 Heritability of Gene Expression Divergence

Gene transcription rates can vary among genotypes such that it is a heritable phenotype. Both the magnitude and rate of changes in gene transcription level in response to selection will depend on the heritability of gene expression (Falconer and Mackay 1996; Gibson and Weir 2005). What proportion of the transcriptional variation in a population is attributable to genetic variation among individuals? Estimation of the heritability of gene expression is
likely to be complicated because sources of transcriptional variation can vary tremendously among tissues within individuals, among individuals, and among populations (Whitehead and Crawford 2006a). While several studies have discovered gene expression differences between diverging populations, significant transcriptional differences need not reflect heritable genetic variation (Roff 2007).

A few studies have formally detected heritable gene transcription differences between populations. These studies quantified genetic differences using common garden experiments which directly quantify levels of gene expression in the absence of environmental variation (Lai et al. 2008). For example, St-Cyr et al. (2008) quantified variation in gene expression for almost 4000 genes in species pairs of lake whitefish (*Coregonus clupeaformis*) from North American lakes under common garden conditions and found that 14% exhibited differences in transcription. These differences are therefore the heritable component of gene expression divergence. Remarkably, genes differentially expressed between species pairs in the common environment were similar to what had been previously identified in the wild. The collective results suggest a predominantly genetic control of differential transcription between these species pairs (Derome and Bernatchez 2006).

In other studies, heritability of gene expression within a population has been estimated using parent-offspring or sibling regressions. Studies of human gene expression have found that approximately 30% of genes have a significant heritable component (Schadt et al. 2003; Monks et al. 2004). Estimating
heritability for wild populations is also possible using REML ‘animal models’ applied to multigenerational data from natural populations (Kruuk 2004). When applied to pedigrees with multiple generations and low immigration rates, these models can reduce bias due to shared environment effects (Kruuk and Hadfield 2007). Roberge et al. (2007) applied this approach to estimate heritability of gene expression in the Atlantic salmon genome, discovering that 16% of 6500 gene transcripts had a heritable component of gene expression, on average explaining 40% of the variation in transcription profiles. These results compare to other species that have found the median heritability among genes with heritable transcription profiles ranged from 0.11 (in mice, Chesler et al. 2005) to 0.84 (in yeast, Brem et al. 2002). Notably, studies estimating heritability using such approaches need to account for the fact that heritability within a population does not equate to heritable differences between populations. Thus, while there are no studies that have quantified the heritability of transcription profile differences underlying ecological speciation, these results suggest that the heritable component of gene expression exists, but is highly variable.

4.7.2 eQTL Mapping

Analyses on the genetic architecture of transcriptome variation offers to further our understanding of the genetic basis of gene expression and adaptive divergence (Rockman and Kruglyak 2006). By ‘genetic architecture' we mean quantifying the number, location, and effect sizes of genes contributing to adaptive divergence (Rieseberg and Noyes 1998). In studies of genetic architecture, a quantitative trait locus (QTL) is defined as a region of the genome
containing one or more genes that affect variation in a quantitative trait, identifiable by its linkage or association to polymorphic marker loci (Gilad et al. 2008; Mackay et al. 2009). Traditional or ‘phenotypic’ QTL (pQTL) uncover associations between genetic regions and traditional phenotypic traits such as morphology. Expression QTL (eQTL) map transcript abundance in the same manner as pQTL map ‘traditional’ traits. eQTL is emerging as a useful technique for localizing genomic regions contributing to gene expression divergence (Gilad et al. 2008). eQTL studies are generally characterized by large numbers of phenotypes (e.g., the number of transcripts on a microarray), but the mapping is typically performed with fewer individuals, due to the still prohibitive cost of running the arrays. Although eQTL studies are still in their infancy, two general patterns have been observed: (1) the predominance of cis-localized eQTL, and (2) the existence of genomic regions associated with the expression level of many transcripts (so called eQTL “hotspots”). We consider each in turn.

The segregation of eQTL has a local genomic context because there are two ways, denoted as cis or trans, that the level of transcript variation may map onto the genome (Hubner et al. 2005; Kirst et al. 2005; Mackay et al. 2009), with each providing a different interpretation about genetic architecture. If the transcription profile maps within the gene region for the transcript in question, this association is referred to as cis or proximal eQTL. In contrast, if the transcription profile maps to another gene or genomic region it is referred to as a trans or distal eQTL (Mackay et al. 2009). Cumulatively, the distribution of cis versus trans eQTL on the genome has shown that cis eQTL seem to have larger genetic
effect sizes than trans eQTL and that there are more cis than trans eQTL in the genome (Wentzell et al. 2007), although the biological interpretation of this pattern remains obscure (Hubner et al. 2005; Landry et al. 2005).

Another emerging pattern is the existence of eQTL “hotspots”: genomic regions that are associated with the expression level of many transcripts (Gibson and Weir 2005; Whiteley et al. 2008). These hotspots may involve the distribution of eQTLs as well as transcriptional covariation between individuals in the mapping family (Rockman and Kruglyak 2006; Mackay et al. 2009). What do these hotspots tell us about the genetics of ecological speciation? First, they show that pQTL and eQTL can map to the same genomic regions. For example, recent studies mapped both eQTL and pQTL for morphological, life history and behavioral traits in dwarf and normal lake whitefish species pairs (Rogers and Bernatchez 2007; Derome et al. 2008; Whiteley et al. 2008). Of 261 white muscle eQTL distributed over 24 linkage groups, 15 eQTL localized with overlapping pQTL (Derome et al. 2008; Bernatchez et al. 2010). Strikingly, almost 90% of eQTL-pQTL co-localizations involved growth rate and condition factor, two traits central to the adaptive divergence of these species pairs (Rogers and Bernatchez 2007). Of course, a caveat about overlapping eQTL and pQTLs is that the sizes of the QTL regions are often quite large, such that the apparent colocalization of these two types of QTL need not imply a functional relationship. Nonetheless, the genes within these regions harboring both eQTL and pQTL are arguably strong candidates for genes involved in ecological speciation.
Additionally, eQTL studies indicate that genomic regions involved in ecological speciation can be nonrandomly distributed across the genome. For example, in the same lake whitefish species pairs noted above, 50% of 249 eQTL identified in the brain were associated with only 12 hotspots distributed over eight linkage groups (Whiteley et al. 2008). A similar pattern was observed in muscle, where 41% of eQTL mapping to six hotspots across four linkage groups (Derome et al. 2008). These findings hint at the existence of localized ‘genomic islands’ of expression divergence, as sometimes reported for islands of genetic differentiation in population genomic studies (Turner et al. 2005; Nosil et al. 2009; but see Michel et al. 2010).

Finally, eQTL have also informed us about the actual mechanisms of speciation, for example confirming that mapped genomic regions differentiated via divergent selection. The direction of additive eQTL reported by in the Whiteley et al. (2008) and Derome et al. (2008) were predominantly in one direction, suggesting a role for directional selection (Orr 1998; Schluter 2000). eQTL hotspots have also been associated with molecular signatures of selection in natural populations. For example, in the whitefish species pairs, ten loci were identified whose genetic divergence in nature exceeds neutral expectations. These are so-called “outlier loci” subject to divergent selection (Beaumont et al. 2005; Nosil et al. 2009). Three of these outlier loci also corresponded to eQTL hotspots (Bernatchez et al. 2010). Finally, eQTL hotspots may be an indication that co-expression involves a regulatory network such that speciation involves complex interactions between genes (Wentzell et al. 2007). Overall, eQTL
studies can thus be used to infer the genomic distribution of expression profiles (Gibson and Weir 2005) with eQTL distributions potentially informing the mechanisms of gene regulation (Dalziel et al. 2009), and providing insight into the process of speciation.

4.7.3 Other Approaches to Studying Heritability of Gene Expression

The previously discussed approaches to studying the heritability of gene expression divergence may be thought of as top-down or a forward genetics approach: they start with the entire transcriptome and work towards narrowing down to regions or genes implicated in adaptive divergence and ecological speciation. However, few QTL maps or genome scans exhibit sufficient resolution to find the exact functional genes or regulatory elements that contain the polymorphisms that are under selection (Beaumont et al. 2005; Nielsen 2005). Moreover, mapping approaches may not be feasible in some organisms. Another approach which relies on sequence comparison of functional polymorphisms may be described as a gene expression approach to “reverse ecology” (Li et al. 2008): after differentially expressed functional genes are identified, sequences of the differentially expressed transcripts are compared by direct sequencing effort that may uncover nonsynonymous mutations in the coding regions of the genes, or genetic polymorphisms in regulatory regions (Cassone et al. 2008). Both of these steps may now be accomplished simultaneously with next generation sequencing (Wolf et al. 2010). Thus, screening the transcriptome for gene expression differences, even in the
absence of a QTL map or genome scan, could simultaneously start the search for functional polymorphisms.

### 4.7.4 Genetic Component of Gene Expression: Conclusions

Common garden and eQTL studies clearly demonstrate that gene expression divergence can have a heritable component and be associated with adaptive genetic divergence. Although progress has been made in identifying specific differentially expressed genomic regions contributing to adaptive divergence, identification of specific mutations, and characterization of interactions among genomic regions, remains a major challenge for future work.

### 4.8 Functional Links between Adaptive Candidate Gene Expression, Adaptive Genetic Divergence, and Reproductive Isolation

Even after genetic components of gene expression are identified, a major question remains: are these components associated with adaptive divergence and reproductive isolation? Several studies have demonstrated that reductions in hybrid fitness can be due to gene (mis)expression (Landry et al. 2007; for review Ortiz-Barrientos et al. 2007), in some cases linking gene misexpression in hybrids to other factors previously identified as contributing to ecological speciation (Renaut et al. 2009). Along these lines, a growing number of studies have now isolated and characterized specific candidate genes or patterns of gene expression associated with the adaptive divergence which drives ecological speciation (e.g., Lexer et al. 2004; Colosimo et al. 2005; Derome and Bernatchez 2006; Steiner et al. 2007; Whiteley et al. 2008). Of these, none have
demonstrated an actual association between gene expression and reproductive isolation (but see Renaut et al. 2009), usually because reproductive isolation itself was not explicitly considered, underlying mutations have not been identified, or mutations causing adaptive divergence lie in cis-regulatory rather than coding regions of genes (Stern and Orgogozo 2008; Chan et al. 2009). To compensate, we work under the assumption that genes whose expression is associated with adaptive divergence might also impact the fitness of immigrant and hybrids and thus make a contribution to ecologically based reproductive isolation (‘immigrant inviability’ and extrinsic postmating isolation), albeit of unknown magnitude. Testing this assumption represents a major avenue for future research. The examples below thus illustrate both the promise and the difficulties associated with linking gene expression to ecological speciation.

