Genome-wide analysis to investigate patterns of ecotype divergence, population structure and life history changes in deep-spawning sockeye salmon, *Oncorhynchus nerka*

by

Farida Samad-zada

B.Sc. (Hon.), University of Toronto, 2018

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)

March 2021

The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis/dissertation entitled:

Genome-wide analysis to investigate patterns of ecotype divergence, population structure and life history changes in deep-spawning sockeye salmon, *Oncorhynchus nerka*

submitted by Farida Samad-zada in partial fulfillment of the requirements of

the degree of Master of Science.

Dr. Michael Russello, Department of Biology, Irving K. Barber Faculty of Science

Supervisor

Dr. Brett van Poorten, Resource and Environmental Management, Simon Fraser University

Supervisory Committee Member

Dr. Sepideh Pakpour, School of Engineering, Faculty of Applied Science

Supervisory Committee Member

Dr. Adam Wei, Department of Earth, Environmental, and Geographic Sciences, Irving K. Barber Faculty of Science

University Examiner

Abstract

Sockeye salmon (*Oncorhynchus nerka*) have been a classic study system for investigating ecotype divergence due to their tremendous life history variation. Multiple independent lineages of freshwater *O. nerka* (kokanee) have evolved from anadromous sockeye, and these migratory forms are further divided into reproductive ecotypes, depending on spawning location (shore-, stream- and deep-spawners). Deep-spawning *O. nerka* in Canada and Japan share unique phenotypic traits, but little is known about the origin and genomic basis of this ecotype. Here, we conducted genome-wide analyses of deep-spawning *O. nerka* on multiple scales, from regional populations in British Columbia, Canada, to those that span the pan Pacific distribution. First, we analyzed the Alouette Lake (British Columbia) *O. nerka* population, which (a) consists of both migrant and resident individuals, and (b) is the only known *O. nerka* population where migrants exhibit deep-spawning behaviour, leading to questions regarding the true ecotype of this population. To investigate the genomic basis of life history variation in this system, we collected SNP data (n = 7,709) for migrant and resident Alouette *O. nerka* (n = 163) and analyzed these samples relative to each other and *O. nerka* (n = 149) from known anadromous sockeye salmon and kokanee populations across the Fraser River drainage. Population structure analyses revealed five distinct clusters, primarily associated with geography, and no evidence for differentiation between resident and migrant Alouette *O. nerka* at the neutral loci. However, we identified eight high-confidence outlier loci divergent between migrant and resident Alouette *O. nerka* that were located on sex chromosomes, suggesting an association between migratory behaviour and sex in this system. We conclude that Alouette *O. nerka* likely represents a single stock best characterized as land-locked sockeye salmon, with individuals that retain the ability to migrate. Second, we genotyped deep- and stream-spawning kokanee (n = 167) from Canada and Japan at

9,721 SNPs, revealing a low number of shared ecotype-associated outliers between the two regions. Unlike in British Columbia, population clustering within Japan was best explained by translocation history. Taken together, these data suggest that evolution of the deep-spawning ecotype on two continents likely proceeded through different genetic pathways.

Lay Summary

The rare deep-spawning form of sockeye salmon (*Oncorhynchus nerka)* occurs in Canada and Japan, but little is known about its genomic basis. Here, we use genetic markers to (1) characterize the population structure of migratory and non-migratory deep-spawning *O. nerka* in Alouette Lake and (2) investigate the genomic basis of deep-spawning behaviour by comparing deep-spawning populations across the pan Pacific. Our results show that migratory and non-migratory Alouette *O. nerka* represent a single population but can be differentiated based on several sex chromosome-linked markers, suggesting an association between migratory behaviour and sex in this system. We also show that deep-spawning kokanee populations in Canada and Japan cluster based on geography and translocation history, respectively. We conclude that deep-spawning kokanee on the two continents represent distinct populations that likely evolved through different genetic pathways. This work contributes to our knowledge of *O. nerka* diversity and has direct implications for fisheries management.

Preface

Many individuals have contributed to this thesis.

For Chapter 2, Dr. Michael Russello and I were responsible for the study design, with contributions from Dr. Brett van Poorten. Samples were provided by the British Columbia Ministry of Environment and Climate Change Strategy. Shannon Harris, Heather Vainionpaa, Jennifer Sarchuk, Allison Hébert, and many others were responsible for sample collection. Additional samples were provided by Dr. Lyse Godbout (Department of Fisheries and Oceans Canada). I was responsible for sample processing, genetic data collection, data analysis and manuscript preparation. A manuscript based on Chapter 2 is currently in preparation, with Dr. Brett van Poorten, Shannon Harris, Dr. Lyse Godbout and Dr. Michael Russello as co-authors. Jennifer Sarchuk provided feedback for the initial draft of the manuscript. This manuscript will be submitted for review to *Evolutionary Applications*.

For Chapter 3, Dr. Michael Russello and I were responsible for the study design. Samples were provided by Dr. Tetsuji Nakabo and Dr. Kouji Nakayama at Kyoto University. I was responsible for sample processing, genetic data collection, data analysis and manuscript preparation. The manuscript is currently in draft.

Dr. Brett van Poorten and Dr. Sepideh Pakpour provided suggestions and feedback both for the thesis proposal, and the final version of this thesis.

Dr. Michael Russello provided guidance throughout the entire project, suggestions for genetic data collection and data analysis, and feedback for the proposal, manuscripts, and all versions of this thesis.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Michael Russello, for giving me the opportunity to work on this project, for creating a working environment that encouraged critical thinking and continuous learning, and also for his constant guidance and support. I would also like to thank my committee member, Dr. Brett van Poorten, for contributing expertise and providing feedback on many aspects of this project. I thank Shannon Harris for providing feedback and for answering my questions on the technical aspects of data collection. I am very grateful to Heather Vainionpaa, Jennifer Sarchuk and Allison Hébert, for helping me navigate the data sources and fisheries terminology. I would like to thank my committee member, Dr. Sepideh Pakpour, for her practical suggestions regarding the technical details of genetic data collection.

The sample collection was coordinated through the British Columbia Ministry of Environment and Climate Change Strategy and Alouette River Management Society and involved many people. I'm very lucky that I had a chance to participate and meet many of them in person. I'm very grateful to everyone who was involved in sample collection: Shannon Harris, Heather Vainionpaa, Jennifer Sarchuk, Allison Hébert, Kevin Gould, Marvin Rosenau, Dave Scott and many others. I'm also very thankful to Dr. Lyse Godbout for the provided samples and for constructive feedback on Chapter 2.

A special thank you goes to Barb Lucente for invaluable help with all the complex administrative procedures, and for always being very helpful and supportive.

Furthermore, I would like to thank all ECGL members, for being wonderful colleagues and friends. I am very fortunate that throughout the entire duration of my degree I was surrounded by such a supportive group of people. Thank you to Yoamel and Luke for helping me navigate the early stages of molecular work; Danielle, for making me feel welcome before I even arrived at UBCO; Bryson, for sharing bioinformatics tips and tricks; Chris, for our coffee runs and interesting discussions; Sarah, for our shared meals and random late-night conversations; and Brock, for always believing in me and being a friend I could always count on, be it for moral support or honest feedback: I am truly grateful for our friendship.

I would like to thank Sultana, for being a great friend and for making sure I had hobbies outside of work: our food adventures and aikido classes have been very crucial for my work-life balance. A big thank you to Dr. Karen Hodges for her advice and support. I also thank Dr. Brendan D'Souza and Dr. Bob Lalonde for their guidance and valuable scientific discussions.

Next, I would like to acknowledge the funding sources that made this work possible: funding for this project was provided through the Freshwater Fisheries Society of BC, The Natural Sciences and Engineering Research Council of Canada and the University of British Columbia.

Last but not least, I would like to thank my friends and family, for always being there for me, despite the distance and time difference. Thank you to my grandmother for her encouragement and wisdom, my sisters for always being only a phone call away and my parents for always supporting me. I'm also very grateful to my partner, for his constant encouragement, kindness and being my main source of motivation and support.

Chapter 1: Introduction

Freshwater habitats support more than 100,000 described species globally and provide crucial ecosystem services to humans (Dudgeon et al., 2006; Strayer & Dudgeon, 2010). Historically, human civilizations relied on freshwater systems for irrigation, transportation, sustenance, power production and more, and with exponential population growth, human dependence on freshwater habitats continues to increase (Meybeck, 2003; Strayer & Dudgeon, 2010). Consequently, such heavy exploitation resulted in freshwater habitats having the highest rates of biodiversity declines, compared to those documented in terrestrial and marine environments (Jenkins, 2003). Due to the high concentration of human activities around areas with freshwater resources, these ecosystems suffer from a range of anthropogenic stressors, including climate change, pollution, overexploitation and habitat fragmentation, all of which are detrimental for freshwater biodiversity (Dudgeon et al., 2006).

Freshwater fishes include all fish species that spend their entire life cycle or critical parts in streams or estuaries (Arthington, Dulvy, Gladstone, & Winfield, 2016). Fish constitute over half of all aquatic vertebrates and are fundamental in regulating food webs and nutrient cycling, with fish population collapses linked to inefficient nutrient cycling and changes in community composition (Holmlund & Hammer, 1999; Moore et al., 2011). Nonetheless, it is estimated that current extinction rate of freshwater fishes is much higher than the background extinction rate and it is only projected to increase over the next decades (Burkhead, 2012). Among the factors that contribute to fish declines, human-mediated landscape alterations are among the most detrimental. Anthropogenic activities often lead to fragmentation and habitat loss that in turn alter phenotypic and life history traits (Palkovacs, Kinnison, Correa, Dalton, & Hendry, 2012), change population structure (Whiteley et al., 2013) and reduce genetic diversity (Coleman et al.,

2018). In addition to demographic impacts, habitat fragmentation influences selection and adaptive potential of isolated populations. Following a disturbance event, previously connected populations can experience different selective pressures that over time may result in adaptive divergence (Cheptou, Hargreaves, Bonte, & Jacquemyn, 2017). Implementation of effective conservation practices often encompasses taking into account adaptative variation that might not be immediately obvious at the morphological and/or behavioural levels but requires availability of genetic data (Scribner et al., 2016).

1.1 Applications of Genotyping-by-Sequencing tools in fisheries management

The rapid development of next-generation sequencing technologies has greatly advanced the field of evolutionary biology. Traditional neutral markers have been widely used in population and conservation genetics for decades; however, they do not fully capture diversity of genomes and are not suited for identifying genes under selection (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013). In contrast, reduced representation genome sequencing has enabled researchers to look at a much larger proportion of the genome and identify variants of interest (Narum et al., 2013). Genotyping-by-Sequencing (GBS) is a cost-efficient approach to identify thousands of single nucleotide polymorphisms (SNPs) and genotype hundreds to thousands of individuals in a relatively short period of time. Restriction site associated DNA sequencing (RADseq) is a GBS tool used to decipher fine-scale population structure (Rodríguez-Ezpeleta et al., 2016), identify genes under divergent selection (Veale & Russello, 2017b), and address phylogenetic questions (Díaz-Arce, Arrizabalaga, Murua, Irigoien, & Rodríguez-Ezpeleta, 2016). The potential of GBS extends far beyond theoretical research and has direct implications for informing conservation planning.

