INVESTIGATING THE DIVERGENCE OF REPRODUCTIVE ECOTYPES IN KOKANEE SALMON

by

Karen Kiyomi Frazer

B.Sc., Dalhousie University, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The College of Graduate Studies

(Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

June 2012

ABSTRACT

Investigating the role of natural selection in driving adaptive diversification has become a central theme in evolutionary ecology as advancements in genome typing technologies provide new approaches for identifying the genetic basis of phenotypic diversity in non-model organisms. I used population-based genome scans to investigate the recent divergence of shore- and stream-spawning ecotypes in kokanee salmon (*Oncorhynchus nerka*). These ecotypes co-exist in many post-glacial lakes throughout their range, and exhibit distinct reproductive behaviors, spawning habitat preference, and life history traits. Several algorithms were used to test for statistical outliers across five replicate ecotype pairs of kokanee salmon. Among 50 expressed sequence tag (EST)-linked and anonymous microsatellite loci, signatures of directional selection were observed at 15 loci, including four loci that exhibited outlier behaviour across two or more lakes. The inconsistency of parallel patterns suggests that either several different genes or genetic pathways underlie ecotype divergence or outlier-detection methods are prone to Type II error when selection is weak. Nonetheless, population structure and differentiation at outlier loci is distinct from that of neutral loci, which infers that outliers may be under selection. Annotations of EST-linked outliers suggest that energy metabolism and pathogen resistance may be involved in initiating and maintaining barriers to gene flow between these two reproductive ecotypes.

Within a kokanee fisheries management context, markers associated with adaptive genetic variation would be very useful since neutral microsatellite markers cannot distinguish these recently diverged ecotypes (<10,000 years). Currently, the absolute abundance of shore-spawning kokanee cannot be determined using conventional methods and ecotypes cannot be determined for angled fish to estimate ecotype-specific harvest rates. Here, I assess the accuracy and power of outlier loci in distinguishing shore- and stream-spawning kokanee using mixed stock analyses and individual assignment tests. In general, outlier loci outperform neutral loci and simulations suggest that management-relevant levels of accuracy (>90%) may be achieved with sufficient baseline sampling and ecotype differentiation. Thus, genome scans can be useful in identifying informative markers for recently diverged stocks.

PREFACE

Multiple individuals have contributed to Chapter 2 and 3 of this thesis. Manuscript versions of these two chapters will be co-authored when submitted for publication. My contribution to the identification and design of this research is shared with my co-authors. I have taken lead responsibility for performing the research, data collection, data analysis, and manuscript preparation.

Some data from Chapter 2 was included in a research article published in Evolutionary Applications:

- Russello, MA, Kirk, S, Frazer, KK, and Askey, P (2012) Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications* **5**:39-52.

I collected a portion of the data and provided comments on the manuscript. Michael Russello and Stephanie Kirk designed the study, collected data, analyzed the data and wrote the manuscript. Paul Askey facilitated sample collection and also provided comments on the manuscript.

The data presented in this thesis were collected from fish sampled according to the animal care protocol of the University of British Columbia Research Ethics Board (# A07-0088 and # A11-0127) and a fish collection permit (# SM10-66091) issued by the Ministry of Environment of the Province of British Columbia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

This thesis was completed under the supervision of Dr Michael Russello. I thank him for his advice and guidance over the last two and a half years. For their contribution on previous drafts of this thesis, I would like to my other committee members, Bob Lalonde and Paul Askey. I thank fellow members of the Conservation Genomics Laboratory, Philippe Henry and Matt Lemay, for sharing their knowledge of population genetics, genomics and statistics. For her contributions in initiating this work, I thank Stephanie Kirk.

I thank the many fishery biologists with the BC Ministry of Environment and private consultant agencies for their assistance in identifying and evaluating lakes for inclusion in this study. For their volunteered assistance in acquiring tissue samples, I thank the Christina Lake Stewardship Society, Joe DeGisi (MOE), Paul Askey (MOE), Andy Morris (MOE), Brian Jantz (MOE), Matt Lemay (UBCO), Dan Rissling (UBCO) and Matthew Abbott. I thank Eric Taylor at UBC for also providing samples for this project. For their assistance in generating genetic data, I would like to thank Christine Braun, Stephanie Kirk, as well as Mahsa and Michaela at Fragment Analysis and DNA Sequencing Services (FADSS). For general assistance in the lab, I thank Matt Lemay, Philippe Henry, Carolina Mino and Yoamel Milián-García.

This research was funded by a the Canadian Foundation for Innovation, the Habitat Conservation Trust Fund and a National Science and Engineering Research Council (NSERC) discovery grant (# 341711-07) awarded to Michael Russello. I was also supported in part by University of BC Graduate Fellowships, an NSERC Canada Graduate Scholarship and a Pacific Century Scholarship.

Finally, I would also like to thank my family for their endless support.

DEDICATION

To my parents.

CHAPTER 1.0 GENERAL INTRODUCTION

1.1 The study of adaptation

1.1.1 Natural selection vs neutral evolution: a classical debate

Understanding the role of natural selection in generating and maintaining biological diversity has been of central interest to evolutionary biologists since Darwin's seminal work in 1859 (Darwin, 1859). According to his theory of natural selection, if traits are polymorphic and there is competition for resources, individuals will experience differential survival. Groups will become more optimally suited to their niche as small, continually arising variations accumulate due to selection, causing a gradual shift in traits. Eventually, these adaptive differences can lead to speciation. Numerous examples of artificial selection (e.g. breeding livestock), specialized morphological traits correlated with resources use (e.g. mockingbirds and Darwin's finches in the Galapagos Islands; Schluter, 2000) and parallel patterns of phenotypic traits in similar, but isolated, environments (Schluter and Conte, 2009) supported the adaptationist view, thus it became widely accepted.

In the 1930's and 1940's, the significance of natural selection came into question. Based on principles of Mendelian inheritance, mathematical models of evolutionary processes revealed that species are more aptly defined as reproductively isolated units than simply groups sharing phenotypic resemblance. Thus, Mayr proffered the Biological Species Concept (1942), which describes a species as a groups of individuals that actually, or potentially, interbreed in nature to produce viable offspring. Therefore, the study of speciation should be focused on identifying the origin and nature of mechanisms that reinforce reproductive barriers among divergent populations (e.g. positive assortative mating caused by divergence in ecology, behavior, or time of reproduction) such that hybrids exhibit reduced fitness and, eventually, reduced viability and/or sterility (Coyne and Orr, 1998; Dobzhansky, 1951; Mayr, 1942). Classical models also suggested that the number of selective deaths (i.e. genetic load) required for selection to initiate species divergence in the face of genetic recombination and ongoing gene flow among closely related, co-existing populations (i.e. sympatric

speciation) was rarely achievable in nature (Haldane, 1949; Kimura, 1995). Only some extrinsic, physical barrier was thought to be capable of sundering gene flow among populations (Felsenstein, 1981; Johannesson, 2001; Kondrashov, 1986; Mayr, 1963). According to neutral (Kimura, 1983), or nearly neutral (Ohta, 1973), theory, evolutionary change (i.e. shifts in population allele frequencies) occurs over long periods of time through the fixation of mutant alleles by random genetic drift, which have little or no effect on the individuals fitness, particularly through founder events and population bottlenecks. Thus, neutral evolution in allopatric populations became widely accepted as the dominant mode of speciation in nature because it provided the most parsimonious explanation for observed patterns in biological diversity at the time (Coyne and Orr, 2004; Provine, 1971). However, empirical support for this was lacking. Several isolating mechanisms can arise simultaneously when populations diverge in allopatry, which makes it impossible to disentangle the roles of selection and neutral evolution in initiating and maintaining reproductive barriers. As a result, little progress was made in assessing the relative merits of these two theories because evolutionary biologists were faced with untestable predictions about a process that they believed could not be witnessed within a lifetime (Hendry, 2009).

More recently, some of the assumptions underlying the allopatric paradigm have been overturned. Empirical and theoretical studies demonstrate that evolution can occur rapidly (i.e. on ecological time-scales; Reznick and Ghalambor, 2005) and in large populations as long as they possess ample genetic variation and selection is strong (Hendry and Kinnison, 1999; Kinnison and Hendry, 2001; Reznick and Ghalambor, 2001; Stockwell and Weeks, 1999). Also, 'divergence-with-gene-flow' (Rice and Hostert, 1993) is now supported by strong theoretical models (Dieckmann and Doebeli, 1999; Kondrashov and Kondrashov, 1999) and empirical studies (Huber *et al.*, 2007; Jiggins, 2008; Ogden and Thorpe, 2002; Quesada *et al.*, 2007; Schliewen *et al.*, 1994). Together, this suggests that natural selection could have a major role in generating reproductive barriers, which has renewed interest in the study of adaptive speciation.

1.1.2 Testing for the role of selection in initiating population divergence

Historically, studies focused on phenotypic divergence after speciation had already taken place. Many claimed to exemplify environment-driven speciation (termed ecological speciation), but rarely provided robust evidence (Hendry, 2009). Studies demonstrated strong correlations between distinct phenotypes and resource use, but no evidence suggested that these phenotypes were involved in generating reproductive isolation. The assumption that extant phenotypic traits are adaptated for their contemporary purpose is often incorrect (Gould and Lewontin, 1979). Other studies demonstrated the inability of closely related yet ecologically distinct species to interbreed, but there was no evidence that adaptation was the cause. Neutral forces (e.g. founder effects) could have initiated divergence and phenotypic disparities corresponding to resource use evolved secondarily (Via, 2009). For example, non-adaptive structures can arise through a developmental correlation with selected features (e.g. pleiotropy, material compensation, mechanically forces correlation or allometry; Gould and Lewontin, 1979). To provide robust inferences for the action of natural selection in driving population divergence, and potentially speciation, it is necessary to link early stages of reproductive isolation with divergent phenotypic traits at the genetic level (Barrett and Hoekstra, 2011). Only with recent advancements in genome-typing technology has it become feasible to investigate the genetic basis of reproductive isolation in natural populations *in situ*.

1.2 Population genomics

Population genomics employs traditional population genetic approaches, but uses increasingly efficient and cost-effective genotyping technologies to achieve genome-wide sampling (i.e. often thousands of loci; Luikart *et al.*, 2003). This approach explicitly acknowledges that neutral evolutionary processes (e.g. genetic drift, gene flow, and mutation) influences the entire genome while natural selection only has locus-specific effects. Using a genome scan approach, neutrality tests are used to screen large numbers of loci and identify statistical outliers (i.e. loci that do not conform to expectations under a neutral model of evolution) so that the truly neutral loci can be segregated from

loci putatively under selection (i.e. outlier loci; Antao *et al.*, 2008; Black *et al.*, 2001; Storz, 2005). Since the inclusion of loci under selection would bias allele frequency distributions in neutral datasets, the elimination of outliers allows for more accurate estimates of population demography and evolutionary history than population genetics approaches would achieve alone (Luikart *et al.*, 2003; Nielsen *et al.*, 2009a). Also, further investigation of outlier loci could provide insights into the ecological mechanism driving adaptive population divergence and the genetic architecture of adaptations. This bottom-up approach for detecting adaptive variation provides unprecedented opportunities to empirically test for the action of natural selection at the genetic level in natural populations of non-model organisms.

The genome-scan approach is commonly equivalated to 'looking for a needle in a haystack', because a very small portion of the genome is under selection (Campbell and Bernatchez, 2004). Most polymorphisms occur in the neutral regions of the genome, therefore studies using anonymous markers such as amplified fragment length polymorphisms (AFLPs) and microsatellites tend to have low detection rates (2-8%; Bonin, 2008; Bonin *et al.*, 2006; Campbell and Bernatchez, 2004; Tice and Carlon, 2011; Wilding *et al.*, 2001). Alternatively, single nucleotide polymorphisms (SNPs) recovered from the transcriptome or microsatellite markers linked to expressed sequence tags (ESTs) can target the functional part of the genome, where polymorphisms are more likely to be maintained by selection (Bonin, 2008; Bouck and Vision, 2007), to achieve higher detection rates (13-20%; Namroud *et al.*, 2008; Oetjen and Reusch, 2007; Shikano *et al.*, 2010; Vasemagi *et al.*, 2005). In addition, convergent evolution in closely related populations may arise from i) the exact same mutation, ii) a different mutation in the same gene, or iii) mutations in different genes within a common genetic pathway. Since EST-linked microsatellites are neutral markers embedded in the flanking regions of a gene and may bear the signature of selection due to 'genetic hitchhiking' (Barton, 2000; Maynard-Smith and Haigh, 1974), there is a greater likelihood of detecting parallel patterns at the gene-level using EST-linked markers than at the nucleotide-level using SNPs (Arendt and Reznick, 2008; Vasemagi *et al.*, 2005). The EST-linked markers are also transferrable among closely related species, and EST libraries continue to grow rapidly due to tremendous advances in next-generation sequencing technologies

(Bouck and Vision, 2007). The use of EST-linked markers is expected increase because they provide a cost-effective approach for organisms with otherwise limited genetic resources.

Detecting robust signatures of selection in natural populations also relies upon using an appropriate study system. Several criteria have been proposed for selecting such a system; i) populations must be closely related to ensure that reproductive isolation is entirely environment-dependent; ii) divergence must be recent (<12,000 years ago) such that processes underlying phenotypic divergence and reproductive isolation are still actively maintained and genetic signatures of selection have not yet deteriorated via genetic recombination; iii) divergence has occurred in the presence of ongoing gene flow (i.e. in sympatry) such that the genome is homogenized except at adaptive loci; iv) populations must have a known phylogeographic history to ensure that divergence was initiated and maintained in sympatry; and v) ideally multiple replicate systems should be available to test for parallel patterns in divergence, because parallel patterns of genetic divergence can provide one of the strongest forms of evidence for adaptive divergence achievable using population genomic approaches (Hendry, 2009). Generally, post-glacial fishes meet these criteria due to the recent glacial history of North America (Rambaut and Schluter, 1996), and so far some of the strongest evidence for adaptive divergence using population genomic approaches has been revealed in this group (Bernatchez *et al.*, 2010; Colosimo *et al.*, 2004).

1.3 Post-glacial fishes for studying evolutionary processes

1.3.1 Recent glacial history of North America and its consequences

A series of glaciation events during the Pleistocene epoch altered the global climate, sea level, and land ice-cover, which significantly impacted mammalian, avian, and teleost phylogeography (Avise *et al.*, 1998). During the Wisconsin glaciation, sheets of ice extended into North America, Asia, and Europe. The Cordilleran and Laurentide ice sheets covered Canada and northern USA, destroying any pre-existing freshwater habitat and displacing freshwater and anadromous fishes into habitats skirting the

glacial fronts (Pielou, 1991). Substantive declines in abundance caused large reductions in genetic diversity (Bernatchez and Wilson, 1998). During the last glacial retreat (8,000-15,000 years ago), meltwater formed small proglacial lakes (i.e. small basins of water contained by glaciers) and meltwater created outflows that served as temporary corridors and allowed access to dispersers from refugial populations in the Bering and Columbia Refugia. As a result, several fishes had colonized post-glacial lakes by the time the ice had withdrawn and the water level dropped to current levels.

An explosion in intraspecific variation followed the colonization of these lakes (Pielou, 1991). Habitats were resource-rich and there were few competitors and predators because opportunities to colonize these new lakes were brief (Rambaut and Schluter, 1996). This presented many ecological opportunities and allowed resource polymorphism to arise in sympatry. Rapid evolution of assortative mating based on niche differentiation, despite a history of gene flow, reinforces associated changes in morphology, physiology, and behavior. This drove intraspecific differentiation of post-glacial freshwater fishes in the northern hemisphere such that they are now considered to be one of the fastest evolving groups of taxa.

1.3.2 Studying evolution in salmonids

Salmonids are ideal candidates for studying adaptation and species divergence because most extant populations of Pacific salmon are descendent from refugial populations (Hendry and Stearns, 2004). As a result, they appear evolutionarily young and appear to have accumulated a remarkable level of diversity in life history strategies and morphologies in a short time (Cossins and Crawford, 2005). Their rapid divergence is facilitated by natal homing (Quinn, 2005). Through chemical imprinting during early life stages, salmon return to the natal spawning grounds to reproduce. This allows for some reproductive isolation among geographically proximate populations and adaptation to local conditions to produces new ecological forms, or ecotypes (Mayr, 1963). Adaptation can occur rapidly in salmonids because they possess a significant amount of genetic variation due to a historical tetraploidization event >25 million years ago (Allendorf and Thorgaard, 1984) and phenotypic

variation due to phenotypic plasticity (Pfennig *et al.*, 2010). Phenotypic plasticity (when a single genotype can produce multiple phenotypes in response to variation in the environment) creates opportunities for selection to act upon novel phenotypes, releases cryptic genetic variation, increases the chances of survival for advantageous phenotypes (Pfennig *et al.*, 2010). Although much of the observed phenotypic variation could be attributed to plasticity, the frequency of failed transplant experiments suggests that salmon are especially well adapted to the local ecological conditions of their natal environment (Fraser *et al.*, 2011; Taylor, 1991). In addition, Pacific salmonids are excellent candidates for genetic studies because many have well-described population structures, substantial and rapidly accumulating genomic resources, and an abundance of archived samples have been collected due to long-term monitoring and research initiatives by commercial and recreational fisheries (Hauser and Seeb, 2008; Wenne *et al.*, 2007).

1.3.3 Common patterns of phenotypic divergence

Among post-glacial fishes, certain species and ecotype pairs commonly re-occur in lakes throughout their distribution. Much of the variation appears to be driven by intra-specific competition for space or resources (Schluter, 1996; Schluter and McPhail, 1992) and has resulted in niche partitioning associated with either trophic position (Landry *et al.*, 2007; Ostberg *et al.*, 2009; Peichel *et al.*, 2001; Saint-Laurent *et al.*, 2003), spawning timing (Creelman *et al.*, 2011; McGlauflin *et al.*, 2011), habitat preference (Lecomte and Dodson, 2004), or anadromy (Shikano *et al.*, 2010; Theriault *et al.*, 2007; Wood *et al.*, 2008). Divergence in trophism generally results in the co-existence of a planktivore that feeds on pelagic zooplankton and a benthivore that feeds on benthic invertebrates or larger prey from littoral zone (Schluter, 1996). In salmonids, often two or more life history types can be found within a single geographic area. For example, many river drainages support both spring-run and fall-run Chinook salmon (Bernier *et al.*, 2008), early- and late-run coho salmon, and summer-run and winter-run steelhead (Waples *et al.*, 2001). However, few studies have investigated the mechanisms driving divergence in reproductive behaviour and habitat preference, which accounts for a considerable amount of the intra-specific diversity in salmonids (Mehner *et al.*, 2011). These phenotypes are

difficult to measure compared to morphological traits, but with recent advancements facilitating the study of non-model organisms in natural environments, investigation of these traits will likely be important role in generalizing our knowledge of environmental drivers of adaptive population divergence and speciation (Bernatchez *et al.*, 2010).

1.4 Study system: kokanee salmon

1.4.1 The origin of non-anadromous *Oncorhynchus* nerkids

Kokanee salmon are a polyphyletic group of obligate freshwater populations that have diverged from anadromous sockeye salmon multiple times since the last glaciation (McPhail and Lindsey, 1970; Taylor *et al.*, 1996; Wood *et al.*, 2008). Similar to juvenile sockeye, kokanee inhabit the limnetic-pelagic zone and feed primarily on crustacean macro-zooplankton zooplankton (Chipps and Bennett, 2000; Clarke *et al.*, 2004). Both kokanee and sockeye spawn in rivers, streams tributary to lakes, or shoreline areas (Quinn, 2005) often associated groundwater seepage (McPhail and Lindsey, 1970). However, through isolation, the non-anadromous kokanee have evolved several morphological, reproductive, and genetic differences (Wood and Foote, 1996). Kokanee exhibit slower growth rates, much smaller size, earlier age at maturity (3-4 years), and they possess significantly more gill rakers (Foote *et al.*, 1999). Sexual dimorphism and secondary sex characteristics (i.e. humped back, bright coloration, hooked jaw) are less pronounced in kokanee and sexual selection for red breeding coloration has led to a divergence in the regulation of carotenoid sequestering in their tissue (Craig and Foote, 2001). Sockeye are now genetically distinct from their lacustrine counterpart, as assessed by mitochondrial, minisatellite, and allozymes frequencies (Taylor *et al.*, 1996; Wood and Foote, 1996). Gene flow is estimated at 0.1 - 0.8% in those tributaries where kokanee and anadromous sockeye spawning grounds overlap, which is much lower than that of different tributaries (Wood and Foote, 1996). Genetic distinction has also been demonstrated through a series of controlled breeding experiments (Foote *et al.*, 1992; Foote *et al.*, 1989; Wood and Foote, 1990). However, due to the plurality of divergent non-anadromous populations, overall ecological similarity, and the ability of non-

anadromous kokanee to revert back to anadromy (Godbout *et al.*, 2010), they are not considered to be distinct species, or even subspecies (McPhail and Lindsey, 1970; Taylor, 1999).

1.4.2 Current distribution

Kokanee salmon are naturally distributed in lakes throughout the coastal regions of North America and northeastern Asia rimming the Pacific Ocean (McPhail and Lindsey, 1970). In North America, kokanee inhabit lakes between the Klamath River, California and Point Hope, Alaska, and in Asia between northern Hokkaido, Japan and Anadyr River, Russia. Kokanee have also been introduced throughout central and southeastern Canada and northwestern USA through experimental stocking programs (Crawford and Muir, 2008; Crossman, 1991). In lakes still accessible from the Pacific Ocean, kokanee and anadromous sockeye salmon share common spawning grounds.

1.4.3 Two reproductive ecotypes in kokanee salmon

In many post-glacial lakes, two reproductive ecotypes co-exist: stream-spawners and shore-spawners (Taylor *et al.*, 1997). The stream-spawners exhibit the ancestral life history form and utilize streambeds for spawning. The shore-spawners use the shoreline adjacent to streams or in other regions of the lake, sometimes in upwelling zones, and exhibit distinct reproductive behaviours. Outside of the spawning period, ecotype pairs are ecologically and morphologically indistinguishable (Taylor *et al.*, 1997; Winans *et al.*, 2003). If divergence has occurred while in sympatry, environment-mediated selection pressures are likely driving local adaptation.

1.4.3.1 Morphology, life history and behavioural differences

During the spawning period, shore and stream-spawning kokanee ecotypes exhibit distinct reproductive strategies and form spatially and temporally discrete spawning aggregations. Similar to sockeye, stream-spawners (Figure 1.1) engage in traditional up-stream migrations in the fall where a female will excavate a redd by beating her tail while on her side (Table 1.1). She evaluates secondary

sex characteristics to select a fit mate and they spawn in unison. The female then dislodges upstream gravel which covers the nest to prevent scour and will continue to defend the nest until fatal exhaustion. Shore-spawners generally form spawning aggregations 2 to 6 weeks (and up to 2 months) later than stream-spawners at specific areas along the shoreline. They tend to select a very narrow depth range within the water column (< 1m) and may use a variety of larger substrates (Shephard, 2000). They do not form mating pairs or defend their nests. In fact, they abandon the shoreline habitat during the day in several lakes. In some lakes, they build redds at greater depths near stream deltas or areas with groundwater seepage (Andrusak and Jantz, 2002), but many only clean off the rocks prior to egg deposition. In a few lakes, body size, egg size, and post-hatching growth rate are slightly lower in shore-spawners than stream-spawners (e.g. Okanagan Lake; Taylor *et al.*, 2000).

Table 1.1 Physical attributes of the spawning habitat and morphological, life history, and behavioural attributes of kokanee ecotypes in British Columbian Lakes.

Category	Phenotypic trait	Stream-spawners	Shore spawners	Reference
Environment	Spawning location	tributaries	shoreline	(de Zwart *et al.*, 2011; Taylor *et al.*, 1997)
	Spawning substrate	Rounded gravel <5cm diameter	Large, angular rocks >5cm diameter	(Shephard, 2000)
	Spawning depth	Shallow	Deeper (up to 6 m)	(Andrusak and Andrusak, 2011; Shephard, 2000)
Morphology	Nuptial coloration	Bright red body and green head	Dark red body and green head	(Dill, 1996)
	Secondary sex characteristics	pronounced	less pronounced	(Dill, 1996)
Life history	Peak spawning time	Fall (Sept – Nov)	2-6 weeks later in fall	(Shephard, 2000)
	Time of emergence	Spring (Mar–Jun)	Spring (Mar-Jun)	(Shephard, 2000)
Behaviour	Mate selection	Courting behaviour and long-term pairing	No obvious pairing	(Dill, 1996)
	Parental care	Female builds redds and defends nest	No nest defense	(Dill, 1996)
	Time of day for spawning	Day-time	Night-time or day-time	(Andrusak and Andrusak, 2011; de Zwart *et al.*, 2011)

Figure 1.1 Photographs of a deceased (A) female and (B) male stream-spawning kokanee found along the shore of Sandners Creek, tributary to Christina Lake, BC. A tissue sample has been taken from the operculum of the female kokanee.

1.4.3.2 Genetic differentiation of ecotypes

Patterns of genetic variation in natural populations are shaped by gene flow, genetic drift, mutation and natural selection. In previous studies of Okanagan Lake kokanee, low levels of neutral genetic differentiation were detected in the frequency of mitochondrial DNA haplotypes (Taylor *et al.*, 1997), five nuclear microsatellite loci (Taylor *et al.*, 2000), and 74 allozyme loci (Winans *et al.*, 2003). This is not surprising given the high potential for gene flow among stocks, however these findings suggest that kokanee ecotypes are not a single panmictic population. Reductions in gene flow may be the result of local adaptation to different spawning habitats. In sockeye salmon, stream-spawners that stray onto shore-spawning grounds are numerous (39%) but appear to have low fitness given the high level of differentiation among ecotypes (Hendry *et al.*, 2000). To investigate the locus-specific effects of selection, Russello *et al.* (2012) conducted a genome-wide scan of 243 EST-linked microsatellites for Okanagan Lake kokanee and used neutrality tests to identify outlier loci. The eight EST-linked markers showing outlier behaviour had a 93% success rate in assigning individuals back to their source

ecotypes compared to the 59% success rate using eight putatively neutral markers. The significant difference in patterns of genetic variation suggests that natural selection may be involved in ecotype divergence, but since the study was limited to a single lake these outlier loci could not be further validated.

1.4.3.3 Habitat differences

Salmon populations are specifically adapted to local environmental conditions, which appear to vary between stream and shore habitats substantially (Dill, 1996; Shephard, 2000). Habitats differ water depth, velocity, temperature, substrate, and dissolved oxygen content and therefore may differentially impact evolutionary trade-offs in life history and physiological traits (Hendry and Stearns, 2004). For example, egg size is a heritable trait linked to egg survival (Hendry and Stearns, 2004) and has been positively correlated with mean geometric size of spawning/incubating gravel for Alaskan sockeye populations (which can vary 30-fold; Quinn *et al.*, 1995; Taylor *et al.*, 2000). A similar trend may be present in kokanee since shore-spawners often use larger substrates than stream-spawners. Also, studies suggest that timing of spawning can differ systematically between habitat types due to temperature differences. In Okanagan Lake, shore-spawning peaks one month after stream-spawning, but they acquire the same number of Accumulated Thermal Units (ATU) and hatch at the same time (Taylor *et al.*, 2000). Therefore, spawning time is probably an evolutionary response that ensures emergence is synchronized with spring algal blooms. Spawning time has already been linked to the *CLOCK* gene in other salmonids (Leder *et al.*, 2006). Biotic interactions may be generating distinct selection pressures, including predation, pathogens, and intra-specific competition for mates or resources (Andrusak and Jantz, 2002). For example, the expression of secondary sex characteristics (e.g. breeding coloration, dorsal hump, teeth, hooked jaw) and reproductive behaviours (e.g. mate choice, nest defence, parental care) are maintained by sexual selection, but environment-mediated selection may lead to the loss of these traits altogether due to a trade-off between attaining high quality mates and high energetic costs or predation pressure (Craig and Foote, 2001). Evidence for adaptive divergence in mate recognition traits have been detected in three-spine stickleback (Rundle

et al., 2000) and cichlids (Seehausen et al., 2008). Finally, habitat preference is also thought to be critical for matching locally adapted phenotypes within heterogeneous landscapes (Davis and Stamps, 2004). These are just a few examples of the traits potentially under selection due to the differences in selection regimes associated with the shore and stream habitats.

1.4.3.4 Population declines and recovery objectives

Kokanee are considered a keystone species because they are an important forage fish for many of species at higher trophic levels, including piscivorous trout and char (Andrusak and Parkinson, 1984). Kokanee are also a popular recreation fish in Canada and the United States (Shephard, 2000). Hence, the substantive declines that have been observed in numerous lakes throughout their native distribution are of great concern. In Okanagan Lake, BC they have declined by 99% since the 1960's. Similar trends have been observed in Kootenay Lake, BC (Anders *et al.*, 2007) and Kathleen Lake, Yukon (L. Freese, *pers. comm*.). In Seton and Anderson Lakes, BC, kokanee once occurred in large numbers, but are now described as "severely depressed" by First Nations to whom they are culturally significant and an important supplementary component of their diet (Morris *et al.*, 2003). Substantive declines have been observed in and Pend O'reille, Idaho (Paragamian and Bowles, 1995). In Samammish Lake, Washington, kokanee are currently being reviewed for listing as threatened or endangered under the US Endangered Species Act (USFWS, 2008).