4.8.1 Bmp4: Beak Shape and Speciation in Darwin’s Finches

Darwin’s finches arose via adaptive radiation on the Galapagos Islands (Grant 1986). Beak morphology diverged adaptively among populations and species in response to divergent selection stemming from competition and use of seeds of differing size and hardness (Schluter and Grant 1984; Grant and Grant 2006). Beak morphology might also contribute to reproductive isolation via song divergence (Podos 2001) or due to selection against immigrants and (intermediate) hybrids (Grant and Grant 2008; Hendry et al. 2009). Among species, higher levels of the bone morphogenetic protein 4 (Bmp4) expression are correlated with deeper beak shapes. Also, manipulation of expression level of Bmp4 in chicken embryos altered beak development in the predicted direction.
(Abzhanov et al. 2004). These results provide compelling evidence that gene expression variation from \textit{Bmp4} affects morphological divergence among species of Darwin’s finches (Fig. 4.3). Similar results occur for another gene, calmodulin (\textit{CaM}, Abzhanov et al. 2006). However, due to a lack of common garden or mapping studies, there is as of yet no evidence that heritable differences in beak morphology are affected by \textit{Bmp4} or \textit{CaM}. The mutations underlying beak size differences in Darwin’s finches have not been identified. Thus, although there is good evidence that regulatory changes underlie morphological divergence among species of Darwin’s finches, the ultimate link between gene expression and genetically based reproductive isolation (i.e., speciation) is yet to be made.

4.8.2 \textit{Pitx1}: Pelvic Reduction and Speciation in Threespine Stickleback

Recently derived postglacial fish populations are amongst the most extensively studied systems of ecological speciation in nature (reviewed in McKinnon and Rundle 2002; Rogers and Bernatchez 2007; Hendry 2009). One such example is the threespine stickleback (\textit{Gasterosteus aculeatus}) complex in which ecological divergence drove speciation between limnetic and benthic pairs within freshwater lakes, and between marine and freshwater populations (Schluter and McPhail 1992; McKinnon et al. 2004). Ancestral marine and most derived freshwater stickleback have a robust pelvic apparatus, while at least 24 independent freshwater populations exhibit a greatly reduced or completely absent pelvic structure (Bell 1987; Gow et al. 2008; Chan et al. 2009). Repeated parallel evolution is itself an indication that divergent selection drove evolution, with evidence pointing to predation and differences in ion concentration as the
mechanisms of selection (Reimchen 1980; Bell et al. 1993; Reimchen and Nosil 2002; Vamosi 2002).

Recent studies have examined the genetic basis of pelvic reduction. QTL studies repeatedly identified a single chromosomal region explaining more than two-thirds the phenotypic variance in pelvic size (Cresko et al. 2004; Shapiro et al. 2004; Coyle et al. 2007). Yet, similar to Bmp4 in finches, the regulatory mutation contributing to differences in expression remained unknown until recently. Chan et al. (2009) reported that a small (501bp) tissue specific enhancer (Pel) drives expression of the gene implicated in pelvic reduction (the Pitx1 gene Chan et al. 2009). Remarkably, small deletions functionally inactivated Pel in nine of 13 tested pelvic reduced populations. These regions exhibiting recurrent deletions, rather than the Pitx1 gene itself, appear to have been subject to positive selection (Chan et al. 2009). These results demonstrate that genetically based expression divergence contributed to adaptive divergence in pelvic morphology. However, direct links to between expression divergence and reproductive isolation remain to be established. The ability to conduct manipulative experiments in seminatural ponds (e.g., Schluter 1994; Barrett and Schluter 2008) indicates that linking gene expression at Pitx1 to reproductive isolation (i.e., reduced fitness of immigrants and hybrids) is a distinct possibility.

4.8.3 Other Examples

There are many other examples of studies of gene expression and adaptation, but few make links to adaptive genetic divergence, and thus few pertain directly to ecological speciation. For instance, cichlid fish species have
adapted to divergent light environments within lakes, via the effects of gene expression on the tuning of visual perceptual sensitivity (Carleton and Kocher 2001). In this case, changes in gene expression contribute to sensory diversification in replicate radiations of cichlid fishes in the clear waters of Lake Malawi versus the turbid waters of Lake Victoria, and functional substitutions contributing to expression divergence were identified (Hofmann 2009). These studies demonstrate important findings with respect to the molecular basis of ecologically driven sensory diversification, but again a direct demonstration that this contributed to reproductive isolation does not yet exist.

Mimetic wing coloration in Heliconius butterflies gives rise to wing patterns that show repeated convergence between species and have adaptive value in mimicry and mate choice, thus potentially associated with ecological speciation (Mallet and Gilbert 1995; Kapan 2001; Jiggins 2008; Ferguson and Jiggins 2009). Comparative gene expression between two species, H. erato and H. melpomene, found that cinnabar expression correlated with the forewing band, providing good evidence that the expression of this gene gives rise to the red-banded phenotype in both species (Ferguson and Jiggins 2009). Chamberlain et al. (2009) report similar associations between wing color and gene expression, but within polymorphic populations. Differences in the actual traits in these studies (wing color and pattern) are heritable, but once again functional mutations contributing to reproductive isolation are lacking.

On the other hand, recent genome wide analyses of the transcriptome have demonstrated that complex patterns of gene misexpression may underlie
reproductive isolation mechanisms in hybrids. Renault et al. (2009) contrasted gene expression divergence at key early developmental stages in species pairs of normal and dwarf whitefish and their F1 hybrids to identify the main mode of action responsible for gene transcription and to discover key genes misexpressed in hybrids. While only five of 5000 transcripts differed in mean expression level between parentals and hybrids at the embryonic stage, 617 out of 5300 transcripts differed significantly for 16 week old juveniles. Remarkably, significant gene misexpression in backcross hybrids involved several genes, most notably the disruption of three key developmental genes involved in protein folding and mRNA translation. Overall, direct demonstrations of how gene expression causes reproductive isolation remains a major missing link in connecting the role of gene expression to ecological speciation. Once such demonstrations are made, it will be necessary to test whether, and how, expression divergence actually reduces gene flow between natural populations.

4.9 Conclusions and Future Directions

Gene expression is likely to be important for the two events required for ecological speciation: population persistence and the evolution of genetically based reproductive isolation. Studies of plasticity and population persistence have yet to address gene expression explicitly. When it comes to adaptive genetic divergence and reproductive isolation, gene expression divergence has been shown to be heritable and to contribute to adaptive genetic divergence, but links to the evolution of reproductive isolation remain indirect (see Table 4.3 for a summary of what is known, and what needs to be done next). Our review
suggests that establishing this link will be challenging because the genetic architecture of ecological speciation can be controlled by gene networks and regulatory regions, rendering an understanding of the functional association between gene expression and adaptive divergence difficult. This implies that it may be difficult to make predictions about the likelihood of ecological speciation based on gene expression profiles until we have a better idea about the genetic architecture of ecological speciation, and how it compares to other mechanisms of speciation (Wu 2001; Rogers and Bernatchez 2007). Nonetheless, isolating the mutations contributing to variation in adaptive traits, and then studying their effects on reproductive isolation, is a necessary task for understanding how gene expression affects ecological speciation (Chan et al. 2009) and whether expression changes are associated with the causes or consequences of ecological speciation. Such goals will likely to be best achieved by integrating multiple molecular techniques with experimental studies of how different mutations (alleles) affect fitness and reproductive isolation (Barrett and Schluter 2008; Dalziel et al. 2009; Schluter and Conte 2009; Storz and Wheat 2010).

4.10 Acknowledgements

We thank Dolph Schluter for discussions about gene expression and speciation. We would like to thank three anonymous reviewers for their constructive comments that greatly improved the manuscript. During part of the writing of this manuscript, SP, HC, and PN were hosted by the Institute for Advanced Study, Wissenschaftskolleg, Berlin. SMR was funded by a Discovery
Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