Fisheries management has been an early adopter of GBS to address problems ranging from resource allocation to estimating impacts of climate change on fish populations (Valenzuela-Quiñonez, 2016). GBS has been successfully used for informing restocking initiatives in several hatcheries in the Pacific Northwest (Davis, Garza, & Banks, 2017), resolving population structure of fisheries stocks (Chen et al., 2020), and preparing targeted SNP panels aimed to differentiate between closely related populations (McKinney et al., 2020). Moreover, GBS can be applied to conduct genome scans to identify loci divergent between different stocks, which can be particularly useful for differentiating between populations that have only recently been isolated and for which neutral variation has not yet coalesced (Russello, Kirk, Frazer, & Askey, 2012).

1.2 Sockeye salmon *Oncorhynchus nerka*

1.2.1 Life history

Sockeye salmon (*Oncorhynchus nerka)*, within the family Salmonidae, are one of the five Pacific salmon species. The distribution of *O. nerka* spans from western North America to East Asia from Russia (Kamchatka) to Japan (Kaeriyama, 1996; Taylor, Foote, & Wood, 1996; Quinn, 2005). Sockeye salmon are comprised of two major forms: anadromous (migratory) and freshwater resident (known as kokanee).

Anadromous sockeye (hereafter referred to as "sockeye salmon") return to their natal freshwater streams to spawn after spending two to three years in the ocean, where they grow and accumulate most of their biomass. After hatching, juveniles spend some time in freshwater and then migrate to the ocean, and the cycle repeats (Quinn, 2005). Transition to the marine stage is accompanied by physiological changes, known as smolting, where dark-brown juvenile parr transform into silvery smolts in preparation for the salt-water environment. Smolting is a

metabolically costly process characterized by increased osmoregulatory ability, elevated growth rate and tendency to decrease aggression in favour of schooling behaviour (Folmar & Dickhoff, 1980).

In contrast, kokanee spend their entire life cycle in freshwater. Sympatric sockeye and kokanee populations can be differentiated from each other using genetic markers. The two ecotypes also differ in their habitat preferences, physiological traits, morphology, life history and behaviour (Foote, Wood, & Withler, 1989; Lemay & Russello, 2015; Taylor et al., 1996; Veale & Russello, 2017b). Mature sockeye salmon are larger than kokanee of the same age (Taylor et al., 1996), have a larger number of vertebrae, slimmer bodies and longer caudal regions, and demonstrate superior swimming performance and velocity (Taylor & Foote, 1991).

1.2.2 Migration: phenotype to genotype

Rather than representing a binary system, migration in sockeye salmon is better characterized as a gradient of behaviours (Quinn & Myers, 2004). For instance, the progeny of sockeye salmon that remain in freshwater are known as residuals, and the number of residuals within an aquatic system depends on environmental conditions, such as lake productivity (Ricker, 1938; Kaeriyama, 1996). Similarly, resident *O. nerka* that have been landlocked for multiple generations, can revert to anadromy when presented with an opportunity to migrate (Godbout et al., 2011; Samarasin, Shuter, & Rodd, 2017). The genetic basis of migratory behaviour in sockeye salmon is not yet well understood. Studies that aimed to identify divergent loci between resident and migrant populations conclude that majority of variation is geography-, rather than ecotype-specific (Nichols, Kozfkay, & Narum, 2016; Veale & Russello, 2017b). For *O. nerka*, certain loci associated with migration have been identified in more than one geographic location: examples include polymorphisms within different genes, such a heat-shock

gene HSP90a, an angiopoietin-4 gene (Veale & Russello, 2017b), and a growth differentiation factor-3 (Nichols et al., 2016) but no single overarching region responsible for migratory behaviour has been identified. On the other hand, studies investigating premature migration in steelhead (*O. mykiss)* and chinook salmon (*O. tshawytscha)* found that in both species, this phenotype is associated with polymorphisms within a single genomic region – GREB1L – that is highly conserved across vertebrates (Prince et al., 2017). Furthermore, transcriptome studies in *O. mykiss* demonstrated that genes involved in osmoregulation and phototransduction are differentially expressed between anadromous and resident populations (Aykanat, Thrower, & Heath, 2011; McKinney et al., 2015). The discovery of different candidate genes associated with anadromy indicate that this life history trait is likely a product of many genes with smaller effects distributed across several genomic regions (Hecht, Campbell, Holecek, & Narum, 2013). These studies have given a start to investigation of genomic basis of a complex life history phenotype, and combined with the availability of the recently published sockeye salmon genome (Christensen et al., 2020), provide a solid foundation for future research on the genetic basis of migratory behaviour in *O. nerka.*

1.2.3 Ecotype divergence

Multiple independent lineages of kokanee have evolved from sockeye salmon across the entire range of the species' pan Pacific distribution, following the last glaciation event (Taylor et al., 1996). Studies that used genetic markers to evaluate evolutionary origin of this ecotype showed that kokanee populations are polyphyletic and are more closely related to sympatric populations of sockeye salmon than to other isolated kokanee populations (Beacham & Withler, 2017; Taylor et al., 1996). The two migratory forms can be further subdivided into reproductive ecotypes, depending on spawning location (shore-, stream- and deep-spawning), which have

repeatedly evolved in different parts of the world (Moreira & Taylor, 2015; Taylor, Harvey, Pollard, & Volpe, 1997; Veale & Russello, 2017b).

Among these ecotypes, deep-spawning kokanee (also known as black kokanee or kunimasu) can be distinguished from the more common shore- and stream-spawning kokanee based on morphology, behaviour and pigmentation at maturity. While stream-spawners and, to a lesser extent, shore-spawners transition to bright or brown-red coloration upon reaching maturity, deep-spawning kokanee remain olive-black, even during spawning (Moreira & Taylor, 2015; Nakabo, Nakayama, Muto, & Miyazawa, 2011; Nakabo et al., 2014). An unusual spawning behaviour further characterizes this ecotype: spawning occurs more than 50 meters below lake surface in the lower part of the profundal zone and continues throughout the winter months. These unique traits are shared between geographically isolated deep-spawning kokanee populations from western Canada and Japan (Moreira & Taylor, 2015; Nakabo et al., 2011), leading to questions about the genomic basis of this reproductive form.

There also exists a single known instance of deep-spawning behaviour in migratory *O. nerka*, which has been documented in the Alouette River system (Fraser River Basin, British Columbia, Canada) (Hébert, 2019). Here, construction of a dam in 1928 effectively landlocked anadromous sockeye salmon until experimental water releases in 2005 and 2006 created conditions for migration (Baxter & Bocking, 2006), whereupon downstream migration of juvenile *O. nerka*, and, two years later, upstream migration of adult *O. nerka* has been documented (Balcke, 2009). Mitochondrial and nuclear microsatellite DNA analyses suggested that returning adult migrants originated in the reservoir and that juvenile migrant, adult migrant and resident "in-lake" individuals shared similar allele frequencies (Godbout et al., 2011). Despite the fact that experimental water releases have continued, not all Alouette Lake *O. nerka*

smolt and migrate to the ocean: many reach maturity and spawn without leaving the lake, and currently proportion of juveniles that smolt remains low. Van Poorten et al. (2018) suggested that Alouette population's pre-impoundment origin could be either sockeye salmon that are now landlocked by the dam, or kokanee that previously coexisted with sockeye salmon. Therefore, it remains unclear whether Alouette Lake is home to two populations (sockeye salmon that smolt and kokanee that do not) or whether the difference in life history represents variation within a single population. Determining whether the contemporary Alouette Lake *O. nerka* population is comprised of one or multiple ecotypes has implications for fisheries management, particularly related to appropriateness and ultimate success of the sockeye salmon restoration efforts.

1.3 Thesis overview and objectives

Ecotype variation of *O. nerka* is a known phenomenon that has been studied for at least several decades but relatively little is known about the genetic basis that underlies ecotype divergence and life history variation in this species. Alouette Lake *O. nerka* is a unique system, in which life history has been altered in response to human-mediated habitat modification and thus provides an ideal framework for studying the impact of artificial impoundments on population structure and life history on a recent timescale. Furthermore, Alouette Lake may represent the only known system, where migrant individuals exhibit deep-spawning behaviour. Fine-scale population analysis of Alouette Lake *O. nerka* adds to the growing literature on the subject of ecotype divergence of salmonids, and at the same time helps to address management questions that have direct conservation implications. Given this background, Chapter 2 uses genome-wide analysis to inform fisheries management by:

1. Investigating the genetic distinctiveness and demographic history of resident and migrant forms of Alouette Lake *O. nerka* relative to each other and to sockeye salmon and kokanee populations across the Fraser River drainage
2. Testing for evidence of adaptive population divergence between resident and migratory forms in the Alouette system to specifically investigate if there is a genetic basis to migratory behaviour

Furthermore, the genetic basis of deep-spawning behaviour in *O. nerka* is currently understudied. Chapter 3 aims to fill this knowledge gap by:

1. Reconstructing fine-scale population structure of deep-spawning kokanee in Canada and Japan
2. Conducting genomic scans to identify outlier loci between paired deep-spawning and stream-spawning populations in Canada and Japan to detect putative genetic signatures of parallel evolution
3. Investigating the level of hybridization between deep- and stream-spawning kokanee in Saiko Lake, one of the only two locations in the world where the two ecotypes have been documented to co-occur

Chapter 2: Genome-wide analysis reveals demographic and life history changes associated with habitat modification in a deep-spawning land-locked sockeye salmon (*Oncorhynchus nerka)* population

2.1 Background

Life history traits, genetic diversity and structure of wild populations are frequently influenced by anthropogenic stressors, such as human-induced landscape modifications, habitat loss and fragmentation (Almeida-Gomes & Rocha, 2015; Arantes, Fitzgerald, Hoeinghaus, & Winemiller, 2019; Boyle, Zartman, Spironello, & Smith, 2012; Haag et al., 2010; Roberts, Angermeier, & Hallerman, 2013). In freshwater ecosystems, water control structures such as dams can restrict spatial habitat connectivity leading to a broad range of consequences, both at the inter- and intra-specific levels (Cooke, Paukert, & Hogan, 2012). For instance, the loss of top predators due to river impediments can inhibit nutrient cycling between different habitats, as well as disrupt the trophic cascade within the lacustrine system (Mattocks, Hall, & Jordaan, 2017). In addition to ecological impacts, connectivity disruption can: 1) lower effective population sizes, increase inbreeding, decrease genetic diversity (Coleman et al., 2018) and cause genetic homogenization (Baggio, Araujo, Ayllón, & Boeger, 2018); 2) skew reproductive success (Maekawa & Koseki, 2001); 3) influence life history strategies (Morita, Yamamoto, & Hoshino, 2000); 4) alter population structure (Whiteley et al., 2013); 5) lead to local adaptation (Fraser, Debes, Bernatchez, & Hutchings, 2014); and 6) result in extirpation (Morita et al., 2019) and the loss of biodiversity (Liermann, Nilsson, Robertson, & Ng, 2012). Species that exhibit anadromy are dependent on upstream and downstream migrations to complete their life cycles and are therefore especially vulnerable (Junge, Museth, Hindar, Kraabøl, & Vøllestad, 2014).

The Pacific Northwest is home to many anadromous species, including several salmonids, among which sockeye salmon (*O. nerka*) is particularly notable for its life history variation. *O. nerka* is comprised of two major migratory forms: anadromous sockeye salmon (hereafter referred to as "sockeye salmon") and non-anadromous, resident kokanee (hereafter referred to as "kokanee"), which are further subdivided into ecologically divergent reproductive ecotypes (Taylor et al., 1996; Quinn, 2005). Due to their migratory lifestyle, sockeye salmon provide marine-derived nutrients to riparian ecosystems that are linked to increases in lake productivity and terrestrial vegetation (Chen et al., 2011; Gende, Edwards, Willson, & Wipfli, 2002; Quinn, Helfield, Austin, Hovel, & Bunn, 2018; Willson & Halupka, 1995). Moreover, Pacific salmon are deeply valued by the First Nations, as for thousands of years *Oncorhynchus spp*. have been a traditional source of sustenance and trade while serving important cultural and spiritual roles within the communities (Garner & Parfitt, 2006; Jacob, McDaniels, & Hinch, 2010). In addition, Pacific salmon constitute exceptionally valuable fisheries, contributing $4.8 billion annually in total economic output in the United States and Canada alone (Gislason, Lam, Knapp, & Guettabi, 2017). Despite the ecological, cultural and economic importance of *O. nerka*, the species has experienced significant declines, with many populations currently at risk of extirpation (Gustafson et al., 2007, Rand et al., 2012).