Generally, declines have been attributed to the introduction of non-native species that compete for food resources (e.g. opossum shrimp, *Mysis relicta*) or predators (e.g. lake trout, *Salvelinus namaycush*), habitat degradation through stream channelization, hydroelectric dam construction, shoreline development, lake draw-down, overfishing, competition with increasing sockeye and rainbow trout populations (Sebastian *et al.*, 2003), and decreased lake productivity due to anthropogenic factors that decrease nutrient inputs (e.g. dams and logging). Also, lakes throughout their range are supplemented with eggs from foreign kokanee stocks to increase recreation opportunities in northwestern USA (Parametrix, 2003) and southern British Columbia with unknown genetic consequences. Several of

these factors specifically impact either shore or stream habitats, therefore independent management of shore- and stream-spawners is required (Andrusak and Jantz, 2002).

1.5 A genetics-based approach for managing recently diverged stocks

A recent study estimated that ~30% of historical salmon populations in the Pacific Northwest have been extirpated in the past 200 years and 50% of extant populations are being listed as threatened or endangered under the USA's Endangered Species Act (see http://www.nwr.noaa.gov/ESA-Salmon-Listings/Index.cfm; Gustafson *et al.*, 2001). Therefore, fine-scale delineation of evolutionarily and ecologically unique populations is needed to conserve the diversity of remaining stocks for the future. A major obstacle to maintaining sustainable kokanee fisheries is the diversification of populations into sympatric ecotypes, which often leads to mixed stock fisheries. Managing mixed stock kokanee fisheries is a challenge, because it is difficult to manage harvest levels to specific stocks with varying levels of stock productivity (and sustainable harvest). In kokanee, ecotypes (or further divided sub-stocks) are visually indistinguishable at time of harvest, but each stock will sustain different levels of harvest because of the inherent productivity of the stock's spawning habitat. Furthermore, kokanee enumeration programs rely on visual counts conducted while in their spawning aggregations. The dark coloration, seasonal foul weather and extreme densities of aggregated fish, particularly at depths up to ten meters, make shore-spawners very difficult to count with accuracy. Other methods include hydroacoustics and trawl surveys, but these methods only measure the aggregate abundance of all ecotypes, and can be subject to several biases. Stream-spawners can potentially be accurately censused by adding fences to spawning tributaries, however, when the stock is distributed among many tributaries, the labour and costs become prohibitive (P. Askey, *pers. comm*.). Genetics-based approaches for monitoring and assessing kokanee stocks are now being sought because of its well-demonstrated success in other species (Beacham *et al.*, 2006; Beacham *et al.*, 2008; Hauser and Seeb, 2008). However the conventional use of neutral markers is not effective in detecting patterns recently diverged ecotypes because neutral differences have not yet accumulated. So far, few have attempted to test the power of putatively adaptive markers (i.e. markers linked to traits of adaptive significance) for

genetic stock identification (GSI) in recently diverged ecotypes (but see Ackerman *et al.*, 2011; Creelman *et al.*, 2011). Since a primary goal in conservation biology is to preserve maximum genetic diversity (neutral and adaptive) and thereby species' capacity to endure disturbance and continually adapt to changing fitness landscapes (Hilborn *et al.*, 2003), the integration of adaptive markers within a fisheries management plan will be important for achieving this goal.

1.6 Thesis objectives

Kokanee salmon represent an ideal system for investigating the genetic basis of ecotype divergence relating to habitat-use and for evaluating the potential for outlier loci to inform fisheries management. The many lakes throughout BC that contain shore- and stream-spawning ecotype pairs of kokanee, which share a common geological history, enable a replicated experimental approach within a natural setting. Using the analytical tools of population genomic and genetic approaches, Chapter 2 reconstructs evolutionary relationships among multiple ecotype pairs to determine if shore-spawning behaviour evolved independently within multiple lakes or has descended from a single source population. I also test outlier loci detected from genome-scans for unique patterns of genetic differentiation compared to neutral loci to assess the likelihood that natural selection is involved in the divergence of these reproductive ecotypes and identify the expressed genes associated with each outlier marker. In Chapter 3, I evaluate outlier loci for their ability to consistently distinguish ecotypes from multiple lakes throughout British Columbia and evaluate the extent to which a genetics-based approach that utilizes these outlier loci may improve the accuracy of range-wide kokanee stock identification and management.

CHAPTER 2.0 INVESTIGATING THE GENETIC BASIS OF ECOTYPE DIVERGENCE IN KOKANEE SALMON ACROSS MULTIPLE LAKES

2.1 Background

The relative role of natural selection and neutral evolutionary processes in generating and maintaining biological diversity has long been debated. Recently, tools to test for environment-genotype relationships have become available, allowing researchers to attain more robust evidence for the action of natural selection in generating reproductive barriers among locally adapted populations (Schluter, 2001). Salmonids provide ideal systems for studying evolution driven by ecological processes because they exhibit an extensive range in morphological, physiological, and behavioral traits at the intra-specific level that have arisen over a short period of time (<10,000 years; Hendry and Stearns, 2004). The plurality of certain phenotype-environment correlations among geographically discrete populations and their rapid evolution and persistence in sympatry suggest that natural selection is driving population divergence rather than neutral evolutionary processes (Schluter, 1996). Extensive phenotypic plasticity in salmonids provides the variability upon which selection may act when individuals disperse to new environments in response to high competition for limited resources in their natal environment (Pfennig *et al.*, 2010). While plasticity alone could explain much of the phenotypic variation in salmonids, many failed transplant experiments suggest that many locally adapted populations are not ecologically exchangeable (Fraser *et al.*, 2011; Miller *et al.*, 2001). These failures underscore the need to identify those traits that have significant fitness consequences and thereby achieve a better understanding of extrinsic and intrinsic factors that promote and constrain adaptation (Schluter, 2001).

In kokanee salmon, stocks utilizing distinct spawning habitats (e.g. shorelines and streams) within the same lake often exhibit different reproductive traits (e.g. mating behavior, secondary sex characteristics, parental care, and spawning time; see Table 1.1), yet show no morphological or ecological differences prior to maturation (Taylor *et al.*, 1997; Taylor *et al.*, 2000; Winans *et al.*, 2003).

In general, rather than exhibiting the ancestral traits of stream-spawning sockeye, the shore-spawners form large spawning aggregations along the shoreline a couple weeks or months following sympatric stream-spawners. They do not exhibit courting behaviour, mate selection, redd excavation, or nest defence (de Zwart *et al.*, 2011; Dill, 1996; Shephard, 2000). Shore habitats are typically deeper (0.1-10 meters), warmer (at the same time of year), low-flow environments with larger rocky substrates. The specific fitness-related traits initially involved in generating reproductive isolation among these two ecotypes are unknown, but are possibly related to habitat preference, mating behavior, energy metabolism, and/or life-history ecology that increase reproductive success in adult spawners and survival at early life stages (Lecomte and Dodson, 2004).

Ecological diversification via habitat partitioning is commonly observed in post-glacial fishes (Berner *et al.*, 2010; Lecomte and Dodson, 2004; Ostberg *et al.*, 2009; Rogers and Bernatchez, 2006; Saint-Laurent *et al.*, 2003) and has lead to speciation in more deeply divergent shore-spawning lineages of marine fish including smelt and silversides (Martin and Swiderski, 2001). Presently, it is unclear if there is truly an adaptive basis for shore-spawning behaviour in kokanee salmon (and sockeye salmon) or if it is simply a plastic response to resource availability. Sympatric ecotypes may represent two ecologically and evolutionarily distinct populations if unique phenotypes exhibited by shore-spawners confer a strong fitness advantage in shoreline habitats such that gene flow is reduced at genes underlying these adaptations. Alternatively, sympatric ecotypes may represent a single panmictic population if individual stocks (and ecotypes) are being maintained through natal homing, but remain entirely undifferentiated due to straying among habitats. A substantial amount of the intraspecific diversity in salmonids is a consequence of niche partitioning within lakes, and the relative importance of evolutionary versus plastic responses to distinct spawning habitats needs to be teased apart (Hendry and Stearns, 2004). To determine if shore-spawning kokanee are uniquely adapted to shoreline habitats, the genetic basis of their divergence needs to be revealed.

Recent advancements in genome-typing technologies and analytical tools allow us to simultaneously investigate the role of various evolutionary processes at the genetic level in non-model organisms

while in their natural environment (Nosil *et al.*, 2009; Storz, 2005). Population-based genome scans can identify gene regions of adaptive significance by screening a large number of markers distributed throughout the genome and segregating those that correspond with neutral expectations from those assessed to be statistical outliers (Black *et al.*, 2001; Luikart *et al.*, 2003; Nielsen, 2005; Stinchcombe and Hoekstra, 2008; Storz, 2005). Only the locus-specific effects of selection can explain such patterns of genetic diversity. This 'bottom-up' strategy allows us to assay genetic variability among ecotypes with no *a priori* assumptions about the specific phenotypic traits under selection.

Several types of markers have been widely used in genome scan studies of non-model organisms, including amplified fragment length polymorphisms (AFLP), single nucleotide polymorphisms (SNP), and microsatellites (Luikart *et al.*, 2003). Amplified fragment length polymorphisms have broad genomic coverage, produce many markers at low cost, and do not require prior sequence information (Bonin *et al.*, 2007), but these dominant markers contain less information and any locus exhibiting signatures of selection will be anonymous. Single nucleotide polymorphisms are also broadly distribution throughout the genome, have low genotyping error rates, and better-understood mutation models. However, a large number of these co-dominant loci are needed (2-6 times as many as polymorphic loci), which makes their identification and application expensive and laborious in non-model organisms without sufficient genomic resources (Morin *et al.*, 2004). Alternatively, microsatellites are highly variable and the primers work in closely related species. Expressed sequence-tag (EST) libraries are available for many species, which offers a much more cost-effective and efficient approach for non-model organisms because they enable users to target the functional portion of the genome (Bouck and Vision, 2007; Hauser and Seeb, 2008; Vasemagi *et al.*, 2005; Wiehe *et al.*, 2007).

Expressed sequence tag-linked microsatellites are sequences of tandem repeat units (Bouck and Vision, 2007). They are found in the introns flanking expressed genes and are unlikely to be broken up by recombination due to their close proximity to the gene. The spread of a beneficial mutation through a population reduces variability at the selected gene as well as its flanking regions through hitch-hiking

effects (Slatkin, 1995; Smith and Haigh, 1974). Positive selection acting on the gene can be inferred when patterns of significantly reduced variability are detected at the linked microsatellite marker.

This approach has previously been applied to divergent ecotypes in kokanee from Okanagan Lake, British Columbia (Russello *et al.*, 2012). Over 11,000 EST-linked markers were scanned for polymorphism, resulting in a panel of 57 markers (49 EST-linked and 8 anonymous). Using three different outlier-detection approaches, eight putative outliers were identified including three loci that were detected by multiple approaches. However, this study was based on a single lake, which precluded validation of the role of selection in driving adaptive divergence in this system.

Since outliers can be difficult to distinguish from background selection or neutral variation when selection is weak, several methods have been proposed to eliminate false positives and ensure that detected outliers are robust. Environmental-genotype correlations can be tested when environmental data is available (Bonin *et al.*, 2006; Holderegger *et al.*, 2008; Joost *et al.*, 2007; Landry *et al.*, 2007; Wilding *et al.*, 2001), but this approach is most appropriate when populations are collected along an environmental gradient and agents of selection are known *a priori*. In general, the use of multiple outlier-detection approaches is commonly advocated (Nunes *et al.*, 2011) as well as testing for parallel patterns in outlier behaviour across multiple independent samples (Colosimo *et al.*, 2004; Hohenlohe *et al.*, 2010; Oetjen and Reusch, 2007; Schlötterer, 2003). Local adaptation is the most parsimonious explanation for the repeated evolution of particular phenotypes in geographically discrete populations experiencing similar ecological conditions (Johannesson, 2001; McKinnon *et al.*, 2004). If a common genetic basis (e.g. same genes) can be identified across multiple ecotype pairs, strong evidence for the action of natural selection can be inferred since the probability of detecting spurious patterns of parallel evolution is very low (Hendry, 2009; Johannesson, 2001; McKinnon *et al.*, 2004).

Here, I test for evidence of directional selection driving adaptive divergence in sympatric ecotypes (shore and stream-spawners) of kokanee salmon in five British Columbian lakes. Four conceptually different outlier-detection approaches are used to evaluate patterns of gene diversity and

differentiation at 57 EST-linked and anonymous microsatellite loci. Recovered patterns of neutral and adaptive genetic variation are used to test hypotheses regarding the origin of shore-spawning behaviour (e.g. either a polyphyletic group arising independently in each lake or a para- or monophyletic group with a common ancestor) and identify candidate genes associated with local adaptations that may be restricting gene flow among shore- and stream-spawning kokanee.

2.2 Methods

2.2.1 Study sites

Six lakes containing sympatric populations of shore- and stream-spawning kokanee were included in this study: Wood, Okanagan, Kootenay, Duncan, Christina, and Tchesinkut Lakes (Figure 2.1). The lakes vary in size and time since isolation. They all have records of minimal stocking, little or no overlap with wild sockeye populations, and are distributed across the Columbia and Fraser River drainage systems. Wood and Okanagan Lake are located in the central interior of BC. Kootenay, Duncan, and Christina Lakes are located in the eastern interior of BC. Tchesinkut Lake is located in northern BC (Figure 2.1). Each lake is characterized by deep, cold, clear water with rocky shores and low productivity (i.e. oligotrophic conditions), except for Wood Lake (Table A.1). Productivity has increased in Wood Lake due to longer water residence time and increased nutrient loads over the last 20 years. As a result, the kokanee grow larger than usual and sustain the highest angling pressure in BC. Wood Lake is the first of five valley bottom lakes in the Okanagan River basin, which is tributary to the Columbia River. Okanagan Lake is the third lake in this chain and separated from Wood Lake by Kalamalka Lake. These three lakes contain both stream- and shore-spawning ecotypes, but historical gene flow was probably low owing to the marked difference in nutrient water chemistry of Kalamalka Lake. Starting in the 1920's, dam construction above and below Okanagan Lake has eliminated the upstream migration of sockeye salmon from most of the Okanagan River and any migration of kokanee between Okanagan and Wood Lake (Long, 2003). Despite the presence of native kokanee, all five Okanagan River lakes were stocked from Kootenay Lake to some extent.

Figure 2.1 The geographic location of the six lakes sampled in British Columbia, Canada. Okanagan and Wood Lakes are part of the Okanagan River Chain, Duncan and Kootenay Lake are part of the Kootenay River Chain, and Christina Lake feeds into the Kettle River, all of which are tributary to the Columbia River in Washington state, USA. Tchesinkut Lake is part of the Fraser River Drainage system, which feeds into the Pacific just above the Canada-USA border. This map was generated by Natural Resources Canada.

In the east, Kootenay is the largest lake and consists of three geochemically distinct arms that experience very limited biotic exchange (Anders *et al.*, 2007). Our study sites are concentrated at the distal end of the West Arm, which is riverine in shape and flow. Since North Arm kokanee (especially Meadow Creek) have been extensively used to stock other lakes in BC, I included two sites from this region (Meadow Creek and Lower Duncan River) to evaluate its influence on the genetic composition

of populations in recipient lakes (including the West Arm, Okanagan, Wood, and Christina Lakes). In 1967, Duncan Dam was constructed just upstream of the North Arm of Kootenay Lake, which created a small reservoir now known as Duncan Lake. Shore-spawners were first observed in this area following the construction of the dam and shore sites are found in close proximity to tributaries. Both Duncan and Kootenay Lakes are part of the Kootenay River chain and are tributary to the Columbia River. Below Kootenay Lake lies Bonnington Falls, a natural feature that has blocked the passage of sockeye since the lake was formed. Christina Lake is part of the Kettle River system, which runs adjacent to the Kootenay River and also feed into the Columbia River. This small lake is very deep, but warm, and kokanee ecotypes exhibit the greatest divergence in spawning timing (~3 months). Despite low fishing pressure, the lake was stocked from the North and West Arms of Kootenay Lake in the 1980's.

In northern BC, Tchesinkut Lake is a small but deep lake in the headwaters of the Fraser River system. The small catchment area of this lake causes smaller tributaries (e.g. Drew creek) to be prone to drought. The shore-spawning site encompasses a small (0.5 km^2) island in the center of the lake, which can be subject to considerable wave action due to wind. Beaver dams are thought to make this lake inaccessible to anadromous sockeye because there are no records of sockeye ever entering this lake (J. DeGisi, *pers. comm*.). Tchesinkut Lake has no record of stocking from foreign lakes.

2.2.2 Site selection & sample collection

Between 2007 and 2011, tissue samples were collected from 16 to 48 mature kokanee (i.e. exhibiting bright breeding coloration) from each sampling site during the peak of their respective spawning period (Table A.2). Adipose fins were taken from live spawners caught with dip nets in Drew Creek of Tchesinkut Lake. At the shore sites of Christina, Tchesinkut, and Wood Lakes, a 3-4 cm gillnet was set parallel to the shoreline over night to catch kokanee because carcasses rarely wash onshore or are quickly consumed by scavengers at these sites. At all other sampling sites, a single hole-punch was used to collect operculum tissue from fresh carcasses found along the banks adjacent to shore- and

stream-spawning aggregations. All tissue samples were preserved in 2 ml vials containing 100% ethanol and stored at -20°C for subsequent DNA analysis.

Given the many spawning streams in Okanagan Lake, only the primary stream- and shore-spawning stocks were sampled. In the West Arm of Kootenay Lake, the two streams that were selected had the greatest potential for gene flow with the shore-spawners based on the populations size and proximity to the shore site. Inclusion of multiple stocks of the same ecotype in most lakes should lead to a more conservative assessment of outlier behavior (i.e. reduces the probability of Type I error).

2.2.3 Data collection

Genotypic data was previously collected for three shore- (*n*=72) and four stream-spawning stocks (*n*=72) in Okanagan Lake in 2007 and 2010 (Table A.2; Russello *et al.*, 2012). For 488 fish sampled Ffrom the other five lakes, total genomic DNA was extracted with the NucleoSpin Tissue kit (Macherey Nagel) according to the manufacturer's suggested protocol for 96-well plates. All samples were polymerase chain reaction (PCR)-amplified at 49 EST-linked microsatellite loci and 8 anonymous microsatellite loci known to be polymorphic among ecotypes in Okanagan Lake kokanee (Russello *et al.*, 2012), except for kokanee from the North Arm of Kootenay Lake. Shore-spawners are not found in the North Arm but 56 stream-spawners were amplified at the 8 anonymous loci for phylogeographic analyses.

Each PCR contained 1.25 µl of 10x PCR buffer, 1.25 µl of 2 mM dNTP mix, 0.5 µl of 1 mM forward primer, 0.5 µl of 10 mM M13 fluorescent labeled primer, 0.5 µl of 10 mM reverse primer, 0.5 Units of *Taq* polymerase and 20 to 80 ng of DNA template for a total reaction volume of 12.5 µl. To allow for multiplex genotyping, all forward primers were modified to incorporate the M13 sequence [5'-TCCCAGTCACGA-CGT -3'] at the 5'-end of the PCR amplicon (Schuelke, 2000) so that the M13 primer that was labeled with one of four fluorescent dyes: 6-FAM (Integrated DNA Technologies) VIC, NED, or PET (Applied Biosystems) could be incorporated. PCR products for four markers, one with each

fluorescent tag, were then combined on a single panel for genotyping. Each reverse primer was modified to include a GTTT-tail for improved scoring quality (Brownstein *et al.*, 1996). All reactions use KAPA *Taq* DNA polymerase (Kapa Biosystems), except for markers *EV170, OMM5008, OMM5067*, and *One14*. Ampli*Taq* Gold DNA polymerase (Applied Biosystems) was used to promote amplification of these markers.

Amplification of targeted loci was achieved using a touchdown cycling program on a Veriti thermal cycler (Applied Biosystems). The program started with an initial denaturation at 94 ˚C for 2 minutes (or 10 minutes for reactions using Ampli*Taq* Gold), followed by 20 cycles at 94 ˚C for 30 seconds, 60 ˚C for 30 seconds, and 72 ˚C for 30 seconds with the annealing temperature decreasing by 0.5 ˚C per cycle. The annealing temperature is held at 50 ˚C for 15 more cycles and then there is a final extension at 72 ˚C for 2 min. DNA fragments were analyzed on an Applied Biosystems 3130XL DNA automated sequencer using the GS500 LIZ size standard to determine fragment length. Two independent investigators manually scored all alleles in GENEMAPPER 4.0 (Applied Biosystems) based on peak topography and intensity. Universal marker bins were used to improve the consistency of allele calls across the entire dataset. At this point, the raw genotypic data generated for the Okanagan Lake kokanee by Russello *et al.* (2012) was incorporated with the newly generated data.

2.2.4 Data quality and definition of genetic units

Loci were tested for null alleles using MICROCHECKER (Van Oosterhout *et al.*, 2004), deviations from Hardy-Weinberg Equilibrium (HWE) using the Markov chain–Monte Carlo (MCMC) approximation of Fisher's exact test (using 1,000 batches with 1,000 iterations; Guo and Thompson, 1992) and linkage disequilibrium (LD) was assessed for all possible marker combinations using simulated exact tests as implemented in GENEPOP 3.3 (Raymond and Rousset, 1995). In tests of LD and HWE, statistical significance (α) was adjusted for the number of simultaneous tests k (α/k for $\alpha = 0.05$) using a sequential Bonferroni procedure (Rice, 1989) to reduce Type I errors. Since the action of selection can generate patterns of LD and violates a critical assumption of HWE, ecotype groups from each lake were

evaluated separately and loci were only removed if LD or Hardy-Weinberg disequilibrium was detected for both ecotypes.

Descriptive statistics for the final dataset were generated in GenAlEx version 6.2 (Peakall and Smouse, 2006), including locus-, site- (Table A.3), and ecotype-specific (Table 2.1) mean number of alleles (N_A), sample size (N), observed heterozygosity (H_o), expected heterozygosity (H_e; Nei, 1987), and the percentage of polymorphic loci in each lake. In lieu of markers known to be truly neutral, a preliminary assessment of population structure was conducted using eight anonymous, highly variable microsatellite markers that are frequently used in population genetic studies of *Oncorhynchus nerka* (Olsen *et al.*, 1996; Scribner *et al.*, 1996; Wright *et al.*, 2008). Given the extensive stocking history of kokanee in BC and unknown strength of philopatry, I tested the correspondence of geographically separated stocks and ecotypes as discrete genetic units by calculating pair-wise estimates of differentiation (F_{ST}) in ARLEQUIN 3.5 (Excoffier and Lischer, 2010) and using the Bayesian clustering method of Pritchard *et al.* (2000) implemented in STRUCTURE 2.3.3 (see population genetics section for details of the analysis).

2.2.5 Outlier locus detection and annotation

A number of different statistical approaches are available for outlier-detection, each with a different algorithm and associated assumptions regarding gene flow, effective population size, and population structure (Storz, 2005; Vasemagi and Primmer, 2005). I tested for signatures of directional selection at polymorphic loci based on patterns of heterozygosity (*ln*RH; Kauer *et al.*, 2003), F_{ST} (DetSel; Vitalis *et al.*, 2003), and both F_{ST} and heterozygosity (Lositan Selection Workbench; Antao *et al.*, 2008; BayeScan; Foll and Gaggiotti, 2008) for each ecotype pair, separately.

In general, selection on a linked marker should appear as a selective sweep, i.e. an increase in homozygosity at the selected gene and its flanking regions as the gene sweeps through the population (Barton, 2000; Kaplan *et al.*, 1989). This process is inferred by a heterozygosity deficit. The *ln*RH test

compares genetic diversity among ecotype pairs by calculating the ratio of expected heterozygosity for each locus (Kauer *et al.*, 2003). Monomorphic loci were assumed to have one allele that differed from the others to avoid dividing by zero. The *ln*RH estimates were standardized to a mean of 0 and a standard deviation of 1, so that 90%, 95% and 99% of the loci are expected to have values of ± 1.64, ± 1.96, and ± 2.58, respectively. Loci with values outside these boundaries were considered significant at the respective level.

Under a pure divergence model (i.e. an ancestral population splits into two daughter populations), outlier behaviour was assessed in a pair-wise fashion based on population specific *F*-statistics using the probabilistic approach implemented in DETSEL 1.0 (Vitalis *et al.*, 2003). Coalescent simulations were used to generate a joint distribution of expected F_{ST} values under a neutral model of evolution (i.e. a confidence envelope) against which outlier behaviour was evaluated. For all post-glacial ecotype pairs, null distributions were generated assuming that the ancestral population had a constant effective population size (N_e) of 500, 1000, or 10,000, prior to a bottleneck when population size (N_o) declined to 500 individuals for a duration of 50, 100, or 1000 non-overlapping generations (T_o). The mutation rate (μ) was assumed to be 0.0001 or 0.00001 and time since the population split (t) was assumed to be 100 generations. Since Duncan Lake was formed only 45 years ago, nuisance parameters were adjusted for this lake (N_e = 100, 500, and 5000; N_o= 50; T_o =5, 10, or 20; t = 1). Outliers were determined based on an empirical *P*-value for each locus at the 90%, 95% and 99% levels using two-dimensional arrays of 50 × 50 square cells (Vitalis *et al.*, 2001). Loci falling outside of the confidence envelope were identified as putatively outliers.

LOSITAN and BAYESCAN both implement the FDIST2 approach of Beaumont and Nichols (1996), which simulates the expected relationship between F_{ST} and H_e under a neutral model of evolution against which outlier behaviour can be assessed. The approach implemented in LOSITAN SELECTION WORKBENCH (Antao *et al.*, 2008) assumes an island model of migration (i.e. a set of populations with constant and equal subpopulation sizes that are connected by gene flow), and uses coalescent simulations to generate the null distribution to identify loci displaying exceptionally high (or low) F_{ST} values. An

infinite alleles model (e.g. assuming each mutation that arises is unique) was used because the dynamics of microsatellite mutation is more complex than is reflected by the step-wise mutation model (i.e. it assumes slippage during replication can cause single repeat unit changes), especially when differences in the type and length of the repeat motifs in our dataset are considered (Ellegren, 2000). In the initial 50,000 simulations, loci outside a 95% confidence interval (CI) were removed so that a more accurate, mean neutral F_{ST} could be calculated. In a second set of 50,000 simulations, this mean neutral F_{ST} was forced to calculate the probability of each locus being under selection based on 90%, 95%, and 99% CI. To assess the influence of population substructure, if any, the same algorithm was implemented in ARELQUIN (Excoffier and Lischer, 2010), but a hierarchical island model of migration was incorporated such that stocks within ecotypes could exchange migrants at a higher rate than among ecotypes, because strays may prefer their native type of spawning habitat (Gomez-Uchida *et al.*, 2011). Results from these two methods were compared, but ultimately the results from ARLEQUIN were not reported owing to its propensity for Type I errors in this study and in other literature (Narum and Hess, 2011).

The FDIST2 approach implemented in BAYESCAN 2.0 (Foll and Gaggiotti, 2008) is modified to use Bayesian-based simulations. For each locus, posterior probabilities are estimated for two alternative models (one with and one without the locus-specific effects of selection) using a reversible jump MCMC approach. The Bayes Factor is calculated from the ratio of posterior probabilities for these two models, which provides the scale of evidence. For the MCMC algorithm, 20 pilot runs of 5000 iterations were conducted followed by 100,000 iterations with a burn-in of 50,000. A Bayes Factor threshold of >3 was used to identify outlier loci. According to Jeffrey's scale of evidence, posterior probabilities of >0.76, >0.91, >0.97 are interpreted as 'substantial', 'strong', and 'very strong' support for the action of selection.

Controlling for multiple testing to reduce the number of false positives can be achieved using a Bonferonni correction following the *ln*RH test and analyses in DETSEL (Schlötterer, 2003) and the false discovery rate (FDR) is often applied following outlier-detection analyses in LOSITAN and BAYESCAN

(Benjamini and Hochberg, 1995). The FDR is defined as the expected proportion of false positives. However, it was difficult to predict which FDR value (e.g. 0.01, 0.05, 0.10) was a suitable threshold for minimizing Type I error without increasing Type II error since selection is weak in this system (Narum and Hess, 2011) and only polymorphic EST-linked markers were included in this dataset. Instead, I used the repeated detection of outlier behaviour by multiple algorithms as evidence for robustness (Bonin *et al.*, 2006; Foll and Gaggiotti, 2008). Therefore, outlier-detection results from all four approaches at 90%, 95% and 99% CIs were compared and 'true outliers' (as referred to hereon) were identified as any locus exhibiting outlier behaviour in at least two of the four approaches. Any locus exhibiting outlier behaviour in only one approach was considered a false positive ('false outliers'). If similar selective forces are driving divergence at the same genes in all five closely related ecotype pairs, the same genes regions may be involved in ecotype divergence in all five lakes (Campbell and Bernatchez, 2004; Colosimo *et al.*, 2004). Therefore, 'true outliers' identified in two or more lakes were identified as 'repeat-outliers'. 'Repeat-outliers' represent the most promising candidates to be under selection. Any locus not detected by any outlier-detection approach in any lake was considered to be truly neutral (referred to as 'neutral loci' hereon).

Sequence similarity searches were conducted for all 'true outlier' loci within the consortium for Genomics Research on All Salmon (cGRASP) and the salmonidae database in BLAST to identify the expressed genes linked to 'true outlier' loci (Salem *et al.*, 2010). The functional annotations of each locus are discussed to shed light on some possible mechanisms underlying barriers to gene flow among reproductive ecotypes.

2.2.6 Population genetic analyses

Since loci known to be free of selection can provide a more accurate picture of neutral population structure, I assessed the nature of the origin of shore-spawning populations by constructing a discriminate analysis of principal components plot (DAPC) using only 'neutral loci'. This ordination method assumes no model of evolution and plots individuals using linear combinations of allelic data

(synthetic variables) that maximize differences observed among pre-defined groups while minimizing within-group variation (Jombart *et al.*, 2010). I plotted all individuals using the first two principal components while retaining 90% of the variation. Lines connect members of each ecotype group to a central point with a 95% confidence envelope around each ecotype group. If greater overlap is exhibited among groups from the same lake than groups of the same ecotype, these populations likely have polyphyletic origins and ecotypes likely diverged in sympatry within each lake.