4.11 Literature Cited


4.12 Tables
Table 4.1 Comparisons of gene expression methodologies. Technique denoted with (1) do not require a priori, whereas techniques denoted with (2) require that candidate genes are known ahead of time.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Approach</th>
<th>Pros</th>
<th>Cons</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDRT-PCR (1) rarely used</td>
<td>-A subset of differentially expressed genes</td>
<td>-Inexpensive</td>
<td>-Only good for genes in high abundance</td>
<td>(Reiland and Noor 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Generates differentially expressed genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT-qPCR (2)</td>
<td>-Individual candidate genes are compared</td>
<td>-Precise, less expensive than some other methods</td>
<td>-Candidate genes are a prerequisite</td>
<td>(Kobayashi et al. 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Great follow up on transcriptome-wide techniques</td>
<td>-Primer development can be difficult</td>
<td></td>
</tr>
<tr>
<td>Microarrays (1)</td>
<td>-Up to tens of thousands of genes assayed at one time</td>
<td>-Used to generate candidate genes</td>
<td>-Only available for some taxa</td>
<td>(Podrabsky and Somero 2004; Derome et al. 2006; Whiteley et al. 2008)</td>
</tr>
<tr>
<td>SAGE (1)</td>
<td>-Many genes assayed and sequenced at one time</td>
<td>-Less development cost than microarrays</td>
<td>-Fairly expensive but getting cheaper</td>
<td>(Bernier et al. 2008; Molina et al. 2008)</td>
</tr>
<tr>
<td>Suppression Subtractive Hybridization (1)</td>
<td>-Identify genes differentially expressed</td>
<td>-Inexpensive</td>
<td>-Requires sequencing to identify the physiological function of the differentially expressed genes</td>
<td>(Shaw and Danley 2003; Jones et al. 2006)</td>
</tr>
<tr>
<td>Rarely used</td>
<td></td>
<td>-Generates differentially expressed genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Blot (2)</td>
<td>-Probes for a single gene</td>
<td>-The original gene expression tool</td>
<td>-One gene at a time, limited utility in quantification</td>
<td>(Berger and Kimmel 1987)</td>
</tr>
<tr>
<td>RNA-Sequencing via Next Generation Sequencing (1)</td>
<td>-Sequence all transcripts</td>
<td>-The ultimate tool, price coming down</td>
<td>-Expensive, computationally demanding</td>
<td>(Elmer et al. 2010; Ferguson et al. 2010; Goetz et al. 2010; Renaut et al. 2010; Wolf et al. 2010)</td>
</tr>
</tbody>
</table>
Table 4.2. Examples of studies showing that gene expression is affected by ecological stress. Under the assumption that gene expression allows populations to better persist in stressful environments, these studies indicate that differential gene expression can promote the colonization of, and subsequent persistence in, novel environments (table starts on the next page and continues to the following horizontally).
<table>
<thead>
<tr>
<th>Organism</th>
<th>Study</th>
<th>Method</th>
<th>Environmental Stressor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Killifish</td>
<td>(Podrabsky and Somero 2004)</td>
<td>cDNA microarray</td>
<td>Daily and seasonal temperatures regimes</td>
</tr>
<tr>
<td>Austrofundulus limnaeus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bivalve</td>
<td>(Zapata et al. 2009)</td>
<td>Suppression subtractive ESTs library + quantitative RT-PCR (candidate genes)</td>
<td>Copper tolerance</td>
</tr>
<tr>
<td>Argopecten purpuratus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazilian flounder</td>
<td>(Meier et al. 2009)</td>
<td>Semiquantitative RT-PCR (candidate gene approach)</td>
<td>Hyperosmosis</td>
</tr>
<tr>
<td>Penalichthys orbignyanus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coral fish</td>
<td>(Kassahn et al. 2007)</td>
<td>cDNA microarray</td>
<td>Prolonged heat hypoxia</td>
</tr>
<tr>
<td>Pomacentrus moluccensis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthropod</td>
<td>(Roelofs et al. 2009)</td>
<td>cDNA microarray</td>
<td>Cadmium</td>
</tr>
<tr>
<td>Orchesella cincta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antarctic nematode</td>
<td>(Adhikari et al. 2009)</td>
<td>Suppression subtractive hybridization/ ESTs library + quantitative RT-PCR</td>
<td>Desiccation resistance</td>
</tr>
<tr>
<td>Plectus murrayi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thale cress</td>
<td>(Kreps et al. 2002)</td>
<td>Oligo microarray</td>
<td>Salt, osmotic regulation, temperature</td>
</tr>
<tr>
<td>Arabidopsis spp.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black cottonwood</td>
<td>(Ralph et al. 2006)</td>
<td>cDNA microarray</td>
<td>Herbivory</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic salmon</td>
<td>(Roberge et al. 2007)</td>
<td>cDNA microarray</td>
<td>Pathogens</td>
</tr>
<tr>
<td>Salmo salar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycorrhizal involutus</td>
<td>(Le Quere et al. 2006)</td>
<td>cDNA microarray</td>
<td>fungus Pillus Host specificity</td>
</tr>
<tr>
<td>Organism</td>
<td>Proportion of genes affected by treatment</td>
<td>Major Physiological Function Affected</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Killifish <em>Austrofundulus limnaeus</em></td>
<td>11%</td>
<td>Molecular chaperones, cholesterol and fatty acids synthesis, membrane structure, solute carrier, carbohydrate metabolism, nitrogen metabolism, intermediary metabolism, cytoskeleton elements, protein turnover, complement and innate immunity, cell growth and proliferation</td>
<td></td>
</tr>
<tr>
<td>Bivalve <em>Argopecten purpuratus</em></td>
<td>8%</td>
<td>Cell differentiation, cellular communication, cytoskeleton, development and differentiation, energetic metabolism, protein regulation, respiratory chain, stress protein, translation and post-translations processing, cellulose hydrolysis, ribosomal protein</td>
<td></td>
</tr>
<tr>
<td>Brazilian flounder <em>Penalichys orbignyanus</em></td>
<td>2 candidate genes up-regulated</td>
<td>Growth</td>
<td></td>
</tr>
<tr>
<td>Coral fish <em>Pomacentrus moluccensis</em></td>
<td>2% (down-regulation mostly)</td>
<td>Cell adhesion, cell cycle and growth, cytoskeleton, metabolism, protein processing, stress proteins, signal transduction, transcription, translation, transport</td>
<td></td>
</tr>
<tr>
<td>Arthropod <em>Orchesella cincta</em></td>
<td>14% (down-regulation mostly)</td>
<td>Translation, signal transduction, stress protein, redox state, general metabolism, chromatin remodeling, proteolysis digestion</td>
<td></td>
</tr>
<tr>
<td>Antarctic nematode <em>Plectus murrayi</em></td>
<td>6%</td>
<td>Carbohydrate metabolism, amino acid metabolism, lipid metabolism, xenobiotic metabolism, membrane transport, signal transduction, transcription, translation, replication, cell growth and death, cell communication</td>
<td></td>
</tr>
<tr>
<td>Thale cress <em>Arabidopsis spp.</em></td>
<td>12 to 25% (up-regulation mostly)</td>
<td>Oxidative stress, membrane transport, phosphoregulation, transcription, circadian clock, fatty acid metabolism, stress protein, cytoskeleton, membrane protein, carbohydrate metabolism</td>
<td></td>
</tr>
<tr>
<td>Black cottonwood <em>Populus trichocarpa</em></td>
<td>5% (up-regulation mostly)</td>
<td>Photosynthesis, general metabolism, transport, transcription, octadecanoid and ethylene signaling, detoxification and redox processes, secondary metabolism</td>
<td></td>
</tr>
<tr>
<td>Atlantic salmon <em>Salmo salar</em></td>
<td>17% (up-regulation mostly)</td>
<td>Immunity related genes, extracellular matrix component, electron and ion transport chain, signal transduction, transcription, metal binding protein, pyrimidine biosynthesis, protein degradation, localization and folding, DNA replication, cell structure and adhesion</td>
<td></td>
</tr>
<tr>
<td>Mycorrhizal <em>involutus</em></td>
<td>16%</td>
<td>Electron transport, lipid and fatty acid metabolism, transcription, sex determination, regulation of cell cycle, glycolysis, stress protein, protein biosynthesis, aromatic compounds metabolism</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3 Summary of what is known about gene expression and ecological speciation, what is missing (i.e., future directions), and how these gaps in our knowledge might be addressed (table continues on the following page).

<table>
<thead>
<tr>
<th>Category</th>
<th>What is Known</th>
<th>What is Missing</th>
<th>How to Address Gaps in Our Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population persistence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1. Role of plasticity</strong></td>
<td>-Phenotypic plasticity can promote population persistence.</td>
<td>-To what extent does this involve gene expression?</td>
<td>-Add gene expression data to studies of phenotypic plasticity and population persistence.</td>
</tr>
<tr>
<td>Heritability of expression divergence</td>
<td><strong>1. Common garden and/or animal model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Expression divergence between populations can have a genetic basis, and can involve parallel evolution across independent populations.</td>
<td>-How important is heritable gene expression divergence relative to other forms of genetic divergence (i.e., coding region changes)?</td>
<td>-Integrate studies of gene expression with studies examining functional mutations affecting trait divergence.</td>
</tr>
<tr>
<td></td>
<td><strong>2. eQTL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-eQTL hotspots exist, exhibit signatures of divergent selection, and provide candidate gene regions for ecological speciation. -Networks of gene interactions may be implicated in adaptive divergence.</td>
<td>-What role do eQTL hotspots have in adaptive divergence and/or reproductive isolation? -To what extent can we establish a mechanistic understanding of gene networks? -Can this inform us about the genetics of ecological speciation? -Does ecological speciation have a genetic architecture that is different from other types of speciation?</td>
<td>-Perform genome wide studies to understanding how gene expression affects ecological speciation.</td>
</tr>
<tr>
<td>Category</td>
<td>What is Known</td>
<td>What is Missing</td>
<td>How to Address Gaps in Our Understanding</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Gene expression and reproductive isolation    | -Gene expression divergence is known to affect adaptive phenotypic divergence, and in some cases has been tied to adaptive genetic divergence.  
                                                      -Few underlying mutations have been identified. | -To what extent does expression divergence actually generate reproductive isolation, either ecologically based or other forms?  
                                                      -To what extent can experimental studies of gene expression add to our understanding of the mechanisms of ecological speciation?  
                                                      -Can predictions be about the likelihood of ecological speciation based on gene expression profiles?  
                                                      -Is divergence in gene expression associated with the causes of ecological speciation or the consequences? | -Quantify the extent to which expression divergence contributes to reproductive isolation of all forms. |
4.13 Figures
Figure 4.1 Conceptual diagram of the different ways the genetic and environmental components of gene expression might contribute to ecological speciation. Both components might contribute to population persistence, which is required for eventual speciation. The genetic components of gene expression could contribute to the adaptive genetic divergence which drives ecological speciation. See text for details.
Figure 4.2. The effects of gene expression mediated phenotypic plasticity (GMPP: y-axis) on colonization of new environments, and subsequent population persistence. Genotype A has a lesser breadth of GMPP compared with genotype B. Both genotypes have high persistence in the source habitat, but genotype A has no potential to persist in a colonized habitat. Genotype B’s GMPP allows persistence in the colonized habitat, allowing time for adaptive genetic divergence.
Figure 4.3 An example of the effects of gene expression in two genes (bone morphogenetic protein 4, Bmp4, and calmodulin, CaM) on phenotypic traits of likely importance for ecological speciation: in Geospiza Darwin’s finches. A) Evidence for divergent selection on beak depth from reconstructions of adaptive landscapes. Lines depict the expected population density of a solitary granivorous finch species on two Galápagos Islands (similar results were observed on 13 other islands). Dots depict mean log beak depths of actual populations for each curve. Distinct peaks in the adaptive landscape indicate divergent selection, as supported by the observation that actual beak depths differ among populations and tend to correspond to peaks in the landscape. Thus, selection against migrants between environments and intermediate hybrids would likely cause reproductive isolation. Modified from Schluter and Grant (1984) and reprinted with permission of the American Society of Naturalists. B) Summary of the evidence that Bmp- and CaM-dependent signaling regulates growth along different axes of bill morphology, facilitating the evolution of distinct beak morphologies in Darwin’s finches. A beak of the sharp-beaked finch reflects a basal morphology for Geospiza. Abbreviations: C, caudal; D, dorsal; R, rostral; V, ventral. Modified from Abzhanov et al. (2006) and reprinted with permission of Nature.
CHAPTER 5: ECOLOGICAL TRANSCRIPTOMICS IN LAKE-TYPE AND RIVERINE SOCKEYE SALMON (ONCORHYNCHUS NERKA)

Author list: Scott A. Pavey, Ben J.G. Sutherland, Jong Leong, Troy R. Hamon, Ben F. Koop, and Jennifer L. Nielsen.

Author contribution: Scott A. Pavey planned the study, conducted the field and lab work, ran the analysis, and was the primary writer of the paper.