Alouette Lake, located in the lower Fraser River drainage in British Columbia, Canada (Figure 2.1), historically supported a population of anadromous sockeye salmon, but the construction of a dam in 1928 to divert water for hydroelectricity blocked passage to the ocean, functionally landlocking *O. nerka* in the newly formed reservoir (Foerster, 1930; Hirst, 1991). The last records of sockeye salmon date back to the 1930's and the population was first described as kokanee in 1951 (Godbout et al., 2011). Following detection of *O. nerka* juvenile

downstream migrants (JDM) in November 2005 during an intentional experimental water release over the spillway, an initiative to restore sockeye salmon in Alouette Lake was proposed (Baxter & Bocking, 2006). In 2007 and 2008, *O. nerka* adult upstream migrants (AUM) were discovered at the base of the Alouette Dam for the first time since initial extirpation (Balcke, 2009). Mitochondrial and nuclear microsatellite DNA analyses in combination with otolith microchemistry showed that returning adults were the progenies of resident *O. nerka* from Alouette Reservoir (Godbout et al., 2011). Additionally, the low diversity at nuclear microsatellites and the fixation of a single mitochondrial DNA haplotype suggested evidence for a recent population bottleneck (Godbout et al., 2011). A subsequent microsatellite-based study also indicated that Alouette Lake *O. nerka* underwent a recent reduction in effective population size in contrast to what was found for populations in neighbouring watersheds (Samarasin, Shuter, & Rodd, 2017). Interestingly, both resident and migratory individuals in the Alouette watershed are distinguished morphologically and behaviourally from typical *O. nerka* found elsewhere; they exhibit a characteristic black or dark olive colouration and occur at a depth of 10-105 m (34 m median depth) below lake surface (Hébert, 2019). Furthermore, resident *O. nerka* have been detected spawning at these depths; although no migrant *O. nerka* were observed in the process of spawning, detection of migrant individuals at the same depth during peak spawning period suggests that migrant *O. nerka* in Alouette Lake are likely deep-spawning as well (Hébert, 2019).

In 2016, the Fish and Wildlife Compensation Program (FWCP) identified sockeye salmon restoration in Alouette Lake to be of critical importance (Borick-Cunningham, 2018). However, one persistent challenge is the low proportion of juveniles that smolt. It remains unclear whether Alouette Lake is home to two populations (sockeye salmon that smolt and

kokanee that do not), or whether the life history difference represents variation within a single population. This uncertainty arises due to the lack of records on ecotype variation in Alouette Lake *O. nerka* prior to dam construction. Van Poorten et al. (2018) suggested that this population's pre-impoundment origin could be either sockeye salmon that are now landlocked by the dam, or kokanee that previously coexisted with sockeye salmon. Determining whether the contemporary Alouette Lake *O. nerka* population is comprised of one or multiple ecotypes has implications for fisheries management, particularly related to the appropriateness and ultimate success of the sockeye salmon restoration efforts.

To help fill the existing knowledge gaps, we used genotyping-by-sequencing (GBS) of in-lake, JDM and AUM individuals to investigate the genetic distinctiveness and demographic history of resident and migrant forms of Alouette Lake *O. nerka* relative to each other and to sockeye salmon and kokanee populations across the Fraser River drainage (Figure 2.1). In addition, we tested for evidence of adaptive population divergence between resident and migratory forms in the Alouette system to specifically investigate if there is a genetic basis to migratory behaviour. Together, these two objectives afforded broader insights regarding how artificial impoundments may shape evolutionary trajectory, life history traits and population structure of recently landlocked *O. nerka,* while providing information and tools for guiding fisheries management.

2.2 Methods

2.2.1 Study site

Alouette Lake (49.3337° N, 122.4181° W), located in British Columbia, Canada, is a small oligotrophic system [area: 16.6 km^2, maximum depth point: 152 m, dam present] that is

comprised of two connected basins, where the southern basin flows into the Alouette River (Plate, Mathews, & Bocking, 2014) (Figure 2.1). Construction of the Alouette Dam in 1928 at the mouth of the Alouette River isolated the basin, creating the reservoir and preventing salmonid migration. The reservoir has been subject to a nutrient restoration program beginning in 1999, resulting in substantial population growth of *O. nerka* (Harris et al., 2011; Scott, Harris, Hébert, & van Poorten, 2017; van Poorten, Harris, & Hébert, 2018; Vainionpaa, Sarchuk, Andrusak, & Harris, 2020).

2.2.2 Sample collection

Alouette deep-spawning resident individuals (n = 68) were sampled in September 2018 by the British Columbia Ministry of Environment and Climate Change Strategy (BC ENV) from 11 stations located across the reservoir. Fish were caught in September 2018 with overnight sets with either depth-stratified standard RISC gill nets (0, 10, 15, 20 m) in the pelagic habitat or with RISC gill net and experimental meshed gill nets in the nearshore habitat. Captured fish were measured [fork length (FL) and mass], sexed and assessed for life stage. Operculum punches were taken and stored in vials with 100% ethanol. In addition, Alouette River *O. nerka* AUM (n = 85) and JDM (n = 26) individuals were provided by the Department of Fisheries and Oceans that were originally sampled from 2009-2019 under contract with the Katzie First Nation and LGL Ltd. These samples consisted of a combination of muscle tissues, operculum tissues and fins preserved in 100% ethanol.

To allow for broader comparative analyses, we also obtained samples from other *O. nerka* populations of known ecotype from locations across the Fraser River drainage, including deep-spawning kokanee [East Barrière Lake (28.3 km^2 area, 100 m maximum depth), Seton Lake (24.3 km^2 area, 151 m maximum depth), Anderson Lake (28.3 km^2 area, 215 m maximum

depth)], stream-spawning kokanee [Nicola Lake (62.2 km^2 area, 55 m maximum depth)], and lake-type sockeye salmon (Scotch Creek, Portage Creek).

East Barrière Lake deep-spawning kokanee individuals (n = 31) were sampled by BC ENV in November 2019. Fish were caught in the pelagic zone, using either 2" mesh or RISC gill nets, and in the nearshore, using 2-2.5" mesh and RISC gill nets, starting with short sets that were later changed to overnight sets. Captured fish were measured (FL and weight), sexed and assessed for life stage. Operculum punches were taken and stored in vials with 100% ethanol.

Scotch Creek sockeye salmon individuals (n = 25) were sampled as deadpitch in September 2019 by the Little Shuswap Lake Band and provided by Marvin Rosenau (Fish Wildlife and Recreation, British Columbia Institute of Technology). Muscle tissue was obtained from carcasses and preserved in 100% ethanol.

Nicola Lake kokanee tissue samples (n = 25) were provided by Andy Morris (British Columbia Ministry of Forests, Lands, Natural Resource Operations and Rural Development). These samples were collected by trawl at the time of spawning in the Upper Nicola River in September-October 2012. The full sample distribution is further summarized in Figure 2.1 and Table A.1.

For Portage Creek sockeye salmon samples, and Anderson and Seton deep-spawning kokanee samples, we used previously published data from Veale & Russello (2017b).

2.2.3 Library preparation

Genomic DNA was extracted from operculum or muscle tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol with the addition of 4 µl of 100 mg/ml 7000U RNase (Qiagen) prior to ethanol precipitation. We used restriction site-associated DNA sequencing (RADseq) to simultaneously identify and genotype single nucleotide

polymorphisms (SNPs) within the processed *O. nerka* samples. Specifically, we employed a RADseq protocol following Baird et al. (2008) as modified in Lemay & Russello (2015) in order to ensure direct connectivity with a broader dataset generated by Veale & Russello (2017b). Overall, we constructed six libraries that included 260 unique individuals, in addition to 12 within library and seven between library replicates (Table A.1). Replicates were added to allow for estimation of genotyping error rate and potential batch effects (Tintle, Gordon, Van Bruggen, & Finch, 2009). Genomic DNA was digested using the *Sbf1* restriction enzyme and each individual in a library was assigned a unique 6 nucleotides long barcode. Shearing was performed using a Bioruptor® NGS (Diagenode). Sheared aliquots were cleaned using 1.5X Solid Phase Reversible Immobilization (SPRI) beads and then size-selected using a Pippin Prep™ (Sage Science) to retain fragments of approximately 500 base pairs. Libraries were PCR-amplified in parallel by repeating the reaction for 14 cycles. After the final clean up and size-selection, libraries were sent to the McGill University and Génome Québec Innovation Centre and sequenced using one lane each of Illumina HiSeq 2500 PE125 or Illumina HiSeq 4000 PE150 sequencing (six lanes total).

2.2.4 Genotyping and SNP ascertainment

We combined the newly generated raw sequence reads with those previously collected by Veale & Russello (2017b) for individuals from Anderson Lake (n = 22), Seton Lake (n = 23) and Portage Creek (n = 23) (Table 2.1). Raw paired-end reads were demultiplexed and trimmed to 100 bp via the *process radtags* command in STACKS v2.41 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011) to ensure consistency with previously sequenced reads from Veale & Russello (2017b). Identical reads generated due to PCR amplification were removed using the *clone filter* command in STACKS v2.41 (Catchen et al., 2011). Processed and filtered reads were

interleaved and aligned to a reference genome (*Oner_1*, GenBank Assembly Accession ID: GCA_006149115.1, (Christensen et al., 2020) using the *bwa mem* algorithm in BWA (Li & Durbin, 2009). The resulting bam files were sorted using samtools v1.9 (Li et al., 2009) and used to generate loci and call SNPs via the *gstacks* command in STACKS v2.41 (Catchen et al., 2011). Next, we processed the resulting loci through the *populations* module in STACKS v2.41 (Catchen et al., 2011), calculated mean coverage data per individual using VCFtools v0.1.15 (Danecek et al., 2011), and removed individuals with mean coverage lower than 6x. We then performed a sensitivity analysis on the retained individuals by running the *populations* module in STACKS v2.41 (Catchen et al., 2011) with a varying set of parameters to determine the optimal set for SNP ascertainment. Based on the sensitivity analysis (Table A.2), we only retained loci observed in 80% (r80) or more individuals within a population and present in all eight populations, with a minimum allele frequency of 0.05 and maximum observed heterozygosity of 0.50. Additionally, --*write-single-snp* flag was used to only retain one SNP per locus to decrease the effects of linkage disequilibrium *(--r 0.8, --p 8, --min-maf 0.05, --max-obs-het 0.5, --write-single-snp*). We further processed this dataset through VCFtools v0.1.15 (Danecek et al., 2011) to only include sites with a minimum mean depth of 10x and a maximum mean depth of 100x and exclude sites with more than 10% missing data (--*max-missing 0.9*). Putative paralogs were identified using *HDplot* function available from https://github.com/gjmckinney/HDplot (McKinney, Waples, Seeb, & Seeb, 2017) and removed using VCFtools v0.1.15 (Danecek et al., 2011). To control for potential batch effects, we estimated genotyping error rates using a custom python script and then removed replicate individuals from the dataset.