Most outlier-detection approaches use allele frequency distributions to evaluate outlier behaviour, therefore false positives can result when heterozygosity excess is generated through the rapid loss of alleles characteristic of a population bottleneck (Teshima *et al.*, 2006; Wiehe *et al.*, 2007). Knowing several lakes included in this study have undergone substantive population declines, I tested for recent population bottlenecks using the one-tailed Wilcoxon test (Cornuet and Luikart, 1996) and mode-shift indicator test implemented in BOTTLENECK (Piry *et al.*, 1999) under a two-phase model of mutation (80% step-wise mutation model; Dirienzo *et al.*, 1994) to determine if demographic history influenced outlier-detection results.

Loci truly under selection are expected to show distinct patterns of genetic variation compared to neutral loci (Storz, 2005). Therefore, to further validate outlier loci, I calculated pair-wise estimates of population differentiation (Weir and Cockerham, 1984) with 95% CIs at 'repeat-outliers', 'true outliers', and 'neutral loci' for each ecotype pair in ARLEQUIN 3.5 (Excoffier and Lischer, 2010). The overall amount of genetic variation occurring at the ecotype level was also assessed across all lakes using a hierarchical analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992) as implemented in ARLEQUIN 3.5 (Excoffier and Lischer, 2010). If outliers are truly linked to adaptive gene regions, 'repeat-outliers' and 'lake-specific true outliers' are expected to show greater differentiation at the ecotype level overall and among each ecotype pair compared to 'neutral loci'. If ecotypes have diverged by a common genetic mechanism in multiple lakes, 'repeat outlier' will show greater differentiation at the ecotype level than the 'true outliers'. Alternatively, if ecotype divergence has a unique genetic basis in each lake, the 'lake-specific true outliers' will have much higher pair-wise F_{ST} estimates than the

'repeat-outliers'. A student t-test is used to determine if 'lake-specific true outliers' had significantly higher F_{ST} estimates across all ecotype pairs (P<0.05).

Finally, the Bayesian clustering algorithm implemented in STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) was used to visualize differences in the level of admixture between ecotypes using the 'true outliers' and 'neutral loci'. Assuming an admixture model and correlated allele frequencies between clusters (*K*), I used a burn-in period of 500,000, then 1,000,000 MCMC replicates. Since structuring is weak in these lakes, sampling habitats (stream and shore) were used as prior information to assist in clustering (Hubisz *et al.*, 2009). In each run, samples were assigned to the cluster that they had the greatest probability of originating from. Then the Δ*K* method (Evanno *et al.*, 2005) was used to infer the most likely number of genetically distinct units within each lake. The number of clusters was varied from 1 to 7 with 5 iterations per value of *K* to confirm the consistency of log-likelihood probabilities. If ecotype pairs are identified as two genetically discrete clusters (*K*=2) by 'true outliers' but not 'neutral loci' (*K*=1), this will further suggest that gene regions linked to 'outlier loci' are truly under divergent selection.

2.3 Results

2.3.1 Data quality

A total of 688 individuals and 50 loci were retained following assessments of data quality. Twelve individuals with 21.6-68.4% missing data were removed owing to poor DNA quality. Overall, the final dataset contained 1.4% missing data and less than 2.9% at any single spawning site. Loci *EV103, EV626* and *One109* exhibited false alleles in both shore- and stream-spawners from Okanagan Lake, and the latter two loci also deviated from HWE in Okanagan Lake. Three pairs of loci consistently exhibited patterns of LD in Okanagan and Wood Lake (*OMM5099* & *Ots29, Ca687* & *EV712*, and *Ca613* & *Ots14*) and two of which were identified in Christina Lake (*OMM5099* & *Ots29* and *Ca687* & *EV712*). Loci *Ca613, Ca687*, and *Ots29* were removed because they had more missing data. Locus *Ssa85* was

also removed because primers sequences used throughout data were inconsistent. Consequently, the final dataset was reduced from 57 to 50 loci.

Based on eight anonymous loci, two discrete genetic units corresponded with the major drainage systems represented in this study were inferred (K=2; Figure 2.2). Tchesinkut Lake in the Fraser River system made up one cluster and the other five lakes from the central and eastern interior of BC that are tributary to the Columbia River made up the other cluster. A small secondary peak inferred five clusters (K=5) corresponding with: i) Tchesinkut lake, ii) lakes in the Okanagan River Chain (Okanagan and Wood), iii) Duncan Lake and nearby stocks from the North Arm of Kootenay, iv) the West Arm of Kootenay Lake and stream-spawners from Christina Lake, and v) Christina Lake shore-spawners. Therefore, stream-spawners in Christina Lake are not native and the entire lake had to be removed from all subsequent analyses. Assessments of demographic history showed no evidence of heterozygote excess and a normal L-shaped distribution of alleles, therefore no populations were removed due to a severe bottleneck despite observations of population declines in the recent past. Estimates of neutral genetic differentiation among stocks within each ecotype group were low on average (F_{ST} = 0.008) ranging from -0.003 to 0.021 (Table A.4), allowing for spawning sites with the same habitat type to be pooled and thereby increase sample sizes without generating population substructure.

The level of polymorphism in our dataset varied across lakes, but was high overall (Table 2.1 and A.3). It ranged from 80% to 100%, and was lower in lakes outside of the Okanagan River Chain. Eight monomorphic loci were found in Tchesinkut Lake (*EV291, EV691, OMM5032, OMM5037, OMM5099, OMM5121, One8, Ots06*), three in Duncan Lake (*EV723, EV911, Ots06*) and one in Kootenay Lake (*Ots06*). The number of alleles per locus ranged from 1 to 23, with a mean of 5.35. Heterozygosity ranged from 0 to 0.93 across loci with an overall mean of 0.46. Okanagan Lake kokanee had the most alleles (mean 7.32) and gene diversity (mean 0.518). No significant trends were observed among shore- and stream-spawners in N_A or H_E, although H_E was slightly higher in the stream ecotype in four out of five lakes.

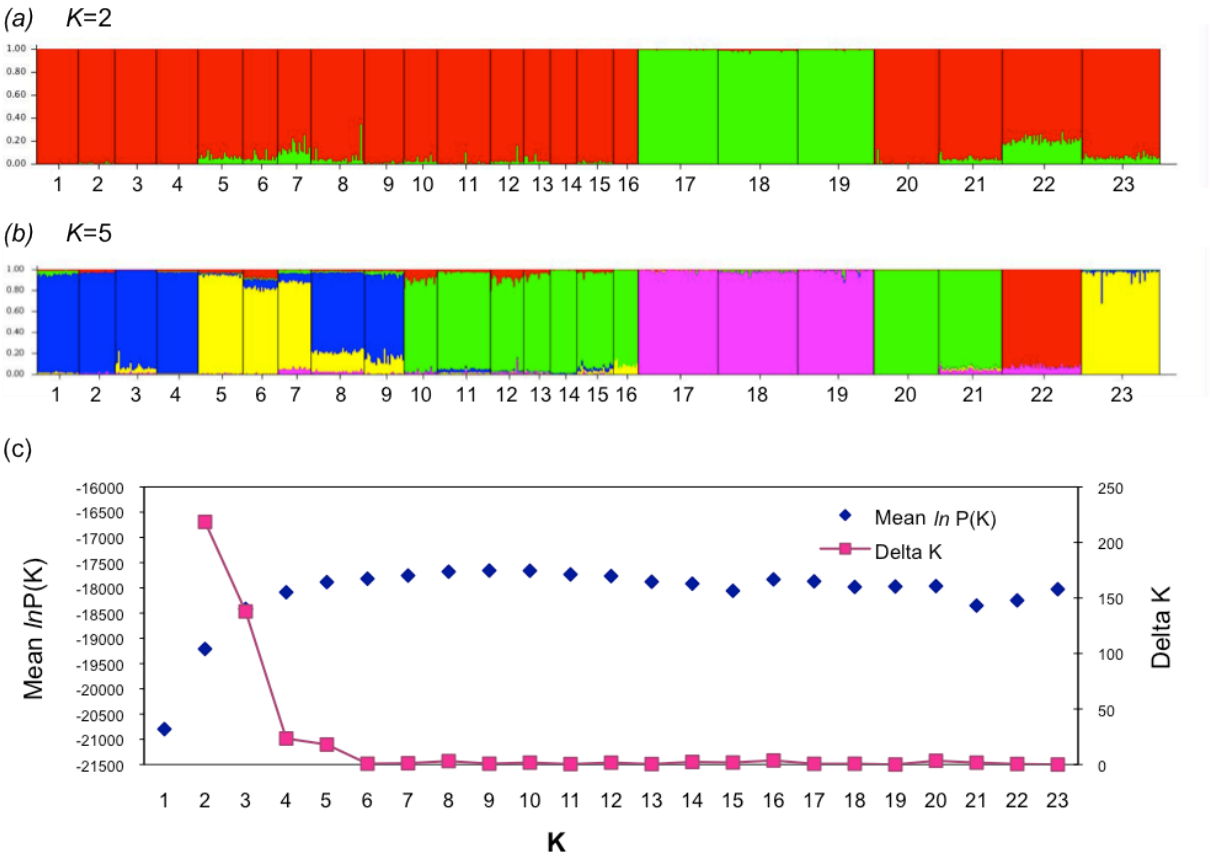

Figure 2.2 STRUCTURE plots depict 23 sampled stocks as (a) 2 and possibly (b) 5 discrete genetic clusters (*K*), based on (c) estimates of *ΔK* (Evanno *et al.*, 2005) for a range of *K* values (1-23). For each individual, the probability of membership to each cluster is represented by the color composition of a vertical bar. Sampled stocks include Duncan Lake (1-4): 1) Griz shore, 2) Little Glacier shore, 3) SOB Creek and 4) Upper Duncan River; West Arm of Kootenay Lake (5-7): 5) Six Mile shore, 6) Six Mile Creek, 7) Harrop Creek; North Arm of Kootenay Lake (8-9): 8) Lower Duncan River, 9) Meadow Spawning Channel; Okanagan Lake (10-16): , 10) Northeast shore, 11) Northwest shore, 12) Southeast shore, 13) Peachland Creek, 14) Penticton Creek, 15) Mission Creek, 16) Powers Creek; Tchesinkut Lake (17-19): 17) the island shore, 18) Drew Creek, 19) Tchesinkut Inlet Creek; Wood Lake (20-21): 20) the shore,21) Middle Vernon Creek; and Christina Lake (22-23): 22) the shore, 23) Sandners Creek.

Table 2.1 Estimates of population genetic parameters for each ecotype within each lake using all 50 loci, including: sample size (N), mean number of alleles per locus (N_A), range in the number of alleles, mean expected heterozygosity (H_e), mean observed heterozygosity (H_o), and the percentage of polymorphic loci (%poly).

Lake	Ecotype	N	N_A	N_A range	H_e	H_o	% poly
Okanagan	Shore	72	7.451	2-23	0.507	0.513	100%
	Stream	69	7.196	2-23	0.516	0.507	100%
Wood	Shore	39	5.196	2-19	0.505	0.525	100%
	Stream	38	5.137	1-17	0.526	0.535	96%
Duncan	Shore	47	6.000	1-21	0.515	0.527	90%
	Stream	50	6.280	1-22	0.529	0.538	92%
Kootenay	Shore	27	4.320	1-14	0.438	0.449	92%
	Stream	41	5.220	1-17	0.467	0.474	98%
Tchesinkut	Shore	48	3.451	1-13	0.326	0.337	80%
	Stream	94	3.843	1-14	0.317	0.326	80%

2.3.2 Outlier locus detection and dataset definition

Thirty out of 42 EST-linked microsatellite loci (71.4%) and three of out eight anonymous microsatellite loci (37.5%) were identified as putative outliers by at least one approach in at least one lake (without correcting for multiple comparisons; Table 2.2). Based on patterns in outlier behaviour, each locus is classified as 'neutral', a 'false outlier', or a 'true outlier'. 'Neutral loci' included the 17 loci that were polymorphic in all lakes but exhibited no outlier behaviour at all. The 'false outliers' included the 18 loci that were only detected by one algorithm. The 'true outliers' included 15 loci that were detected by two or more algorithms in at least one lake (Table 2.2; Table A.3). From the 'true outliers', a subset of four loci detected in multiple lakes were defined as 'repeat-outliers' (*EV358, OMM5003, OMM5067, TAP2*), although none showed parallel patterns across all five ecotype pairs. I also defined the 'lake-specific true-outliers' dataset for situations when only the 'true outliers' for each respective lake (3 to 6 loci) were used in an analysis when lakes were analyzed independently. Although 33% to 66% of the 'lake-specific true outliers' were included in the 'repeat-outliers' for each lake, this dataset allows for loci with inconsistent but strong outlier behaviour in a lake to be included (e.g. *OMM5125* in Okanagan and Ots06 in Wood Lake) while excluding the 'true outliers' that are monomorphic in that lake (e.g. *Ots06, OMM5037, OMM5121, EV691*).

Table 2.2 Loci detected as outliers in four different algorithms (BAYESCAN/LOSITAN/DETSEL/*lnRH*) with a probability of <0.010, <0.05, and <0.01 are identified in each of five British Columbian Lakes. Loci detected by only one approach are listed as false outliers ('False'). Loci detected by at least two approaches in at least one lake are 'true outliers' (Outlier), and those 'true outliers' that are detected in two or more lakes are 'repeat outliers' (R Outlier). Each marker is identified as either an EST-linked microsatellite marker (EST) or an anonymous microsatellite marker (Anon).

Marker type	Locus	Okanagan	Wood	Duncan	Kootenay	Tchesinkut	Status
EST	*TAP2*	-/-/*/-			-/**/*/-	-/*/**/-	R Outlier
EST	*EV358*	-/**/*/**	-/**/*/-	-/-/**/**		-/-/*/-	R Outlier
EST	*OMM5003*		-/**/*/-	-/*/**/-		*/***/**/-	R Outlier
EST	*OMM5067*		-/**/**/**	-/-/**/-	-/**/*/**	-/-/*/-	R Outlier
EST	*OMM5125*	***/***/**/-					Outlier
EST	*Ots06*	-/-/-/***	-/***/**/***	MONO	MONO	MONO	Outlier
EST	*EV862*	-/-/**/-	-/*/*/-				Outlier
EST	*EV170*	-/-/-/**			-/*/-/*		Outlier
EST	*EV642*	-/**/**/-	-/*/-/-		-/-/*/-		Outlier
EST	*EV685*		-/-/-/*		-/**/-/-	-/-/*/***	Outlier
EST	*OMM5037*		-/**/-/**			MONO	Outlier
EST	*EV691*		-/**/-/-	-/**/-/*		MONO	Outlier
EST	*EV740*			-/***/-/***			Outlier
EST	*OMM5033*			-/-/-/**	-/*/-/*		Outlier
EST	*OMM5121*				-/**/*/-	MONO	Outlier
EST	*EV536*	-/-/*/-					False
EST	*EV188*	-/-/*/-	-/**/-/-				False
EST	*OMM5124*		-/**/-/-				False
EST	*Ca983*		-/*/-/-				False
EST	*EV291*		-/-/-/*			MONO	False
EST	*EV597*		-/-/*/-				False
Anon	*One8*		-/-/*/-	-/*/-/-		MONO	False
EST	*OMM5058*		-/*/-/-	-/*/-/-			False
EST	*EV475*			-/-/**/-			False
EST	*OMM5053*		*/-/-/-	-/-/-/*	-/*/-/-		False
EST	*OMM5032*				-/*/-/-	MONO	False
EST	*EV365*				-/**/-/-		False
EST	*EV634*				-/*/-/-		False
Anon	*One108*				-/*/-/-		False
EST	*EV220*				-/-/-/**		False
Anon	*Ots14*				-/-/-/*		False
EST	*EV484*					-/-/-/*	False
EST	*EV911*			MONO		-/-/-/**	False

MONO indicates that the locus is monomorphic in that lake
* P<0.10, ** P<0.05, *** P<0.01

Assessments of outlier behaviour for 50 loci in each of five lakes (a total of 250 possible detections) yielded positive detection rates from 1.2% to 13.0% across the four approaches (Table 2.3). LOSITAN and DETSEL had the highest detection rates and BAYESCAN had the lowest. Type I error rates were much higher than Type II across all approaches except BAYESCAN. Three loci were detected by all three of the other approaches, demonstrating that BAYESCAN is under-sensitive to patterns of outlier behaviour.

However, all other approaches exhibited relatively high Type I error rates (33% to 45%). LOSITAN, DETSEL and *ln*RH detected the most loci that were subsequently identified as 'false outliers' (13, 10, and 9 respectively). DETSEL, *ln*RH, and particularly LOSITAN demonstrated over-sensitivity to outlier behaviour because they detected loci determined to be 'false outliers' (Table 2.3). Although, there were four instances when a 'repeat outlier' was detected by only one approach in a lake, therefore it is possible some of these 'false outliers' are actually 'true outliers' and/or some 'true outliers' are consistent across more lakes than I indicate here. Overall, a large proportion of loci identified as 'false outliers' were only detected in either Kootenay (6 loci) or Wood Lake (4 loci), which showed the greatest neutral divergence among shore and stream spawning sites at the eight anonymous loci (F_{ST}=0.039 and 0.015, respectively; Table A.4). When the 18 loci identified as 'false outliers' are discarded, the overall detection rate of 'true outliers' is 30% (15 out of 50 loci).

Table 2.3 A summary of the total number of loci detected as statistical outliers by each of four algorithms, as well as the number of false positive detections and false negative detections to assess the sensitivity of each algorithm. False positives are the loci detected by the present algorithm that were not detected in any other lakes by any other algorithm. The false negatives are loci not detected by the present algorithm when all three other algorithms detected it as an outlier.

Outlier-detection approach	No. outliers detected (out of 50)	False positive	False negative
BAYESCAN	3	0	3
LOSITAN	33	7	0
DETSEL	27	3	0
*ln*RH	20	5	2

2.3.3 Neutral and adaptive population divergence

At the 15 'neutral loci', shore- and stream-spawners sampled from the same lake appeared more genetically similar to one another than they were to groups of the same ecotype in other lakes based on the DAPC analysis (Figure 2.3). Lakes from the same geographical region were clustered together. There was considerable overlap among groups from Okanagan and Wood Lakes, and some overlap among groups from Kootenay and Duncan Lakes as well. Tchesinkut showed the least genetic similarity to any of the other groups. Based on the organization of groups, the first principal

component appears to correspond to a north-south axis, which captures the majority of the variation, and the second principal component corresponds to an east-west axis.

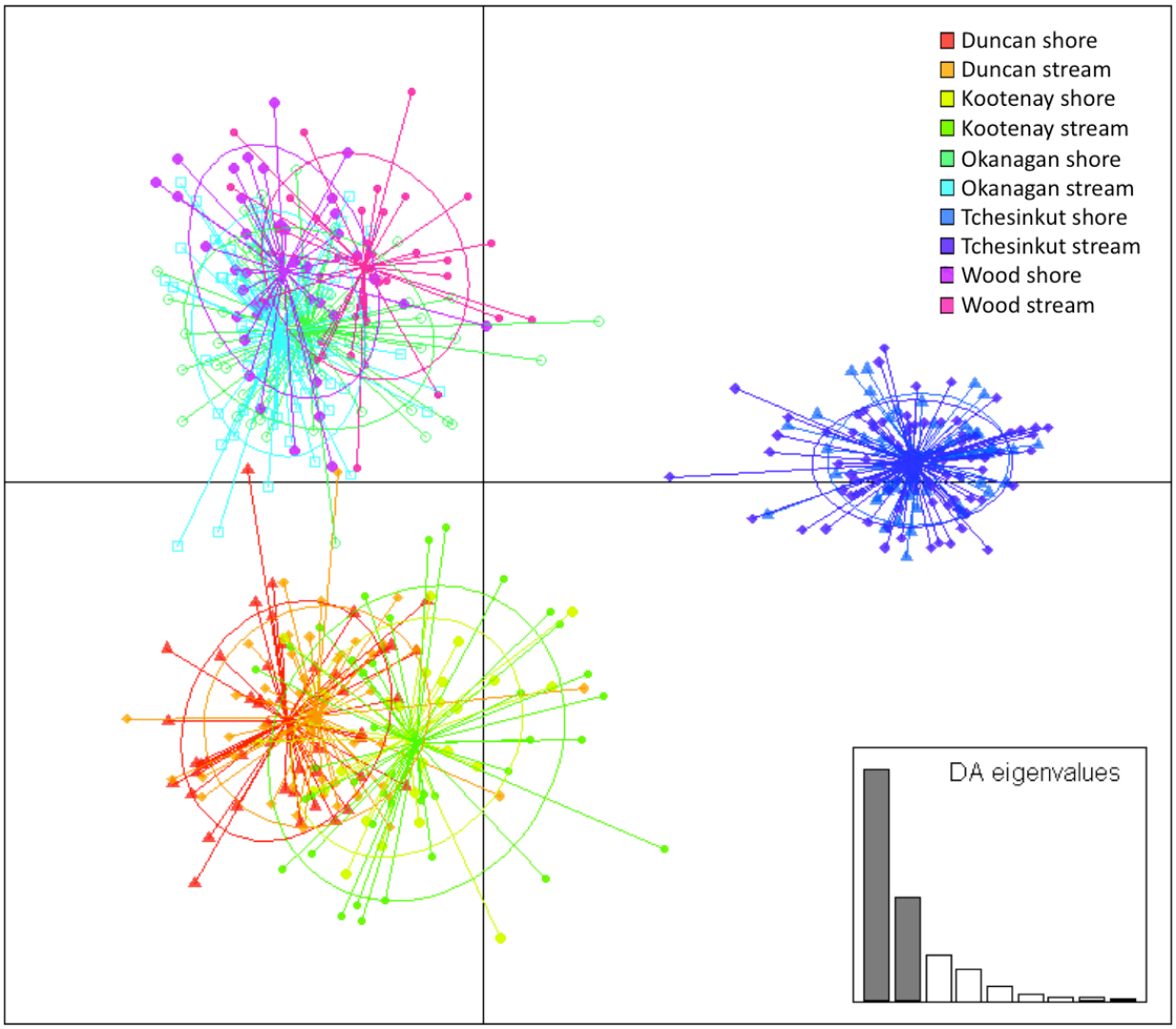

Figure 2.3 A discriminate analysis of principal components (DAPC) plot depicting the relationships among individuals from each ecotype group in each lake based on genetic similarity. Similarity is estimated from polymorphism frequency data at 15 'neutral loci' and displayed on the first two principal components (x- and y-axes).

In a global analysis of the hierarchical organization of genetic variation, among-ecotype variation putatively due to selection (outlier loci) generally exceeded the variation due to drift (neutral loci; Table 2.4). Genetic variation occurring among ecotype groups based on the four 'repeat-outliers', all 15 'true outliers', and 15 'neutral loci' was 3.33%, 2.49%, and 0.72% respectively. Overall, there was a

significant difference in patterns of variation revealed at 'repeat-outliers' compared to 'neutral loci' (Chi-squared test, *P*=0.002), but not at 'true outliers' compared to the 'neutral loci' (Chi-squared test, *P*=0.332).

Table 2.4 The percentage of genetic variation occurring among lakes and within lakes among ecotypes as assessed by a hierarchical analysis of molecular variance (AMOVA) at the 'repeat-outliers', all 'true outliers' and 'neutral loci'.

Level	4 Repeat-outliers (%)	15 True outliers (%)	15 Neutral loci (%)
Among lakes	12.71*	19.14*	14.98*
Among ecotypes	3.33*	2.49*	0.72*
Within ecotypes	83.96	78.37	84.30

* indicates significance of P<0.05

Similarly, genetic differentiation among ecotypes was consistently higher using outlier loci than neutral loci in pair-wise F_{ST} comparisons for each lake (Table 2.5). The mean F_{ST} for the 'repeat-outliers', 'lake-specific true outliers' and 'neutral loci' were 0.044 (range of 0.013 to 0.104), 0.074 (range of 0.017 to 0.126), and 0.009 (range of -0.001 to 0.032), respectively. A significant difference between the 'repeat-outliers' and lake-specific 'true outliers' were observed in Kootenay and Tchesinkut Lakes, and particularly in Okanagan Lake. However, only marginal levels of differentiation were found in Duncan Lake by any dataset. Wood Lake ecotypes were highly differentiated using both outlier datasets (F_{ST}=0.126), but also using 'neutral loci' (F_{ST}=0.032). Overall, the level of ecotype differentiation revealed at 'lake-specific true outliers' was significantly greater than 'repeat-outliers' (paired t-test; *P*=0.018), suggesting that the one or two 'true outliers' that were not identified in any other lakes can substantially increase amount of genetic variation that reflects ecotype differences within individual lakes.

Table 2.5 The proportion of genetic variation occurring among ecotype pairs from five lakes as assessed by pair-wise F_{ST} using 4 'repeat-outliers', 'lake-specific true outliers' lake (number of loci indicated in parentheses), and 15 'neutral loci'.

Lake	4 Repeat-outliers	Lake-specific true outliers (n_{loci})	15 Neutral loci
Duncan	0.013*	0.017* (4)	-0.001
Kootenay	0.044*	0.071* (5)	0.008*
Okanagan	0.017*	0.081* (3)	0.005*
Wood	0.104*	0.126* (6)	0.032*
Tchesinkut	0.041*	0.074* (3)	0.000

* indicates significance of P<0.05

Distinct patterns of populations structuring were revealed by STRUCTURE plots for three of the five lakes when using the lake-specific 'true outliers' compared to 'neutral loci' (Figure 2.4). Using the outliers, two clusters (K=2) corresponding to shore- and stream-spawning ecotypes were inferred in Okanagan, Wood, Tchesinkut, and Kootenay Lakes. Using the neutral loci, two clusters were only inferred in Wood Lake. Complete admixture was inferred by both datasets in Duncan Lake (K=1). Overall, these results are consistent with previous analyses in demonstrating differential levels of genetic differentiation among kokanee ecotypes using outliers versus neutral loci in several lakes.

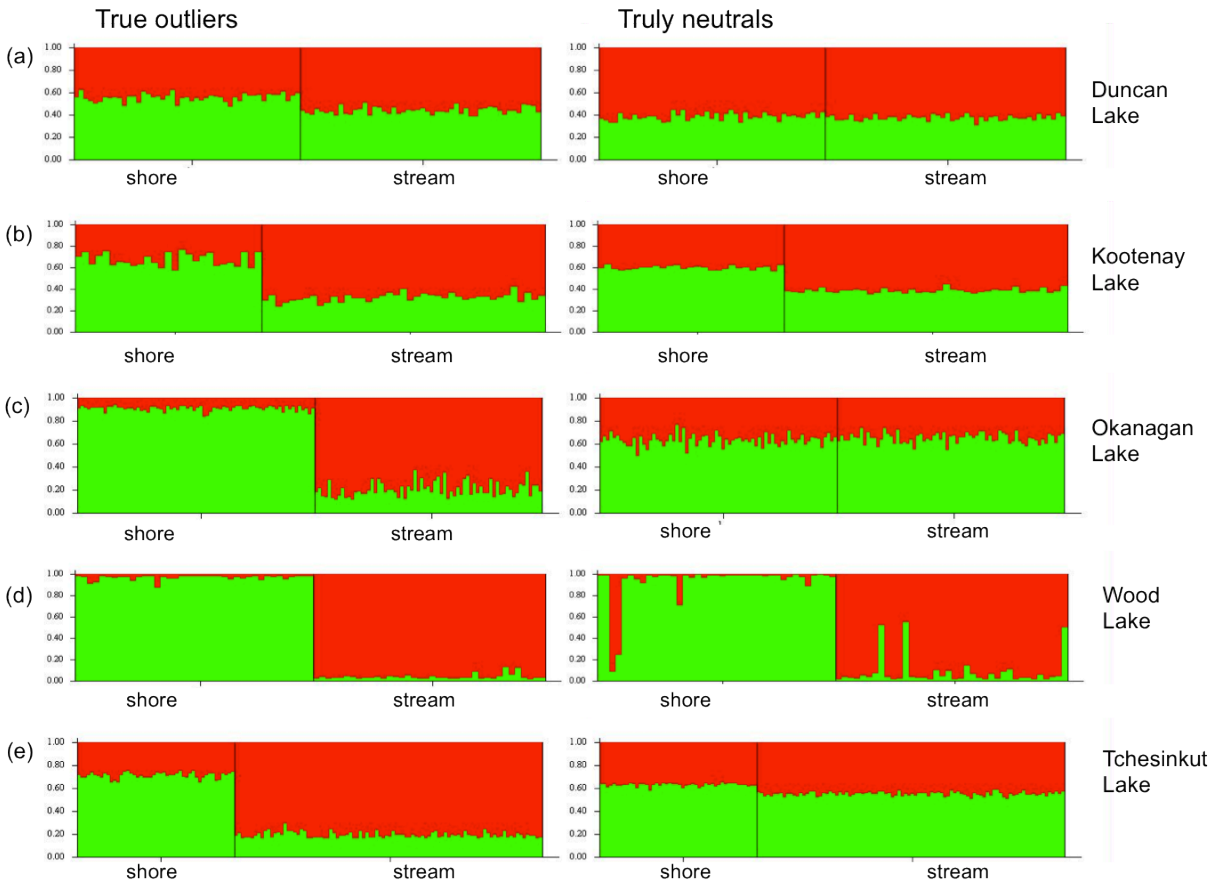

Figure 2.4 STUCTURE plots indicate the probability of membership to a cluster for each individuals (vertical bars) when K=2 is forced. Green represents one cluster and red represents the second cluster. Plots were constructed using the neutral loci (n_{loci}=15; right) and lake-specific outliers (n_{loci}=3-6; left) for (a) Duncan Lake, n_{loci}=4, (b) Kootenay Lake, n_{loci}=5, (c) Okanagan Lake, n_{loci}=3, (d) Tchesinkut Lake, n_{loci}=3, and (e) Wood Lake, n_{loci}=6.