5.1 Abstract

There are a growing number of genomes sequenced with tentative functions assigned to a large proportion of the individual genes. Model organisms in laboratory settings form the basis for the assignment of gene function; the ecological context of gene function is lacking. This has left missing pieces in gene annotation because many genes may be in part or solely ecological in function. This paper provides an ecological context to expressed genes of sockeye salmon muscle tissue. We compared natural juvenile sockeye salmon ecotypes relating to river and lake habitats. Based on significantly divergent morphology, feeding strategy and predation in association with these distinct environments, we expect that burst swimming is favored in riverine life history and continuous swimming is favored in lake-type. In turn, we predict that genes associated with anaerobic metabolism are favored in riverine sockeye while genes related to aerobic metabolism are favored in lake-type sockeye. Gene expression patterns were measured by mRNA abundance in a 16k microarray
chip, discovering 141 genes with significant differential expression. Overall, the identity and function of these genes was consistent with our predictions. In addition, gene ontology (GO) enrichment analyses with a larger set of differentially expressed genes found the “biosynthesis” category enriched for the riverine population and the “metabolism” category enriched for the lake-type population. This study provides a starting point for more extensive, targeted studies determining the ecological function of genes in sockeye salmon.

**Keywords:** microarray, gene expression, ecological genomics, ecological speciation, population persistence.

### 5.2 Introduction

The field of genomics is expanding rapidly with full genome sequencing and Expressed Sequence Tag (EST) transcriptome sequencing, including many nonmodel species. Annotating these genomes continues to pose a challenge (Reeves et al. 2009). Due to sequence conservation of functional genes and the rapidly growing molecular knowledge base of model organisms, basic local alignment search tools (e.g. BLAST) facilitate the initial annotation of nonmodel genomes (Rokas and Abbot 2009). However, the function of many genes remains unknown, even in model organisms. For example, of 6000 genes in yeast, 1000 remain uncharacterized (Pena-Castillo and Hughes 2007). One reason may be that model organisms have rarely been studied outside of a laboratory (Kammenga et al. 2007; Landry and Aubin-Horth 2007; McKay and Stinchcombe 2008). Thus, genes that function primarily in natural settings have less representation.
Studies of fishes are leading the way in providing an ecological context to genomes [e.g., lake whitefish (Coregonus clupeaformis, Derome and Bernatchez 2006; Derome et al. 2006), Atlantic salmon (Salmo solar, Aubin-Horth et al. 2005a; Aubin-Horth et al. 2005b; Roberge et al. 2007), killifish (Fundulus heteroclitus, Oleksiak et al. 2002; Whitehead and Crawford 2006; Scott et al. 2009), and threespine stickleback (Gasterosteus aculeatus, Coyle et al. 2007; Chan et al. 2009)]. These studies have employed three basic methods to relate transcriptomes to ecological systems (Pavey et al. 2010a). First is in situ gene expression analysis (i.e. Derome et al. 2006). Sampling occurs in the ecological context of interest in nature; fish capture and RNA preservation occurs in the field. This method measures both genetic and environmental effects on the transcriptome and it is often not possible to assign gene expression differences to either source. A study design with replication reporting parallel expression differences between two systems reduces population and perhaps environment specific gene expression (e.g. Derome et al. 2006). In general, this method is applicable to many species including large or long-lived species where laboratory rearing or genetic crosses are not practical.

Second, one can remove the environmental effect and only test for genetic effects on gene expression in common garden experiments. This strategy compares the transcriptomes of genetically distinct ecotypes in controlled conditions (e.g. St-Cyr et al. 2008). This method is generally applied to species that can be reared artificially. Third, one can perform eQTL analyses by crossing genetically distinct ecotypes in a laboratory setting and mapping gene expression
phenotypes on linkage groups (e.g. Derome et al. 2008; Whiteley et al. 2008). This method requires artificial rearing and is only practical for species with short generation times. This is the only method able to determine the genetic architecture of gene expression. The latter two methods do not capture plastic responses to the environment and may miss heritable expression that requires certain environmental conditions to manifest.

Juvenile sockeye salmon exhibit a life history dichotomy in their freshwater rearing environments. Lake-type populations rear in lakes for 1-2 years before traveling to the ocean to feed and riverine populations rear in river habitats for 0-2 years (“sea-type” individuals going to sea before the first winter Wood 1995; Beacham et al. 2004). Foraging, water current, and predation differ between habitats (Murphy et al. 1989; Pavey et al. 2010c). Body shape differs between these life history types in association with the environment. In southwest Alaska, riverine sockeye exhibit a deep robust body whereas lake-type sockeye are more fusiform (Pavey et al. 2010c). This may be the result of both predation regime and a foraging strategy favoring burst swimming in riverine and continuous swimming in lake-type habitats. Similar morphological and behavioral differences are apparent within and among different species of Pacific salmon (Hoar 1954; Scott and Crossman 1973; Taylor and McPhail 1985; Swain and Holtby 1989).

A set of recent studies characterized the transcriptome in ecotypes of another salmonid, the lake whitefish; employing both the *in situ* and common garden approaches in dwarf and normal ecotypes (Derome et al. 2006; St-Cyr et al. 2008). The primary ecological trade-off between increased growth and
fecundity in the normal ecotype is increased energetics in the dwarf ecotype (Trudel et al. 2001; Rogers et al. 2002; Derome et al. 2006). The dwarf ecotype exhibits continuous swimming for feeding and is subjected to high predation compared to the normal ecotype (Rogers et al. 2002; Derome et al. 2006). Therefore, both continuous swimming for plankton foraging and burst swimming for predator avoidance are likely favored in the dwarf ecotype, resulting in energy expenditure for metabolism at the expense of growth (Lu and Bernatchez 1999; Trudel et al. 2001). This trade-off results in great differences in growth rate, age at maturity, body shape, and maximum lifespan.

Phenotypically, lake whitefish ecotypes have a drastically different size at age (Lu and Bernatchez 1999). The size distribution of the populations in this study overlap at this life stage but the riverine sockeye are longer and have a more robust body shape (Pavey et al. 2010c). With less extreme morphological differences in ecotypes of sockeye salmon juveniles, we expect the molecular trade-offs to be different from the lake whitefish studies. We expect genes differentially expressed to reflect the differing emphasis on continuous swimming for lake-type and burst swimming for riverine.

In this study, we compare the in situ transcriptome of two sockeye salmon populations in the same drainage that exhibit these divergent life histories. Differences foraging strategy and predation may have led to genetic (Pavey et al. 2007) and morphological (Pavey et al. 2010c) differences between these populations. We expect the transcriptome to reflect the functional molecular trade-offs driven by the ecological differences in these life histories. A greater
understanding of the molecular mechanisms that relate to functional ecology will enhance our understanding of the phenotypic diversity of this species as well as increase the knowledge base of the ecological function of genes.

5.3 Methods

5.3.1 Study Site and Field Collection

Aniakchak National Monument and Preserve (ANMP) in southwest Alaska provides a unique system to study these sockeye life history strategies (Fig. 5.1). The ANMP has undergone several recent geologic events (Fig. 5.1). A massive volcanic eruption 3,650 years ago formed a large caldera (Aniakchak Caldera) that filled with water creating a lake (McGimsey et al. 1994; Pearce et al. 2004). Approximately 1,800 years ago (VanderHoek and Myron 2004) the caldera wall collapsed resulting in a large flood and the formation of the Aniakchak River, which connects the remainder of the caldera lake (Surprise Lake; elevation 321 m) with the Pacific Ocean through “The Gates”, a chasm opened through the caldera wall by the flood (Hubbard 1931). A large fluvial plain was established when the passing flood dropped sediment as it exited the caldera. Several smaller eruptions have occurred, including well documented events approximately 500 and 79 b.p. (McGimsey et al. 1994). Sometime after the 500 b.p. eruption, lake-type sockeye salmon colonized Surprise Lake (SL) and used the lake for juvenile rearing (Pavey et al. 2010b). A riverine sockeye population rear in Albert Johnson Creek (AJC), the largest tributary of Aniakchak River (Pavey et al. 2007). Albert Johnson Creek is a low gradient stream that meets Aniakchak River at the base of the volcano in the large fluvial plain that was the
result of the caldera draining flood. Thus, current populations representing each of lake-type (SL) and riverine (AJC) life history types coexist in the same drainage.

Juvenile sockeye salmon were sampled on August 8\textsuperscript{th} 2007. The time of sampling for Albert Johnson Creek was 1535h to 1703h and Surprise Lake was 1832h to 2057h. The entire sampling effort took place within 5.5 hours including transportation from Albert Johnson Creek to Surprise Lake by a Cessna 185 airplane. A beach seine was used to capture fish and a strict sampling protocol including sampling time was enforced to reduce fish-to-fish sampling bias. We sampled fish of similar lengths from each site. Mean fork length was 45.9mm (SD: 3.5mm) for AJC and 45.0mm (SD 5.6mm) for SL. One fish from each seine haul was placed in a lethal solution of MS-222 (100 mg/l). An incision was made in the body cavity with a scalpel and the entire fish was placed in RNA\textsubscript{later}\textsuperscript{TM} (Ambion). The maximum time between netting a fish to RNA preservation was five minutes. The samples were kept cool in the field and transported, then frozen and stored at -20°C to -80°C until RNA extraction.

5.3.2 RNA Preparation

The samples were thawed and blotted with a Kimwipe\textsuperscript{®}. All muscle tissue was removed from each fish. Total RNA was extracted with a modified protocol of the Invitrogen TRIZOL\textsuperscript{®} Plus RNA purification kit using PureLink\textsuperscript{TM} Micro-to-Midi\textsuperscript{TM} columns. Disruption and homogenization were achieved with a MixerMill MM301 (Retsch). The manufacturer’s protocol was followed with the exception that 150 µl of chloroform and 150 µl of low Ph phenol were used to dissociate
tissue. The quality of all RNA samples was verified on a 1% agarose gel. All samples were quantified with a Spectrophotometer ND-1000 (NanoDrop).

5.3.3 cDNA and aRNA Synthesis and Labeling

cDNA was synthesized with Invitrogen SuperScript™ III Indirect cDNA labeling system kits per manufacturer’s instructions. In brief, 10 µg of total RNA was combined with a master mix cocktail including the retro transcriptase enzyme and oligo (dT) primers. This reaction was incubated for three hours at 46°C to synthesize single-stranded cDNA. The samples were cleaned following the S.N.A.P™ column purification procedure.

An RNA reference pool composed of juvenile muscle, juvenile liver and adult sockeye brain, muscle and liver was created. The reference RNA was amplified using an Amino Allyl MessageAmp™ II aRNA amplification Kit (Ambion) as per the manufacturers instructions. Briefly, single-stranded cDNA was synthesized from the RNA pool whereupon the second strand was synthesized with DNA polymerase. This product was purified through columns, and then amino allyl-modified aRNA was transcribed from the double-stranded cDNA. Sufficient aRNA was created such that every sample on the microarrays was compared to the same reference pool.