2.2.5 Outlier detection

Given the high false-positive rates associated with outlier detection approaches and the hierarchical population structure of our dataset, we employed three different analyses, including the Fst-based approaches implemented in Arlequin v3.5 (Excoffier & Lischer, 2010) and Bayescan v2.0 (Foll & Gaggiotti, 2008), and the principal component analysis (PCA)-based approach implemented in *pcadapt* (Luu, Bazin, & Blum, 2017). For Arlequin, we used the hierarchical island model (Slatkin & Voelm, 1991) that allows a higher migration rate between populations within a group than between groups. We performed 20,000 simulations, with the number of simulated demes set to 100, and the number of simulated groups set to 10, and simulated derived allele frequency set to 0.05. We considered all loci with *p*-values < 0.01 in the first and last quantile as candidate loci under divergent or balancing selection, respectively. For BayeScan v2.0 (Foll & Gaggiotti, 2008), we used a pairwise approach comparing allele frequencies between all possible pairs of populations, resulting in 28 pairwise comparisons. Additionally, we used this method to detect outliers between combined ecotype datasets using the following comparisons (1) all sockeye salmon populations (Scotch Creek, Portage Creek) versus all kokanee populations (Anderson, Seton, Nicola, East Barrière) and (2) all deep-spawning kokanee (Anderson, Seton, East Barrière) versus stream-spawning kokanee (Nicola). Alouette resident and migrant individuals were excluded from the ecotype comparisons given the uncertainty regarding the true ecotype of this population. The analyses were run for 100,000 iterations with 50,000 burn-in period with Prior Odds set to 10, and loci with q-value less than 0.05 were marked as outliers and were removed from the dataset. Lastly, we inferred genetic clusters through analyses of principal components (PCs) using *pcadapt* v4.1.0 (Luu et al., 2017) and Cuttel's rule to infer the most likely number of PCs that explain the genetic structure within

the dataset. The resulting *p*-values were corrected for multiple comparisons using the method of Benjamini-Yekutieli (2001), and loci with adjusted *p*-values of less than 0.05 were deemed outliers. We functionally annotated loci that were found in the comparisons of sockeye salmon versus kokanee, as well as deep-spawning versus stream-spawning kokanee. Specifically, we used the *blastn* function in BLAST v2.9.0 (Altschul, Gish, Miller, Myers, & Lipman, 1990) to compare the locus sequences including outlier SNPs to the *nr* database within the *Oncorhynchus* taxon, accepting hits with an e-value lower than $1e^{-28}$ and keeping the hits with the lowest e-value.

We also conducted outlier detection directly for the Alouette migrant versus resident individuals using BayeScan v2.0 (Foll & Gaggiotti, 2008) and *pcadapt* v4.1.0 (Luu et al., 2017), and the same parameters as for the full dataset (see above). In addition, we conducted a GWAS analysis to investigate the relationship between SNPs and phenotype (i.e. resident or migrant) using the Mixed Linear Model implemented in the R program GAPIT v3.0 (Lipka et al., 2012). This analysis was conducted on a reduced dataset of 6,775 SNPs that successfully mapped to linkage groups (*Oner_1*, GenBank Assembly Accession ID: GCA_006149115.1, Christensen et al., 2020), rather than to unplaced scaffolds ("UN"). Relatedness between pairs of individuals was accounted for by calculating a kinship matrix (VanRaden, 2008), however, the number of PCs was set to 0 given the absence of population structure in Alouette Lake *O. nerka* based on STRUCTURE and PCA analyses (see Figure 2.2). The FDR-corrected threshold was set to 0.05, and all SNPs below that threshold were considered significant.

To explore the relationship between population structure and sex within the Alouette system, we again implemented BayeScan v2.0 (Foll & Gaggiotti, 2008) using the same parameters as above to identify outliers that were divergent between male and female *O. nerka.*

This analysis was completed only for Alouette resident and East Barrière resident individuals, as these were the only populations for which sex data were available. Given that our original sample of Alouette resident individuals contained a disproportionately higher number of males than females, we randomly removed 15 males to ensure a 1:1 sex ratio (23 males/23 females); similarly, we randomly removed 7 males from the East Barrière dataset (12 males/12 females). We mapped and functionally annotated all outliers that were detected by both BayeScan (Foll & Gaggiotti, 2008) and *pcadapt* (Luu et al., 2017) using *blastn* (Altschul et al., 1990) and the same parameters as above.

2.2.6 Population genetics analyses

To construct a putatively neutral dataset, we removed any locus identified as an outlier in any comparison. Following outlier removal, we removed loci that significantly (-h 0.05) deviated from Hardy-Weinberg Equilibrium in 50% or more of the populations using the *filter_hwe_by_pop.pl* script available from https://github.com/jpuritz/dDocent/tree/master/scripts. Using the resulting putatively neutral SNP dataset, we calculated inbreeding coefficients (F_{IS}), observed (H_o) and expected (H_e) heterozygosities per locus following Nei (1987) and averaged across loci for each population using the *basic.stats* command within the R package *hierfstat* v0.04-22 (Goudet & Jombart, 2015). We also estimated effective population sizes (N_e) for each population using the Linkage Disequilibrium method (Waples & Do, 2008) as implemented in NeEstimator v.2 (Do et al., 2014), with the minimum allele frequency set to 0.05. To calculate levels of population differentiation, we calculated Weir & Cockerham's (1984) θ between all pairs of populations using 1000 permutations in Genetix v.4.05.2 (Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme,

2004). We visualized obtained pairwise θ values via NeighbourNetwork (Bryant & Moulton, 2004) reconstruction using SPLITSTREE v4.0 (Huson & Bryant, 2006).

To evaluate the number of genetic clusters present in our dataset, we used the Bayesian method of Pritchard et al. (2000) as implemented in STRUCTURE v3.4 (Pritchard, Stephens, & Donnelly, 2000). Run length was set to 100,000 Markov chain Monte Carlo replicates after a burn-in period of 100,000 using correlated allele frequencies under an admixture model using the LOCPRIOR option. We varied the number of clusters (K) from one to ten, with ten iterations of each. The resulting output was then summarized using STRUCTURE HARVESTER (Earl & vonHoldt, 2012). To infer the optimal K value, we employed a combination of the ΔK method (Evanno, Regnaut, & Goudet, 2005) and the plotting of the log probability of the data (Pritchard et al., 2000) to assess where ln Pr(X|K) plateaued (see STRUCTURE manual) and then used CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) to visualize the results. Additionally, we conducted a PCA analysis to visualize the relationships among populations using *SNPRelate* v1.14.0 (Zheng et al., 2012).

2.3 Results

2.3.1 Dataset quality

After demultiplexing, trimming and quality filtering, we retained a mean of 7,201,813.24 reads per sample. Across samples, an average of 94.22% of reads was successfully mapped to the reference genome. After full filtering, 7,709 SNPs were retained for 312 individuals (17 individuals were removed due to insufficient coverage < 6x), with a mean depth of 25.27x and mean missing percentage of 3.84%. Mean within- and among-library genotyping error rates were 5.46% and 3.99%, respectively.

2.3.2 Outlier loci detection, mapping and annotation

For the Fraser River drainage-wide analyses, Arlequin identified 154 high F_{ST} outliers and 118 low F_{ST} outliers. BayeScan detected 253 outliers with a *q*-value lower than 0.05 across all pairwise population comparisons. The *pcadapt* analysis identified 473 loci with a *p*-value lower than 0.05 after the Benjamini-Yekutieli correction. The first and second principal components showed six distinct clusters, largely associated with geography, where PC1 separated the two Alouette populations from the rest of the lakes, and PC2 divided the populations into four clusters. Of the identified outliers, 54 loci were detected in common by all three methods. Outlier detection between sockeye salmon-kokanee ecotypes resulted in 14 outliers, while outlier detection between deep-spawning and stream-spawning kokanee resulted in four outliers. Mapping to the *O. nerka* genome assembly showed that all outlier loci, including the 14 loci divergent between sockeye salmon and kokanee ecotypes were distributed across different linkage groups. Of the 14 sockeye salmon-kokanee outlier loci, 12 produced significant annotations (Table 2.2), six of which overlapped with those found in previous studies (Veale & Russello, 2017b, Nichols, Kozfkay, & Narum, 2016).

For Alouette Lake specifically, pairwise BayeScan analysis between migratory and resident individuals detected nine outliers, whereas *pcadapt* analysis identified 48 outliers associated with PC2; eight outliers were found in common between the two analyses. All detected outliers mapped to Linkage Groups 9a and 9b (Figure 2.3). GWAS did not detect any outliers after the FDR correction; however, the Manhattan plot generated by this analysis showed that SNPs distributed across Linkage Groups 9a and 9b had higher log values (Figure A.1). Of the eight outliers detected by both methods, four produced significant annotations (Table 2.3),

but none overlapped with the 14 detected in the basin-wide sockeye salmon-kokanee outlier analysis.

In addition, 21 outliers were detected between male and female *O. nerka* within the Alouette resident dataset, however, no sex-specific outlier SNPs were detected in East Barrière Lake. All 21 sex-specific outliers detected in Alouette Lake mapped to Linkage Groups 9a and 9b, seven of which overlapped with those detected in the Alouette-specific migrant-resident outlier scan. Of the 21 sex-specific outliers, 16 produced significant annotations (Table 2.3).

Within the Alouette population, migration- and sex-associated loci exhibited uneven distributions of genotypes across the samples, with resident females and migrant individuals exhibiting similar frequencies to each other relative to resident males (Figure 2.4). Moreover, resident females and migrant individuals exhibited clear heterozygote deficits (H_o = 0.00-0.21; migratory mean H_o = 0.17, resident female mean H_o = 0.05) across all seven common outliers detected in the migrant-resident and sex-specific scans, unlike the levels detected in resident males (H_o = 0.83-0.97; mean H_o = 0.93).

2.3.3 Population genetics

We removed all SNPs that were identified as outliers (n = 696) by any of the three above-mentioned analyses. One additional locus was found to deviate from HWE in more than 50% of the populations and was also removed. Based on this putatively neutral dataset of 7,012 SNPs, H_o, H_e, and F_{IS} values were similar across the eight populations, with both H_o and H_e, ranging from 0.20 to 0.27 (Table 2.1). None of the F_{IS} values significantly differed from 0 across all eight populations (Table 2.1). Nicola Lake kokanee had the highest N_e [2962.4 (2011.6-5608.2)], while both migrant and resident Alouette populations had the lowest N_e [794.6 (778.4-811.4), 564.1(550.9-578.0), respectively] (Table 2.1).

The STRUCTURE analysis revealed evidence for five clusters that best explained the genetic variation within our dataset, largely conforming to geography (Figure 2.2b; Figure A.2; Table A.3). Alouette Lake was identified as a distinct cluster starting from $K = 2$, with both resident and migrant individuals belonging to the same cluster, even with increasing values of *K*. East Barrière kokanee separated from the remaining populations at $K = 3$ and Nicola Lake separated at $K = 4$. At $K = 5$, Anderson and Seton deep-spawning kokanee formed a cluster, while Portage Creek and Scotch Creek sockeye salmon formed a separate cluster. The Portage and Scotch Creek sockeye salmon did not separate into distinct clusters at any higher values of *K*, despite the large geographic distance between the two localities. None of the populations demonstrated any evidence for further sub-structure and no admixture was detected between the populations.