2.3.4 Putative functional annotations of outlier loci

Annotations for nine of fifteen EST-linked outlier loci were recovered from the cGRASP database (*Salmo salar*) or Genbank via BLAST (using 'salmonidae'), including three of the four 'repeat-outliers' (Table 2.6). Locus *EV358* showed high coverage and strong sequence similarity to the chemokine receptor type 4 (CXCR4) protein-coding gene in Atlantic salmon (*Salmo salar*; Genbank # NP_001158765) and rainbow trout (*Oncorhynchus mykiss*; Genbank # CAA04493), which is a rhodopsin-like G protein receptor involved in immunological functioning. Locus *OMM5003* aligned with a plasminogen activator inhibitor-1 RNA-binding (PAIRB) protein-coding gene (Genbank #

BT044770) in *Salmo salar* (cGRASP#gn1|UG|Ssa#S47626048). Its function is unknown in salmonids. Locus *TAP2* exhibited similarity to the transport-associated protein-coding gene in *Salmo salar* (Genbank # CAB05916). This protein is found in the ATP-binding cassette (ABC) transporter transmembrane region superfamily, which hydrolyzes ATP to export substrates (e.g. noxious substances and extracellular toxins) across cellular membranes. Locus *EV170* showed similarity to a malate dehydrogenase protein-coding gene found in *Salmo salar* (Genbank # ACN10417.1), which is an important part of the citric acid cycle and therefore involved in energy metabolism. Locus *EV642* exhibited weak similarity to a transcription elongation factor B (ELOB) polypeptide 2 protein-coding gene in *Salmo salar* (cGRASP # gn1|UG|Ssa#S47726589). Locus EV740 showed similarity to a secretory carrier-associated membrane protein (SCAMP)-coding gene in *Salmo salar* (Genbank # ACI66888.1) as well as a 3-oxo-5-alpha-steroid 4-dehydrogenase 2 protein-coding gene in *Salmo salar* (Genbank # ACM08278.1), which is found in the phospholipid methyltransferase (PEMT) region. Locus EV685 showed similarity to a procollagen-proline, 2-oxoglutarate 4-dioxygenase, alpha 1 polypeptide precursor in *Salmo salar* (Genbank NP_001167096.1). Locus *EV691* was similar to a Ras-related protein Rab-14 coding gene found in *Salmo salar* (Genbank # ACN58615.1.1), which is part of the P-loop-containing nucleoside triphosphate hydrolase (NTPase) region. Locus *EV862* showed some similarity to an LBH protein-coding gene in *Salmo salar* (Genbank # ACI67592.1), which is a highly conserved transcription activator for the mitogen-activated protein kinase-signaling pathway in the heart. Annotations were not found for most of the OMM- markers originally described in rainbow trout (*Oncorhynchus mykiss*) by Rexroad *et al* (2005).

Table 2.6 A description of the gene annotations for each of the 15 EST-linked microsatellite loci exhibiting 'true outlier' behaviour. For each annotated gene, the name, function of the protein, location of expression and the lake(s) in which outlier behaviour was detected are given.

Locus	Gene	Protein function	Location of expression	Lake
EV358	chemokine receptor 4 (CXCR4) gene	transmembrane receptor	immune system	DUN/WOO
OMM5003	1 RNA-binding protein (PAIRB or PAI1) gene	receptor	blood	DUN/TCH/WOO
TAP2	transporter 2 ATP-binding cassette	antigen peptide transporter	immune system	KOO/TCH
OMM5067	-	-	-	WOO/KOO
EV170	Malate dehydrogenase	enzyme in citric acid cycle	mitochondria	KOO
EV642	Transcription elongation factor B polypeptide 2 (ELOB) gene	RNA transcription regulator	n/a	OKA
EV685	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide 2 (P4HA2) gene	enzyme	n/a	TCH
EV691	Ras-related protein Rab-14	protein transporter	n/a	DUN
EV740	secretory carrier-associated membrane protein	transporter	n/a	DUN
EV862	LBH protein	transcription regulator	heart	WOO
OMM5037	-	-	-	WOO
OMM5121	-	-	-	WOO
Ots06	-	-	-	WOO
OMM5125	-	-	-	OKA
OMM5033	-	-	-	KOO

OKA- Okanagan Lake, WOO- Wood Lake, DUN-Duncan Lake, KOO- Kootenay Lake, TCH- Tchesinkut Lake

2.4 Discussion

A recent surge of studies are attempting to uncover the genetic basis of phenotype-environment correlations using genome scans, particularly in non-model organisms, in an effort to better understand the role of natural selection in generating biodiversity (Campbell and Bernatchez, 2004; Namroud *et al.*, 2008). Loci exhibiting parallel patterns of outlier behaviour across multiple ecotype pairs have the greatest potential to be under the action of natural selection and involved in initiating and maintaining barriers to gene flow in divergent ecotypes (Hendry, 2009). Some success has arisen from the study of post-glacial fishes, which generally meet the necessary criteria to ensure that

statistical outliers truly reflect the locus-specific effects of selection rather than neutral processes (i.e. divergence of ecotypes among closely related populations with a well understood phylogeographic history; Hendry, 2009). In global assessments of population structure using anonymous loci, patterns of variation appear to reflect the known colonization and phylogeographic history of *Oncorhynchus nerka* in British Columbia (Taylor *et al.*, 1996). Clustering analyses with (STRUCTURE plots; Figure 2.2) and without an inferred model of evolution (DAPC; Figure 2.3), as well as population differentiation (pair-wise F_{ST}; Table 2.5) and trends in polymorphism (Table 2.1) demonstrate that Tchesinkut Lake is distinct from the other lakes. This corresponds with studies that suggest sockeye from the Bering Refugia colonized post-glacial lakes in northern BC and Alaska and sockeye from the Columbia Refugia colonized southern BC and northern USA (Avise *et al.*, 1987; Foote *et al.*, 1989; Taylor *et al.*, 1996). Lakes sampled in the north, central interior, and east interior suggest that there is limited connectivity among populations in the Kettle, Kootenay, Okanagan, and Fraser River systems (i.e. since the water in BC reached current levels ~ 10,000 years ago), but populations have a more recent shared lineage within these river systems. Strong genetic similarities observed in Okanagan and Wood Lakes and Duncan and Kootenay Lakes are not surprising given the connectivity between them prior to dam construction. The DAPC demonstrated that shore- and stream-spawners sampled from the same lake were more similar to each other than groups of the same ecotype from other lakes, even in geographically proximate and recently connected lakes such as Okanagan and Wood Lakes (Figure 2.3). Overall, results suggest that kokanee populations are closely related, divergence was likely initiated and maintained in sympatry (although differentiation is still quite weak), and it is unlikely that dispersal or stocking from other lakes founded any of the stocks in this study, except for Christina Lake. Christina Lake was eliminated because stream-spawners were genetically much more similar to kokanee from West Arm of Kootenay Lake (Figure 2.2b) than the shore-spawners in Christina Lake, suggesting that stocked fish founded the Sandners Creek stock. The remaining populations appear to conform to all the criteria initially set out for ensuring that neutral evolutionary processes would not confound signals of selection detected in this study.

2.4.1 Evidence of parallel patterns

No locus exhibited outlier behaviour in all five ecotype pairs, but four loci were detected in three or four ecotype pairs. In general, the strength of the hitchhiking effect is determined by the strength of selection, recombination rate, initial frequency of the advantageous allele, and time to fixation (Kaplan *et al.*, 1989; Smith and Haigh, 1974; Storz, 2005). Therefore, the lack of congruency across all pairs may reflect a lack of power to detect signatures of selection (Stinchcombe and Hoekstra, 2008; Teshima *et al.*, 2006) because the environment (and selection pressure) is heterogeneously distributed, traits under selection are not pleiotropic (e.g. morphological traits are more likely to have contrasting fitness consequences in alternative environments than physiological traits; Liao *et al.*, 2010), or linkage between the neutral marker and the fitness-relevant mutation may be broken up by recombination if loci are not tightly associated with each other or sufficient time has passed (Nielsen, 2005; Stinchcombe and Hoekstra, 2008). If adaptive traits have arisen from standing variation (i.e. a soft sweep), the gene associated with the favoured trait will have existed in the population at a low frequency for some time, generating polymorphism at the linked microsatellite marker through neutral processes (i.e. mutation and drift). When the population encountered the new environment, several different alleles may have hitchhiked as the selected gene moved towards fixation (Barrett *et al.*, 2011; Hermisson and Pennings, 2005). A simulation study by Teshima *et al.* (2006) demonstrates that many loci under selection may be missed by genome scans, especially if the locus under selection was previously neutral. Signatures of selection may be obscured by other confounding factors including inflated differentiation at neutral loci due to random chance, sampling bias (i.e. ascertainment bias; Thornton and Jensen, 2007), the inclusion of unidentified strays (Excoffier *et al.*, 2009). Biologically significant levels of differentiation were not detected among spawning sites of the same ecotypes, therefore straying among spawning sites must maintain gene flow despite natal homing. Alternatively, there may be a biological reason for the lack of parallel patterns in outlier behaviour. Many other empirical studies have failed to detect parallel patterns at the genetic level (Campbell and Bernatchez, 2004; DeFaveri *et al.*, 2011; Egan *et al.*, 2008; Renaut *et al.*, 2011; Tice and Carlon, 2011) and different mutations have been identified to produce the same phenotype in some closely related populations

(Hoekstra *et al.*, 2006) bringing into question how often parallel mechanisms underlie observed replicated phenotype-environment correlations in natural populations (Arendt and Reznick, 2008; Elmer and Meyer, 2011).

2.4.2 Outlier locus detection

Several challenges are associated with the statistical detection of loci that exhibit signatures of selection. Models of evolution used to estimate population parameters that do not accurately reflect the complex demographic history of natural populations (Akey *et al.*, 2004) cause all outlier-detection approaches to be susceptible to high Type I and II error rates, especially when selection is weak (Excoffier *et al.*, 2009; Narum and Hess, 2011). Since five ecotype pairs cannot be expected to fit all of the assumption associated with any one algorithm, I chose to use several algorithms to avoid making spurious conclusions due to violated assumptions (Bonin *et al.*, 2006; Egan *et al.*, 2008; Luikart *et al.*, 2003; Vasemagi and Primmer, 2005; Wilding *et al.*, 2001). All four algorithms used a different combination of i) model of evolution and consideration of neutral population structure (e.g. pure divergence model, island model, hierarchical island model, no model), ii) metric for assessing signatures of selection (i.e. H_e to detect a selective sweep or F_{ST} to detect excessive population differentiation), iii) approach for assessing outlier behaviour (i.e. null hypothesis testing or comparing alternative models; Foll and Gaggiotti, 2008), and iv) assumptions regarding population sizes (i.e. equal or constant; Narum and Hess, 2011). The importance of controlling for false positives by using the most stringent criteria (Egan *et al.*, 2008; Wilding *et al.*, 2001) and correcting for multiple testing (Schlötterer, 2003; Storz and Nachman, 2003) have been underscored in several studies. However, FDRs and Bonferroni corrections were not used here because selection is likely weak in this system and choosing a threshold that would reduce Type I error without inflating Type II error is difficult to determine, especially since genome scans using EST-linked markers are expected to have higher detection rates (13-20%; Namroud *et al.*, 2008; Shikano *et al.*, 2010; Vasemagi *et al.*, 2005) than those using SNPs or other anonymous markers (2–10%; Bonin *et al.*, 2006; Campbell and Bernatchez, 2004; Stinchcombe and Hoekstra, 2008; Wilding *et al.*, 2001) that are found most frequently in non-coding

regions. I feel that the requirement for detection by two or more approaches is adequately conservative given that 18 of 33 markers were eliminated as false positives. Foll and Gaggiotti (2008) found that detection of loci across multiple replicate ecotype-pairs is an adequate approach for reducing Type I and II errors when data on *L. saxatilis* (Wilding *et al.*, 2001) was reanalyzed. Although it is impossible to determine if any true outliers were classified as 'false outliers' (for this reason, they were not retained in the 'truly neutral' dataset) or if false positives were classified as 'true outliers' at this point, the four 'repeat-outliers' are very promising candidates for loci under selection.

Outlier-detection approaches varied in their sensitivity to outlier behaviour. BAYESCAN had very low detection rates (1.2%) and exhibited under-sensitivity to true outliers. While BAYESCAN is generally reported to be the most stringent approach, it is also commonly described as the most robust to complex demographic scenarios, which is somewhat contradictory to our findings. Under-sensitivity to outlier behaviour was greatest in the *ln*RH approach, probably because soft sweeps will not produce differences in gene diversity (H_e) as they will population differentiation (Hohenlohe *et al.*, 2012). The detection of false positives was much more common than false negatives for three of the four outlier-detection approaches. DETSEL, LOSITAN and *ln*RH exhibited over-sensitivity to outlier behaviour, mostly in Kootenay and Wood Lakes. In these lakes, complex demographic histories may be influencing outlier-detection. High detection rates are frequently reported for the FDIST2 approach implemented in LOSITAN. (Narum and Hess, 2011). Tchesinkut Lake kokanee exhibited the lowest neutral differentiation among spawning sites and only two false positives were identified, while all four 'repeat-outliers' were detected by at least one approach in this lake. This supports the notion that despite their propensity for high Type I error rates, a genome scan approach can identify loci under selection when replicate lakes are used to identify false positives.

2.4.3 Testing for a unique signal at outlier loci

Tests for distinct patterns of genetic differentiation and structuring among ecotypes at outliers compared to neutral loci was used to further validate outliers because only natural selection can

generate barriers to gene flow at specific loci (outliers) without affecting the entire genome (neutral loci). Estimates of differentiation at 'repeat-outliers' were significantly higher that those estimated at 'neutral loci' across all lakes and for each ecotype pair (Tables 2.4 and 2.5), which further supports 'repeat-outliers' as loci truly under selection. Although there is a chance that focusing on 'repeat-outliers' may result in overlooking loci of interest (Stinchcombe and Hoekstra, 2008). Even if the same selective forces are favouring the same phenotype in multiple populations, different mutations within the same gene, mutations in different genes, and mutations in genes belonging to different pathways could produce a similar phenotype (Shikano *et al.*, 2010). Variation in loci identified as 'true outliers' across lakes suggests that different genes may underlie ecotype divergence in different lakes. For example, some loci were detected as strong outliers by two or three different algorithms in only a single lake (i.e. *OMM5125* in Okanagan, and *Ots06* in Wood Lake) and some 'true outliers' were monomorphic in other lakes. To investigate the validity of these outliers, I compared patterns of variation at 'lake-specific true outliers', all 'true outliers' and 'repeat-outliers'. The 'lake-specific true outliers' exhibited the highest levels of ecotype differentiation, which suggests that different genes may be involved in ecotype divergence. This trend could also be produced simply by ascertainment bias (Thornton and Jensen, 2007), but the 2.5-fold increase in F_{ST} observed in Okanagan Lake likely has some biological significance. Still, there is considerable overlap between 'repeat-outlier' and 'lake-specific true outliers' (Table 2.2) and the difference in the number of discrete genetic units inferred by the 'lake-specific true outliers' compared to the 'neutral loci' suggests that patterns of genetic variation at outliers are distinct from that of neutral loci and possibly due to the locus-specific effects of selection. Repeat-outliers represent the best candidates for further validation, but in cases of high congruence among detection approaches, some 'lake-specific true outliers' are certainly worthy of further investigation as well.

2.4.4 Traits putatively under selection

Using a genome-scan approach allowed us to identifying loci putatively under selection with no *a priori* knowledge of the phenotypes involved. While additional validation is necessary, preliminary

inferences can be made about the possible traits involved based on the function of the most closely linked genes. Since outliers are closely linked to genes that code for transporters (*TAP2, EV691, EV740*), receptors (*EV358*, *OMM5003*) and enzymes (*EV170, EV685*), selection is most likely acting on physiological traits rather than morphological traits (which are generally associated with transcriptional regulators; Liao *et al.*, 2010). More specifically, the annotations recovered for nine outliers infer that the divergence of stream- and shore-spawning ecotypes may involve differences in immune responses to pathogens and energy metabolism. For example, locus *EV358* is a strong candidate and linked to an immunological gene, CXCR4. This gene has shown differential expression in response to exposure to sea lice in Atlantic salmon (Skugar *et al.*, 2008) and saprolegniasis (a fungal infection (Roberge *et al.*, 2007). CXCR4 has potent chemotactic activity for lymphocyctes and has been shown to inhibit haematopoietic stem cell proliferation (Nie *et al.*, 2008). The *TAP2* transporter protein has been identified as part of the immune system in brown trout and European trout (*Salmo trutta*; Abele and Tampe, 2006; Jensen *et al.*, 2008; Keller *et al.*, 2011). Some viruses can suppress the functioning of this protein, which suggests there that there is potential for strong pathogen-mediated selection. Studies of Alaskan sockeye have identified variation at SNPs in the major histocompatibility complex (MHC) that are linked to immune function and disease resistance and correlated with different spawning sites (Creelman *et al.*, 2011) and habitat types (McGlauflin *et al.*, 2011). Locus *EV170* is associated with an enzyme in the citric acid cycle, and is therefore crucially involved in energy metabolism throughout the body. Selection on this gene may reflect the reduced metabolic demands of shore spawners, since the migration to shore sites is far less demanding than stream sites. This gene may relate to differences in size observed in some ecotype pairs since an over-expression of genes with key roles in the citric acid cycle (including malate dehydrogenase) have been implicated in the divergence of dwarf and normal lake whitefish (St-Cyr *et al.*, 2008). The annotations of the strongest outliers identified here offer highly plausible mechanisms driving ecotype divergence and some of which known to be involved in the divergence of ecotype pairs in other systems. This further supports the notion that genome scans can successfully link reproductive isolation to gene regions under selection and shed new light on the possible mechanisms underlying divergence.

2.5 Summary

Using genome scans to identify genes of adaptive significance has several inherent challenges, yet identifying parallel patterns in outlier behavior across multiple ecotype pairs can provide the most robust evidence for the action of natural selection on a gene (and linked regions). Loci identified here as 'repeat-outliers' represent the best candidates for further investigation and validation. Although, parallel patterns were not detected across all five ecotype pairs, there is still strong evidence for ecotype differentiation at outlier loci in multiple lakes based on AMOVA and STRUCTURE results, potentially reflecting the locus-specific effects of selection. It is possible that differences in selection pressure or standing variation may be obscuring signatures of selection in some lakes or that different genes in the same pathways are involved in the divergence of shore- and stream-spawning kokanee, which is not unprecedented in the literature (Hoekstra, 2006).

These outlier loci are strong candidates for selection and plausible mechanisms can be inferred based on the most closely linked gene, but further validation is necessary. This was only the first step towards uncovering the genetic basis of ecotype divergence in kokanee salmon. Ultimately, this work may identify those environmental factors that drive divergence in salmonids utilizing different spawning habitats and factors that promote and constrain progress towards speciation in natural populations. Such insights could substantially enhance our ability to identify population boundaries, designate management units, predict future evolutionary trajectories of wild stocks, and recover populations of conservation concern (Fraser *et al.*, 2011; Hauser and Seeb, 2008).

CHAPTER 3.0 A GENOME-SCAN APPROACH FOR IDENTIFYING INFORMATIVE MARKERS FOR LANDSCAPE-LEVEL FRESHWATER FISHERIES

3.1 Background

Population genetics provides significant insight into the migrational behaviour, mating systems, demographics, and particularly the population structure of fishes, which would otherwise be very difficult to assess (Hauser and Seeb, 2008). Over the last six decades, genetic approaches have been increasingly used for genetic stock identification (GSI) in commercial fisheries (Shaklee *et al.*, 1999) as well as informing decisions on recovery initiatives (Allendorf *et al.*, 2010; Hauser and Seeb, 2008; Waples and Hendry, 2008). Presently, the delineation of management and conservation units heavily relies on neutral population structure, reflecting historical patterns of reproductive isolation. Substantial effort has been dedicated to the development of increasingly accurate and more powerful markers to define evolutionary significant units on increasingly finer spatial and temporal scales (Narum *et al.*, 2008). However, studies suggest that neutral markers are only effective for resolving population boundaries on relatively large scales (e.g. 100 km in Chinook salmon; Beacham *et al.*, 2006). At smaller scales, niche partitioning and local adaptation promote reproductive isolation among proximate populations (Gomez-Uchida *et al.*, 2011), generating much of the intraspecific diversity (e.g. life history forms) that has arisen since the last glaciation (Fraser and Bernatchez, 2005). Many failed transplant experiments suggests that these populations are uniquely adapted to local conditions (Fraser *et al.*, 2011) and underscores the need to identify, delineate, and preserve locally adapted stocks to maintain ecological complexity and thereby stock productivity and stability in the future (Fraser *et al.*, 2011; Harmon *et al.*, 2009). Conventional neutral genetic markers cannot distinguish recently diverged ecotypes because selection only acts on fitness-related genes and neutral differences have not yet accumulated or are lost due to the homogenizing effects of gene flow. In such cases, adaptive genetic markers, i.e. markers directly linked to polymorphisms in genes under selection, are needed to conduct assessments for recently diverged stocks and inform fishery management decisions.

In the past, crossbreeding experiments and quantitative trait loci (QTL) studies were required to investigate the genetic basis of adaptive traits. However, these approaches can be prohibitively expensive and time consuming, especially for species with longer generation times (Storz, 2005). The use of population-based genomic scans represents a faster and relatively inexpensive alternative for identifying informative loci (i.e. outlier loci) for distinguishing recently evolved ecotypes with no knowledge of the phenotypes under selection. If outliers truly reflect adaptive variation, they could be used for genetic stock identification (GSI)-based individual assignment (IA) and mixed composition (MC) analysis in conspecific populations exhibiting the same patterns of phenotypic divergence, in theory (Andre *et al.*, 2011; Nielsen *et al.*, 2009a). Few have investigated the utility of outliers versus truly neutral markers within a management context (Ackerman *et al.*, 2011; Andre *et al.*, 2011; Freamo *et al.*, 2011; Nielsen *et al.*, 2009b; Russello *et al.*, 2012; VanDeHey *et al.*, 2009), but the approach seems quite promising (Helyar *et al.*, 2011).

Currently, there is a need to develop adaptive markers in kokanee salmon, *Oncorhynchus nerka*. Kokanee are an economically and ecologically important freshwater fish that exhibit two distinct reproductive ecotypes: shore- and stream-spawners (Shephard, 2000; Taylor *et al.*, 1997; Taylor *et al.*, 2000). These ecotypes exhibit distinct reproductive behaviours and spawning habitat preferences (see Table 1.1), but are ecologically and morphologically indistinguishable outside of the spawning period. In the last 30 years, lake-wide kokanee populations have declined in abundance by more than 90% in several BC and northwestern US lakes (Andrusak and Andrusak, 2011; Paragamian and Bowles, 1995; Shephard, 2000). Managers have been urged to treat shore- and stream-spawners as discrete management units based on their geographical and temporal discreteness, modest levels of neutral genetic variation, and possible divergence in some phenotypic traits (Taylor *et al.*, 2000). Since they use distinct spawning habitats, ecotypes are also differentially vulnerable to various anthropogenic activities (e.g. lake draw-down, shoreline development). Obtaining ecotype-specific estimates of absolute abundance will be critical for the future recovery of at-risk populations, however standard

methods do not allow for independent estimates of absolute abundance or harvest rates for each ecotype when they occur in sympatry.

Visual counts of kokanee aggregations during the peak spawning time is the standard approach for obtaining instantaneous abundance estimates for individual stocks. Absolute abundance can be accurately determined for stream-spawners using more labour-intensive counting methods (e.g. a counting fence) or applying the provincial standard expansion factor to visual counts (x1.5), which takes into account the residence time of fish (by movement or death) and the proportion of fish that are generally visible (Ashley *et al.*, 1998). This information is easy to obtain in streams because fish are in a closed environment and they tend to hold their position in the stream to defend their nest. Visual counts of shore-spawner abundance have simply been used as an index for monitoring long-term trends because the absolute abundance cannot be calculated (P. Askey, *pers. comm*.). Expansion factors developed for stream-spawners are not suitable because their residence time is unknown, and visibility is much lower due to wave action, less colouration, and depth of spawning (Ashley *et al.*, 1998). Other methods for estimating absolute abundance (e.g. mark-recapture or counting fences) cannot be used because shorelines are open areas and fish are free to come and go. Therefore, genetic markers capable of accurately distinguishing shore- and stream-spawning kokanee are needed to (i) calculate the absolute abundance of shore-spawners and (ii) identify the source ecotype of fish that are otherwise indistinguishable (e.g. mixed samples). The relative ecotype proportions could be determined from mixed samples obtained from gillnet, trawl survey and then the absolute abundance of stream-spawners can be used to estimate that for shore-spawners. Similarly, the ecotype-specific harvest rates could be calculated from estimates of absolute abundance and the number of angler-caught fish for each ecotype. This could be achieved by assigning each fish sampled in an angler survey to an ecotype using a reference sample. With this information, fishery managers would be better able to evaluate the ecotype-specific impacts of management decisions (e.g. harvest rate) and prioritize conservation efforts.

Here, loci identified as outliers in a genome scan (see Chapter 2) are assessed for their ability to identify the ecotype of individuals with unknown origins and estimate the relative proportions of each ecotype in mixed samples for kokanee from five British Columbian lakes. Specifically, I assessed the bias and accuracy of five datasets consisting of different combinations of outlier and neutral loci in MC and IA tests. Since sampling is often limited by time and resources (Hauser and Seeb, 2008), the minimum sample sizes needed to achieve management-relevant levels of accuracy (>90%) are simulated for each lake. Finally, I discuss the prospect of using these outlier loci in other lakes distributed throughout BC and northwestern USA.

3.2 Methods

3.2.1 Data collection & dataset definition

Baseline information for this study was comprised of genotypic data collected at 42 EST-linked microsatellite loci and 8 anonymous microsatellite loci for 525 individuals across five lakes, including Okanagan, Wood, Kootenay, Duncan, and Tchesinkut Lakes (see Chapter 2 for sampling and genotyping details). Five datasets were defined based on marker type and behaviour in outlier-detection analyses: 'repeat-outliers', 'true outliers', 'lake-specific true outliers', 'anonymous loci', and truly 'neutral loci' (see Chapter 2). The 'repeat outlier' dataset consisted of loci exhibiting outlier behaviour in multiple lakes using multiple algorithms (n_{loci}=4), and had the greatest potential to be informative in lakes throughout the native range of kokanee. The 'true outlier' dataset consisted of all loci detected by two or more algorithms in any one lake (n_{loci}=15). This dataset was used to assess the added benefit of incorporating strong outliers that were only detected in only one lake and evaluate the possibility that multiple genes are involved in ecotype divergence. For a similar purpose, I included the 'lake-specific outlier' dataset. This dataset consisted of five different sets of loci corresponding to those detected by at least two algorithms within each respective lake (Duncan n_{loci}=4, Kootenay n_{loci}=5, Okanagan n_{loci}=3, Wood n_{loci}=6, Tchesinkut n_{loci}=3), and therefore eliminated the influence of uninformative markers in each lake. The 'anonymous loci' dataset (n_{loci}=8) contained highly variable microsatellite loci that are

commonly used in genetic studies of *O. nerka*. Finally, the 'neutral dataset' consisted of any polymorphic locus that showed no outlier behaviour at all (n_{loci}=15). This dataset acted as a baseline for comparison when evaluating the power of putatively adaptive markers in distinguishing recently evolved shore- and stream-spawning ecotypes.

3.2.2 Individual assignment tests

The ability of each dataset to assign individuals to the most likely ecotype of origin was assessed using the method of Rannala and Mountain (1997) implemented in the GSI program, ONCOR (Kalinowski *et al.*, 2007). First, realistic fishery simulations were used to generate multi-locus genotypes (Anderson *et al.*, 2008). Both genotype frequencies and mixture proportions were used to calculate the ecotype with the highest probability of producing the given genotype. The percentage of individuals correctly assigned was calculated for each ecotype in all five lakes. The leave-one-out individual assignment test was also used to assess how well individuals from the baseline were assigned back to their ecotype of origin (Kalinowski *et al.*, 2007).

3.2.3 Mixed composition analyses

Mixed-stock proportions were estimated using the conditional maximum likelihood-based approach implemented in ONCOR (Kalinowski *et al.*, 2007). This algorithm avoids overly optimistic assessments of power by simulating genotypes from the existing baselines to create mixture samples and then estimating their probability of occurrence in the baseline populations (Anderson *et al.*, 2008). Two series of simulation analyses were conducted to evaluate the utility of our baseline datasets in MC analysis. In the first series, ecotype proportions were skewed to test for any systematically bias in the estimation of ecotype contributions in a mixed sample. Using the realistic fishery simulation method in ONCOR, six mixture scenarios were defined for each lake (shore : stream): 90:10, 75:25, 50:50, 25:75, 10: 90 and 0:100. Mixture samples of 200 multi-locus genotypes were generated by drawing fish from the baseline sample, with replacement, to achieve the specified proportions of each ecotype (Anderson and Slatkin, 2007). I ran 1000 simulations and five replicates per scenario with the reporting groups

defined by ecotype. The proportion of shore spawners estimated by ONCOR was subtracted by the actual (pre-defined) proportion of shore-spawners to calculate the residual from each mixture scenario in each lake. A positive residual indicated an overestimate of shore-spawners and a negative residual indicated an under-estimate of shore-spawners relative to the true value. The residuals were used to assess any bias in observed estimates and determine the accuracy of different datasets in MC analyses. The absolute sum of the residuals for each lake were compared to assess the accuracy of the five datasets.