Sample and reference material was coupled with mono-reactive CyDye™ packs (GE Healthcare). In short, the reference aRNA was coupled with Cy3 and the sample cDNA with Cy5 dyes for one hour at 4°C. The samples were then purified to remove all uncoupled dye using S.N.A.P.™ columns as per
manufacturer’s instructions. The dye coupled sample and reference were stored at 4°C in the dark until commencement of hybridization.

5.3.4 Microarray Hybridization

We used the cGRASP 16k cDNA microarray to compare the transcriptome of these life histories. This microarray consists of 16,000 ESTs chosen from 300,000 Atlantic salmon and rainbow trout (Oncorhynchus mykiss) cDNA libraries (von Schalburg et al. 2005). The ESTs represent a variety of tissue types, developmental stages, and conditions. ESTs were chosen for minimum overlap, sequence quality, and other criteria detailed in von Schalburg et al. (2005). cDNA array elements are longer in size than in oligo arrays and are therefore superior when used for different species than which they were designed (von Schalburg et al. 2008).

We followed an established hybridization protocol for the 16k cDNA array (von Schalburg et al. 2005). Both 250 ng of reference aRNA and 500 ng of sample cDNA were collected in a single tube and kept dark. The mixture was concentrated with a speed vacuum and brought up to 23 μl with RNase free water (Gibco). Hybridization buffer #3 (Ambion) was heated to 65 °C while occasionally mixing for one hour. The heated buffer was then added to the collected sample, with LNAdt blocker (Genesphere) as per manufacturer’s instructions. Before the sample injection, the programmed Tecan washed the microarrays with several solutions containing first 1xSSC, then 0.1xSSC 0.014% SDS, then 5x SSC, 0.01% SDS, and 0.2% BSA. Samples were heated to 80°C for 5 minutes, then kept at 65°C until injected onto the pre-washed arrays in a
Tecan HS 4800 Pro, as per manufacturers’ instructions (Tecan). Microarrays were hybridized for 16 hours, and the full protocol for the hybridization can be viewed in Table 5.1.

Posthybridization, arrays were rinsed in the Tecan modules with increasingly stringent SSC and SDS solutions, starting with 2xSSC, 0.014%SDS for four washes incrementally decreasing temperature, then one final wash of 0.2xSSC at 23 °C. Finally, slides were dried with 37 psi nitrogen gas and kept dark until scanned. Current protocols for cGRASP microarrays are available at: (http://web.uvic.ca/grasp/microarray/protocols/tecan_hybridization_protocol.pdf)

5.3.5 Scanning and Quantifying

All microarrays were scanned immediately after hybridization was complete using a ScanArray Express (Perkin-Elmer). The microarray images were quantified manually with ImaGene 5.6.1 (BioDiscovery). Spots with unusual morphology, offset, or other poor quality parameters were flagged as marginal and excluded from downstream analyses.

5.3.6 Array Normalization and Statistical Analysis

We performed all analyses in GeneSpring GX 7.3 (Agilent). The arrays were normalized as per typical two-color experiments by performing an array-wide intensity dependent lowess normalization, followed by a per gene normalization, which normalized each spot to a median value. The average base/proportionate value was calculated to be an intensity of 72, so we filtered data to retain only entities with average raw signal expression values greater
than 72 in at least one of the populations. This became our base expression data for analysis. Our GeneSpring analysis was performed in two ways. First, the dataset was filtered to retain only the genes that were differentially expressed by \( \geq 1.5 \) fold. This list was then further filtered to exclude genes that did not meet a significance threshold of \( P = 0.05 \) in a t-test assuming unequal variances after a Benjamini and Hochberg False Discovery Rate to control for multiple tests (Benjamini and Yekutieli 2001). The tentative gene descriptions from the cGRASP 16k annotation file [current annotation files available at: http://web.uvic.ca/grasp/microarray/array.html (Koop et al. 2008)] were used to associate ESTs on the array with gene descriptions.

5.3.7 Gene Ontology Analysis

In order to gain an understanding of functional trends in gene expression, a less stringent filtering on the base gene list was performed. We included genes differentially expressed by any amount as identified by a t-test \( (P<0.05) \) but without any multiple test correction. This was less stringent in order to include a larger subset of entities to ensure a useable amount of gene ontology (GO) annotation. We then performed a GO enrichment analysis on this list of genes using the GO browser in GeneSpring. We produced GO categories significantly enriched \( (P<0.05) \) in the entity lists for GO slim categories of both “biological process” and “molecular function.”

We created a heat map and expression clustering tree of all individuals in GeneSpring using the same gene list as for the GO analysis. The similarity
measure is based on Pearson’s correlation coefficients with an average linkage clustering algorithm.

5.4 Results

One hundred and forty one ESTs exhibited significant differential expression (≥ 1.5 fold) between the riverine and lake populations (t-test with Benjamini Hochberg FDR P < 0.05). Of these, 81 were over-expressed in riverine compared with lake-type (Table 5.2) while 60 were over-expressed in lake-type compared with riverine (Table 5.3). The fold differences were modest, most of which were below two fold (< 2.1 fold over-expressed in AJC and < 2.9 fold over-expressed in SL; Tables 5.2, 5.3). In AJC, the genes with the highest fold change with corrected P values were: type II keratin E1 (CB510619; fold 2.1; P = 0.0317); CCAAT/enhancer-binding protein (CA050914; fold 1.9; P = 0.0442); and 60S ribosomal protein L9 (CA045500; fold 1.9; P = 0.0517). In SL the genes with the highest fold change were: structural maintenance of chromosomes protein 1B (CB488712; fold 2.9; P = 0.024); CD81 antigen (CA039936; fold 2.76; P = 0.033); collagen alpha-2(I) chain precursor (CB515159; fold 2.36, P = 0.028); ferritin, heavy subunit (CB505866; fold 2.33, P = 0.0438); and troponin I, slow skeletal muscle (CB509964; fold 2.29; P = 0.0390).

Using the less stringent filter to produce an entity list for GO enrichment analysis, we found 1026 genes significantly differentially expressed in muscle (P < 0.05, without correction for multiple tests). Of these, 498 genes were over-expressed in AJC compared with SL and 140 of which had GO annotations. In biological process category, biosynthesis (GO:9058) was the only GO category
disproportionately represented with 30 genes (P = 0.015). In the molecular function ontology, the only enriched category was structural molecular activity (GO:5198; 26 genes, P < 0.001).

We found 528 genes significantly over-expressed in SL muscle compared to AJC (P < 0.05, without correction for multiple tests). Of this list of 209 annotated genes, metabolism (GO:8152) was the only biological process category significantly enriched with 156 genes (P = 0.0044). There are five significant categories enriched from the molecular function GO slim category (Table 5.4).

The heat map based on all individuals, and grouped according to Pearson’s Correlation coefficients, matched the habitat type with the exception of a single individual from AJC that grouped with the rest of the SL samples (Fig. 5.2).

5.5 Discussion

We discovered differentially expressed genes and enriched functional categories associated with different life history types of sockeye salmon in two habitats. This work represents the first characterization of the molecular phenotype of juvenile sockeye in these common habitat types.

Sockeye from AJC are larger than SL and riverine sockeye in general and have a deep, robust body compared with lake-type (Pavey et al. 2010c) and some patterns in expression profiles in the present study reflect these phenotypes. For example, in AJC, nine ribosomal proteins were over-expressed
compared with SL and one of these (CA045500) was among the highest over-expressed in AJC (Table 5.2). In comparison, we did not identify any ribosomal proteins over-expressed in SL compared with AJC. Ribosomal proteins stabilize the structure composed mostly of ribosomal RNA (Alberts et al. 2002). Thus, the differential expression of these ribosomal proteins may indicate more protein synthesis in the muscle tissue of AJC sockeye. In addition, five genes associated with cell division and DNA replication were over-expressed in AJC. These patterns are consistent with faster growth and more muscle mass associated with the deeper body morphology in AJC sockeye (Pavey et al. 2010c). The GO category of biosynthesis (GO:9058) is defined as “The chemical reactions and pathways resulting in the formation of substances; typically the energy-requiring part of metabolism in which simpler substances are transformed into more complex ones” (http://amigo.geneontology.org). This was the sole enriched category in AJC of the GO slim biological process ontology which further underscores that the expression profile in AJC corresponds to increased production.

Creatine kinase (CB503498) was over-expressed in AJC compared with SL. This gene is potentially important in both aerobic metabolism in the pathway of oxidative phosphorylation, as well as anaerobic metabolism in glycolysis (Wallimann et al. 1992). However, in both processes this gene buffers the amount of cellular ATP and therefore facilitates fluctuating energy demands (Wallimann et al. 1992). This may be important to the riverine “wait and burst” feeding style of AJC, which may involve more variable levels of feeding activity.
In the SL gene enrichment analysis, the sole significant generic GO slim ontology is metabolism (GO:8152). Many of the individual genes in the over-expressed list relate to energy metabolism, mitochondria, and muscle contraction regulation. This is compatible with increased metabolism, especially for continuous swimming. Two of these genes may be particularly important for the continuous swimming strategy of lake-type sockeye; troponin 1 slow twitch (two separate ESTs) was present in replicate in the over-expression list (CB509964 and CB510901), increasing the reliability of this transcript’s abundance. Troponin is a gene that regulates muscle contraction and the “slow twitch” label of the annotation indicates that this EST is specific to slow twitch or aerobic muscle fiber (http://www.ncbi.nlm.nih.gov/protein/NP_003272.3). This could be the result of either increased red muscle fibers present, increased recruitment of red muscle fiber or both. Furthermore 5-aminolevulinitic acid synthase (CA 058106) is involved in heme biosynthesis (http://www.ncbi.nlm.nih.gov). In addition, 72 kDa type IV collagenase precursor (CB510651), implicated in blood vessel maturation (http://www.ncbi.nlm.nih.gov), is over-expressed which may lay the infrastructure for increased aerobic needs. Another SL over-expressed gene, selenoprotein K (CA054647), is a response to oxidative stress (http://www.uniprot.org/ uniprot/Q9Y6D0). The relatively large over-expression of ferritin, heavy subunit (CB505886; fold 2.3; P = 0.044) further demonstrates the importance of aerobic metabolism in SL. A different EST annotated as ferritin was over-expressed to a lesser extent in AJC (CB494485; fold 1.5; P = 0.047). We investigated this situation further, and found that the EST sequence was
incorrectly assembled into a ferritin contig. We blasted this EST directly and found that it is more likely to be another ribosomal protein. This demonstrates that problems do arise in the use of assembly programs, especially with a large number of ESTs and recent genome duplication (Li and Khuri 2004; Koop et al. 2008).