The PCA on the neutral dataset also demonstrated evidence for five clusters, with Alouette migrant and resident individuals belonging to the same cluster, regardless of which eigenvectors were plotted (Figure 2.2a). Eigenvector 1 explained 37% of the variation and separated Alouette migrant and resident individuals from the rest of the populations; eigenvector 2 explained 12.2% of the variation and separated East Barrière. Similar to the STRUCTURE analysis, Anderson Lake and Seton Lake deep-spawning kokanee populations clustered together regardless of the eigenvectors used. Portage Creek and Scotch Creek sockeye salmon clustered close together when eigenvectors 1 and 2 were plotted; however, they formed two distinct clusters when this was extended to eigenvectors 3 and 4 (Figure 2.2a).

The phylogenetic network based on the neutral dataset did not show any clear separation by ecotype but provided further evidence for geographic differentiation (Figure 2.5a). As in the STRUCTURE and PCA analysis, Alouette Lake migrant and resident individuals clustered

together, as did Portage Creek and Scotch Creek sockeye salmon (Figure 2.5a). Kokanee populations that were not geographically close to a sampled sockeye salmon population (Nicola Lake stream-spawning and East Barrière Lake deep-spawning kokanee) were more isolated, as indicated by longer branch lengths (Figure 2.5a).

2.4 Discussion

Our genome-wide analyses provide clear evidence that Alouette Lake resident and migrant individuals are genetically distinct from other *O. nerka* populations in the Fraser River drainage included in this study but are not differentiated from each other at neutral loci, likely constituting a single population. Although our findings cannot unequivocally assign Alouette Lake *O. nerka* to a specific migratory ecotype (i.e. anadromous or resident) in the absence of the dam, identified outlier loci revealed a potential genetic basis underlying migratory behaviour, which may involve interactions with sex-associated loci. These results provide interesting insights into the genetic basis of *O. nerka* life history variation, in general, and have direct implications for informing on-going fisheries management in the Alouette watershed.

2.4.1 Geographic differentiation

Our results demonstrated that population structure of sockeye salmon and kokanee across the Fraser River Basin was largely associated with geography rather than ecotype, consistent with previous findings in this system (Beacham & Withler, 2017; Veale & Russello, 2017b). For example, East Barrière Lake deep-spawning kokanee and Nicola Lake stream-spawning kokanee each formed distinct clusters, likely due to the geographic and temporal isolation of these resident populations (Wood, Bickham, Nelson, Foote, & Patton, 2008). Our results align with previous findings that demonstrated that within the drainage, kokanee inhabiting different lakes are genetically isolated from each other and more closely related to sympatric sockeye salmon

populations if present (Beacham & Withler, 2017). Here, the only group within which sockeye salmon and kokanee populations were in direct geographic proximity was Anderson Lake and Seton Lake that are connected by Portage Creek. In this system, deep-spawning kokanee from Anderson Lake and Seton Lake grouped together and were most closely related to Portage Creek sockeye salmon, consistent with previous studies (Moreira & Taylor, 2015; Veale & Russello, 2017b). Moreover, Portage Creek and Scotch Creek sockeye salmon displayed high genetic affinity despite being located more than 300 km away from each other, exhibiting no evidence of pairwise differentiation (Figure 2.5b), while forming a single STRUCTURE cluster (Figure 2.2b) and largely overlapping PCA clusters (Figure 2.2a). The lack of differentiation at the neutral data may be explained by Portage Creek hatchery supplementation from the Lower Adams River that aimed to restore the declining sockeye salmon population in the 1950's (Withler, Le, Nelson, Miller, & Beacham, 2000). Overall, these patterns were consistent with previous findings by Wood et al. (2008), which showed that within drainages, among population differentiation was lower for lake-type sockeye salmon than for kokanee.

Alouette Lake *O. nerka,* in particular, showed clear separation from the rest of the populations in this study, exhibiting the highest and lowest levels of differentiation from Nicola Lake kokanee and Portage Creek sockeye salmon, respectively (Figure 2.5a). Yet, while Portage Creek and Scotch Creek sockeye salmon showed little to no evidence of genome-wide differentiation (Figures 2.2 and 2.5), Alouette Lake *O. nerka* was still significantly differentiated from each. These reconstructed patterns are not necessarily indicative of the *O. nerka* ecotype in Alouette Lake in the absence of the dam, however, as the substantial pairwise genetic differentiation relative to all other populations in the study could be due to extreme drift (Perrier, Bourret, Kent, & Bernatchez, 2013) given the genetic bottleneck that has been previously

reported in this system (Godbout et al., 2011; Samarasin et al., 2017). The role of drift may be further evidenced by the lower N_e in Alouette Lake compared to all other populations in this study (Table 2.1), and also consistent with the low census adult population size (N_c = ~20, 000) reported prior to the start of fertilization program in 1999 that has subsequently increased to ~ 200,000 individuals in 2019 (Sarchuk, Harris, & Johner., in prep).

At a finer-level, within-Alouette Lake analyses of migratory and resident individuals revealed no evidence for differentiation based on neutral genome-wide data (Figures 2.2 and 2.5), consistent with a single population. These results agree with those from a previous microsatellite study which revealed that Alouette Lake individuals formed a single genetic cluster, regardless of their migratory tendencies (Godbout et al., 2011). Understanding whether migratory behaviour of Alouette individuals has an underlying genetic basis, however, cannot be deciphered using neutral data alone.

2.4.2 Outlier loci and ecotype identification

Identifying genetic mechanisms responsible for parallel evolution can help our understanding of repeatability of evolution, as well as molecular processes that shape phenotypic variation, local adaptation and life history traits (Lee, Gould, & Stinchcombe, 2014). Outlier locus detection is frequently used for investigating molecular drivers behind parallel phenotypic divergence (Deagle et al., 2012; Perrier et al., 2013; Westram et al., 2014). Moreover, outlier loci can be useful for differentiating between populations that have only recently been isolated and for which neutral variation has not yet coalesced (Russello et al., 2012). Comparison of allele frequencies between all sockeye salmon and kokanee populations within our dataset revealed 14 loci that were significantly differentiated between the two ecotypes. Functional annotations were available for 12 of these SNPs (Table 2.2), however, we limit our discussion to six robust loci

that were also identified as candidates under divergent selection in previous population genomic studies of sockeye salmon and kokanee (Nichols et al., 2016; Veale & Russello, 2017b), or have been detected in multiple comparisons in this study.

SZNR01010580.1_848156 maps to the leucine-rich repeat-containing protein 9 (LRRC9) gene, at which specific genotypes have been previously found to be associated with spawning location (GG = shore-/beach-spawning; TT/GT = stream-/river-spawning) in both migrant and resident *O. nerka* across the entire distribution (Veale & Russello, 2017a, 2017b). Genotypes at this locus were entirely consistent with the previously known reproductive ecotypes, including shore/deep-spawning kokanee (East Barrière Lake: G allele frequency = 1.00; Anderson Lake: G allele frequency = 0.95; Seton Lake: G allele frequency = 1.00), stream-spawning kokanee (Nicola Lake: G allele frequency = 0.22); and stream-spawning sockeye salmon (Portage Creek: G allele frequency = 0.17; Scotch Creek: G allele frequency = 0.06). In Atlantic salmon, LRRC9 is located 142 kb away from the *six6* gene that exhibits signature of divergent selection with respect to spawning ecotypes and has been associated with age at maturity (Barson et al., 2015) and marine diet specialization (Aykanat et al., 2020).

SZNR01024871.1_9385 is located in the *O. nerka* heat shock protein HSP 90-alpha gene and was previously identified as an outlier in sockeye salmon-kokanee comparisons in the Okanagan and Anderson-Seton-Portage systems (Veale & Russello, 2017b). Heat shock proteins are molecular chaperones that assist protein folding and stabilization to help cells combat thermal stress; HSP90, in particular, is a highly interactive protein, involved in numerous molecular pathways (Saibil, 2013). Transcriptomic studies found that expression of HSP90 was increased in the gills of chinook salmon, *O. tshawytscha*, in response to increased water temperatures (Tomalty et al., 2015). More broadly, the debilitating effect that increasing water

temperature can have on migratory salmon (Crossin et al., 2008) has been associated with changing the expression of HSP90 at various periods of migration (Miller et al., 2009).

Two other outliers were annotated to regions likely associated with diet. SZNR01007172.1_549647 is located in coho salmon, *O. kisutch*, stearoyl-CoA desaturase 5 gene (SCD-5). However, this gene annotation was predicted computationally and taken together with studies that show that SCD-5 has been lost in teleost fishes, it is possible that this SNP is located in another SCD gene (Castro et al., 2011). The exact function of SCD genes in fish is unknown, but previous research found an influence of dietary intake on the expression of some SCD genes, which might differ based on the availability of dietary fatty acids in the fish rearing habitat (Castro et al., 2011). Likewise, SZNR01007191.1_179523 annotated to *O. nerka* partitioning defective 6 homolog alpha-like mRNA (*par-6*); *par-6* homolog expression has been demonstrated to change in Atlantic salmon liver following a dietary switch (Leaver et al., 2008).

The last two outliers are found in genes that have been related to transition from the marine environment to freshwater, and tissue regeneration. SZNR01004638.1_505442 is located in *O. tshawytscha* follicle stimulating hormone beta subunit (FSHbeta) gene. The FSH hormone belongs to the Glycoprotein Hormone Family (GPH), and expression of FSHbeta changes upon transition to freshwater in adult chum salmon (*O. keta*; Kim et al., 2013). In addition, SZNR01002048.1_136609, which mapped to *O. kisutch* proteoglycan 4, has been associated with wound healing (Hirose, Narita, Asano, & Nakane, 2018).

We also specifically examined the genotypes of Alouette Lake migratory and resident *O. nerka* individuals at the 14 sockeye salmon-kokanee outliers to further investigate specific associations with different migratory and reproductive behaviours. Overall, migratory and resident individuals demonstrated similar allele frequencies across the 14 loci, further consistent

with the hypothesis that Alouette Lake *O. nerka* comprise a single population. Of particular note, all resident and migrant individuals were fixed for the GG genotype at SZNR01010580.1_848156 (LRRC9), diagnostic of shore spawners (Veale & Russello, 2017a). Interestingly, Hirst (1991) indicated that Gold Creek, a tributary of Alouette Lake (Figure 2.1), was the original spawning location for returning sockeye salmon, which was still accessible after the dam was constructed. However, dams often alter not only accessibility to the spawning grounds but water temperature, food web dynamics and quality of available habitat (Angilletta et al., 2008; Sheer & Steel, 2006). Human-mediated changes to the Alouette watershed may have acted as a selection pressure or promoted plasticity in *O. nerka* spawning location, potentially driving a life-history shift to deep-spawning along the shoreline. In addition to spawning location, genotypes at SZNR01024871.1_9385 (HSP90) revealed all Alouette Lake individuals possessed the "G" allele, which has previously been reported for more than 95% of sockeye salmon distributed across multiple catchments (Columbia, Fraser) within British Columbia (Veale & Russello, 2017b).