In the second series, the 100% simulation method of Anderson and Slatkin (2007) was used to explore the influence of baseline sample sizes on accuracy of MC estimates using only the 'repeat-outlier' dataset. Baseline samples consisting of 50, 100, 150, 200, and 250 multi-locus genotypes were generated by drawing alleles from reduced variance estimates of allele frequencies in the baseline dataset for each lake (Kalinowski *et al.*, 2007). Samples were generated based on a pre-defined composition of stream- and shore-spawners (25:75) to better reflect realistic proportions and minimize bias in the results since ONCOR does not perform as well when proportions are skewed (see results). Again, five replicates were run for each sample size. Accuracy was plotted against baseline sample size to determine the minimum number of samples needed to attain the desired threshold of 90% accuracy for each lake.

3.3 Results

3.3.1 Individual Assignment Tests

Assignment accuracy varied substantially across lakes and datasets (29.4 % to 95.9%; Table B.2) when using the realistic fishery simulation approach in ONCOR. Stream-spawners were assigned to the correct ecotype more often than shore-spawners by 1-6%, except in Wood Lake (shore was 2% higher; Figure 3.1). Overall, mean IA accuracy across all five lakes was 71.2% using the 'repeat-outliers', 75.6% using all 15 'true outliers', 74.7% using the 'lake-specific true outliers', 69.2% using the

'anonymous loci', and 67.3% using the 'neutral loci'. The three outlier datasets outperformed the anonymous and neutral datasets, but none achieved >80% assignment accuracy, aside from Wood Lake. The IA accuracy based on 'anonymous' and 'neutral' datasets produced similar results overall, although the 'anonymous loci' were 5.4% more accurate in both Duncan and Tchesinkut Lakes. The 'repeat outliers' outperformed better than the 'anonymous' and 'neutral' datasets in three out of five lakes, and was best out of all five datasets in Tchesinkut Lake (Figure 3.1). Using the leave-one-out method in ONCOR, estimates of IA accuracy were very similar, but on average they were 1% higher for all datasets (data not shown).

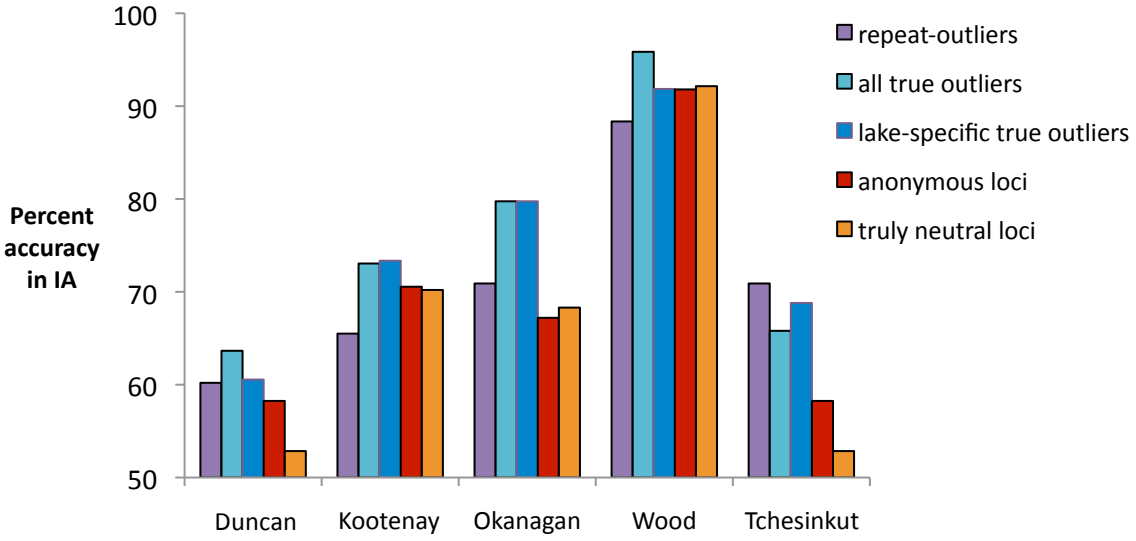

Figure 3.1 The percentage of genotypes accurately assigned to the ecotype of origin as assessed in individual assignment (IA) tests using the realistic fishery simulation approach implemented in ONCOR. Mean accuracy is shown for each lake, using each of five different datasets, including repeat outliers (n_{loci}=4), true outliers (n_{loci}=15), lake-specific true outliers (Duncan n_{loci}=4, Kootenay n_{loci}=5, Okanagan n_{loci}=3, Wood n_{loci}=6, Tchesinkut n_{loci}=3), anonymous microsatellite loci (n_{loci}=8), and truly neutral loci (n_{loci}=15).

3.3.2 Mixed Stock Analyses

In mixed-stock analyses, the estimated ecotype proportions differed from the true proportions by 0.4% to 51.6% across the six mixture scenarios and five lakes (Figure 3.2). In general, all estimates were most accurate when the true proportions of shore- and stream-spawners were equal (50:50). Weakly

contributing stocks were consistently over-estimated when the ecotype proportions were highly skewed. Estimates were also slightly more accurate when the proportions were skewed in favour of stream-spawners, except in Duncan Lake. In the 0:100 mixture scenario (no shore-spawners), zero was included in the confidence interval (CI) when using all five datasets in Wood Lake, only 'lake-specific' and 'repeat-outlier' datasets in Tchesinkut, and only the 'lake-specific outliers' in Kootenay and Okanagan Lakes.

Differences in the accuracy of outliers compared to neutral datasets in MC analyses were most evident in Okanagan, Tchesinkut Lakes, and Kootenay Lakes. All five datasets produced similar results in MC analyses conducted for Duncan and Wood Lakes. All estimates were within 3% of the true proportion in Wood Lake. The 15 'true outliers' performed the best, only deviating by a mean of 0.5% across the six scenarios. In Duncan Lake, the estimated proportion of shore spawners differed from the true proportions by 20.9% on average and never deviated from 50:50 by more than 15.3% (range of 30.6%). The 'neutral dataset' deviated from 50:50 the least across all scenarios (range of 7.8%). The 'true outliers' deviated the most (range of 43.1%), but the 95% CIs did not include the true proportions for five of the six scenarios in Duncan Lake.

Using 'repeat outlier' dataset, mixed-stock estimates deviated from true proportions by 9.2% overall (Table B.1). Deviations were less than 5% in 19 out of 30 scenarios across all five lakes. The 'lake-specific outliers' and all 'true outlier' datasets performed slightly better than the 'repeat-outliers' with mean deviations of 5.7% and 8.9% respectively. All three outlier datasets performed significantly better than the 'anonymous' and 'neutral' datasets with mean deviations of 13.9% and 16.6%, respectively. In Kootenay Lake, the 'lake-specific outliers' and 'true outliers' performed the best, deviating from true proportions by a mean of 4.1% and 9.5%, respectively. Kootenay was the only lake where estimates based on 'repeat-outliers' (14.8%) deviated more than that of the 'anonymous' (12.8%) and 'neutral' datasets (13.3%). In Tchesinkut and Okanagan Lake, estimates based on 'neutral' and 'anonymous loci' datasets showed little deviation from 50:50, while the three outlier datasets were within 4.9% and 6.5%, of the true proportions across all six scenarios in Okanagan and

Tchesinkut, respectively. The 'lake-specific outliers' performed the best out of the three outlier datasets, with a mean deviation of 2.9% and 2.4%, respectively.

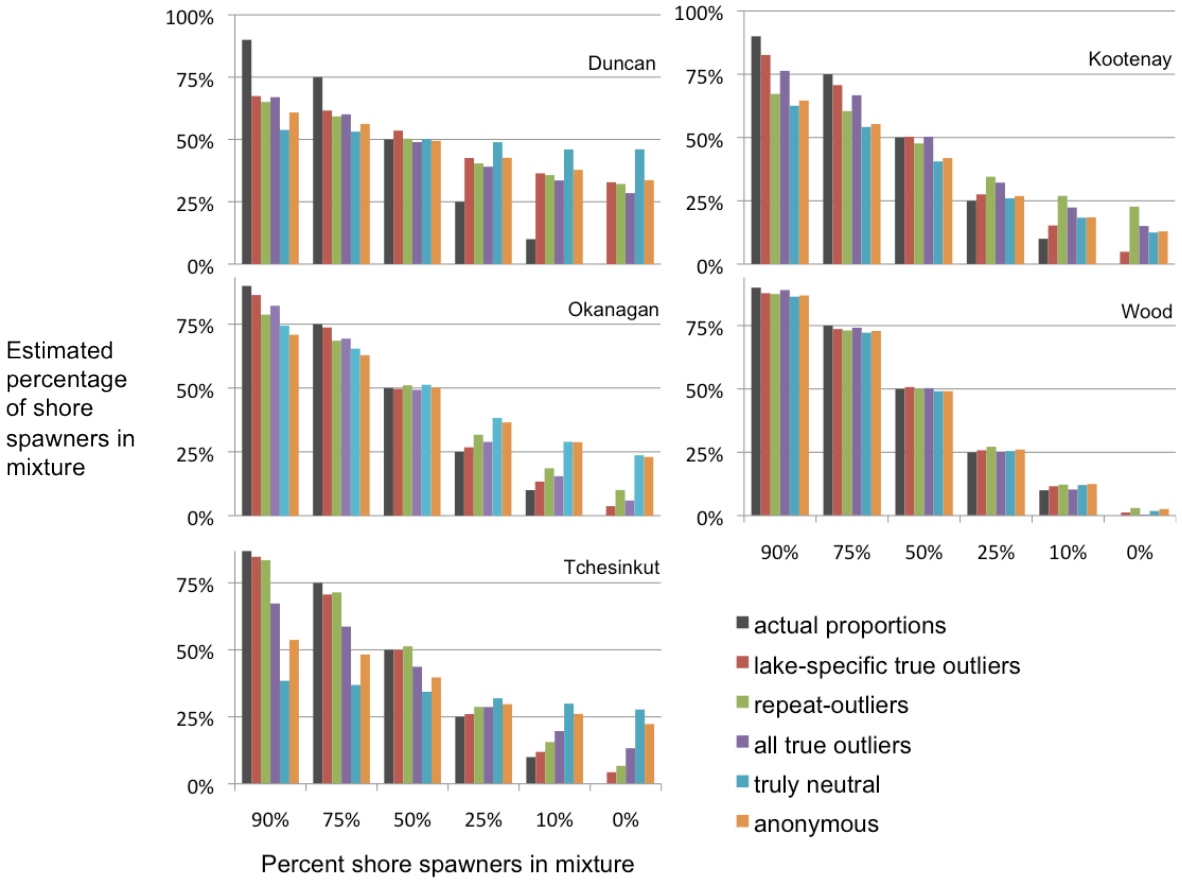

Figure 3.2 The percentage of shore-spawners estimated by ONCOR for six mixture scenarios with pre-defined mixtures of shore-spawners (blue bar). Estimates for all scenarios were calculated using five different datasets, including 'lake-specific outliers' (Duncan n_{loci}=4, Kootenay n_{loci}=3, Okanagan n_{loci}=3, Wood n_{loci}=6, Tchesinkut n_{loci}=3), 'repeat outliers' only (n_{loci}=4), all 'true outliers' (n_{loci}=15), truly 'neutral loci' (n_{loci}=15), and the 'anonymous microsatellite loci' (n_{loci}=8).

3.3.3 100% Simulations using 'repeat-outliers'

Using the four 'repeat-outliers', the proportion of simulated genotypes correctly assigned to the ecotype of origin increased when the baseline sample size was increased from 25 to 250 individuals (Figure 3.3). The number of samples needed to obtain 90% accuracy was <25 individuals in Wood Lake, ~100 individuals in Tchesinkut Lakes, and ~230 individuals in Kootenay and Okanagan Lakes.

Sufficient accuracy in MC analyses (≥90%) cannot be achieved in Duncan Lake regardless of sample size (within reasonable limits).

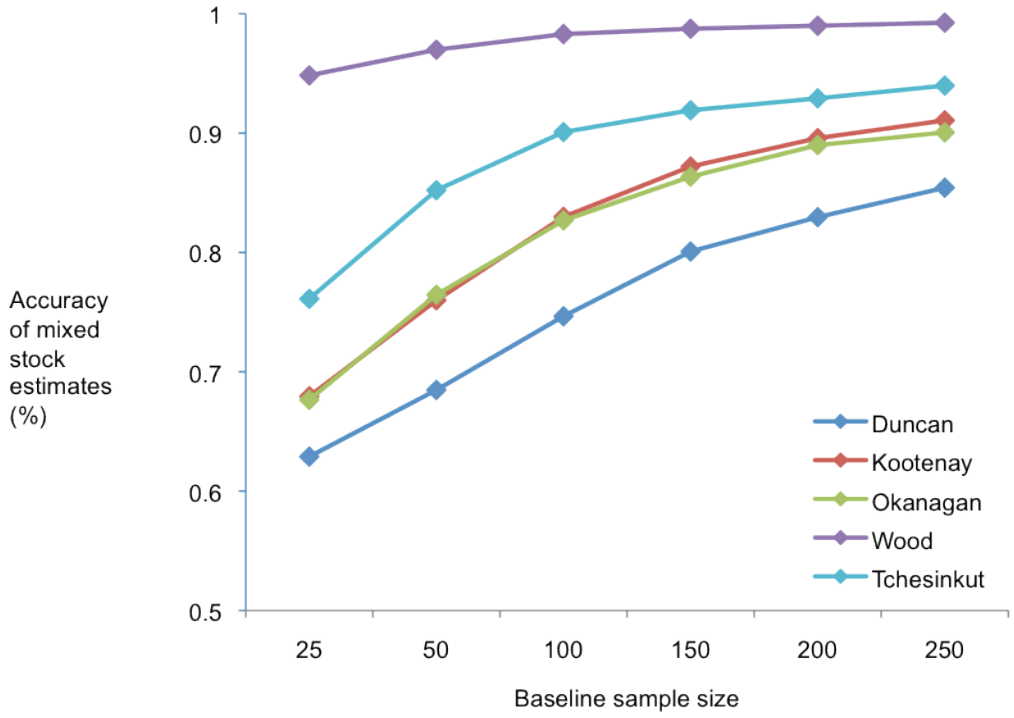

Figure 3.3 The effect of increasing the baseline sample size for the 'repeat-outlier' dataset on the accuracy of mixed stock estimates in 100% simulations implemented in ONCOR. The mixture proportions were pre-defined as 75% shore- and 25% stream-spawners for all simulations.

3.4 Discussion

Substantial intra-specific diversity has arisen in salmonids since the last glaciation, presenting a challenge to fishery managers. A genetics-based approach for conducting stock assessments based on contemporary adaptations to local environments and future evolutionary potential, rather than historical patterns of gene flow could be implemented using markers linked to adaptive genes. Previous work suggested that eight outliers detected in a genome-scan may be the best class of markers for GSI in Okanagan Lake, where shore- and stream-spawning ecotypes of kokanee co-exist (Russello *et al.*, 2012). After applying this approach to multiple lakes, the observed level of accuracy and bias in IA and MC analysis suggests that this is a promising approach for identifying informative

genetic markers and obtaining accurate estimates of shore-spawner abundance, however it may not be effective in all lakes.

3.4.1. Power and accuracy of outlier loci in IA and MC analyses

Given the range in ecotype differentiation observed across our five lakes (see Chapter 2), I was able to empirically assess the power of outlier loci to distinguish ecotypes in a range of natural scenarios within a management context. Ecotypes were very weakly differentiated in Duncan Lake kokanee, moderately differentiated at outlier loci but not neutral loci in Okanagan and Tchesinkut Lake kokanee, weakly differentiated throughout the genome in Kootenay Lake kokanee, and strongly differentiated throughout the genome in Wood Lake kokanee (Table A.4). Overall, IA tests demonstrated that reproductive ecotypes were more genetically discrete at outlier loci compared to neutral loci, though IA success for four out of five lakes was still relatively low (<80%; Figure 3.1). Consistent with estimates of F_{ST}, the contrast in the performance of outlier and neutral loci was most evident in Okanagan and Tchesinkut Lake kokanee. In both cases, the outlier datasets produced the most accurate estimates of true ecotype proportions in MC analyses. Simulations suggest that increasing the baseline for Okanagan and Kootenay Lakes to ~230 individuals could boost the accuracy of MC analysis estimates to a management-relevant level (≥90%; Figure 3.3). Other studies of Atlantic salmon, trout and herring have achieved similar levels of accuracy in MC analyses using markers that exhibit similarly low levels of differentiation, F_{ST} =0.008 (Bekkevold *et al.*, 2011; Hansen *et al.*, 2000; Koljonen *et al.*, 2005). However, in more recently isolated lakes (45 years), where ecotypes exhibit very little differentiation at outlier and neutral loci (e.g. F_{ST} ≤0.001 in Duncan Lake), signatures of divergence appear too weak to achieve adequate levels of accuracy for MC and IA, regardless of the number of loci or number of individuals used (Figure 3.3).

The use of outlier loci identified in genome scans renders more accurate estimates of ecotype proportions from a mixed sample than highly variable neutral loci. Outlier loci exhibited superior performance in IA and MC with fewer loci and alleles, compared to neutral datasets. Generally,

increasing the number of alleles increases the power to resolve population differences in allele frequencies (Beacham *et al.*, 2006). Yet, the outlier datasets revealed greater ecotype differentiation with fewer alleles (ranging between 137 and 18 alleles across 15 'true outliers' and 3 'lake-specific outliers' in Tchesinkut Lake) than anonymous (146 alleles across 8 markers) and neutral datasets (187 alleles across 15 markers). It is difficult to determine which of the three outlier datasets will be the most accurate and reliable markers for distinguishing ecotypes in lakes beyond those included in this study. Generally, the 'true outliers' and the 'lake-specific outliers' yielded slightly more accurate estimates of mixed-stock sample than 'repeat-outliers', but it remains unknown whether the superior IA success was the result of retaining additional, informative markers for different lakes or simply high-grading bias (Anderson, 2010).

High grading bias can be generated when the samples used to select potentially informative markers (Chapter 2) are also used to evaluate their performance (Anderson, 2010). Although 'true outliers' and 'lake-specific true outliers' were able to distinguish ecotypes with high accuracy here, we cannot determine that differentiation observed at these markers is entirely due to true population differences in allele frequencies and not stochastic sampling effects to some extent (Waples, 2010). Even in the absence of selection, loci throughout the genome are expected to exhibit a range in the level of differentiation, which can become inflated by sampling error, especially when sample sizes are small. Therefore, 'lake-specific true outliers' and 'true outlier' datasets may be producing upwardly biased estimates, although this is not a concern for the 'repeat-outliers' because they were selected based on strong outlier behaviour observed in multiple independent lakes. Although using a low number of polymorphic loci (n=57) and reasonable sample size (n=27-96) reduces the potential for high-grading bias, additional independent samples need to be analyzed to cross-validate the power of these markers in the future (e.g. using the Simple Training and Holdout method; Anderson, 2010).

In Okanagan Lake, estimates of individual assignment accuracy using outlier loci were much higher in a previous study by Russello *et al* (2012) than is reported here. Self-assignment accuracy estimated from the leave-one-out analysis implemented in ONCOR was 92.0%. The decreased accuracy reported

here for Okanagan Lake may be explained by the inclusion of false outliers. They used all eight loci that showed any evidence of outlier behaviour in any one of three detection approaches. In this study, only 79.8% of individuals were correctly assigned using the best performing outlier dataset, the 'lake-specific true outliers'. I used nearly the same sample (i.e. nine individuals were retained here compared to the previous study), and the same three loci that were detected by multiple methods in both studies, but I did not include the five outliers only identified only by DETSEL in the Russello *et al* (2012) study. Therefore, the discrepancies in estimates of IA accuracy are either a consequence of the inclusion of false outliers by the previous study or exclusion of weak outlier loci in this study. In either case, these results underscore the importance of cross-validation when evaluating the utility of outlier loci for GSI.

3.4.2 Bias in estimates of ecotype proportions

Simulations using different ecotype ratios in MC analyses demonstrated that there was no bias in estimated proportions in favor of one ecotype or the other. Therefore, abundance estimates from visual counts of stream-spawners could be combined with MC estimates using mixed samples from a trawl, gillnet or creel survey to determine shore-spawner abundance. This approach could be used each year or, if mixed samples are unavailable, an expansion coefficient could be calculated for shore-spawning kokanee. Estimates of shore-spawner abundance based on visual counts would then be more accurate and allow for early detection of population decline. However, ONCOR demonstrated a strong bias in estimates when the proportions were highly skewed (Figure B.1) such that estimated proportions tended to be biased towards 50:50 or $1/k$ (where k is the number of stocks; Bekkevold *et al.*, 2011; Kalinowski *et al.*, 2007). This trend was most pronounced in lakes showing the lowest levels of differentiation at outlier loci (F_{ST} <0.007), e.g. Duncan and Kootenay Lake kokanee. Therefore, if ecotypes are poorly differentiated and their ratio in the lake is highly skewed, MC analyses may not alert managers to nearly depleted levels of abundance. In mixed stock fisheries, all stocks are subject to the same harvesting process, but the weaker stock is typically the one that gets overharvested (P. Askey, *pers. comm*.). These results suggest that biases will cause managers to overestimate the

contribution of the weak stock to the catch, and potentially overestimate its abundance. Therefore, a bias correction factor may need to be estimated for each lake to obtain more accurate abundance estimates in these scenarios. By plotting the relationship between the residuals and the proportions in the simulated mixtures and inversing the equation that defines the line of best fit, this bias could be corrected for these lakes. If applying this method to a new lake, the F_{ST} of the new ecotype pair should be calculated and compared to those described here to determine the most appropriate correction. Using this correction factor, more accurate abundance estimates may be achieved when ecotype proportions are highly skewed (however, this is not recommended when differentiation is very weak).

3.5 Summary

In conclusion, outlier loci detected by genome scans are a promising tool for GSI. In most post-glacial lakes, a management-relevant level of accuracy (>90%) in MC analysis can be achieved with adequate baseline sampling (e.g. ~200 individuals). However, this genetics-based approach may not be reliable when applied to lakes where ecotype proportions are highly skewed and/or ecotypes are poorly differentiated at outlier loci (e.g. Duncan Lake). Still, an expansion factor can be calculated for shore spawners in suitable lakes, which will allow for the calculation of absolute abundance in shore-spawners using visual counts in the future. It appears that these markers may be effective in other lakes where sympatric ecotypes exist, although they will not be useful for detecting shore-spawners in lakes where their existence is currently unknown if those stocks are poorly differentiated. Once mixed samples are obtained from these lakes, an expansion factor can be calculated to allow fishery managers to better track fluctuations in true abundance and achieve a better understanding of the health of the whole-lake population in each lake. This will be important for guiding management decisions pertaining to habitat quality and exploitation as managers strive to recover depleted kokanee populations in British Columbia.

CHAPTER 4.0 GENERAL CONCLUSIONS

4.1 Research findings and significance

Few studies have investigated the genetic basis of divergent ecological forms using EST-linked markers on a comparable scale to this work. Fewer still have used replicate divergent pairs to identify genes potentially under selection. Results presented here provide strong empirical evidence for the action of natural selection. Greater differentiation at outlier loci relative to neutral and anonymous loci suggests that there are locus-specific barriers to reproduction, likely due to fitness consequences associated with spawning in a habitat distinct from that to which it is locally adapted. Expressed genes linked to outlier loci suggest possible mechanisms of divergence. Patterns of outlier behaviour also suggest that shore-spawning behaviour may have arisen via multiple genetic pathways throughout the native range of kokanee. Finally, I demonstrated the utility of population-based genome scans in identifying informative loci for distinguishing recently diverged ecotypes with considerable accuracy.

While the use of genome scans has been widely criticized for their inability to link anonymous outlier loci with the functional basis of adaptations (e.g. genes and genetic pathways responsible for adaptations), the use of EST-linked microsatellites allowed me to identify genes linked to ecotype divergence with no *a priori* knowledge of phenotypes involved in Chapter 2. By restricting our search to the functional portion of the genome and using multiple ecotype pairs, I achieved high outlier-detection rates and produced strong evidence for selection acting on kokanee ecotypes (Bonin, 2008; Namroud *et al.*, 2008). Based on the annotations of nine strong candidate loci, including four that showed consistent outlier behaviour across multiple lakes, I was able to make preliminary inferences about possible ecological mechanisms driving ecotype divergence. Differences in pathogen resistance and energy metabolism offer plausible advantages in alternative spawning environments. Two genes (*CXCR4 and TAP2*) have already been linked to immunological responses to infection in other fish species. With a better understanding of those environmental characteristics to which populations are locally adapted and that reinforce reproductive isolation, we will be able to better predict evolutionary

responses to human disturbances (e.g. climate change, biological invasions, artificial propagation, habitat alteration, and harvesting; Waples and Hendry, 2008), which will significantly contribute to the success of management and recovery programs (Tucker *et al.*, 2009).

The lack of parallel patterns detected across all ecotype pairs may add to the increasing number of studies that suggest parallelism in natural populations is not as common as previously thought (Arendt and Reznick, 2008; Elmer and Meyer, 2011; Schluter *et al.*, 2004). Despite expectations based on the relatedness of kokanee populations and the speed at which shore-spawning has evolved (Taylor *et al.*, 1997), our findings suggest that ecotypes may have arisen via distinct evolutionary pathways in some traits critical to their fitness in the shore-spawning environments. However, the lack of congruency may also be a limitation of the outlier-detection approaches employed (i.e. high Type II error). The detection of outlier behaviour by a single method in one lake for loci classified as a repeat-outlier based on other lakes suggests that this may be an issue. Also, the power of current approaches to reliably detect true outliers is known to be low when selection is weak. Like many before me, I caution researchers to thoroughly validate detected outliers and recommend the use of multiple divergent pairs whenever possible (Bonin, 2008; Oetjen and Reusch, 2007; Shikano *et al.*, 2010).

If convergence in recently evolved phenotypes does not have a common genetic basis, we may be underestimating the amount of biological diversity within systems consisting of multiple ecotype pairs (Taylor, 1999). This may have significant implications for designating management and conservation units. Presently, distinct ecological forms of salmon are taxonomically unrecognized because they are generally considered to be easily replaceable (i.e. populations are often founded by the same species and presumably by a common genetic mechanism; Colosimo *et al.*, 2004; Shapiro *et al.*, 2004). Shore-spawners have arisen within a few generations in some lakes where they have been stocked, yet there are many lakes where one ecotype exists without the other. Clearly, there are constraints to the conversion of stream- to shore-spawning behaviour and visa versa, but we have yet to determine what these constraints are. Kokanee ecotypes do not represent distinct biological species (Mayr, 1942), but evidence for adaptive differentiation presented here, as well as previous findings of weak neutral

divergence and ecological uniqueness (Taylor *et al.*, 2000), warrants the independent monitoring and maintenance of both ecotypes if managers wish to preserve the productivity and future stability of kokanee populations (Fraser and Bernatchez, 2001; Fraser *et al.*, 2011). The divergence of shore- and stream-spawning kokanee has facilitated the use of a broader spectrum of habitats within each lake, which could act to buffer against whole-lake population declines if some spawning habitats are more susceptible to degradation or loss compared to others (Andrusak *et al.*, 2000; Taylor *et al.*, 2000).

One of the overarching goals of this work was to identify adaptive markers to facilitate the management of sympatric shore- and stream-spawning populations throughout their native range in BC, and potentially the Pacific Rim. In Chapter 3, I selected markers based on patterns of outlier behaviour and evaluated their accuracy in GSI applications (e.g. individual assignment and missed stock composition tests). This is one of the first studies to demonstrate that outlier loci are more powerful for distinguishing recently diverged ecotypes than the highly variable microsatellite markers that are commonly used in BC and northwestern USA for *O. nerka* GSI. While neutral microsatellite markers have considerable accuracy and precision in GSI at large spatial scales, they appear to be much less effective in distinguishing ecological forms on small spatial and temporal scales. With further validation, these loci may be a promising tool for conducting stock assessments with greater accuracy than previously achieved using traditional fisheries methods. These results should encourage other investigators to explore the utility of genome scans for identifying putatively informative markers for GSI in other non-model organisms of conservation concern (Schwartz *et al.*, 2007). Genetic resources are rapidly accumulating due to advancements in genomic technologies and ESTs are easily transferrable between closely related species (Bonin, 2008; Hauser and Seeb, 2008). This approach may be particularly useful for identifying informative markers in other recently diverged life history forms of salmon (e.g. run timing, age at maturity) given the importance of delineating evolutionarily significant units in this group.