An unanticipated discovery was the increased expression of immune function genes in AJC, including two entities annotated with MHC II function (CB492871 and CA048654; Table 5.2). This may indicate differing immunity challenges in the river and lake rearing habitats of this study. This finding is a good example of the utility of large scale data production tools, such as microarrays in hypothesis generation. As microarrays facilitate screening of a large number of genes, they may uncover unexpected traits that are difficult to measure, even if identified as potential traits of interest during experimental design (Pavey et al. 2010a).

We detected significantly over-expressed transcription factors in AJC. Pro-B-cell leukemia transcription factor 2 (CB499801) and CCAAT/enhancer-binding protein delta (CA050914) regulate transcription (http://www.uniprot.org/uniprot/Q6IR52 and http://www.uniprot.org/uniprot/Q00322). These regulatory genes could have cascading effects in gene expression. Regulatory proteins such as these may interact with DNA sequences thousands of base pairs up or downstream of promoter regions for one or many genes. Regulatory proteins are part of the very complex interactions of transcriptional regulation (Alberts et al. 2002).
Our results yield both similarities and differences when compared to the gene expression work on lake whitefish (Derome et al. 2006; St-Cyr et al. 2008). The morphological and expected ecological differences in juvenile sockeye salmon are not nearly as extreme as those observed in lake whitefish, which are drastically different in growth rate and age of maturity. However, like the present study, the fold change differences between ecotypes in the lake whitefish work are modest, suggesting this may be the norm for ecological transcriptomic differences in natural populations. Unlike lake whitefish, sockeye salmon are anadromous, and our study populations move to the ocean after freshwater rearing where feeding environments and access to them may be similar (Burgner 1991). Therefore, differences at the juvenile rearing stage may be limited because this is only one part of a complex life history, and the life history types may developmentally converge for the ocean feeding stage.

In lake whitefish, parvalbumin beta (AF538283) was the only gene involved with muscle contraction regulation that was consistently over-expressed in the dwarf ecotype. We did not find evidence of over-expression of this gene in SL, but another gene involved with muscle contraction, the slow twitch isomer of troponin, was significantly over-expressed in SL in two separate positions on the microarray. It is expected that feeding strategy promotes continuous swimming in dwarf lake whitefish (Lu and Bernatchez 1999). In addition, dwarf lake whitefish are under high predation compared with the normal ecotype, an ecological attribute responsible for increased burst swimming. This should favor both aerobic and anaerobic metabolism in the same ecotype resulting in selection that
favors overall increased metabolism and muscle contraction (Lu and Bernatchez 1999). In the present study, high predation and a burst swimming feeding strategy are expected only in riverine AJC, whereas a continuous swimming strategy and low predation should occur in lake-type SL. The relatively large over-expression of ferritin, heavy subunit (CB505886; fold 2.3; P = 0.044) further demonstrates the importance of aerobic metabolism in SL. This results in finer partitioning of swimming energetics in sockeye salmon compared with lake whitefish.

The two Heat map groups corresponded to their respective populations with the exception of a single individual (Fig. 5.2). If most of the genes in this study are primarily under genetic control as found in liver tissue of whitefish ecotypes (St-Cyr et al. 2008), this may indicate that the individual is the offspring of a lake-type migrant. The heat map demonstrates the large amount of gene expression variability, even among individuals within populations. This is likely to be the norm in natural populations (Oleksiak et al. 2002), particularly for populations in non equilibrium conditions (Haas and McPhail 2001) and is an important consideration in any gene expression experiment.

We have developed the first dataset characterizing gene expression differences between two populations of sockeye salmon representing lake-type and riverine life history. Although this represents a first step in viewing the ecological transcriptomic differences of juvenile sockeye, we have already identified clear patterns relating to the divergent ecological phenotypes of these populations. The relative contributions of genetic and environmental effects are
unknown and an opportunity for future study. The generality of these results may be determined through testing more populations in a similar manner. Our findings represent correlations at present; assigning specific genes to ecological function will require more directed experiments. These relationships represent good starting points for formulating hypotheses and experiments that rigorously test the relationship between gene function and ecology.

5.6 Acknowledgments

We thank the Evolution Discussion Group (FAB*-Lab) at Simon Fraser University and the Vancouver Evolution Group (VEG) for critical discussions of this work. We thank J.E. Krebs, R. Rumelhart and the Staff of Katmai National Park for field assistance. We thank W.D. Noon and R. Wood for help with graphics. We thank the entire lab of Dr. B. Koop, including G. Cooper and K. von Schalburg for training, lab space, use of equipment and discussions about microarrays. This work was funded by the U.S. Geological Survey, the National Park Service, and Natural Sciences and Engineering Research Council operating grant to F. Breden, Simon Fraser University.

5.7 Literature Cited


156


5.8 Tables
Table 5.1 Tecan HS 400 Pro hybridization protocol.

<table>
<thead>
<tr>
<th>Step</th>
<th>Action</th>
<th>Temp. °C</th>
<th>Solution</th>
<th>Wash/Agitation Time</th>
<th>Soak Time</th>
<th>Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wash</td>
<td>23</td>
<td>0.1xSSC, 0.014% SDS</td>
<td>0:00:30</td>
<td>0:00:30</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>wash</td>
<td>23</td>
<td>0.02xSSC</td>
<td>0:01:00</td>
<td>0:00:30</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5xSSC, 0.01% SDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>wash</td>
<td>49</td>
<td>0.2% BSA</td>
<td>0:01:00</td>
<td>0:00:00</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>hybridize</td>
<td>49</td>
<td></td>
<td>1:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>wash</td>
<td>46</td>
<td>2xSSC, 0.014% SDS</td>
<td>0:01:00</td>
<td>0:00:30</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>inject</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>sample</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>hybridize</td>
<td>49</td>
<td></td>
<td>1:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>hybridize</td>
<td>53</td>
<td></td>
<td>1:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>hybridize</td>
<td>49</td>
<td></td>
<td>4:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>hybridize</td>
<td>53</td>
<td></td>
<td>1:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>hybridize</td>
<td>49</td>
<td></td>
<td>4:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>hybridize</td>
<td>53</td>
<td></td>
<td>1:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>hybridize</td>
<td>49</td>
<td></td>
<td>4:00:00</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>wash</td>
<td>49</td>
<td>2xSSC, 0.014% SDS</td>
<td>0:01:00</td>
<td>0:00:00</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>hybridize</td>
<td>49</td>
<td></td>
<td>0:03:00</td>
<td>0:00:00</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>wash</td>
<td>49</td>
<td>2xSSC, 0.014% SDS</td>
<td>0:00:30</td>
<td>0:00:00</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>wash</td>
<td>39</td>
<td>2xSSC, 0.014% SDS</td>
<td>0:00:20</td>
<td>0:00:00</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>wash</td>
<td>30</td>
<td>2xSSC, 0.014% SDS</td>
<td>0:00:20</td>
<td>0:00:00</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>wash</td>
<td>23</td>
<td>1xSSC</td>
<td>0:00:20</td>
<td>0:00:00</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>wash</td>
<td>23</td>
<td>0.02xSSC</td>
<td>0:00:30</td>
<td>0:00:00</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>dry</td>
<td>23</td>
<td></td>
<td>0:02:30</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Table 5.2 Genes significantly over-expressed in Albert Johnson Creek sockeye salmon muscle compared to Surprise Lake sockeye salmon muscle (table continues on the following two pages).