Overall, the lack of genetic distinctiveness between migrant and resident *O. nerka* in Alouette Lake at genome-wide neutral loci, together with genotyping information at outlier loci, suggest that Alouette Lake *O. nerka* represent a recently landlocked sockeye salmon population, as previously proposed (Godbout et al., 2011, Samarasin et al., 2017). This finding is significant, as it identifies Alouette Lake as the only known location where anadromous sockeye salmon may exhibit deep-spawning behaviour. Moreover, many of the characteristics exhibited by Alouette Lake *O. nerka* are shared with residualized (i.e. resident progeny of anadromous parents) and partially migrating salmonids, particularly as they relate to sex ratios. First, Ricker (1938) documented the sex ratio of residuals as heavily skewed towards males relative to those observed

in co-occurring kokanee. A similar pattern was observed in Alouette Lake, where mature residents collected in 2018 were predominantly male (150M:29F). However, the unequal sex ratio could also be due to the timing of sampling, as male *O. nerka* in Alouette Lake tend to inhabit littoral regions prior to females (Hébert, 2019). Although not as common in *O. nerka*, skewed sex ratios are frequently reported in facultatively anadromous salmonid species, such as *O. mykiss*, where anadromy has been demonstrated to be maternally linked (Berejikian, Bush, & Campbell, 2014). Similarly, in brown trout (*Salmo trutta*), resident populations tend to be predominantly male and anadromous predominantly female; however, populations that do not have access to migration due to natural impediments have a more equal sex ratio (reviewed in Ferguson, Reed, McGinnity, & Prodöhl, 2017).

Other characteristics shared by Alouette Lake *O. nerka* and residualized sockeye salmon are associated with spawning, including morphology and behaviour. For example, Ricker (1938) described Cultus Lake residuals as exhibiting a dark olive/black colouration during the spawning period, similar to that observed in both Alouette Lake resident and returning migratory adults. Cultus Lake residuals also used the same redds as returning sockeye salmon during the spawning period (Ricker, 1938). Likewise, in Alouette Lake, telemetry of returning AUM and targeted netting of residents suggests both resident and migratory individuals spawn at the same depth (Hébert, 2019).

2.4.3 Migratory behaviour and sex-associated outlier loci

We identified eight high confidence outlier loci between resident and migratory individuals in Alouette Lake, all of which mapped to Linkage Groups 9a and 9b that correspond to sockeye salmon sex chromosomes (Christensen et al., 2020). Unfortunately, sex information was not available for the migratory individuals genotyped in this study. However, outlier locus

detection between Alouette Lake resident males and females revealed 21 outliers (including seven of the eight detected between migratory and resident individuals), all of which also localized to Linkage Groups 9a and 9b. Interestingly, the same comparison between males and females in the kokanee population in neighboring East Barrière Lake revealed no outliers whatsoever. Although by no means definitive, taken together these results hint that different selection optima may exist for females and males with respect to migration in sockeye salmon, which are likely lost in kokanee populations that exhibit a solely resident life history.

Sexual conflict theory predicts the existence of antagonistic genetic variation for populations in which fitness optima differ for the two sexes (Connallon & Chenowith, 2019; Mank, 2017). Early work in this field hypothesized that antagonistic selection primarily acts on loci located on the X chromosome, however, recent studies across many taxa provide evidence for this acting on autosomal chromosomes as well (Fry, 2010). Sex-associated differences in migration, age and growth rates, as well as antagonistic selection between males and females at the genomic regions associated with these traits, have been previously reported in salmonids (Barson et al., 2015; Pearse et al., 2019, Willis et al., 2020). One such genomic region has been associated with migration in rainbow trout (*O. mykiss*), where a chromosomal inversion on *Omy05* exhibited reversed sex-dependent dominance (Pearse et al., 2019). In *O. mykiss* females, the ancestral karyotype that favours migration appears to be dominant, whereas in males, the pattern is reversed (Pearse et al., 2019). Genes represented in the inverted region were determined to involve circadian rhythms, sensitivity to light, adipogenesis, maturity and sex-specific processes (Pearse et al., 2019). Another example of sex-specific antagonistic dominance has been observed in Atlantic salmon (*Salmo salar*) involving the genomic region containing the highly conserved vestigial-like family member 3 gene (VGLL3); in this instance, females were

dominant for the late-maturing allele while males were dominant for the early-maturing allele (Barson et al., 2015). The absence of more complete information on the genomic architecture of sockeye salmon impedes the exploration of the effects that inversions and translocations might have on *O. nerka* phenotype; however, annotation of the outlier loci on LG9a and LG9b detected here in comparisons between migratory and resident individuals of Alouette Lake *O. nerka* provide some potential lines for further inquiry.

In the Alouette-specific analyses, significant functional annotations were produced for four of eight outliers identified in the migrant-resident comparison, and 16 of 21 outliers detected in the male-female comparison, including four of the seven identified independently in both (Table 2.3). Five of the sex-related outliers were annotated to genes involved in inflammatory and wound healing processes (Hirose et al., 2018; Sveen et al., 2019), four of which (including one also found in migrant-resident outlier scan) specifically mapped to *O. kisutch* proteoglycan 4 gene (Hirose et al., 2018) (Table 2.3) and one (SZNR01007180.1_1770356) annotated to *O. kisutch* cell surface hyaluronidase, which has been described in connection with wound healing (Sveen et al., 2019). Migratory salmon are susceptible to infection by novel microorganisms, that coupled together with warming temperatures, can decrease salmon performance and migration success (Crossin et al., 2008; Miller et al., 2014).

One additional sex-associated outlier (SZNR01009854.1_1379846) was located in *O. mykiss* growth hormone secretagogue receptor type 1-like, which may be associated with observed size differences between individuals exhibiting different migratory behaviours in Alouette Lake. A previous study showed Alouette Lake migrants to be twice as large as residents of the same age based on FL measurements (Godbout et al., 2011). Similarly, male and female sockeye salmon exhibit dimorphism with respect to size (Quinn & Foote, 1994). Differences in

maturation and growth rates between the two sexes, combined with the food availability in the environment, can drive divergent selection with respect to migration (Jonsson & Jonsson, 1993). For salmonids in general, female fecundity and reproductive success are directly proportional to size, and ocean-rearing provides more resources for biomass accumulation. Consequently, it is typically in the best interest of females to maximize feeding potential by migrating to the ocean (Jonsson & Jonsson, 1993); this may help explain why anadromous females are normally larger and have higher fecundity than resident females (Kendall et al., 2015). In contrast, alternative strategies employed by males (*e.g.* sneaking) may decrease the importance of reaching a certain size to maximize reproductive success (Foote, Brown, & Wood, 1997).

2.5 Management Implications

Our study provides important information for guiding on-going fisheries management operations for Alouette Lake. Previous research found no genetic differences between migrants and residents in this system (Godbout et al., 2011; Samarasin et al., 2017); while this holds at neutral data, additional information from outlier loci revealed how human-driven landscape modifications can alter not only the structure but also the life history strategies employed by *O. nerka.* In particular, our genome-wide analyses revealed that Alouette Lake *O. nerka* represents a single stock that is likely best characterized as land-locked sockeye salmon, with individuals that retain the ability to migrate. As a consequence, efforts to provide passage to reinforce sockeye salmon in the system appear sound. Importantly, our results suggest that Alouette Lake hosts the only known population of anadromous sockeye salmon that spawn at depth, punctuating the need for a re-assessment of its conservations status and highlighting its value as a system for future investigation of ecological and evolutionary questions associated with the impacts of water control structures on anadromous species.

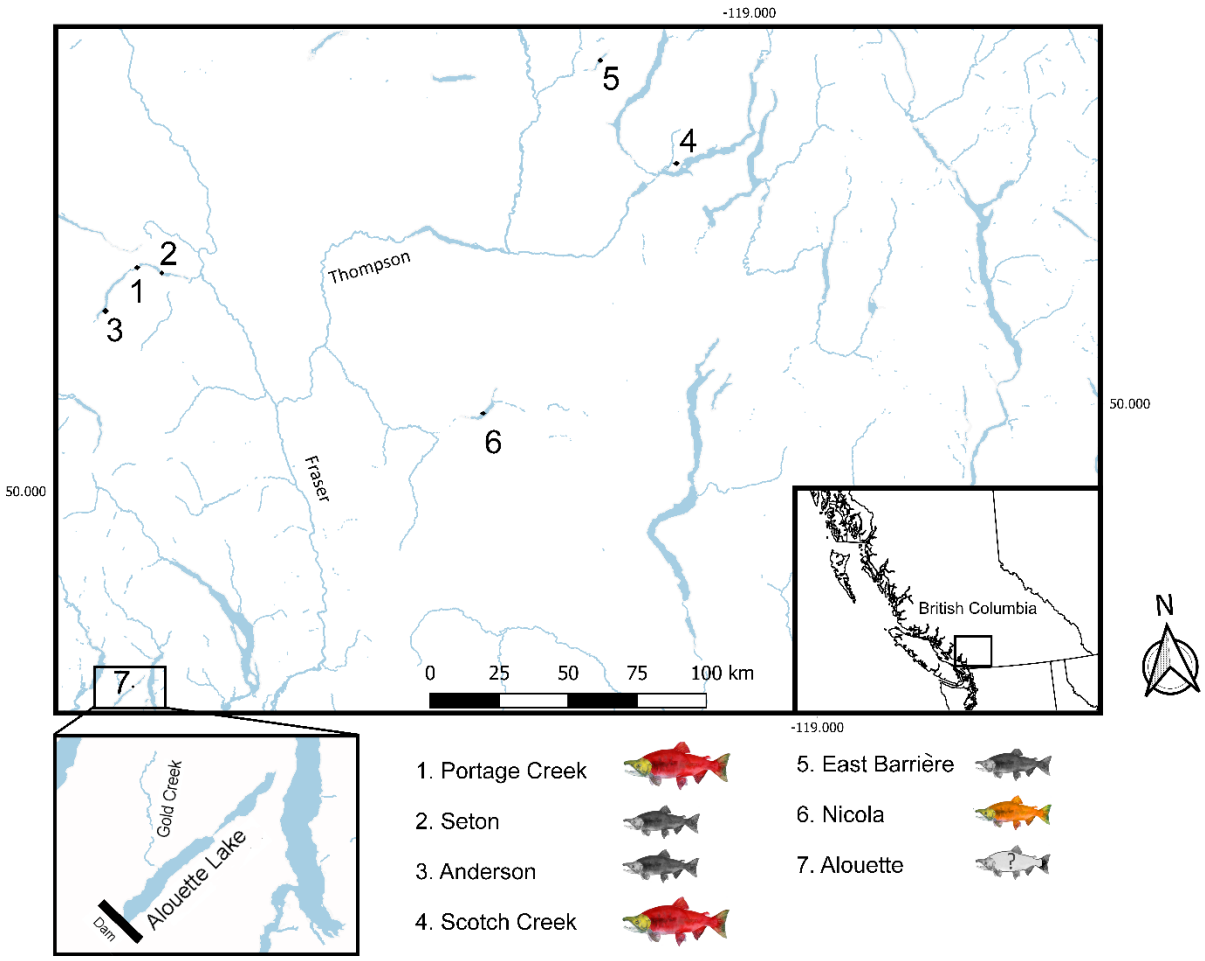

Figure 2.1. Fraser River drainage map, with numbered points indicating lake and creek locations, corresponding to *O. nerka* populations used in this study. Large red fish represent anadromous sockeye, small orange fish represent stream-spawning kokanee, small black fish represent deep-spawning kokanee, colourless fish with a question mark represents the Alouette population. Fish illustrations are a courtesy of Eileen Klatt. Map produced using QGIS.org (2020), QGIS Geographic Information System, Open Source Geospatial Foundation Project (http://qgis.org). Shapefiles were retrieved from B.C. Data Catalogue (https://catalogue.data.gov.bc.ca/dataset).