4.2 Limitations of this study

Some aspects of the study design and evaluation of performance limit our ability to exclude the role of Type I and Type II errors associated with outlier locus detection in influencing our interpretations of the results. This study was not a comprehensive screening of all genes potentially under selection. Coverage was limited by the breadth of the EST libraries and, although 11,390 EST-linked microsatellite loci were tested, the initial screening process conducted in Okanagan Lake may have eliminated truly adaptive loci by chance or due to weak selection in Okanagan Lake. Coverage within the functional genome is also contingent on the size of the linkage groups (Bonin, 2008). The extent of LD can vary between markers and study systems depending upon population history, mating system, recombination rate, and strength of selection (Stinchcombe and Hoekstra, 2008). Next-generation sequencing of the transcriptome for SNP discovery would provide more complete coverage of the functional genome, and is currently underway. This approach is becoming increasingly common as sequencing costs decline and more analytical tools become available for managing massive amounts of data, but screening for polymorphisms consistent with phenotypic variation among ecotypes is still a challenge. I may have only detected a few of the genomic regions that are truly under selection. For example, spawning timing is almost certainly under divergent selection, in at least some lakes. A previous study found that eggs accumulate the same number of thermal units or emerge at the same time despite differences in spawning timing in Okanagan Lake kokanee (Taylor *et al.*, 2000). This suggests that shore-spawners have evolved to spawn later in the year to compensate for differences in over-winter thermal regimes. This ensures that offspring emerge during the spring blooms, which maximizes their chances of survival (Quinn, 2005). Finally, the number of true outliers detected here and prevalence of parallel patterns may be underestimated here because outlier-detection algorithms are susceptible to Type II error. Although, the use of multiple methods, a 90% CI threshold, and no multiple comparison correction should minimize this, other genes or pathways may also be driving ecotype divergence in different lakes.

On the other hand, I cannot conclude that loci identified as 'true outliers' are truly under selection without further investigation. Outlier behaviour in multiple lakes provides strong correlative evidence, but the specific mutations need to be identified to demonstrate a link between genotype and ecologically driven reproductive isolation to infer a mechanism (Luikart *et al.*, 2003). When using EST-linked markers, it is possible that selection that is acting on a distantly linked gene. Therefore, drawing conclusions about the mechanism of selection based on gene annotations could be inaccurate.

The most significant limitation of our assessment of outliers for GSI was the lack of cross-validation using independent samples. Presently, it is unclear whether the superior performance of 'lake-specific true outliers' is a result of true population differences or stochastic sampling effects. The 'repeat-outliers' are robust to sampling effects, but estimates of IA and MC accuracy using the 'lake-specific true outlier' datasets may be upwardly biased. The utility of these markers for province-wide management of kokanee ecotypes is contingent on the hypothesized parallel evolution (i.e. same gene involved) of shore-spawning behaviour. While the performance of 'repeat-outliers' alone suggest a common underlying genetic mechanisms may be involved, the importance of unique genetic mechanisms is unknown. This makes it is difficult to predict how successful these markers will be in other lakes.

4.3 Future work

To address the limitations identified above, future work should be directed towards i) identifying signatures of selection in genes known to be linked to outlier loci at the nucleotide level, ii) then linking genotype and fitness with the phenotypic trait thought to be under selection, and iii) cross-validating estimates of accuracy for outlier loci in GSI. To verify that outliers are truly linked to genes under divergent selection, there is a need to search for SNPs that reflect patterns of divergence observed at linked microsatellite loci. Individuals from all five ecotype pairs will need to be Sanger sequenced at the annotated genes for all 15 true outlier loci. If patterns of differentiation are consistent, it can be confirmed that there is a link between genotypic variation at the candidate gene and reproductive

isolation, which implies the action of selection and local adaptation. Identifying genes of interest for outliers without annotations (e.g. OMM- markers; Scribner *et al.*, 1996) will require a genetic linkage map. Presently, the whole genome is not available for any Pacific salmon species (*Oncorhynchus spp.*), but linkage maps developed for male and female Atlantic salmon (*Salmo salar*) provides a description of SNPs and microsatellites throughout the genome (Moen *et al.*, 2008). Since Atlantic and Pacific salmon diverged over two million years ago (Quinn, 2005), there is a lower chance that proximate genes or linkage group could be determined for the non-annotated loci (Ng *et al.*, 2005; Wenne *et al.*, 2007).

To validate the role of natural selection, a causal link needs to be demonstrated between genotype, phenotype and isolation. Further work will be required to confirm the genes and phenotypic traits under selection. As mentioned, regions of LD can vary widely across species and across the genome (Via and West, 2008). It is difficult to determine if selection is acting directly on the most proximate gene (i.e. the annotated gene) rather than a more distant, unknown gene. The most effective way to rule out pleiotropic effects and validate the adaptive importance of a particular gene is to test for a direct relationship in an experimental study of selection (Barrett and Hoekstra, 2011). Fitness differences between genotypes and phenotypes can be measured in response to treatments that impose a selective force within a controlled environment. For example, survival and changes in allele frequencies at the *TAP2* gene can be measured when experimental populations are exposed to a pathogen commonly found in streams. If a causal link between genotype, phenotype, and fitness in alternative environments can be demonstrated, the adaptive process driving differentiation of shore- and stream-spawning kokanee can be directly inferred.

To implement a genetics-based approach for ecotype-specific monitoring, it needs to be established that linked microsatellite markers can achieve management-relevant levels of accuracy (>90%) in GSI where the two ecotypes co-exist. Baseline samples need to be expanded to verify that simulated estimates of accuracy in Chapter 3 can be achieved. Also, independent mixture samples must be obtained from each lake (and additional lakes) to ensure that estimates of accuracy for MC analyses are

not subject to high-grading bias (Anderson, 2010). If our findings reflect simulated results in Chapter 3, I recommend that these outliers be implemented for estimating the absolute abundance of shore-spawners (with bias correction, if needed). However, if mixed samples are collected from a lake where ecotypes are well differentiated, an expansion factor could be calculated for shore-spawners and applied to lakes province-wide.

As genetic resources grow and progress is made towards identifying the genetic basis of various adaptive traits, managers can pursue the integration of adaptive markers for GSI of commercially exploited species to achieve better resolution of population boundaries at finer spatial scales (Hauser and Seeb, 2008), because this is an aspect of biological diversity that is currently being overlooked. These outlier loci should be tested in shore-spawning populations of other salmonids and I encourage researchers to use a similar framework to uncover the genetic basis of other important life history traits. Achieving a better understanding of ecological attributes that confer a unique fitness advantage in local environments in natural populations is a longstanding goal of conservation biology and new insights generated by the use of population genomic approaches will certainly contribute to more effective management and recovery programs in the future.

REFERENCES

Abele, R, and Tampe, R (2006) Modulation of the antigen transport machinery TAP by friends and enemies. *FEBS Letters* **580**, 1156-1163.

Ackerman, MW, Habicht, C, and Seeb, LW (2011) Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. *Transactions of the American Fisheries Society* **140**, 865-881.

Akey, JM, Eberle, MA, Rieder, MJ, Carlson, CS, Shriver, MD, Nickerson, DA, and Kruglyak, L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology* **2**, 1591-1599.

Allendorf, FW, Hohenlohe, PA, and Luikart, G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709.

Allendorf, FW, and Thorgaard, GH (1984) Tetraploidy and evolution of salmonid fishes. In: *Evolutionary genetics of fishes* (ed. Turner BJ), pp. 1-53.

Anders, P, Faler, J, Powell, M, Andrusak, H, and Holderman, C (2007) Initial Microsatellite Analysis of Kokanee (*Oncorhynchus nerka*) Population Structure in the Kootenai/y River Basin, Idaho, Montana, and British Columbia, Freshwater Fisheries Society of British Columbia and the Kootenai Tribe of Idaho.

Anderson, EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources* **10**, 701-710.

Anderson, EC, and Slatkin, M (2007) Estimation of the number of individuals founding colonized populations. *Evolution* **61**, 972-983.

Anderson, EC, Waples, RS, and Kalinowski, ST (2008) An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences* **65**, 1475-1486.

Andre, C, Larsson, LC, Laikre, L, Bekkevold, D, Brigham, J, Carvalho, GR, Dahlgren, TG, Hutchinson, WF, Mariani, S, Mudde, K, Ruzzante, DE, and Ryman, N (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): Direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity* **106**, 270-280.

Andrusak, G, and Andrusak, H (2011) Observations and analysis of shore spawning kokanee (*Oncorhynchus nerka*) in the West Arm of Kootenay Lake BC Hydro, Fortis BC and Columbia Power Corporation, Redfish Consulting Ltd. Nelson, BC.

Andrusak, G, and Parkinson, EA (1984) Food habits of Gerrard stock rainbow trout in Kootenay Lake, British Columbia. In: *British Columbia Fisheries Technical Circular No. 60*, B.C. Ministry of Environment, Fish and Wildlife Branch.

Andrusak, H, and Jantz, B (2002) Development of Okanagan Lake shore spawning kokanee for the 2001 brood year. Okanagan Nation Fisheries Commission, British Columbia.

Andrusak, H, Sebastian, D, McGregor, I, Matthews, S, Smith, D, Ashley, K, Pollard, S, Scholten, G, Stockner, J, Ward, P, Kirk, R, Lasenby, D, Webster, J, Whall, J, Wilson, G, and Yassien, H (2000) Okanagan Lake Action Plan Year 4 (1999) Report. Fisheries Management Branch, Ministry of Agriculture, Food and Fisheries, Province of British Columbia.

Antao, T, Lopes, A, Lopes, RJ, Beja-Pereira, A, and Luikart, G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F(st)-outlier method. *BMC Bioinformatics* **9**, 323

Arendt, J, and Reznick, D (2008) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution* **23**, 26-32.

Ashley, K, Shepherd, B, Sebastian, D, Thompson, L, Vidmanic, L, Ward, P, Yassien, HA, McEachern, L, MNordin, R, Lasenby, D, Quirt, J, Whall, JD, Dill, P, Taylor, EB, Pollard, S, Wong, C, den Dulk, J, and Scholten, G (1998) Okanagan Lake Action Plan Year 1 (1996-1997) and Year 2 (1997-1998) Report. In: *Fisheries Project Report No. RD 73*. Province of British Columbia, Ministry of Fisheries, Fisheries Management Branch.

Avise, JC, Arnold, J, Ball, RM, Bermingham, E, Lamb, T, Neigel, JE, Reeb, CA, and Saunders, NC (1987) Intraspecific Phylogeography - The mitochondrial DNA bridge between populations genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Avise, JC, Walker, D, and Johns, GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London Series B* **265**, 1707-1712.

Barrett, RDH, and Hoekstra, HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* **12**, 767-780.

Barrett, RDH, Paccard, A, Healy, TM, Bergek, S, Schulte, PM, Schluter, D, and Rogers, SM (2011) Rapid evolution of cold tolerance in stickleback. *Proceedings of the Royal Society B-Biological Sciences* **278**, 233-238.

Barton, NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London Series B* **355**, 1553-1562.

Beacham, TD, Candy, JR, Jonsen, KL, Supernault, J, Wetklo, M, Deng, L, Miller, KM, Withler, RE, and Varnavskaya, N (2006) Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim by use of microsatellite variation. *Transactions of the American Fisheries Society* **135**, 861-888.

Beacham, TD, Wetklo, M, Wallace, C, Olsen, JB, Flannery, BG, Wenburg, JK, Templin, WD, Antonovich, A, and Seeb, LW (2008) The application of Microsatellites for stock identification of Yukon River Chinook salmon. *North American Journal of Fisheries Management* **28**, 283-295.

Beaumont, MA, and Nichols, RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B* **263**, 1619-1626.

Bekkevold, D, Clausen, LAW, Mariani, S, Andre, C, Hatfield, EMC, Torstensen, E, Ryman, N, Carvalho, GR, and Ruzzante, DE (2011) Genetic mixed-stock analysis of Atlantic herring populations in a mixed feeding area. *Marine Ecology Progress Series* **442**, 187-199.

Benjamini, Y, and Hochberg, Y (1995) Controlling the false discovery rate - A practical and powerful approach to multiple testing *Journal of the Royal Statistical Society Series B* **57**, 289-300.

Bernatchez, L, Renaut, S, Whiteley, AR, Derome, N, Jeukens, J, Landry, L, Lu, G, Nolte, AW, Ostbye, K, Rogers, SM, and St-Cyr, J (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B* **365**, 1783-1800.

Bernatchez, L, and Wilson, CC (1998) Comparative phylogeography of nearctic and palearctic fishes. *Molecular Ecology* **7**, 431-452.

Berner, D, Roesti, M, Hendry, AP, and Salzburger, W (2010) Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. *Molecular Ecology* **19**, 4963-4978.

Bernier, JC, Birkeland, SR, Cipriano, MJ, McArthur, AG, and Banks, MA (2008) Differential gene expression between fall- and spring-run Chinook salmon assessed by long serial analysis of gene expression. *Transactions of the American Fisheries Society* **137**, 1378-1388.

Black, WC, Baer, CF, Antolin, MF, and DuTeau, NM (2001) Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology* **46**, 441-469.

Bonin, A (2008) Population genomics: A new generation of genome scans to bridge the gap with functional genomics. *Molecular Ecology* **17**, 3583-3584.

Bonin, A, Ehrich, D, and Manel, S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology* **16**, 3737-3758.

Bonin, A, Taberlet, P, Miaud, C, and Pompanon, F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution* **23**, 773-783.

Bouck, A, and Vision, T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**, 907-924.

Brownstein, MJ, Carpten, JD, and Smith, JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques* **20**, 1004.

Campbell, D, and Bernatchez, L (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution* **21**, 945-956.

Chipps, SR, and Bennett, DH (2000) Zooplanktivory and nutrient regeneration by invertebrate (*Mysis relicta*) and vertebrate (*Oncorhynchus nerka*) planktivores: Implications for trophic interactions in oligotrophic lakes. *Transactions of the American Fisheries Society* **129**, 569-583.

Clarke, LR, Letizia, PS, and Bennett, DH (2004) Autumn-to-spring energetic and diet changes among kokanee from North Idaho Lakes with and without *Mysis relicta*. *North American Journal of Fisheries Management* **24**, 597-608.

Colosimo, PF, Peichel, CL, Nereng, K, Blackman, BK, Shapiro, MD, Schluter, D, and Kingsley, DM (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology* **2**, 635-641.

Cornuet, JM, and Luikart, G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.

Cossins, AR, and Crawford, DL (2005) Opinion - Fish as models for environmental genomics. *Nature Reviews Genetics* **6**, 324-333.

Coyne, JA, and Orr, HA (1998) The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London Series B* **353**, 287-305.

Coyne, JA, and Orr, HA (2004) *Speciation* Sinauer Associates, Sunderland, MA.

Craig, JK, and Foote, CJ (2001) Countergradient variation and secondary sexual color: Phenotypic convergence promotes genetic divergence in carotenoid use between sympatric anadromous and nonanadromous morphs of sockeye salmon (*Oncorhynchus nerka*). *Evolution* **55**, 380-391.

Crawford, SS, and Muir, AM (2008) Global introductions of salmon and trout in the genus *Oncorhynchus*: 1870-2007. *Reviews in Fish Biology and Fisheries* **18**, 313-344.

Creelman, EK, Hauser, L, Simmons, RK, Templin, WD, and Seeb, LW (2011) Temporal and geographic genetic divergence: Characterizing Sockeye salmon populations in the Chignik Watershed, Alaska, using single-nucleotide polymorphisms. *Transactions of the American Fisheries Society* **140**, 749-762.

Crossman, EJ (1991) Introduction fresh-water fishes - A review of the North American perspective with emphasis on Canada. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 46-57.

Darwin, C (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (2nd ed.) John Murray, London.

Davis, J, and Stamps, J (2004) The effect of natal experience on habitat preferences. *Trends in Ecology & Evolution* **19**, 411-416.

de Zwart, I, Andrusak, G, Thorley, J, Irvine, R, and Masse, S (2011) Duncan Dam Project Water Use Plan: Duncan Reservoir fish habitat use monitoring Year 3 (2010) Data Report, Nelson, BC.

DeFaveri, J, Shikano, T, Shimada, Y, Goto, A, and Merila, J (2011) Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution* **65**, 1800-1807.

Dieckmann, U, and Doebeli, M (1999) On the origin of species by sympatric speciation. *Nature* **400**, 354-357.

Dill, PA (1996) A study of shore-spawning kokanee salmon (*Oncorhynchus nerka*) at Bertram Creek Park Okanagan Lake, B.C., 1992-1996. , p. 23 pp. B.C. Ministry of Environment, Penticton, B.C.

Dirienzo, A, Peterson, AC, Garza, JC, Valdes, AM, Slatkin, M, and Freimer, NB (1994) Mutational processes of simple-sequence repeat loci in human popualtions *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3166-3170.

Dobzhansky, T (1951) *Genetics and the Origin of Species* Columbia University Press, London.

Egan, SP, Nosil, P, and Funk, DJ (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* **62**, 1162-1181.

Ellegren, H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics* **24**, 400-402.

Elmer, KR, and Meyer, A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution* **26**, 298-306.

Evanno, G, Regnaut, S, and Goudet, J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier, L, Hofer, T, and Foll, M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.

Excoffier, L, and Lischer, HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Excoffier, L, Smouse, PE, and Quattro, JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - Application to human mitochondrial DNA restriction data *Genetics* **131**, 479-491.

Felsenstein, J (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**, 124-138.

Foll, M, and Gaggiotti, O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977-993.

Foote, CJ, Moore, K, Stenberg, K, Craig, KJ, Wenburg, JK, and Wood, CC (1999) Genetic differentiation in gill raker number and length in sympatric anadromous and nonanadromous morphs of sockeye salmon, *Oncorhynchus nerka*. *Environmental Biology of Fishes* **54**, 263-274.

Foote, CJ, Wood, CC, Clarke, WC, and Blackburn, J (1992) Circannunal cycle of seawater adaptability in *Oncorhynchus nerka* - Genetic differences between sympatric sockeye salmon and kokanee. *Canadian Journal of Fisheries and Aquatic Sciences* **49**, 99-109.

Foote, CJ, Wood, CC, and Withler, RE (1989) Biochemical genetic comparison of anadromous sockeye and kokanee, the anadromous and non-anadromous forms of *Oncorhynchus nerka*. *Canadian Journal of Fisheries and Aquatic Sciences* **46**, 149-158.

Fraser, DJ, and Bernatchez, L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology* **10**, 2741-2752.

Fraser, DJ, and Bernatchez, L (2005) Allopatric origins of sympatric brook charr populations: Colonization history and admixture. *Molecular Ecology* **14**, 1497-1509.

Fraser, DJ, Weir, LK, Bernatchez, L, Hansen, MM, and Taylor, EB (2011) Extent and scale of local adaptation in salmonid fishes: Review and meta-analysis. *Heredity* **106**, 404-420.

Freamo, H, O'Reilly, P, Berg, PR, Lien, S, and Boulding, EG (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources* **11**, 254-267.

Godbout, L, Wood, CC, Withler, RE, Latham, S, Nelson, RJ, Wetzel, L, Barnett-Johnson, R, Grove, MJ, Schmitt, AK, and McKeegan, KD (2010) Sockeye salmon (*Oncorhynchus nerka*) return after an absence of nearly 90 years: A case of reversion to anadromy. *Canadian Journal of Fisheries and Aquatic Sciences* **68**, 1590-1602.

Gomez-Uchida, D, Seeb, JE, Smith, MJ, Habicht, C, Quinn, TP, and Seeb, LW (2011) Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evolutionary Biology* **11**, 17.

Gould, SJ, and Lewontin, RC (1979) Spandrels of San Marco and the Panglossian Paradigm - A critique of the adaptionist program. *Proceedings of the Royal Society of London Series B* **205**, 581-598.

Guo, SW, and Thompson, EA (1992) A Monte-Carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics* **51**, 1111-1126.

Gustafson, RG, Waples, R, Kalinowski, ST, and Winans, GA (2001) Evolution of sockeye salmon ecotypes. *Science* **291**, 251-251.

Haldane, J (1949) The cost of natural selection. *Genetics* **55**, 511-524.

Hansen, MM, Ruzzante, DE, Nielsen, EE, and Mensberg, KLD (2000) Microsatellite and mitochondrial DNA polymorphism reveals life-history dependent interbreeding between hatchery and wild brown trout (*Salmo trutta* L.). *Molecular Ecology* **9**, 583-594.

Harmon, LJ, Matthews, B, Des Roches, S, Chase, JM, Shurin, JB, and Schluter, D (2009) Evolutionary diversification in stickleback affects ecosystem functioning. *Nature* **458**, 1167-1170.

Hauser, L, and Seeb, JE (2008) Advances in molecular technology and their impact on fisheries genetics. *Fish and Fisheries* **9**, 473-486.

Helyar, SJ, Hemmer-Hansen, J, Bekkevold, D, Taylor, MI, Ogden, R, Limborg, MT, Cariani, A, Maes, GE, Diopere, E, Carvalho, GR, and Nielsen, EE (2011) Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources* **11**, 123-136.

Hendry, AP (2009) Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 1383-1398.

Hendry, AP, and Kinnison, MT (1999) The pace of modern life: Measuring rates of contemporary microevolution. *Evolution* **53**, 1637-1653.

Hendry, AP, and Stearns, SC (2004) *Evolution Illuminated: Salmon and Their Relatives* Oxford University Press, New York.

Hendry, AP, Wenburg, J, Bentzen, P, Volk, E, and Quinn, TP (2000) Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science* **290**, 516-518.

Hermisson, J, and Pennings, P (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **162**, 2335-2352.

Hilborn, R, Quinn, TP, Schindler, DE, and Rogers, DE (2003) Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences* **100**, 6564-6568.

Hoekstra, HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**, 222-234.

Hoekstra, HE, Hirschmann, RJ, Bundey, RA, Insel, PA, and Crossland, JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101-104.

Hohenlohe, PA, Bassham, S, Currey, M, and Cresko, WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across three-spine stickleback genomes. *Philosophical Transactions of the Royal Society of London Series B* **367**, 395-408.

Hohenlohe, PA, Bassham, S, Etter, PD, Stiffler, N, Johnson, EA, and Cresko, WA (2010) Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. *PLoS Genetics* **6**.

Holderegger, R, Herrmann, D, Poncet, B, Gugerli, F, Thuiller, W, Taberlet, P, Gielly, L, Rioux, D, Brodbeck, S, Aubert, S, and Manel, S (2008) Land ahead: Using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology & Diversity* **1**, 273-283.

Huber, SK, De Leon, LF, Hendry, AP, Bermingham, E, and Podos, J (2007) Reproductive isolation of sympatric morphs in a population of Darwin's finches. *Proceedings of the Royal Society of London Series B* **274**, 1709-1714.

Hubisz, MJ, Falush, D, Stephens, M, and Pritchard, JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

Jensen, LF, Hansen, MM, Pertoldi, C, Holdensgaard, G, Mensberg, K-LD, and Loeschcke, V (2008) Local adaptation in brown trout early life-history traits: Implications for climate change adaptability. *Proceedings of the Royal Society of London Series B* **275**, 2859-2868.

Jiggins, CD (2008) Ecological speciation in mimetic butterflies. *Bioscience* **58**, 541-548.

Johannesson, K (2001) Parallel speciation: a key to sympatric divergence. *Trends in Ecology & Evolution* **16**, 148-153.

Jombart, T, Devillard, S, and Balloux, F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**.

Joost, S, Bonin, A, Bruford, MW, Despres, L, Conord, C, Erhardt, G, and Taberlet, P (2007) A Spatial Analysis Method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955-3969.

Kalinowski, ST, Manlove, KR, and Taper, ML (2007) ONCOR: a computer program for genetic stock identification, Department of Ecology, Montana State University, Bozeman, MT. Available at http://www.montana.edu/kalinowski/kalinowski_software.htm.

Kaplan, NL, Hudson, RR, and Langley, CH (1989) The hitchhiking effect revisited *Genetics* **123**, 887-899.

Kauer, MO, Dieringer, D, and Schlötterer, C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of *Drosophila melanogaster*. *Genetics* **165**, 1137-1148.

Keller, I, Taverna, A, and Seehausen, O (2011) Evidence of neutral and adaptive genetic divergence between European trout populations sampled along altitudinal gradients. *Molecular Ecology* **20**, 1888-1904.

Kimura, M (1983) *The neutral theory of molecular evolution* Cambridge University Press, Cambridge.

Kimura, M (1995) Limitations of Darwinian Selection in a finite population. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 2343-2344.

Kinnison, MT, and Hendry, AP (2001) The pace of modern life II: From rates of contemporary microevolution to pattern and process. *Genetica* **112**, 145-164.

Koljonen, ML, Pella, JJ, and Masuda, M (2005) Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo salar*) catches from DNA microsatellite data. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 2143-2158.

Kondrashov, AS (1986) Multilocus model of sympatric speciation. III. Computer simulations. *Theoretical Population Biology* **29**, 1-15.

Kondrashov, AS, and Kondrashov, FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**, 351-354.

Landry, L, Vincent, WF, and Bernatchez, L (2007) Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology* **20**, 971-984.

Lecomte, F, and Dodson, JJ (2004) Role of early life-history constraints and resource polymorphism in the segregation of sympatric populations of an estuarine fish. *Evolutionary Ecology Research* **6**, 631-658.

Leder, EH, Danzmann, RG, and Ferguson, MM (2006) The candidate gene, Clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity* **97**, 74-80.

Liao, B-Y, Weng, M-P, and Zhang, J (2010) Contrasting genetic paths to morphological and physiological evolution. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 7353-7358.

Long, K (2003) Okanagan Region Fish Species at Risk Status Report, p. 37. Okanagan Nation Alliance, Fisheries Department, Westbank, BC.

Luikart, G, England, PR, Tallmon, D, Jordan, S, and Taberlet, P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* **4**, 981-994.

Martin, KLM, and Swiderski, DL (2001) Beach spawning in fishes: phylogenetic tests of hypotheses. *American Zoologist* **41**, 526-537.

Maynard-Smith, J, and Haigh, J (1974) The hitch-hiking effect of a favourable gene. *Genetic Resources* **23**, 23-35.

Mayr, E (1942) *Systematics and the origin of species from the viewpoint of a zoologist* Harvard University Press.

Mayr, E (1963) *Animal Species and Evolution* Belknap Press, Cambridge, Massachusetts.

McGlauflin, MT, Schindler, DE, Seeb, LW, Smith, CT, Habicht, C, and Seeb, JE (2011) Spawning habitat and geography influence population structure and juvenile migration timing of sockeye salmon in the Wood River Lakes, Alaska. *Transactions of the American Fisheries Society* **140**, 763-782.

McKinnon, JS, Mori, S, Blackman, BK, David, L, Kingsley, DM, Jamieson, L, Chou, J, and Schluter, D (2004) Evidence for ecology's role in speciation. *Nature* **429**, 294-298.

McPhail, JD, and Lindsey, CC (1970) *Freshwater fishes of northwestern Canada and Alaska* Fisheries Research Board of Canada.

Mehner, T, Freyhof, J, and Reichard, M (2011) Summary and perspective on evolutionary ecology of fishes. *Evolutionary Ecology* **25**, 547-556.

Miller, MR, Brunelli, JP, Wheeler, PA, Liu, S, Rexroad, CE, III, Palti, Y, Doe, CQ, and Thorgaard, GH (2001) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology* **21**, 237-249.

Moen, T, Hayes, B, Baranski, M, Berg, PR, Kjoglum, S, Koop, BF, Davidson, WS, Omholt, SW, and Lien, S (2008) A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. *BMC Genomics* **9**.

Morin, PA, Luikart, G, Wayne, RK, and the SNP Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19**, 208-216.

Morris, A, Braumandl, E, Andrusak, H, and Caverly, A (2003) 2002/2003 Seton and Anderson Lakes kokanee assessment - Feasibility study and study design (ed. British Columbia Conservation Foundation and Ministry of Water LaAP, Kamloops, BC.).

Namroud, M-C, Beaulieu, J, Juge, N, Laroche, J, and Bousquet, J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**, 3599-3613.

Narum, SR, Banks, M, Beacham, TD, Bellinger, MR, Campbell, MR, Dekoning, J, Elz, A, Guthrie, CM, Kozfkay, C, Miller, KM, Moran, P, Phillips, R, Seeb, LW, Smith, CT, Warheit, K, Young, SF, and Garza, JC (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* **17**, 3464-3477.

Narum, SR, and Hess, JE (2011) Comparison of F(ST) outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.

Nei, M (1987) *Molecular evolutionary genetics* Columbia University Press, USA.

Ng, SHS, Artieri, CG, Bosdet, IE, Chiu, R, Danzmann, RG, Davidson, WS, Ferguson, MM, Fjell, CD, Hoyheim, B, Jones, SJM, de Jong, PJ, Koop, BF, Krzywinski, MI, Lubieniecki, K, Marra, MA, Mitchell, LA, Mathewson, C, Osoegawa, K, Parisotto, SE, Phillips, RB, Rise, ML, von Schalburg, KR, Schein, JE, Shin, HS, Siddiqui, A, Thorsen, J, Wye, N, Yang, G, and Zhu, BL (2005) A physical map of the genome of Atlantic salmon, *Salmo salar*. *Genomics* **86**, 396-404.

Nie, YC, Han, YC, and Zou, YR (2008) CXCR4 is required for the quiescence of primitive hematopoietic cells. Journal of *Experimental Medicine* **205**, 777-783.

Nielsen, EE, Hemmer-Hansen, J, Larsen, PF, and Bekkevold, D (2009a) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* **18**, 3128-3150.

Nielsen, EE, Hemmer-Hansen, J, Poulsen, NA, Loeschcke, V, Moen, T, Johansen, T, Mittelholzer, C, Taranger, G-L, Ogden, R, and Carvalho, GR (2009b) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**.

Nielsen, R (2005) Molecular signatures of natural selection. In: *Annual Review of Genetics*, pp. 197-218.

Nosil, P, Funk, DJ, and Ortiz-Barrientos, D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.

Nunes, VL, Beaumont, MA, Butlin, RK, and Paulo, OS (2011) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology* **20**, 193-205.

Oetjen, K, and Reusch, TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology* **16**, 5156-5167.

Ogden, R, and Thorpe, RS (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 13612-13615.

Ohta, T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96-98.

Olsen, JB, Wenburg, JK, and Bentzen, P (1996) Semiautomated multilocus genotyping of Pacific salmon (*Oncorhynchus spp.*) using microsatellites. *Molecular Marine Biology and Biotechnology* **5**, 259-272.