<table>
<thead>
<tr>
<th>Function</th>
<th>Fold Change (Up in AJC)</th>
<th>P Value</th>
<th>Genbank</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell division</td>
<td>1.778</td>
<td>0.0452</td>
<td>CB515443</td>
<td>Nuclear autoantigenic sperm protein</td>
</tr>
<tr>
<td>Cell division</td>
<td>1.678</td>
<td>0.0412</td>
<td>CA060603</td>
<td>Katanin p80 WD40-containing subunit B1</td>
</tr>
<tr>
<td>Cell division</td>
<td>1.518</td>
<td>0.0317</td>
<td>CA059189</td>
<td>Golgi reassembly-stacking protein 2</td>
</tr>
<tr>
<td>Cell structure</td>
<td>2.118</td>
<td>0.0317</td>
<td>CB510619</td>
<td>Oncorhynchus mykiss mRNA for type II keratin E1 (E1 gene)</td>
</tr>
<tr>
<td>Development</td>
<td>1.755</td>
<td>0.0346</td>
<td>CB498195</td>
<td>Reticulon-4</td>
</tr>
<tr>
<td>Development</td>
<td>1.644</td>
<td>0.0306</td>
<td>CA052389</td>
<td>Plexin-A3 precursor</td>
</tr>
<tr>
<td>DNA Replication</td>
<td>1.738</td>
<td>0.0406</td>
<td>CA053921</td>
<td>Girdin</td>
</tr>
<tr>
<td>DNA Replication</td>
<td>1.722</td>
<td>0.0317</td>
<td>CB494447</td>
<td>Poly [ADP-ribose] polymerase reductase</td>
</tr>
<tr>
<td>Energetic Metabolism</td>
<td>1.796</td>
<td>0.0493</td>
<td>CA064415</td>
<td>Methionine-R-sulfoxide reductase</td>
</tr>
<tr>
<td>Energetic Metabolism</td>
<td>1.685</td>
<td>0.0406</td>
<td>CB503498</td>
<td>Creatine kinase, testis isozyme</td>
</tr>
<tr>
<td>Iron Binding</td>
<td>1.657</td>
<td>0.0329</td>
<td>CB501208</td>
<td>Hemoglobin subunit alpha-4</td>
</tr>
<tr>
<td>Iron Binding</td>
<td>1.540</td>
<td>0.0466</td>
<td>CB494485</td>
<td>Ferritin, heavy subunit</td>
</tr>
<tr>
<td>Immune</td>
<td>1.893</td>
<td>0.0317</td>
<td>CA063034</td>
<td>Transient receptor potential cation channel subfamily M member 4</td>
</tr>
<tr>
<td>Immune</td>
<td>1.722</td>
<td>0.0317</td>
<td>CB492871</td>
<td>Oncorhynchus mykiss mRNA for MHC class II alpha (onmy-DAA*02 gene)</td>
</tr>
<tr>
<td>Immune</td>
<td>1.597</td>
<td>0.0449</td>
<td>CA048654</td>
<td>Oncorhynchus mykiss SYPG1 MHCII-alpha and Raftlin-like pseudogenes TIM21-like protein, mitochondrial precursor</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>1.873</td>
<td>0.0412</td>
<td>CB489874</td>
<td>NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>1.528</td>
<td>0.0282</td>
<td>CB498084</td>
<td>NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4</td>
</tr>
<tr>
<td>Protein folding</td>
<td>1.644</td>
<td>0.0393</td>
<td>CA052325</td>
<td>Prefoldin subunit 2</td>
</tr>
<tr>
<td>Protein transport</td>
<td>2.103</td>
<td>0.0282</td>
<td>CB491150</td>
<td>Kinesin-like protein KIF20A</td>
</tr>
<tr>
<td>Protein transport</td>
<td>1.911</td>
<td>0.0306</td>
<td>CA051435</td>
<td>Ras-related protein Rab-14</td>
</tr>
<tr>
<td>Protein transport</td>
<td>1.674</td>
<td>0.0393</td>
<td>CA054426</td>
<td>ADP-ribosylation factor 1</td>
</tr>
<tr>
<td>Signal</td>
<td>1.544</td>
<td>0.0317</td>
<td>CB514723</td>
<td>G-protein coupled receptor APJ homolog</td>
</tr>
<tr>
<td>Transcription factor (neg)</td>
<td>1.652</td>
<td>0.0452</td>
<td>CB499801</td>
<td>Pre-B-cell leukemia transcription factor 2</td>
</tr>
<tr>
<td>Transcription regulation</td>
<td>1.974</td>
<td>0.0442</td>
<td>CA050914</td>
<td>CCAAT/enhancer-binding protein delta</td>
</tr>
<tr>
<td>Transcription regulation</td>
<td>1.903</td>
<td>0.0306</td>
<td>CB498272</td>
<td>Nucleolar protein 5A</td>
</tr>
<tr>
<td>Translation</td>
<td>1.967</td>
<td>0.0317</td>
<td>CA045500</td>
<td>60S ribosomal protein L9</td>
</tr>
<tr>
<td>Translation</td>
<td>1.719</td>
<td>0.0449</td>
<td>CB514402</td>
<td>60S ribosomal protein L19</td>
</tr>
<tr>
<td>Function</td>
<td>Fold Change (Up in AJC)</td>
<td>P Value</td>
<td>Genbank</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------</td>
<td>---------</td>
<td>----------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>Translation</td>
<td>1.711</td>
<td>0.0306</td>
<td>CB493907</td>
<td>40S ribosomal protein S19</td>
</tr>
<tr>
<td>Translation</td>
<td>1.632</td>
<td>0.0384</td>
<td>CB503205</td>
<td>60S ribosomal protein L4-B</td>
</tr>
<tr>
<td>Translation</td>
<td>1.613</td>
<td>0.0282</td>
<td>CA770261</td>
<td>60S ribosomal protein L23</td>
</tr>
<tr>
<td>Translation</td>
<td>1.538</td>
<td>0.0493</td>
<td>CA770402</td>
<td>60S ribosomal protein L15</td>
</tr>
<tr>
<td>Translation</td>
<td>1.538</td>
<td>0.0406</td>
<td>CB493600</td>
<td>40S ribosomal protein S30</td>
</tr>
<tr>
<td>Translation</td>
<td>1.528</td>
<td>0.0282</td>
<td>CA055741</td>
<td>60S ribosomal protein L9</td>
</tr>
<tr>
<td>Translation</td>
<td>1.518</td>
<td>0.0282</td>
<td>CA037622</td>
<td>60S ribosomal protein L36</td>
</tr>
<tr>
<td>Transmembrane</td>
<td>1.536</td>
<td>0.0397</td>
<td>CA041663</td>
<td>pfam07690, MFS_1, Major Facilitator Superfamily</td>
</tr>
<tr>
<td>Many functions</td>
<td>1.502</td>
<td>0.0452</td>
<td>CA052412</td>
<td>Ectodysplasin-A</td>
</tr>
<tr>
<td>Many functions</td>
<td>1.630</td>
<td>0.0452</td>
<td>CA061568</td>
<td>Nucleophosmin</td>
</tr>
<tr>
<td>Many functions</td>
<td>1.614</td>
<td>0.0452</td>
<td>CB515428</td>
<td>Tartrate-resistant acid phosphatase type 5 precursor</td>
</tr>
<tr>
<td>Many functions</td>
<td>1.526</td>
<td>0.0317</td>
<td>CA047582</td>
<td>Somatotropin precursor</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.842</td>
<td>0.0452</td>
<td>CB490454</td>
<td>Zona pellucida sperm-binding protein 3 precursor</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.775</td>
<td>0.0380</td>
<td>CB511161</td>
<td>Voltage-dependent anion-selective channel protein 2</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.762</td>
<td>0.0317</td>
<td>CA044554</td>
<td>Prothymosin alpha</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.752</td>
<td>0.0412</td>
<td>CB511880</td>
<td>Serine/threonine-protein kinase Haspin</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.677</td>
<td>0.0331</td>
<td>CA055729</td>
<td>Transmembrane protein 178 precursor</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.640</td>
<td>0.0317</td>
<td>CB499656</td>
<td>Dual specificity mitogen-activated protein kinase kinase 6</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.630</td>
<td>0.0317</td>
<td>CB492596</td>
<td>Proteasome subunit alpha type 4</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.614</td>
<td>0.0419</td>
<td>CB494074</td>
<td>Oncorhynchus mykiss clone Glan 1 transposon e</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.574</td>
<td>0.0471</td>
<td>CB497649</td>
<td>Nucleoside diphosphate kinase B</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.536</td>
<td>0.0365</td>
<td>CB514435</td>
<td>chromosome region 16 protein homolog</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.528</td>
<td>0.0466</td>
<td>CB498610</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.518</td>
<td>0.0393</td>
<td>CA059713</td>
<td>Probable RNA-directed DNA polymerase from transposon BS</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.513</td>
<td>0.0282</td>
<td>CB514112</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.507</td>
<td>0.0452</td>
<td>CA058231</td>
<td>Antolefinin</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.891</td>
<td>0.0483</td>
<td>CB490094</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.872</td>
<td>0.0282</td>
<td>CA058259</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.816</td>
<td>0.0442</td>
<td>CB500083</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.765</td>
<td>0.0282</td>
<td>CA040487</td>
<td>PREDICTED: similar to Keratin-associated protein 10-1</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.730</td>
<td>0.0452</td>
<td>CB509719</td>
<td>PREDICTED: similar to CC chemokine SCYA103 [Danio rerio]</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.723</td>
<td>0.0317</td>
<td>CA054597</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.715</td>
<td>0.0380</td>
<td>CA051475</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.698</td>
<td>0.0331</td>
<td>CA061924</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.695</td>
<td>0.0282</td>
<td>CB497128</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.693</td>
<td>0.0449</td>
<td>CB510709</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Function</td>
<td>Fold Change (Up in AJC)</td>
<td>P Value</td>
<td>Genbank</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------</td>
<td>---------</td>
<td>-----------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.674</td>
<td>0.0452</td>
<td>CA041505</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.660</td>
<td>0.0420</td>
<td>CB494690</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.657</td>
<td>0.0452</td>
<td>CA039908</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.634</td>
<td>0.0282</td>
<td>CB492905</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.626</td>
<td>0.0306</td>
<td>CB511393</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.619</td>
<td>0.0485</td>
<td>CB513822</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.607</td>
<td>0.0323</td>
<td>CA050842</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.605</td>
<td>0.0449</td>
<td>CK990538</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.604</td>
<td>0.0452</td>
<td>CB514071</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.565</td>
<td>0.0317</td>
<td>CA057262</td>
<td>PREDICTED: similar to Protein C14orf159, mitochondrial precursor [Rattus norvegicus]</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.558</td>
<td>0.0384</td>
<td>CB492336</td>
<td>PREDICTED: similar to Rsbn1 protein [Danio rerio]</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.538</td>
<td>0.0290</td>
<td>CA039963</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.537</td>
<td>0.0452</td>
<td>CA058522</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.501</td>
<td>0.0412</td>
<td>CK991114</td>
<td>UNKNOWN</td>
</tr>
</tbody>
</table>
Table 5.3 Genes significantly over-expressed in Surprise Lake sockeye salmon muscle compared to Albert Johnson Creek muscle (table continues on the following two pages).