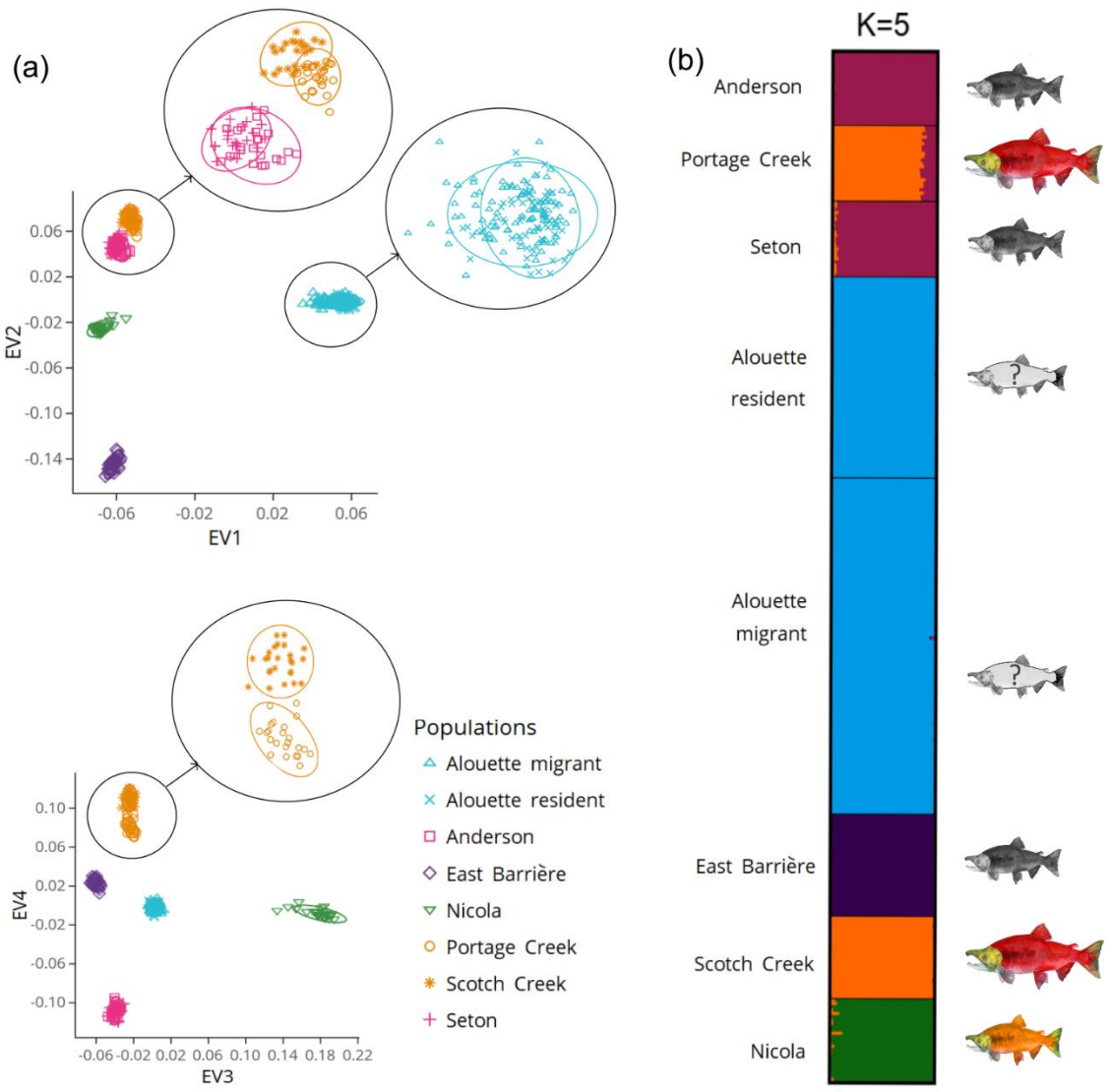

Figure 2.2. (a) Principal component analysis (PCA) for 312 individuals, produced using 7,012 putatively neutral SNPs. This analysis was conducted in *SNPRelate* v1.14.0 (Zheng et al., 2012). EV1, EV2, EV3 and EV4 explain 37.31%, 12.29%, 11.97% and 7.76% of the variation, respectively. (b) Results of Bayesian clustering method, as implemented in STRUCTURE v3.4 (Pritchard et al., 2000). Output results represent the optimal K value (K = 5), as determined by the ΔK method (Evanno et al., 2005), as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). Visualized using CLUMPAK (Kopelman et al., 2015).

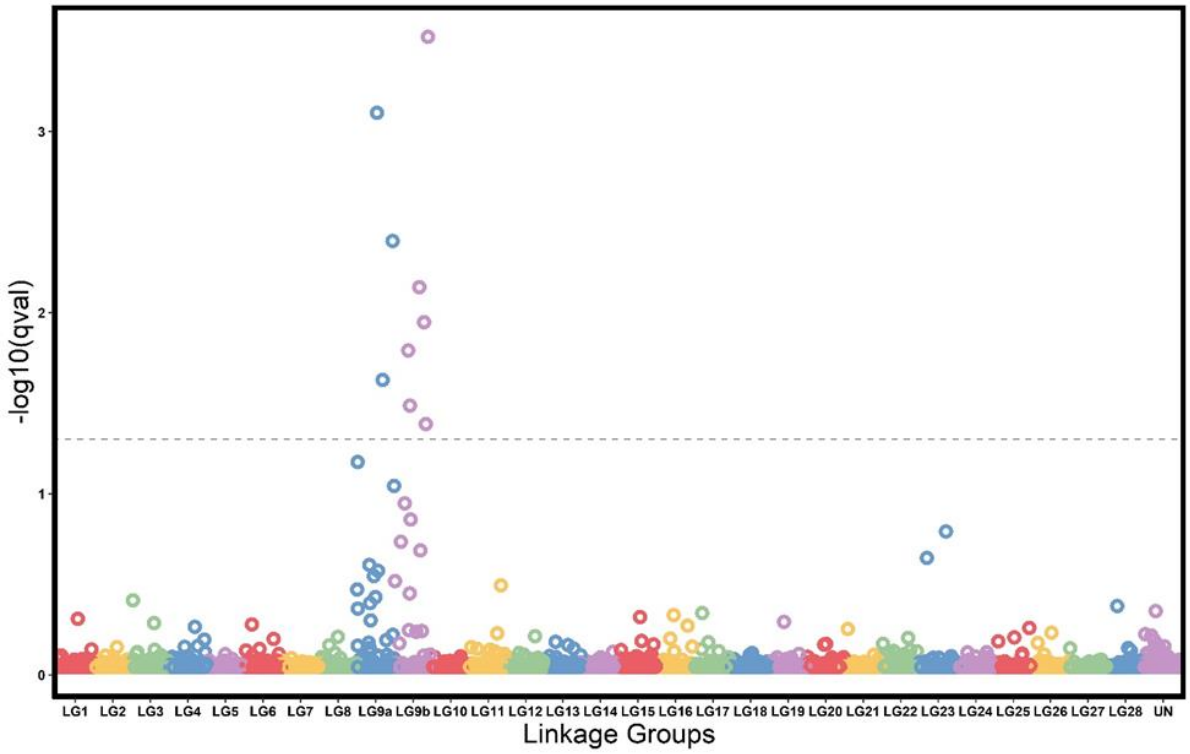

Figure 2.3. Manhattan plot, representing the -log10(qval) of 7709 SNPs as calculated by the BayeScan (Foll & Gaggiotti, 2008) outlier scan between resident and migrant *O. nerka* in Alouette Lake. The dash-line corresponds to the q-value of 0.05.

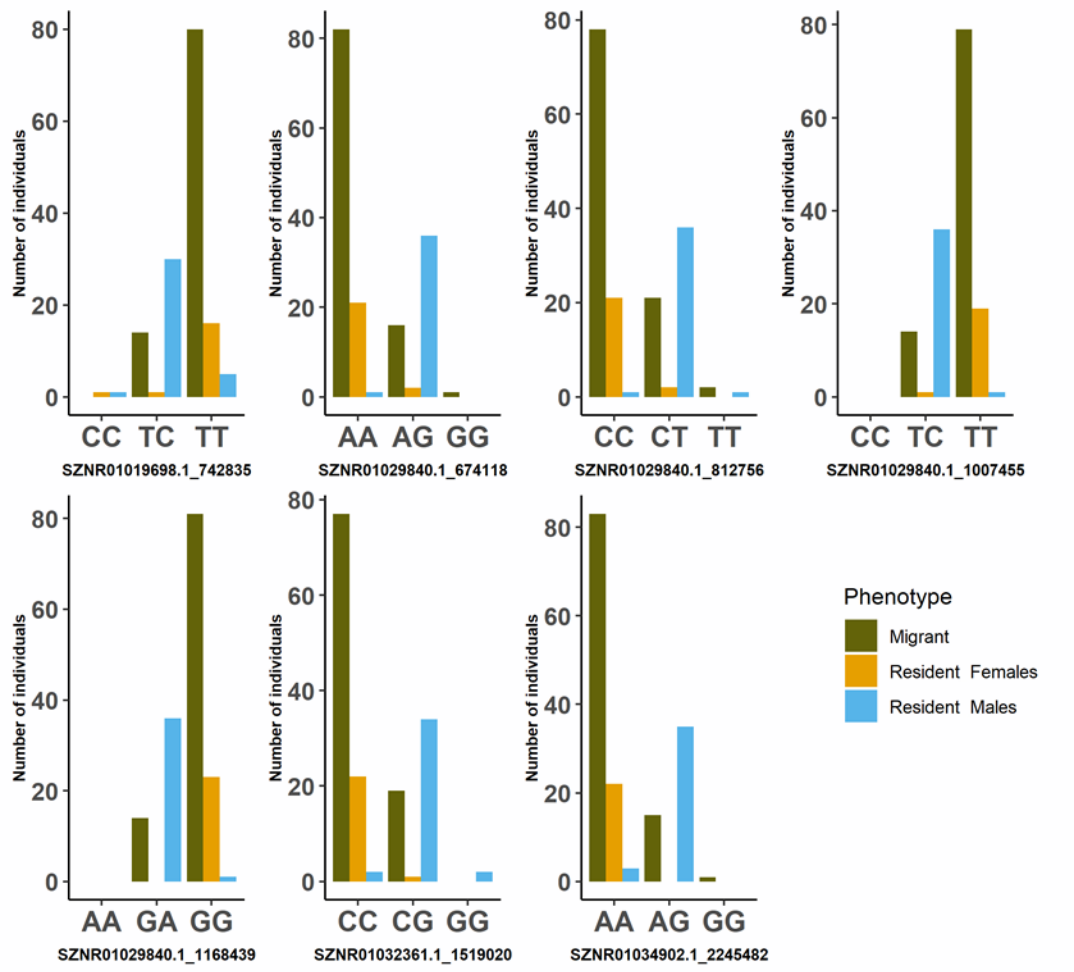

Figure 2.4. Allele frequencies of the 7 high-confidence outliers found in common between Alouette migrant-resident, and Alouette male-female resident outlier scans, compared across three groups: resident female *O. nerka* (n_{total} = 23), resident male *O. nerka* (n_{total} = 38), migrant *O. nerka* (n_{total} = 102). Samples that did not get genotyped at a particular SNP (genotype = NN) are not included in this figure.