Ostberg, CO, Pavlov, SD, and Hauser, L (2009) Evolutionary relationships among sympatric life history forms of Dolly Varden inhabiting the landlocked Kronotsky Lake, Kamchatka, and a neighboring anadromous population. *Transactions of the American Fisheries Society* **138**, 1-14.

Paragamian, VL, and Bowles, EC (1995) Factors affecting survival of kokanee stocked in Lake Pend Orielle, Idaho. *North American Journal of Fisheries Management* **15**, 208-219.

Parametrix (2003) Lake Whatcome and Bellingham Hatcheries production replacement feasibility report. Washington Department of Fish and Wildlife, Resident Native Fish Program and Hatchery Program, Olympia, WA.

Peakall, R, and Smouse, PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Peichel, CL, Nereng, KS, Ohgi, KA, Cole, BLE, Colosimo, PF, Buerkle, CA, Schluter, D, and Kingsley, DM (2001) The genetic architecture of divergence between three-spine stickleback species. *Nature* **414**, 901-905.

Pfennig, DW, Wund, MA, Snell-Rood, EC, Cruickshank, T, Schlichting, CD, and Moczek, AP (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution* **25**, 459-467.

Pielou, EC (1991) *After the Ice Age: The return of life to glaciated North America* The University of Chicago Press, Chicago.

Piry, S, Luikart, G, and Cornuet, JM (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502-503.

Pritchard, JK, Stephens, M, and Donnelly, P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Provine, WB (1971) *The origin of theoretical populatiosn genetics* Chigaco University Press, Chicago.

Quesada, H, Posada, D, Caballero, A, Moran, P, and Rolan-Alvarez, E (2007) Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**, 1600-1612.

Quinn, TP (2005) *The Behaviour and Ecology of Pacific Salmon and Trout* University of Washington Press, Vancouver.

Quinn, TP, Hendry, AP, and Wetzel, LA (1995) The influence of life history trade-offs and the size of incubation gravels on egg size variation in sockeye salmon (*Oncorhynchus nerka*). *Oikos* **74**, 425-438.

Rambaut, A, and Schluter, D (1996) Ecological speciation in postglacial fishes - Discussion. *Philosophical Transactions of the Royal Society of London Series B* **351**, 814-814.

Rannala, B, and Mountain, JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9197-9201.

Raymond, M, and Rousset, F (1995) An exact test for population differentiation. *Evolution* **49**, 1280-1283.

Renaut, S, Nolte, AW, Rogers, SM, Derome, N, and Bernatchez, L (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus spp*.). *Molecular Ecology* **20**, 545-559.

Rexroad, CE, Rodriguez, MF, Coulibaly, I, Gharbi, K, Danzmann, RG, DeKoning, J, Phillips, R, and Palti, Y (2005) Comparative mapping of expressed sequence tags containing microsatellites in rainbow trout (*Oncorhynchus mykiss*). *BMC Genomics* **6**, 54.

Reznick, DN, and Ghalambor, CK (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* **112**, 183-198.

Reznick, DN, and Ghalambor, CK (2005) Selection in nature: experimental manipulations of natural populations. *Integrative and Comparative Biology* **45**, 456-462.

Rice, WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Rice, WR, and Hostert, EE (1993) Laboratory experiments on speciation - What have we learned in 40 years. *Evolution* **47**, 1637-1653.

Roberge, C, Paez, DJ, Rossignol, O, Guderley, H, Dodson, J, and Bernatchez, L (2007) Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular Immunology* **44**, 1374-1383.

Rogers, SM, and Bernatchez, L (2006) The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology* **19**, 1979-1994.

Rundle, HD, Nagel, L, Boughman, JW, and Schluter, D (2000) Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306-308.

Russello, MA, Kirk, S, Frazer, KK, and Askey, P (2012) Detection of outlier loci and their utility for fisheries management *Evolutionary Applications* **5**, 39-52.

Saint-Laurent, R, Legault, M, and Bernatchez, L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). *Molecular Ecology* **12**, 315-330.

Salem, M, Rexroad, CE, III, Wang, J, Thorgaard, GH, and Yao, J (2010) Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. *BMC Genomics* **11**, 564-574.

Schliewen, UK, Tautz, D, and Paabo, S (1994) Sympatric speciation suggested by monophyly of Crater Lake cichlids. *Nature* **368**, 629-632.

Schlötterer, C (2003) Hitchhiking mapping - functional genomics from the population genetics perspective. *Trends in Genetics* **19**, 32-38.

Schluter, D (1996) Ecological speciation in postglacial fishes. *Philosophical Transactions of the Royal Society of London Series B* **351**, 807-814.

Schluter, D (2000) *The Ecology of Adaptive Radiation* Oxford University Press, Oxford.

Schluter, D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372-380.

Schluter, D, Clifford, EA, Nemethy, M, and McKinnon, JS (2004) Parallel evolution and inheritance of quantitative traits. *American Naturalist* **163**, 809-822.

Schluter, D, and Conte, GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9955-9962.

Schluter, D, and McPhail, JD (1992) Ecological character displacement and speciation in sticklebacks. *American Naturalist* **140**, 85-108.

Schuelke, M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233-234.

Schwartz, MK, Luikart, G, and Waples, RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution* **22**, 25-33.

Scribner, KT, Gust, JR, and Fields, RL (1996) Isolation and characterization of novel salmon microsatellite loci: Cross-species amplification and population genetic applications. *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 833-841.

Sebastian, DC, Dolighan, R, Andrusak, H, Hume, J, Woodruff, P, and Scholten, G (2003) Summary of Quesnel Lake kokanee and rainbow trout biology with reference to sockeye salmon. In: *Stock Management Report No. 17*, Province of British Columbia.

Seehausen, O, Terai, Y, Magalhaes, IS, Carleton, KL, Mrosso, HDJ, Miyagi, R, van der Sluijs, I, Schneider, MV, Maan, ME, Tachida, H, Imai, H, and Okada, N (2008) Speciation through sensory drive in cichlid fish. *Nature* **455**, 620-623.

Shaklee, JB, Beacham, TD, Seeb, L, and White, BA (1999) Managing fisheries using genetic data: case studies from four species of Pacific salmon. *Fisheries Research* **43**, 45-78.

Shapiro, MD, Marks, ME, Peichel, CL, Blackman, BK, Nereng, KS, Jonsson, B, Schluter, D, and Kingsley, DM (2004) Genetic and developmental basis of evolutionary pelvic reduction in three-spine sticklebacks. *Nature* **428**, 717-723.

Shephard, BG (2000) A case history: the kokanee stocks of Okanagan Lake. In: *Proceedings of a conference on the biology and management of species and habitats at risk* (ed. Darling LM), pp. 609-616, Kamloops, B.C.

Shikano, T, Ramadevi, J, and Merila, J (2010) Identification of local- and habitat-dependent selection: Scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution* **27**, 2775-2789.

Skugar, S, Glover, KA, Nilsen, F, and Krasnov, A (2008) Local and systemic gene expression responses of Atlantic salmon (*Salmo salar* L.) to infection with the salmon louse (*Lepeophtheirus salmonis*). *BMC Genomics* **9**, 498-516.

Slatkin, M (1995) Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution* **12**, 473-480.

Smith, JM, and Haigh, J (1974) Hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23-35.

St-Cyr, J, Derome, N, and Bernatchez, L (2008) The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus sp.*). *Molecular Ecology* **17**, 1850-1870.

Stinchcombe, JR, and Hoekstra, HE (2008) Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity* **100**, 158-170.

Stockwell, CA, and Weeks, SC (1999) Translocations and rapid evolutionary responses in recently established populations of western mosquitofish (*Gambusia affinis*). *Animal Conservation* **2**, 103-110.

Storz, JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

Storz, JF, and Nachman, MW (2003) Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution* **57**, 2628-2635.

Taylor, EB (1991) A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture* **98**, 185-207.

Taylor, EB (1999) Species pairs of north temperate freshwater fishes: Evolution, taxonomy, and conservation. *Reviews in Fish Biology and Fisheries* **9**, 299-324.

Taylor, EB, Foote, CJ, and Wood, CC (1996) Molecular genetic evidence for parallel life-history evolution within a Pacific salmon (sockeye salmon and kokanee, *Oncorhynchus nerka*). *Evolution* **50**, 401-416.

Taylor, EB, Harvey, S, Pollard, S, and Volpe, J (1997) Postglacial genetic differentiation of reproductive ecotypes of kokanee *Oncorhynchus nerka* in Okanagan Lake, British Columbia. *Molecular Ecology* **6**, 503-517.

Taylor, EB, Kuiper, A, Troffe, PM, Hoysak, DJ, and Pollard, S (2000) Variation in developmental biology and microsatellite DNA in reproductive ecotypes of kokanee, *Oncorhynchus nerka*: Implications for declining populations in a large British Columbia lake. *Conservation Genetics* **1**, 231-249.

Teshima, KM, Coop, G, and Przeworski, M (2006) How reliable are empirical genomic scans for selective sweeps? *Genome Research* **16**, 702-712.

Theriault, V, Bernatchez, L, and Dodson, JJ (2007) Mating system and individual reproductive success of sympatric anadromous and resident brook charr, *Salvelinus fontinalis*, under natural conditions. *Behavioral Ecology and Sociobiology* **62**, 51-65.

Thornton, KR, and Jensen, JD (2007) Controlling the false-positive rate in multilocus genome scans for selection. *Genetics* **175**, 737-750.

Tice, KA, and Carlon, DB (2011) Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *Journal of Evolutionary Biology* **24**, 1814-1825.

Tucker, S, Trudel, M, Welch, DW, Candy, JR, Morris, JFT, Thiess, ME, Wallace, C, Teel, DJ, Crawford, W, Farley, EV, Jr., and Beacham, TD (2009) Seasonal Stock-Specific Migrations of Juvenile Sockeye Salmon

along the West Coast of North America: Implications for Growth. *Transactions of the American Fisheries Society* **138**, 1458-1480.

USFWS (2008) Endangered and threatened wildlife and plants; 90-Day finding on a petition to list kokanee (*Oncorhynchus nerka*) in Lake Sammamish, Washington, as threatened or endangered.

Van Oosterhout, C, Hutchinson, WF, Wills, DPM, and Shipley, P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

VanDeHey, JA, Sloss, BL, Peeters, PJ, and Sutton, TM (2009) Determining the efficacy of microsatellite DNA-based mixed stock analysis of Lake Michigan's Lake Whitefish commercial Fishery. *Journal of Great Lakes Research* **36**, 52-58.

Vasemagi, A, Nilsson, J, and Primmer, CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution* **22**, 1067-1076.

Vasemagi, A, and Primmer, CR (2005) Challenges for identifying functionally important genetic variation: The promise of combining complementary research strategies. *Molecular Ecology* **14**, 3623-3642.

Via, S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9939-9946.

Via, S, and West, J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology* **17**, 4334-4345.

Vitalis, R, Dawson, K, and Boursot, P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811-1823.

Vitalis, R, Dawson, K, Boursot, P, and Belkhir, K (2003) DetSel 1.0: A computer program to detect markers responding to selection. *Journal of Heredity* **94**, 429-431.

Waples, RS (2010) High-grading bias: subtle problems with assessing power of selected subsets of loci for population assignment. *Molecular Ecology* **19**, 2599-2601.

Waples, RS, Gustafson, RG, Weitkamp, LA, Myers, JM, Johnson, OW, Busby, PJ, Hard, JJ, Bryant, GJ, Waknitz, FW, Neely, K, Teel, D, Grant, WS, Winans, GA, Phelps, S, Marshall, A, and Baker, BM (2001) Characterizing diversity in salmon from the Pacific Northwest. *Journal of Fish Biology* **59**, 1-41.

Waples, RS, and Hendry, AP (2008) Special Issue: Evolutionary perspectives on salmonid conservation and management. *Evolutionary Applications* **1**, 183-188.

Weir, BS, and Cockerham, CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Wenne, R, Boudry, P, Hemmer-Hansen, J, Lubieniecki, KP, Was, A, and Kause, A (2007) What role for genomics in fisheries management and aquaculture? *Aquatic Living Resources* **20**, 241-255.

Wiehe, T, Nolte, V, Zivkovic, D, and Schlötterer, C (2007) Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics* **175**, 207-218.

Wilding, CS, Butlin, RK, and Grahame, J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**, 611-619.

Winans, GA, Pollard, S, and Kuligowski, DR (2003) Two reproductive life history types of kokanee, *Onchorynchus nerka*, exhibit multivariate morphometric and protein genetic differentiation. *Environmental Biology of Fishes* **67**, 87-100.

Wood, CC, Bickham, JW, Nelson, RJ, Foote, CJ, and Patton, JC (2008) Recurrent evolution of life history ecotypes in sockeye salmon: implications for conservation and future evolution. *Evolutionary Applications* **1**, 207-221.

Wood, CC, and Foote, CJ (1990) Genetic differences in the early development and growth of sympatric sockeye salmon and kokanee (*Oncorhynchus nerka*) and their hybrids. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 2250-2260.

Wood, CC, and Foote, CJ (1996) Evidence for sympatric genetic divergence of anadromous and nonanadromous morphs of sockeye salmon (*Oncorhynchus nerka*). *Evolution* **50**, 1265-1279.

Wright, JJ, Lubieniecki, KP, Park, JW, Ng, SHS, Devlin, RH, Leong, J, Koop, BF, and Davidson, WS (2008) Sixteen Type 1 polymorphic microsatellite markers from Chinook salmon (*Oncorhynchus tshawytscha*) expressed sequence tags. *Animal Genetics* **39**, 84-85.

APPENDICES

Appendix A: Supplementary Material for Chapter 2

Table A.1 The physical and biological attributes of each British Columbian Lake included in this study.

Lake	Latitude, Longitude	Surface area	Max depth (mean)	When lake formed	Catchment area (km^2)	Residence time	Water level fluctuation	Productivity
Christina	49°07'N, 118°15'W	25km^2	49 m (34 m)	Post-glacial	n/a	n/a	n/a	oligotrophic
Duncan	50°26'N, 116°59'W	73km^2	40 m (n/a)	Man-made (1967)	25	n/a	regulated	n/a
Kootenay, West Arm	49°40'N, 116°60'W	24 km^2	47 m (12 m)	Post-glacial	45,584	5.5 days	regulated	oligotrophic
Kootenay, Main Lake	50°00'N, 116°59'W	389 km^2	154 m (94 m)	Post-glacial		1.5 years	regulated	oligotrophic
Okanagan	50°0'N 119°30'W	351 km^2	230 m (76 m)	Post-glacial	6,188	52.8 years	regulated	oligotrophic
Wood	50°05'N 119°23'W	9 km^2	34 m (22 m)	Post-glacial	190	33.8 years	regulated	eutrophic (last 20 yrs)
Tchesinkut	54°05'N 125°35'W	34 km^2	149 m (62 m)	Post-glacial	small	19 years	~2 m due to beaver dams	oligotrophic

Table A.2 Details of kokanee sample collection are listed, including the number of sites and total number of samples collected for each ecotype group. The capture method used to obtain samples, the type of tissue collected and the timing of tissue collection are also provided for each ecotype group in each lake.

Lake	Ecotype	No. sites	Total no. samples	Capture method	Tissue type	Date
Christina	Shore	1	48	gillnet	operculum	Dec 2011
	Stream	1	48	carcass	operculum	Sept 2009
Duncan	Shore	2	50	carcass	operculum	Sept 2011
	Stream	2	50	carcass	operculum	Sept 2011
Kootenay (West)	Shore	1	27	carcass	operculum	Sept 2010
	Stream	2	41	carcass	operculum	Sept 2010
Kootenay (North)	Stream	2	56	carcass	operculum	Sept 2010
Okanagan**	Shore	3	72	carcass	operculum adipose fin	Oct 2007, 2010
	Stream	4	72	carcass	operculum	Oct 2007, 2010
Wood	Shore	1	40	carcass	operculum	Oct 2007
	Stream	1	40	carcass	operculum	Oct 2007
Tchesinkut	Shore	1	48	gillnet & carcass	operculum	Nov 2010
	Stream	2	96	dip net	adipose fin	Nov 2010

**Samples were genotyped by Russello *et al.* (2012)

Table A.3 A descriptive summary of all 50 microsatellite loci (42 EST-linked and 8 anonymous) used in this study, including marker name, type, Genbank accession number, microsatellite motif, allele size range, total number of alleles, F_{ST}, and locus behaviour for each locus.

Marker	Marker type	Microsatellite repeat motif	Size range (bp)	No of alleles	F_{ST}	Outlier behaviour	Genbank accession
Ca983	EST	(GT)13	263-311	21	0.187	False	CA039983
EV149	EST	(AT)12	204-232	13	0.088	Neutral	EV377149
EV170	EST	(AC)6	214-224	6	0.120	Outlier	EV375170
EV188	EST	(AC)24	231-271	19	0.245	False	EV383188
EV220	EST	(GT)11	199-217	6	0.465	False	EV378220
EV249	EST	(AC)7	212-226	4	0.236	Neutral	EV377249
EV291	EST	(GT)8	253259	4	0.213	False	EV377291
EV358	EST	(AC)33	182-246	29	0.097	R Outlier	EV383358
EV365	EST	(GT)8	226-228	2	0.241	False	EV377365
EV475	EST	(GT)12	225-335	7	0.153	False	EV376475
EV484	EST	(AC)9	169-179	6	0.384	False	EV379484
EV536	EST	(AT)8	237-243	4	0.096	False	EV381536
EV597	EST	(GT)8	184-194	5	0.280	False	EV375597
EV634	EST	(AC)8	188-192	3	0.044	False	EV380634
EV642	EST	(GT)10	232-286	12	0.222	Outlier	EV382642
EV685	EST	(GT)12	222-230	5	0.283	Outlier	EV382685
EV691	EST	(AC)7	224-232	4	0.039	Outlier	EV379691
EV712	EST	(AG)10	212-222	4	0.258	Neutral	EV382712
EV723	EST	(AC)9	233-333	3	0.293	Neutral*	EV380723
EV740	EST	(AT)8	264-272	5	0.387	Outlier	EV374740
EV769	EST	(GT)11	171-177	2	0.116	Neutral	EV377769
EV862	EST	(AC)6	222-230	4	0.285	Outlier	EV376862
EV911	EST	(AG)7	234-256	5	0.740	False	EV375911
OMM5003	EST	(GT)3	196-216	11	0.163	R Outlier	CO805109
OMM5007	EST	(GT)25	187-201	8	0.162	Neutral	CO805113
OMM5008	EST	(GT)19	241-267	11	0.179	Neutral	CO805114
OMM5032	EST	(CA)13	189-207	8	0.356	False	CA349143
OMM5033	EST	(CA)28	263-359	33	0.141	Outlier	CA349148
OMM5037	EST	(CA)15	207-301	10	0.053	Outlier	CA348625
OMM5053	EST	(GT)21	245-333	32	0.063	False	CA349198
OMM5058	EST	(CA)11	234-286	26	0.063	False	CA348781
OMM5067	EST	(CA)13	209-235	6	0.245	R Outlier	CA348790
OMM5075	EST	(GT)12	214-234	7	0.090	Neutral	CA348807
OMM5091	EST	(GA)49(GT)11	229-263	9	0.415	Neutral	CA348850
OMM5099	EST	(CT)24	273-297	10	0.250	Neutral*	CA348959
OMM5108	EST	(GT)12	286-302	8	0.086	Neutral	CA349062
OMM5121	EST	(AG)15	196-200	3	0.380	Outlier	CA349040
OMM5124	EST	(GT10)	286-294	5	0.158	False	CA349048
OMM5125	EST	(CA)13	268-282	5	0.137	Outlier	CO805127
One102	Anonymous	(ATCT)10	222-290	18	0.061	Neutral	AF274518
One105	Anonymous	(TAGA)9	151-195	9	0.149	Neutral	AF274521
One108	Anonymous	(ATCT)21	201-331	28	0.050	False	AF274524
One110	Anonymous	(TAGA)21	239-323	28	0.079	Neutral	AF274526
One112	Anonymous	(ATCT)28	135-263	27	0.059	Neutral	AF274528
One14	Anonymous	(CA)26	148-172	13	0.058	Neutral	n/a
One8	Anonymous	(CA)24	204-330	14	0.411	False	n/a
Ots06	EST	(GTAT)n	174-178	2	0.160	Outlier	CX351581
Ots07	EST	(AT)n	247-275	13	0.065	Neutral	CX351642
Ots14	Anonymous	(GT)n	293-313	9	0.298	False	CX352740
TAP2	EST	(xxxx)n	329-341	2	0.055	R Outlier	Z83326

* identifies a truly neutral locus not included in the 'neutral' dataset because it was monomorphic in ≥1 lake

Table A.4 Population pair-wise F_{ST} comparisons among sampling sites in (a) Duncan, (b) Kootenay, (c) Okanagan, (d) Tchesinkut and (e) Wood Lake, using eight anonymous microsatellite markers. Inter-ecotype comparisons are bolded.

(a) Duncan

Ecotype		Shore		Stream	
	Site	Griz	Little Glacier	SOB	Upper Duncan
Shore	Griz	-			
	Little Glacier	**0.006**	-		
Stream	SOB Creek	0.006*	0.003	-	
	Upper Duncan	0.013*	-0.003	**0.006**	-

(b) Kootenay

Ecotype		Shore	Stream	
	Site	Six Mile	Six Mile	Harrop
Shore	Six Mile	-		
Stream	Duhamel Creek	0.018*	-	
	Harrop Creek	0.011*	**0.004**	-

(c) Okanagan

Ecotype		Shore			Stream			
	Site	North-east	North-west	South-east	Peachland creek	Penticton creek	Mission creek	Powers creek
Shore	Northeast	-						
	Northwest	**0.012***	-					
	Southeast	**0.009**	**0.003**	-				
Stream	Peachland	0.001	0.010*	0.018*	-			
	Penticton	0.000	0.020*	0.021*	**0.008**	-		
	Mission	0.010	0.002	0.005	**0.007**	**0.021***	-	
	Powers	0.006*	-0.001	0.006	**0.002**	**0.001**	**-0.002**	-

(d) Tchesinkut

Ecotype		Shore	Stream	
	Site	Island	Drew Creek	Tchesinkut Inlet
Shore	Island	-		
Stream	Drew Creek	0.001	-	
	Tchesinkut Inlet	0.003	**-0.005**	-

(e) Wood

Ecotype	Shore	Stream
Shore	-	
Stream	0.039*	-

*Significant value P<0.05

Table A.5 Pair-wise population F_{ST} values for all lake-ecotype comparisons using the 15 'neutral loci' (above diagonal) and the 4 'repeat-outlier' loci (below diagonal). Bold indicates comparisons where estimated F_{ST} were greater using the outlier dataset compared to the truly neutral dataset.

Lake		Duncan		Kootenay		Okanagan		Tchesinkut		Wood	
	Ecotype	shore	stream	shore	stream	shore	stream	shore	stream	shore	stream
Duncan	shore	-	-0.001	0.088	0.072	0.139	0.130	0.266	0.288	0.133	0.150
	stream	0.013	-	0.089	0.068	0.135	0.127	0.253	0.276	0.124	0.140
Kootenay	shore	0.122	0.095	-	0.008	0.122	0.123	0.231	0.240	0.149	0.139
	stream	0.143	0.103	0.049	-	0.108	0.108	0.183	0.195	0.121	0.113
Okanagan	shore	0.084	0.130	0.204	0.268	-	0.004	0.193	0.211	0.044	0.044
	stream	0.081	0.115	0.200	0.243	0.017	-	0.194	0.211	0.036	0.039
Tchesinkut	shore	0.103*	0.134*	0.128*	0.162*	0.150*	0.161*	-	0.000	0.204	0.149
	stream	0.177*	0.198*	0.128*	0.188*	0.214*	0.233*	0.041*	-	0.221	0.163
Wood	shore	0.138	0.192	0.267	0.352	0.031	0.052	0.229*	0.282*	-	0.032
	stream	0.067	0.101	0.147	0.190	0.050	0.046	0.095*	0.134*	0.104*	-

* indicates F_{ST} values are significant $P<0.05$ (exact tests)

Table A.6 A summary of the sample size (N), allelic richness (N_a), observed (H_O) and expected heterozygosity (H_E), and fixation index (F) for shore- and stream-spawning kokanee from five British Columbian lakes at all 50 markers.

		Duncan		Kootenay		Okanagan		Wood		Tchesinkut		
Locus		Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Mean (SE)
Ca983	*N*	46	50	26	40	71	68	38	38	48	94	51.9 (6.367)
	N_a	13	12	7	8	13	16	10	8	3	3	9.3 (1.367)
	H_O	0.826	0.640	0.692	0.800	0.803	0.897	0.632	0.789	0.333	0.245	0.666 (0.068)
	H_E	0.853	0.823	0.766	0.712	0.838	0.876	0.714	0.809	0.308	0.242	0.694 (0.072)
	F	0.031	0.222	0.097	-0.124	0.042	-0.024	0.115	0.024	-0.081	-0.010	0.029 (0.031)
EV149	*N*	47	50	27	41	71	68	36	37	47	90	51.4 (6.090)
	N_a	10	9	5	7	11	12	8	9	8	9	8.8 (0.629)
	H_O	0.809	0.840	0.630	0.659	0.676	0.676	0.472	0.865	0.766	0.744	0.714 (0.370)
	H_E	0.816	0.831	0.639	0.753	0.742	0.807	0.754	0.840	0.768	0.732	0.768 (0.019)
	F	0.010	-0.011	0.014	0.126	0.089	0.162	0.374	-0.030	0.002	-0.017	0.072 (0.039)
EV170	*N*	47	50	27	41	72	69	39	38	47	90	52.0 (6.053)
	N_a	4	4	2	2	5	5	5	5	3	3	3.8 (0.389)
	H_O	0.404	0.480	0.074	0.171	0.778	0.623	0.744	0.658	0.383	0.389	0.470 (0.074)
	H_E	0.403	0.508	0.071	0.232	0.713	0.595	0.761	0.735	0.467	0.403	0.489 (0.071)
	F	-0.003	0.054	-0.038	0.265	-0.091	-0.048	0.022	0.105	0.180	0.034	0.048 (0.035)
EV188	*N*	47	50	26	40	72	68	39	38	48	94	52.2 (6.368)
	N_a	10	11	8	6	11	14	7	9	3	6	8.5 (1.003)
	H_O	0.681	0.780	0.769	0.425	0.722	0.735	0.538	0.684	0.458	0.383	0.618 (0.048)
	H_E	0.726	0.735	0.666	0.529	0.725	0.743	0.669	0.736	0.419	0.438	0.639 (0.040)
	F	0.062	-0.061	-0.156	0.197	0.003	0.010	0.195	0.071	-0.094	0.125	0.035 (0.037)
EV220	*N*	46	50	26	41	71	68	39	38	94	92	51.9 (6.178)
	N_a	3	3	1	3	5	6	3	3	6	2	3.1 (0.458)
	H_O	0.435	0.540	0.000	0.122	0.296	0.235	0.231	0.289	0.383	0.109	0.238 (0.051)
	H_E	0.519	0.538	0.000	0.116	0.319	0.	0.207	0.387	0.438	0.103	0.252 (0.058)
	F	0.162	-0.003	n/a	-0.054	0.074	-0.092	-0.116	0.252	0.125	-0.057	0.011 (0.040)
EV249	*N*	47	50	27	41	72	69	39	38	48	92	52.3 (6.186)
	N_a	3	3	2	2	3	3	3	3	2	2	2.6 (0.163)
	H_O	0.468	0.380	0.407	0.488	0.583	0.478	0.641	0.447	0.063	0.076	0.403 (0.061)
	H_E	0.396	0.380	0.431	0.433	0.520	0.534	0.534	0.430	0.061	0.073	0.379 (0.055)
	F	-0.182	-0.001	0.056	-0.126	-0.122	0.104	-0.200	-0.041	-0.032	-0.040	-0.058 (0.031)
EV291	*N*	47	49	27	41	72	69	38	37	48	94	52.2 (6.385)
	N_a	3	4	2	2	2	2	2	2	1	1	2.1 (0.277)
	H_O	0.468	0.469	0.037	0.024	0.514	0.420	0.132	0.432	0.000	0.000	0.250 (0.072)
	H_E	0.458	0.441	0.036	0.024	0.499	0.485	0.123	0.368	0.000	0.000	0.243 (0.071)
	F	-0.022	-0.065	-0.019	-0.012	-0.030	0.133	-0.070	-0.175	n/a	n/a	-0.032 (0.027)
EV358	*N*	46	50	25	41	72	69	38	38	48	93	52.0 (6.384)
	N_a	16	19	13	14	18	14	11	12	6	10	13.3 (1.221)
	H_O	0.848	0.840	0.880	0.756	0.833	0.768	0.842	0.947	0.521	0.505	0.744 (0.047)
	H_E	0.769	0.874	0.838	0.767	0.884	0.832	0.828	0.811	0.480	0.455	0.754 (0.049)
	F	-0.102	0.038	-0.051	0.014	0.057	0.076	-0.017	-0.169	-0.086	-0.111	-0.035 (0.026)
EV365	*N*	47	50	27	41	72	69	39	38	47	94	52.4 (6.339)
	N_a	2	2	2	2	2	2	2	2	1	2	1.9 (0.100)
	H_O	0.404	0.480	0.481	0.488	0.528	0.449	0.513	0.395	0.000	0.011	0.375 (0.063)
	H_E	0.390	0.385	0.401	0.500	0.497	0.489	0.460	0.497	0.000	0.011	0.363 (0.610)
	F	-0.035	-0.247	-0.200	0.024	-0.063	0.082	-0.114	0.206	n/a	-0.005	-0.039 (0.044)
EV475	*N*	46	50	26	41	72	68	37	37	47	93	51.7 (6.367)
	N_a	3	3	2	3	4	3	4	4	3	3	3.2 (0.200)
	H_O	0.630	0.700	0.346	0.366	0.333	0.338	0.459	0.622	0.213	0.247	0.426 (0.054)
	H_E	0.582	0.622	0.375	0.307	0.304	0.364	0.437	0.504	0.209	0.218	0.392 (0.045)
	F	-0.083	-0.126	0.077	-0.193	-0.097	0.072	-0.051	-0.234	-0.018	-0.134	-0.079 (0.032)
EV484	*N*	47	50	26	41	72	68	39	38	46	92	51.9 (6.221)
	N_a	4	3	2	2	3	4	2	2	2	3	2.7 (0.260)
	H_O	0.213	0.380	0.462	0.366	0.236	0.221	0.256	0.079	0.196	0.293	0.27 (0.035)
	H_E	0.334	0.359	0.426	0.404	0.210	0.199	0.224	0.076	0.177	0.282	0.269 (0.035)
	F	0.363	-0.060	-0.083	0.094	-0.126	-0.107	-0.147	-0.041	-0.108	-0.041	-0.026 (0.048)
EV536	*N*	47	50	26	41	71	66	38	37	46	92	51.4 (6.179)
	N_a	3	3	3	3	3	3	3	3	3	3	3.0 (0.00)
	H_O	0.553	0.620	0.538	0.415	0.549	0.545	0.605	0.730	0.435	0.511	0.55 (0.029)
	H_E	0.615	0.638	0.452	0.428	0.619	0.585	0.606	0.606	0.540	0.561	0.565 (0.023)
	F	0.100	0.028	-0.191	0.032	0.112	0.068	0.001	-0.204	0.194	0.090	0.023 (0.041)