<table>
<thead>
<tr>
<th>Function</th>
<th>Fold change (Up in SL)</th>
<th>P</th>
<th>Genbank</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell interactions</td>
<td>2.755</td>
<td>0.0330</td>
<td>CA039936</td>
<td>CD81 antigen</td>
</tr>
<tr>
<td>Cell interactions</td>
<td>2.358</td>
<td>0.0281</td>
<td>CBS15159</td>
<td>Collagen alpha-2(I) chain precursor</td>
</tr>
<tr>
<td>Cell interactions</td>
<td>1.698</td>
<td>0.0281</td>
<td>CB488646</td>
<td>Cysteine-rich protein 1</td>
</tr>
<tr>
<td>Cell interactions</td>
<td>1.543</td>
<td>0.0228</td>
<td>CB498736</td>
<td>Cysteine-rich protein 1</td>
</tr>
<tr>
<td>DNA</td>
<td>2.907</td>
<td>0.0241</td>
<td>CB488712</td>
<td>Structural maintenance of chromosomes protein 1B</td>
</tr>
<tr>
<td>DNA</td>
<td>2.252</td>
<td>0.0051</td>
<td>CB490371</td>
<td>Histone H3.3</td>
</tr>
<tr>
<td>DNA</td>
<td>1.610</td>
<td>0.0199</td>
<td>CA051642</td>
<td>Ribonucleoside-diphosphate reductase M2 subunit</td>
</tr>
<tr>
<td>Energetic Metabolism</td>
<td>1.825</td>
<td>0.0360</td>
<td>CA041073</td>
<td>Calcium/calmodulin-dependent protein kinase type II delta chain</td>
</tr>
<tr>
<td>Energetic Metabolism</td>
<td>1.684</td>
<td>0.0281</td>
<td>CB498472</td>
<td>Selenide, water dikinase 2</td>
</tr>
<tr>
<td>Energetic Metabolism</td>
<td>1.553</td>
<td>0.0199</td>
<td>CA064428</td>
<td>6-phosphogluconate dehydrogenase, decarboxylating</td>
</tr>
<tr>
<td>Energetic Metabolism</td>
<td>1.534</td>
<td>0.0199</td>
<td>CA061459</td>
<td>cAMP-dependent protein kinase, beta-2-catalytic subunit</td>
</tr>
<tr>
<td>Energetic Metabolism</td>
<td>1.508</td>
<td>0.0356</td>
<td>CA049006</td>
<td>SUMO-activating enzyme subunit 2</td>
</tr>
<tr>
<td>Iron Binding</td>
<td>2.326</td>
<td>0.0438</td>
<td>CB505886</td>
<td>Ferritin, heavy subunit</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>2.053</td>
<td>0.0281</td>
<td>CA056342</td>
<td>Carnitine O-palmitoyltransferase 2, mitochondrial precursor</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>1.745</td>
<td>0.0438</td>
<td>CA060901</td>
<td>Adenylate kinase isoenzyme 2, mitochondrial</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>1.623</td>
<td>0.0356</td>
<td>CA062017</td>
<td>Single-stranded DNA-binding protein, mitochondrial precursor</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>2.083</td>
<td>0.0390</td>
<td>CA058136</td>
<td>5-aminolevulinate synthase, nonspecific, mitochondrial precursor</td>
</tr>
<tr>
<td>Aerobic infrastructure</td>
<td>2.208</td>
<td>0.0241</td>
<td>CBS10651</td>
<td>72 kDa type IV collagenase precursor</td>
</tr>
<tr>
<td>Lipid catabolism</td>
<td>1.580</td>
<td>0.0330</td>
<td>CA038195</td>
<td>Phospholipase A2, acidic 1 precursor</td>
</tr>
<tr>
<td>Muscle contraction regulation</td>
<td>2.294</td>
<td>0.0390</td>
<td>CB5099964</td>
<td>Troponin I, slow skeletal muscle</td>
</tr>
<tr>
<td>Muscle contraction regulation</td>
<td>1.613</td>
<td>0.0438</td>
<td>CBS10901</td>
<td>Troponin I, slow skeletal muscle</td>
</tr>
<tr>
<td>Organelle movement</td>
<td>1.672</td>
<td>0.0199</td>
<td>CB498105</td>
<td>Tubulin alpha-1C chain</td>
</tr>
<tr>
<td>Protein breakdown</td>
<td>1.520</td>
<td>0.0371</td>
<td>CK991067</td>
<td>Cathepsin H precursor</td>
</tr>
<tr>
<td>Protein breakdown</td>
<td>1.504</td>
<td>0.0345</td>
<td>CK990590</td>
<td>Trypsin precursor</td>
</tr>
<tr>
<td>Protein transport</td>
<td>1.575</td>
<td>0.0470</td>
<td>CA042407</td>
<td>Protein disulfide-isomerase A6 precursor</td>
</tr>
<tr>
<td>Function</td>
<td>Fold change (Up in SL)</td>
<td>P</td>
<td>Genbank</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------</td>
<td>---------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Protein transport</td>
<td>1.570</td>
<td>0.0199</td>
<td>CA044589</td>
<td>Protein transport protein Sec61 subunit beta serum lectin isoform 3 precursor [Salmo salar]</td>
</tr>
<tr>
<td>Sugar binding</td>
<td>1.515</td>
<td>0.0390</td>
<td>CB505852</td>
<td>Sugar binding 1.515 0.0390 CB505852 serum lectin isoform 3 precursor [Salmo salar]</td>
</tr>
<tr>
<td>Transcription (neg)</td>
<td>1.961</td>
<td>0.0470</td>
<td>CA054647</td>
<td>Selenoprotein K</td>
</tr>
<tr>
<td>Transcription regulation</td>
<td>1.592</td>
<td>0.0390</td>
<td>CB494071</td>
<td>14-3-3-like protein GF14-F</td>
</tr>
<tr>
<td>Translation</td>
<td>1.721</td>
<td>0.0051</td>
<td>CB516915</td>
<td>Eukaryotic translation initiation factor 3 subunit 7</td>
</tr>
<tr>
<td>Translation</td>
<td>1.590</td>
<td>0.0199</td>
<td>CA061402</td>
<td>Pseudouridylate synthase 7 homolog</td>
</tr>
<tr>
<td>Transport</td>
<td>2.037</td>
<td>0.0273</td>
<td>CB496796</td>
<td>Transmembrane emp24 domain-containing protein 3 precursor</td>
</tr>
<tr>
<td>Transport</td>
<td>1.880</td>
<td>0.0438</td>
<td>CB512385</td>
<td>Vacuolar protein sorting-associated protein 41 homolog</td>
</tr>
<tr>
<td>Many functions</td>
<td>1.845</td>
<td>0.0103</td>
<td>CA050496</td>
<td>Transitional endoplasmic reticulum ATPase</td>
</tr>
<tr>
<td>Many functions</td>
<td>1.704</td>
<td>0.0019</td>
<td>CA048664</td>
<td>Protein C-ets-1</td>
</tr>
<tr>
<td>Many functions</td>
<td>1.678</td>
<td>0.0371</td>
<td>CB515873</td>
<td>Dihydropyrimidinase</td>
</tr>
<tr>
<td>Other functions</td>
<td>2.105</td>
<td>0.0374</td>
<td>CA051433</td>
<td>Transmembrane and ubiquitin-like domain-containing protein 2</td>
</tr>
<tr>
<td>Other functions</td>
<td>2.016</td>
<td>0.0345</td>
<td>CA056436</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.976</td>
<td>0.0327</td>
<td>CB498745</td>
<td>KH domain-containing, RNA-binding, signal transduction-associated protein 1</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.855</td>
<td>0.0371</td>
<td>CA044775</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.692</td>
<td>0.0327</td>
<td>CB516919</td>
<td>Extracellular matrix protein 1 precursor Lithognathus mormyrus clone lmos2p08h02 mRNA sequence</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.645</td>
<td>0.0181</td>
<td>CA056626</td>
<td>Kelch-like protein 6</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.570</td>
<td>0.0494</td>
<td>CB516580</td>
<td>Kelch-like protein 6</td>
</tr>
<tr>
<td>Other functions</td>
<td>1.502</td>
<td>0.0199</td>
<td>CA062511</td>
<td>Translocon-associated protein subunit alpha precursor</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.808</td>
<td>0.0330</td>
<td>CB502569</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.789</td>
<td>0.0199</td>
<td>CA041684</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.751</td>
<td>0.0104</td>
<td>CK991281</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.727</td>
<td>0.0199</td>
<td>CB516202</td>
<td>Sterile alpha motif domain-containing protein 9-like PREDICTED: similar to Transforming growth factor, beta-induced [Danio rerio] &gt;</td>
</tr>
<tr>
<td>Function</td>
<td>Fold change (Up in SL)</td>
<td>P</td>
<td>Genbank</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------</td>
<td>------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.536</td>
<td>0.0438</td>
<td>CA040470</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.534</td>
<td>0.0199</td>
<td>CB492594</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.517</td>
<td>0.0438</td>
<td>CB515453</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td>Category</td>
<td>Genes on chip</td>
<td>% on chip</td>
<td>Genes in list</td>
<td>% of Genes in list</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>-----------</td>
<td>---------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>GO:5489: electron transporter activity</td>
<td>125</td>
<td>1.8</td>
<td>10</td>
<td>3.7</td>
</tr>
<tr>
<td>GO:8135: translation factor activity, nucleic acid binding</td>
<td>93</td>
<td>1.4</td>
<td>11</td>
<td>4.1</td>
</tr>
<tr>
<td>GO:8233: peptidase activity</td>
<td>397</td>
<td>5.8</td>
<td>23</td>
<td>8.6</td>
</tr>
<tr>
<td>GO:3824: catalytic activity</td>
<td>2647</td>
<td>38.8</td>
<td>124</td>
<td>46.1</td>
</tr>
</tbody>
</table>
5.9 Figures
Figure 5.1 The two sampling locations: Surprise Lake and Albert Johnson Creek in southwest Alaska.
Figure 5.2 Heat map depicting all of 1026 genes significantly differentially expressed (1.5 fold; \( P = 0.05 \); not corrected for multiple comparisons) between Albert Johnson Creek (AJC) and Surprise Lake (SL).
CHAPTER 6: GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusions

After many years of the biological sciences dividing and subdividing into specialized subdisciplines, the future holds great potential in the reuniting and synthesis of disciplines to explore processes and answer driving questions about ecology and evolution. In this Ph.D. dissertation, I explored life history of sockeye salmon, combining techniques of morphology, genetics, and genomics. The result is a unique survey of the interaction of ecology and evolution that underscores the diversity within species.

I found that sockeye may naturally colonize and rapidly adapt to new habitats (Pavey et al. 2010b). This work demonstrated that phenotypic change in egg size and body depth in predictable ecological directions occurred in less than 100 generations. Moreover, this change occurred despite gene flow between the diverging populations.

I established the first morphological comparison between lake-type and riverine juveniles (Pavey et al. 2010c). Body shape differed in a similar fashion to differences among species of *Oncorhynchus*. Differences in body depth and caudal peduncle shape functionally relate to divergent predation regime, habitat complexity and feeding ecology.
Though our understanding of gene function is expanding exponentially, nearly all of what is known is based on model organisms in a laboratory setting. Gene expression measurement is a tool that facilitates seeing what was previously difficult or impossible to measure. Applying these technologies to diverging populations and species will result in a deeper understanding of the process of ecological speciation as well as the ecology of gene function. I reviewed the state of this application as it relates to ecological speciation (Pavey et al. 2010a).

Finally, I applied gene expression microarrays to the divergent juvenile life history and sampled these populations in their natural divergent environments. I found genes that are differentially expressed in muscle tissue that have a functional relation to feeding ecology and morphology.

The result of this multidisciplinary approach is a holistic characterization of life history differences in sockeye salmon. The combination of these techniques provide novel understanding of the ecology and evolution of sockeye life history.

6.2 Future Directions for the Field

Since the fairly recent resolution of the salmonid phylogeny (Ramsden et al. 2003; Crespi and Fulton 2004), meaningful comparisons among species are now possible. Linkage maps and karyotypes reveal chromosomal fragmenting and fusing in drastic rearrangement among species of Pacific salmon, Atlantic salmon and char (Phillips et al. 2009). Full genome sequences will soon be available to further elevate salmonids as a model for evolution and ecology.
Specifically, there will be opportunities to study the relationship of genome duplications and subsequent rearrangement of genomes on phenotypes, ecology, and diversification.

The application of this and future work will stretch far beyond the bounds of sockeye salmon and even salmonids. Given that much of the diversification present on earth may be the result of ecological speciation, so understanding how genes interact with ecology will have rippling impact across many taxa. It will be imperative that a public database is created that places genes in an ecological context.

Despite the long standing model of salmonids for ecology, evolution, and physiology, recent developments have led to drastic realizations. Anadromy, previously thought to be ancestral in Atlantic salmon, now appears to be derived, with freshwater residency more likely ancestral (Aubin-Horth et al. 2005). This may be ancestral to all salmonids, as recent data suggests pike (Esociformes) is sister to salmonids, and not Osmeroidei as previously thought (Ramsden et al. 2003).

I expect major insights such as these to become more frequent with this multidisciplinary, holistic approach to ecology and evolution. Through collaborative efforts, we will use the salmonid models to understand the interaction of ecology, evolution and genes.
6.3 Literature Cited