		Duncan		Kootenay		Okanagan		Wood		Tchesinkut		
Locus		Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Mean (SE)
EV597	*N*	47	50	27	41	70	69	37	38	48	93	52.0 (6.241)
	N_a	3	3	3	3	5	2	3	2	2	2	2.8 (0.291)
	H_O	0.234	0.320	0.407	0.561	0.171	0.159	0.351	0.447	0.438	0.409	0.350 (0.041)
	H_E	0.212	0.335	0.393	0.487	0.161	0.147	0.305	0.400	0.404	0.393	0.324 (0.036)
	F	-0.104	0.045	-0.037	-0.152	-0.063	-0.087	-0.152	-0.119	-0.082	-0.039	-0.079 (0.019)
EV634	*N*	47	50	27	41	72	69	38	37	48	92	52.1 (6.237)
	N_a	3	3	2	2	3	3	2	2	2	2	2.4 (0.163)
	H_O	0.447	0.500	0.556	0.415	0.611	0.536	0.553	0.405	0.396	0.609	0.503 (0.026)
	H_E	0.486	0.514	0.489	0.381	0.486	0.542	0.472	0.487	0.498	0.496	0.485 (0.013)
	F	0.080	0.028	-0.136	-0.088	-0.258	0.011	-0.171	0.167	0.205	-0.227	-0.039 (0.051)
EV642	*N*	47	50	26	41	72	69	37	38	47	88	51.1 (6.013)
	N_a	6	7	4	6	7	7	4	4	5	5	5.5 (0.401)
	H_O	0.787	0.820	0.538	0.561	0.667	0.623	0.703	0.711	0.553	0.545	0.651 (0.033)
	H_E	0.782	0.715	0.438	0.527	0.647	0.697	0.611	0.713	0.576	0.558	0.625 (0.033)
	F	-0.007	-0.146	-0.230	-0.065	-0.031	0.106	-0.150	0.004	0.039	0.023	-0.046 (0.032)
EV685	*N*	45	50	27	41	72	69	39	36	46	92	51.7 (6.279)
	N_a	3	4	1	2	3	4	2	2	2	2	2.5 (0.307)
	H_O	0.533	0.520	0.000	0.073	0.264	0.246	0.103	0.389	0.043	0.011	0.218 (0.065)
	H_E	0.592	0.627	0.000	0.070	0.256	0.280	0.097	0.313	0.043	0.011	0.229 (0.073)
	F	0.100	0.171	n/a	-0.038	-0.032	0.119	-0.054	-0.241	-0.022	-0.005	0.00 (0.038)
EV691	*N*	47	50	27	41	72	69	39	37	46	92	52.0 (6.231)
	N_a	2	2	2	2	3	2	2	1	1	1	1.8 (0.2)
	H_O	0.085	0.040	0.074	0.171	0.167	0.159	0.077	0.000	0.000	0.000	0.077 (0.022)
	H_E	0.081	0.039	0.071	0.156	0.177	0.147	0.074	0.000	0.000	0.000	0.075 (0.021)
	F	-0.044	-0.020	-0.038	-0.093	0.057	-0.087	-0.040	n/a	n/a	n/a	-0.038 (0.016)
EV712	*N*	47	50	26	41	72	69	39	38	48	92	52.2 (6.232)
	N_a	2	2	2	2	3	2	2	2	2	1	2.0 (0.149)
	H_O	0.532	0.440	0.269	0.512	0.431	0.449	0.615	0.395	0.021	0.000	0.366 (0.066)
	H_E	0.486	0.471	0.411	0.489	0.389	0.449	0.492	0.441	0.021	0.000	0.365 (0.060)
	F	-0.096	0.066	0.344	-0.047	-0.108	0.000	-0.251	0.106	-0.011	n/a	0.00 (0.052)
EV723	*N*	47	50	27	41	72	69	39	38	46	94	52.3 (6.349)
	N_a	1	1	2	2	2	2	3	2	2	2	1.9 (0.180)
	H_O	0.000	0.000	0.037	0.024	0.500	0.420	0.564	0.526	0.413	0.287	0.277 (0.075)
	H_E	0.000	0.000	0.036	0.024	0.497	0.496	0.509	0.465	0.375	0.289	0.269 (0.072)
	F	n/a	n/a	-0.019	-0.012	-0.007	0.153	-0.109	-0.131	-0.101	0.008	-0.027 (0.029)
EV740	*N*	47	50	27	41	72	69	38	37	48	94	52.3 (6.381)
	N_a	1	3	3	2	4	3	2	2	2	2	2.4 (0.267)
	H_O	0.000	0.080	0.185	0.171	0.111	0.159	0.105	0.270	0.479	0.404	0.197 (0.047)
	H_E	0.000	0.077	0.230	0.156	0.107	0.149	0.100	0.234	0.422	0.473	0.195 (0.048)
	F	n/a	-0.034	0.194	-0.093	-0.043	-0.068	-0.056	-0.156	-0.136	0.145	-0.027 (0.038)
EV769	*N*	47	50	27	41	71	69	39	38	48	94	52.4 (6.297)
	N_a	2	2	2	2	2	2	2	2	1	2	1.9 (0.100)
	H_O	0.447	0.620	0.296	0.268	0.268	0.290	0.410	0.237	0.000	0.011	0.285(0.059)
	H_E	0.486	0.476	0.417	0.299	0.252	0.268	0.355	0.209	0.000	0.011	0.277 (0.054)
	F	0.080	-0.303	0.289	0.102	-0.062	-0.082	-0.156	-0.134	n/a	-0.005	-0.030 (0.054)
EV862	*N*	47	50	26	41	72	69	38	37	46	91	51.7(6.229)
	N_a	2	4	2	2	2	2	2	2	2	2	2.2 (0.2)
	H_O	0.532	0.520	0.385	0.512	0.236	0.391	0.605	0.405	0.087	0.132	0.381 (0.056)
	H_E	0.486	0.562	0.355	0.424	0.249	0.364	0.497	0.407	0.122	0.123	0.359 (0.048)
	F	-0.096	0.075	-0.083	-0.208	0.052	-0.075	-0.218	0.003	0.287	-0.071	-0.033 (0.047)
EV911	*N*	46	50	26	41	71	69	39	38	48	94	52.2 (6.352)
	N_a	1	1	1	2	3	2	2	3	2	2	1.9 (0.233)
	H_O	0.000	0.000	0.000	0.024	0.113	0.087	0.333	0.132	0.083	0.160	0.093 (0.032)
	H_E	0.000	0.000	0.000	0.024	0.107	0.083	0.311	0.171	0.080	0.147	0.092 (0.031)
	F	n/a	n/a	n/a	-0.012	-0.053	-0.045	-0.073	0.232	-0.043	-0.087	-0.012 (0.035)
OMM5003	*N*	45	50	27	41	71	68	37	38	48	92	51.7 (6.193)
	N_a	7	7	6	8	9	8	5	4	3	4	6.1 (0.640)
	H_O	0.644	0.680	0.667	0.683	0.620	0.618	0.514	0.553	0.854	0.609	0.644 (0.029)
	H_E	0.656	0.664	0.692	0.705	0.548	0.572	0.474	0.577	0.655	0.555	0.610 (0.024)
	F	0.018	-0.025	0.037	0.031	-0.131	-0.080	-0.082	0.041	-0.303	-0.097	-0.059 (0.034)
OMM5007	*N*	46	50	27	41	72	69	38	38	48	90	51.9 (6.080)
	N_a	6	6	4	5	6	5	2	2	3	3	4.200 (0.512)
	H_O	0.500	0.400	0.519	0.439	0.514	0.536	0.500	0.579	0.354	0.344	0.469 (0.025)
	H_E	0.494	0.377	0.592	0.473	0.547	0.508	0.491	0.478	0.351	0.355	0.466 (0.026)
	F	-0.013	-0.060	0.124	0.071	0.060	-0.056	-0.018	-0.212	-0.010	0.030	-0.008 (0.029)

		Duncan		Kootenay		Okanagan		Wood		Tchesinkut		
Locus		Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Mean (SE)
OMM5008	*N*	47	50	26	41	67	69	37	38	48	92	51.5 (6.123)
	N_a	7	7	4	5	8	8	4	5	3	3	5.4 (0.618)
	H_O	0.617	0.560	0.615	0.488	0.537	0.565	0.676	0.684	0.146	0.109	0.50 (0.065)
	H_E	0.575	0.538	0.570	0.548	0.591	0.543	0.644	0.639	0.137	0.104	0.489 (0.063)
	F	-0.072	-0.040	-0.079	0.111	0.091	-0.040	-0.049	-0.072	-0.068	-0.043	-0.026 (0.022)
OMM5032	*N*	46	50	25	41	71	69	38	37	46	92	51.5 (6.32)
	N_a	7	7	5	7	5	5	4	4	1	1	4.6 (0.702)
	H_O	0.848	0.820	0.680	0.805	0.676	0.507	0.395	0.595	0.000	0.000	0.533 (0.099)
	H_E	0.764	0.763	0.590	0.705	0.645	0.589	0.441	0.640	0.000	0.000	0.514 (0.091)
	F	-0.110	-0.074	-0.153	-0.142	-0.048	0.138	0.105	0.071	n/a	n/a	-0.027 (0.036)
OMM5033	*N*	46	50	27	40	71	66	38	37	46	93	51.4 (6.236)
	N_a	13	19	8	12	21	14	10	10	10	9	12.6 (1.368)
	H_O	0.848	0.960	0.444	0.750	0.859	0.773	0.868	0.892	0.674	0.699	0.777(0.047)
	H_E	0.819	0.897	0.540	0.747	0.851	0.816	0.770	0.795	0.658	0.683	0.758 (0.033)
	F	-0.035	-0.070	0.178	-0.004	-0.009	0.053	-0.128	-0.122	-0.024	-0.023	-0.019 (0.028)
OMM5037	*N*	47	50	27	41	70	68	38	36	48	94	51.9 (6.313)
	N_a	4	3	2	3	6	6	3	5	1	1	3.4 (0.581)
	H_O	0.362	0.360	0.259	0.195	0.257	0.294	0.105	0.444	0.000	0.000	0.228 (0.048)
	H_E	0.303	0.302	0.226	0.178	0.232	0.265	0.101	0.378	0.000	0.000	0.198 (0.041)
	F	-0.193	-0.194	-0.149	-0.095	-0.110	-0.111	-0.045	-0.177	n/a	n/a	-0.134 (0.017)
OMM5053	*N*	45	50	26	40	69	67	37	38	46	92	51.0 (6.166)
	N_a	21	22	11	16	23	23	19	17	13	14	17.9 (1.378)
	H_O	0.867	0.960	0.885	0.850	0.928	0.881	0.892	0.842	0.783	0.739	0.863 (0.020)
	H_E	0.886	0.932	0.827	0.830	0.929	0.912	0.887	0.900	0.795	0.794	0.869 (0.017)
	F	0.022	-0.030	-0.070	-0.024	0.002	0.035	-0.005	0.064	0.015	0.070	0.008 (0.014)
OMM5058	*N*	46	50	27	39	70	68	35	37	45	91	50.8 (6.221)
	N_a	13	13	11	13	17	18	9	10	9	9	12.20(1.031)
	H_O	0.870	0.860	0.815	0.769	0.886	0.912	0.600	0.865	0.800	0.802	0.818 (0.028)
	H_E	0.873	0.878	0.835	0.831	0.852	0.855	0.760	0.845	0.810	0.761	0.830 (0.013)
	F	0.004	0.020	0.024	0.075	-0.040	-0.067	0.211	-0.024	0.012	-0.054	0.016 (0.025)
OMM5067	*N*	47	50	27	41	70	68	38	36	47	90	51.4 (6.030)
	N_a	4	3	2	2	3	3	4	5	3	3	3.2 (0.291)
	H_O	0.681	0.500	0.370	0.171	0.329	0.485	0.132	0.694	0.702	0.533	0.460 (0.066)
	H_E	0.589	0.569	0.417	0.195	0.427	0.482	0.126	0.559	0.652	0.657	0.467 (0.058)
	F	-0.156	0.122	0.112	0.126	0.231	-0.006	-0.047	-0.241	-0.077	0.188	0.25 (0.049)
OMM5075	*N*	45	50	27	41	72	69	38	37	46	92	51.7 (6.275)
	N_a	3	4	3	4	5	6	4	4	4	4	4.1 (0.277)
	H_O	0.267	0.220	0.741	0.585	0.278	0.348	0.158	0.216	0.304	0.391	0.351 (0.057)
	H_E	0.234	0.218	0.612	0.564	0.262	0.326	0.149	0.199	0.299	0.355	0.322 (0.048)
	F	-0.140	-0.008	-0.209	-0.038	-0.061	-0.068	-0.058	-0.086	-0.017	-0.104	-0.079 (0.019)
OMM5091	*N*	47	50	27	41	72	69	38	38	48	92	52.2 (6.211)
	N_a	3	6	3	3	3	4	2	2	2	1	2.9 (0.433)
	H_O	0.404	0.320	0.630	0.659	0.125	0.188	0.237	0.211	0.021	0.000	0.279 (0.072)
	H_E	0.365	0.400	0.628	0.647	0.119	0.199	0.209	0.188	0.021	0.000	0.278 (0.072)
	F	-0.107	0.201	-0.003	-0.017	-0.055	0.052	-0.134	-0.118	-0.011	n/a	-0.021 (0.033)
OMM5099	*N*	47	50	27	40	71	67	39	38	48	93	52.0 (6.188)
	N_a	5	4	3	3	6	8	3	4	1	1	3.8 (.680)
	H_O	0.511	0.460	0.667	0.600	0.507	0.597	0.692	0.500	0.000	0.000	0.453 (0.079)
	H_E	0.512	0.426	0.510	0.599	0.546	0.570	0.506	0.508	0.000	0.000	0.418 (0.071)
	F	0.002	-0.080	-0.306	-0.002	0.071	-0.047	-0.368	0.016	n/a	n/a	-0.089 (0.051)
OMM5108	*N*	47	50	26	41	71	68	38	38	47	92	51.8 (6.2)
	N_a	4	5	4	4	6	7	3	4	2	2	4.10 (0.504)
	H_O	0.277	0.480	0.500	0.537	0.451	0.471	0.579	0.579	0.021	0.022	0.392 (0.067)
	H_E	0.385	0.527	0.399	0.482	0.469	0.458	0.450	0.479	0.021	0.022	0.369 (0.059)
	F	0.282	0.090	-0.254	-0.114	0.039	-0.026	-0.287	-0.209	-0.011	-0.011	-0.05 (0.055)
OMM5121	*N*	47	49	27	41	71	65	37	38	47	92	51.4 (6.103)
	N_a	1	2	2	2	3	2	2	2	1	1	1.8 (0.20)
	H_O	0.000	0.020	0.296	0.512	0.620	0.462	0.486	0.447	0.000	0.000	0.284 (0.08)
	H_E	0.000	0.020	0.384	0.499	0.507	0.499	0.418	0.400	0.000	0.000	0.273 (0.074)
	F	n/a	-0.010	0.229	-0.027	-0.221	0.075	-0.164	-0.119	n/a	n/a	-0.034 (0.048)
OMM5124	*N*	47	50	27	40	72	69	39	38	46	92	52.0 (6.225)
	N_a	5	5	4	4	4	3	2	3	3	3	3.6 (0.306)
	H_O	0.830	0.680	0.519	0.675	0.250	0.232	0.385	0.316	0.652	0.609	0.515 (0.066)
	H_E	0.687	0.629	0.540	0.642	0.222	0.250	0.369	0.336	0.550	0.503	0.473 (0.053)
	F	-0.207	-0.082	0.041	-0.051	-0.126	0.072	-0.044	0.061	-0.185	-0.210	-0.073 (0.034)

		Duncan		Kootenay		Okanagan		Wood		Tchesinkut		
Locus		Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Shore	Stream	Mean (SE)
OMM5125	*N*	45	50	27	41	72	67	38	37	47	92	51.6 (6.208)
	N_a	4	3	2	3	3	3	3	3	2	2	2.8 (0.200)
	H_O	0.711	0.660	0.222	0.220	0.444	0.597	0.737	0.730	0.596	0.554	0.547 (0.061)
	H_E	0.656	0.664	0.198	0.239	0.477	0.588	0.608	0.665	0.500	0.499	0.509 (0.053)
	F	-0.084	0.006	-0.125	0.083	0.067	-0.015	-0.213	-0.097	-0.192	-0.110	-0.068 (0.032)
One102	*N*	47	50	27	41	71	67	39	38	46	89	51.5 (5.896)
	N_a	13	16	9	12	13	13	10	9	7	7	10.9 (936)
	H_O	0.936	0.920	0.778	0.878	0.873	0.821	0.923	0.921	0.761	0.640	0.845 (0.030)
	H_E	0.901	0.897	0.791	0.856	0.884	0.890	0.840	0.853	0.710	0.655	0.828 (0.027)
	F	-0.039	-0.026	0.016	-0.025	0.012	0.077	-0.099	-0.080	-0.072	0.022	-0.021 (0.017)
One105	*N*	47	50	27	41	70	69	39	37	48	89	51.7 (5.931)
	N_a	5	5	5	4	6	8	5	5	6	7	5.6 (0.371)
	H_O	0.574	0.580	0.481	0.537	0.600	0.652	0.615	0.622	0.646	0.652	0.596 (0.017)
	H_E	0.544	0.641	0.508	0.514	0.622	0.654	0.575	0.627	0.672	0.681	0.604 (0.020)
	F	-0.055	0.095	0.053	-0.045	0.036	0.004	-0.070	0.009	0.038	0.044	0.11 (0.017)
One108	*N*	47	50	26	40	69	65	36	38	47	89	50.70 (5.903)
	N_a	14	12	11	14	23	19	15	14	10	13	14.5 (1.222)
	H_O	0.851	0.880	0.846	0.800	0.884	0.908	0.889	0.868	0.809	0.876	0.861 (0.011)
	H_E	0.895	0.868	0.848	0.873	0.922	0.909	0.868	0.821	0.844	0.866	0.871 (0.010)
	F	0.049	-0.014	0.003	0.083	0.041	0.002	-0.024	-0.058	0.042	-0.012	0.011 (0.013)
One110	*N*	45	49	25	41	70	65	38	38	47	89	50.70 (5.935)
	N_a	13	14	13	17	18	16	13	14	8	9	13.5 (1.003)
	H_O	0.911	0.898	0.760	0.902	0.886	0.969	0.921	0.895	0.723	0.584	0.845 (0.037)
	H_E	0.848	0.874	0.751	0.843	0.904	0.911	0.858	0.839	0.658	0.647	0.813 (0.030)
	F	-0.074	-0.028	-0.012	-0.070	0.020	-0.064	-0.073	-0.066	-0.100	0.097	-0.037 (0.019)
One112	*N*	47	50	27	40	71	68	36	36	46	90	51.1 (6.145)
	N_a	21	18	14	15	21	20	18	13	8	11	15.9 (1.402)
	H_O	0.915	0.900	0.963	0.925	0.859	0.882	0.889	0.556	0.804	0.800	0.849 (0.036)
	H_E	0.923	0.911	0.856	0.895	0.888	0.897	0.900	0.869	0.787	0.837	0.876(0.013)
	F	0.009	0.012	-0.125	-0.033	0.033	0.016	0.012	0.361	-0.022	0.044	0.031 (0.040)
One14	*N*	47	50	25	41	72	68	38	37	47	89	51.4 (6.098)
	N_a	6	5	6	7	10	8	8	8	4	5	6.7 (0.578)
	H_O	0.383	0.480	0.560	0.561	0.542	0.456	0.711	0.730	0.277	0.404	0.510(0.045)
	H_E	0.391	0.485	0.574	0.597	0.497	0.480	0.729	0.733	0.282	0.363	0.513(0.047)
	F	0.021	0.009	0.025	0.061	-0.089	0.050	0.025	0.004	0.020	-0.113	0.001(0.018)
One8	*N*	45	50	27	40	72	69	39	37	48	93	52.0(6.326)
	N_a	6	7	3	6	9	9	7	6	1	1	5.5(0.922)
	H_O	0.422	0.540	0.222	0.425	0.514	0.580	0.718	0.486	0.000	0.000	0.391(0.076)
	H_E	0.400	0.488	0.201	0.419	0.514	0.561	0.600	0.515	0.000	0.000	0.370 (0.071)
	F	-0.056	-0.106	-0.106	-0.014	0.001	-0.033	-0.197	0.055	n/a	n/a	-0.057(0.025)
Ots06	*N*	47	50	26	41	72	68	37	38	48	93	52.0(6.325)
	N_a	1	1	1	1	2	2	2	1	1	1	1.3(0.153)
	H_O	0.000	0.000	0.000	0.000	0.014	0.088	0.432	0.000	0.000	0.000	0.053(0.043)
	H_E	0.000	0.000	0.000	0.000	0.014	0.084	0.339	0.000	0.000	0.000	0.044(0.034)
	F	n/a	n/a	n/a	n/a	-0.007	-0.046	-0.276	n/a	n/a	n/a	-0.110(0.046
Ots07	*N*	47	50	27	41	72	67	37	36	46	93	51.6(6.321)
	N_a	7	6	3	5	11	11	7	6	3	3	6.2(0.94)
	H_O	0.447	0.520	0.259	0.610	0.708	0.657	0.757	0.611	0.370	0.333	0.527(0.053)
	H_E	0.485	0.527	0.331	0.540	0.667	0.667	0.686	0.576	0.318	0.330	0.513(0.046)
	F	0.078	0.013	0.217	-0.130	-0.062	0.015	-0.103	-0.062	-0.161	-0.009	-0.020(0.035)
Ots14	*N*	46	50	27	41	71	68	37	37	47	91	51.5 (6.143)
	N_a	5	4	2	3	5	5	2	3	2	2	3.3(0.423)
	H_O	0.783	0.740	0.074	0.195	0.577	0.544	0.432	0.486	0.021	0.022	0.388(0.092)
	H_E	0.601	0.517	0.071	0.217	0.500	0.516	0.500	0.473	0.021	0.022	0.344(0.074)
	F	-0.301	-0.431	-0.038	0.100	-0.155	-0.055	0.135	-0.029	-0.011	-0.011	-0.080(0.055)
TAP2	*N*	47	49	27	41	72	69	38	38	47	93	52.1(6.295)
	N_a	2	2	2	2	2	2	2	2	2	2	2.0(0.0)
	H_O	0.553	0.469	0.593	0.537	0.514	0.406	0.526	0.553	0.511	0.419	0.508(0.019)
	H_E	0.494	0.493	0.444	0.485	0.500	0.470	0.465	0.497	0.482	0.378	0.471(0.012)
	F	-0.119	0.047	-0.333	-0.105	-0.028	0.136	-0.131	-0.112	-0.060	-0.110	-0.082(0.039)

* Multiple populations of the same ecotype were pooled for Okanagan Lake shore- and stream-spawners, Kootenay Lake stream-spawners, Duncan Lake shore-and stream-spawners, Tchesinkut stream-spawners.

Appendix B: Supplementary Material for Chapter 3

Table B.1 Estimated ecotype proportions from mixed-composition (MC) analyses conducted in ONCOR. True ecotype proportions in each mixture scenario were pre-defined. The residual for each scenario (estimated proportion of shore-spawners minus the actual proportion of shore spawners) and the absolute sum of the residuals were calculated for each lake to determine the accuracy of this MC estimates using the 'repeat-outlier' dataset.

Lake	Mixture Scenario	Ecotype	Actual Proportion	Estimated proportion	95% CI	Residual	Absolute sum of residuals
Duncan	90:10	Shore	0.90	0.650	(0.553, 0.752)	-0.249	
		Stream	0.10	0.349	(0.243, 0.444)		
	75:25	Shore	0.75	0.592	(0.492, 0.698)	-0.157	
		Stream	0.25	0.407	(0.289, 0.504)		
	50:50	Shore	0.50	0.503	(0.375, 0.621)	0.003	
		Stream	0.50	0.496	(0.376, 0.618)		
	25:75	Shore	0.25	0.404	(0.284, 0.500)	0.150	
		Stream	0.75	0.596	(0.493, 0.702)		
	10:90	Shore	0.10	0.357	(0.216, 0.449)	0.257	
		Stream	0.90	0.642	(0.539, 0.767)		
	0:100	Shore	0.0	0.321	(0.186, 0.430)	0.322	
		Stream	1.00	0.678	(0.561, 0.784)		
							1.143
Kootenay	90:10	Shore	0.90	0.672	(0.565, 0.754)	-0.228	
		Stream	0.10	0.327	(0.202, 0.433)		
	75:25	Shore	0.75	0.604	(0.485, 0.711)	-0.146	
		Stream	0.25	0.395	(0.285, 0.501)		
	50:50	Shore	0.50	0.476	(0.369, 0.589)	-0.023	
		Stream	0.50	0.523	(0.409, 0.623)		
	25:75	Shore	0.25	0.345	(0.249, 0.433)	0.095	
		Stream	0.75	0.654	(0.558, 0.743)		
	10:90	Shore	0.10	0.269	(0.169, 0.368)	0.169	
		Stream	0.90	0.730	(0.621, 0.822)		
	0:100	Shore	0.0	0.226	(0.131, 0.297)	0.227	
		Stream	1.00	0.773	(0.692, 0.858)		
							0.888
Okanagan	90:10	Shore	0.90	0.787	(0.679, 0.854)	-0.113	
		Stream	0.10	0.213	(0.139, 0.317)		
	75:25	Shore	0.75	0.685	(0.565, 0.792)	-0.065	
		Stream	0.25	0.315	(0.205, 0.423)		
	50:50	Shore	0.50	0.511	(0.418, 0.614)	0.011	
		Stream	0.50	0.490	(0.363, 0.576)		
	25:75	Shore	0.25	0.317	(0.223, 0.425)	0.067	
		Stream	0.75	0.683	(0.565, 0.768)		
	10:90	Shore	0.10	0.186	(0.094, 0.288)	0.086	
		Stream	0.90	0.814	(0.707, 0.902)		
	0:100	Shore	0.0	0.037	(0.001, 0.083)	0.037	
		Stream	1.00	0.963	(0.910, 0.999)		
							0.378

Lake	Mixture Scenario	Ecotype	Actual Proportion	Estimated proportion	95% CI	Residual	Absolute sum of residuals
Tchesinkut	90:10	Shore	0.90	0.834	(0.722, 0.916)	-0.066	
		Stream	0.10	0.166	(0.080, 0.268)		
	75:25	Shore	0.75	0.714	(0.590, 0.810)	-0.036	
		Stream	0.25	0.286	(0.166, 0.397)		
	50:50	Shore	0.50	0.513	(0.397, 0.638)	0.013	
		Stream	0.50	0.487	(0.357, 0.593)		
	25:75	Shore	0.25	0.287	(0.181, 0.382)	0.037	
		Stream	0.75	0.713	(0.596, 0.808)		
	10:90	Shore	0.10	0.156	(0.066, 0.275)	0.056	
		Stream	0.90	0.844	(0.714, 0.927)		
	0:100	Shore	0.0	0.067	(0.000, 0.138)	0.067	
		Stream	1.00	0.933	(0.848, 0.999)		
							0.275
Wood	90:10	Shore	0.90	0.878	(0.817, 0.932)	-0.022	
		Stream	0.10	0.122	(0.066, 0.179)		
	75:25	Shore	0.75	0.737	(0.686, 0.800)	-0.013	
		Stream	0.25	0.263	(0.190, 0.313)		
	50:50	Shore	0.50	0.508	(0.414, 0.585)	0.008	
		Stream	0.50	0.492	(0.413, 0.583)		
	25:75	Shore	0.25	0.258	(0.193, 0.320)	0.008	
		Stream	0.75	0.749	(0.675, 0.800)		
	10:90	Shore	0.10	0.116	0.071, 0.157)	0.016	
		Stream	0.90	0.884	(0.841, 0.925)		
	0:100	Shore	0.0	0.012	(0.000, 0.042)	0.012	
		Stream	1.00	0.988	(0.956, 1.000)		
							0.079