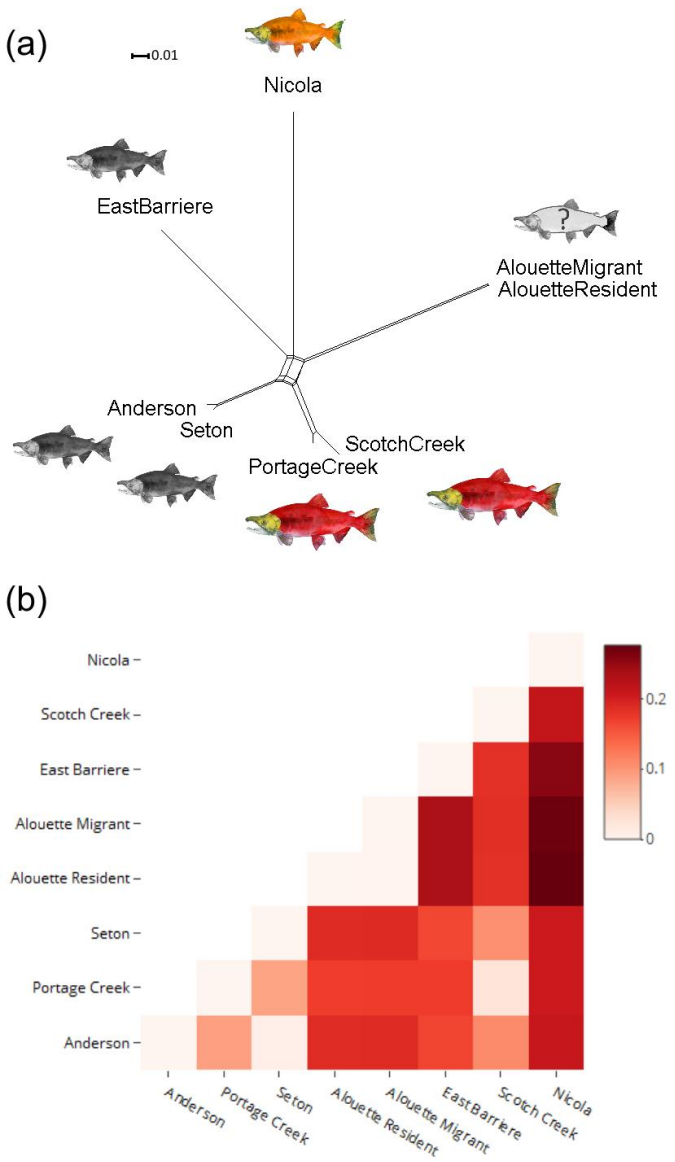

Figure 2.5. (a) NeighborNetwork (Bryant & Moulton, 2004), based on the Weir & Cockerham's (1984) pairwise θ values calculated in Genetix (Belkhir et al., 2004), using 7012 putatively neutral SNPs, visualized using SPLITSTREE v4.0 (Huson & Bryant, 2006). (b) Heat map of the θ matrix produced using R package *plotly* v4.9.0. The colour scale bar represents pairwise θ values.

Table 2.1. Sample size, ecotype and diversity statistics of the eight *O. nerka* populations**.**

Population	Morphs	Spawner type	N	N_e (95% CI)	H_e	H_o	F_{IS}
Anderson	kokanee	deep-spawning	22	2120.8 (1645.0-2982.0)	0.2646	0.2683	-0.0101
Portage Creek	sockeye	stream-spawning	23	892.0 (800.8 – 1006.5)	0.2662	0.2672	-0.0037
Seton	kokanee	deep-spawning	23	1331.9 (1142.8-1595.5)	0.2671	0.2701	-0.0104
Alouette	resident	deep-spawning	61	564.1 (550.9-578.0)	0.2694	0.2731	-0.0060
Alouette	migrant	deep-spawning[±]	102	794.6 (778.4-811.4)	0.2650	0.2460	0.0682
East Barrière	kokanee	deep-spawning	31	909.1 (829.7-1005.3)	0.2337	0.2448	-0.0357
Scotch Creek	sockeye	stream-spawning	25	1856.1 (1530.5-2356.5)	0.2647	0.2647	0.0015
Nicola	kokanee	stream-spawning	25	2962.4 (2011.6-5608.2)	0.2013	0.2029	-0.0052

[±] See Introduction for spawning information of Alouette Lake *O. nerka* migrants.

Abbreviations: sample size (N), confidence interval (CI), expected heterozygosity (H_e), observed heterozygosity (H_o), inbreeding coefficient (F_{IS}), and effective population size (N_e).

Table 2.2. IDs and annotations of migrant-resident outliers detected in this study and corresponding loci ID of outliers detected in Nichols et al. (2016), Veale & Russello (2017b).

SNP	Annotation	Nichols et al. (2016)	Veale & Russello (2017b)
SZNR01010580.1_848156*	*Oncorhynchus nerka* isolate LRRC9_Ok_shore leucine-rich repeat-containing protein 9-like protein gene, complete cds	RADtag_57884	68810_51
SZNR01019686.1_513507	PREDICTED: *Oncorhynchus kisutch* TIR domain containing adaptor protein (tirap), mRNA		
SZNR01002048.1_136609	PREDICTED: *Oncorhynchus kisutch* proteoglycan 4 (LOC109908567), mRNA	RADtag_66595	
SZNR01024871.1_93859	PREDICTED: *Oncorhynchus nerka* heat shock protein HSP 90-alpha 1 (LOC115118567), transcript variant X1, mRNA		40949_10
SZNR01027302.1_1918892	PREDICTED: *Oncorhynchus kisutch* proteoglycan 4 (LOC109908567), mRNA		
SZNR01027302.1_1918732	PREDICTED: *Oncorhynchus kisutch* inactive phospholipid phosphatase 7 (LOC109868517), mRNA	RADtag_18513	112822_83
SZNR01029823.1_2579190	PREDICTED: *Oncorhynchus mykiss* uncharacterized LOC110523490 (LOC110523490), ncRNA		
SZNR01029834.1_1048492	PREDICTED: *Oncorhynchus mykiss* uncharacterized LOC110496934 (LOC110496934), ncRNA		
SZNR01004580.1_453594	PREDICTED: *Oncorhynchus kisutch* proline-rich transmembrane protein 1-like (LOC109908550), mRNA	RADtag_7544	14428_85

SNP	Annotation	Nichols et al. (2016)	Veale & Russello (2017b)
SZNR01004638.1_505442	*Oncorhynchus tshawytscha* follicle stimulating hormone beta subunit (FSHbeta) gene, promoter and complete cds		
SZNR01007172.1_549647*	PREDICTED*: Oncorhynchus kisutch* stearoyl-CoA desaturase 5 (LOC109868414), mRNA		
SZNR01007191.1_179523	PREDICTED: *Oncorhynchus nerka* partitioning defective 6 homolog alpha-like (LOC115134284), mRNA		3833_28
SZNR01010580.1_883229*	NA		
SZNR01022265.1_96641	NA		

* indicates outliers detected both in the ecotype outlier scan, and multiple pairwise comparison scans between sockeye salmon and kokanee populations.

Table 2.3. IDs and annotations of outlier loci detected between migrant and resident Alouette Lake *O. nerka* (migration-associated), and male and female Alouette Lake *O. nerka* residents (sex-associated). Sex-associated SNPs are denoted by "S", migration-associated SNPs are denoted by "M", SNPs detected in both sex- and migration-associated outliers scans are denoted by "SM".

SNP	Annotation	Type
SZNR01007180.1_287818	NA	S
SZNR01007180.1_1770356	PREDICTED: *Oncorhynchus kisutch* cell surface hyaluronidase (LOC109895707), transcript variant X2, mRNA	S
SZNR01009854.1_1379846	PREDICTED: *Oncorhynchus mykiss* growth hormone secretagogue receptor type 1-like (LOC110495701), mRNA 1	S
SZNR01019691.1_1544307	PREDICTED: *Oncorhynchus kisutch* plexin-B2-like (LOC109867097), mRNA	S
SZNR01019691.1_1548960	PREDICTED: *Oncorhynchus nerka* plexin-B2-like (LOC115134416), mRNA	S
SZNR01019691.1_1566159	NA	S
SZNR01019691.1_5451347	PREDICTED: *Oncorhynchus kisutch* G1/S-specific cyclin-D1-like (LOC109889644), transcript variant X2, mRNA	S
SZNR01019698.1_742835	NA	SM
SZNR01019700.1_1210193	PREDICTED: *Oncorhynchus kisutch* uncharacterized LOC109903989 (LOC109903989), transcript variant X1, ncRNA	S
SZNR01019781.1_61293	PREDICTED: *Oncorhynchus kisutch* homeodomain-interacting protein kinase 1-like (LOC109884283), transcript variant X5, mRNA	S
SZNR01029840.1_666180	*Oncorhynchus mykiss* SYPG1 (SYPG1), PHF1 (PHF1), and RGL2 (RGL2) genes, complete cds; DNaseII pseudogene, complete sequence; LGN-like, PBX2 (PBX2), NOTCH-like, TAP1 (TAP1), and BRD2 (BRD2) genes, complete cds; and MHCII-alpha and Raftlin-like pseudogenes, complete sequence	S
SZNR01029840.1_674118	PREDICTED: *Oncorhynchus kisutch* TIR domain containing adaptor protein (tirap), mRNA	SM
SZNR01029840.1_812756	NA	SM

SNP	Annotation	Type
SZNR01029840.1_1007455	PREDICTED: *Oncorhynchus nerka* phosphatidylinositol 3-kinase regulatory subunit gamma-like (LOC115114221), transcript variant X3, mRNA 1	SM
SZNR01029840.1_1168439	PREDICTED: *Oncorhynchus kisutch* proteoglycan 4 (LOC109908567), mRNA	SM
SZNR01032361.1_472043	PREDICTED: *Oncorhynchus kisutch* cyclin-dependent kinase 6-like (LOC109904398), transcript variant X2, mRNA 1	S
SZNR01032361.1_1519020	PREDICTED: *Oncorhynchus kisutch* thioredoxin domain containing 16 (txndc16), transcript variant X1, mRNA	SM
SZNR01032361.1_2562724	PREDICTED: *Oncorhynchus kisutch* proteoglycan 4 (LOC109908567), mRNA 409	S
SZNR01034902.1_1863761	PREDICTED: *Oncorhynchus kisutch* proteoglycan 4 (LOC109908567), mRNA	S
SZNR01034902.1_2245482	NA	SM
SZNR01034902.1_4350081	PREDICTED: *Oncorhynchus kisutch* proteoglycan 4 (LOC109908567), mRNA	S
SZNR01032361.1_2116325	NA	M

Chapter 3: Reconstructing the origin and genomic basis of deep-spawning kokanee (*Oncorhynchus nerka*) ecotype across its pan Pacific distribution

3.1 Background

Parallel evolution, the independent evolution of the same trait in closely related lineages, has been documented across a wide range of taxa (Haldane, 1932; Wood, Burke, & Rieseberg, 2005). For example, short-winged ecomorphs of Japanese scorpionfly have independently evolved in geographically isolated populations as an adaptation to high-altitude conditions (Suzuki, Suzuki, & Tojo, 2019); adult dwarfism has repeatedly emerged in Icelandic charr inhabiting volcanic springs (Macqueen, Kristjánsson, Paxton, Vieira, & Johnston, 2011); and sympatric, microhabitat-associated ecotypes of marine snails have independently originated several times (Quesada, Posada, Caballero, Morán, & Rolán-Alvarez, 2007). Molecular mechanisms underlying parallel evolution have been traditionally demonstrated for phenotypic traits that are regulated by specific candidate genes or sets of genes, the functions of which have been known *a priori.* Examples include insecticide resistance in red flour beetles, where different point mutations in a resistance-conferring gene have evolved in geographically distinct populations (Andreev, Kreitman, Phillips, Beeman, & ffrench-Constant, 1999); and the depigmentation phenotype present in isolated populations of Mexican blind cavefish, resulting from different mutations within the melanocortin receptor gene (Gross, Borowsky, & Tabin, 2009). In contrast, deciphering molecular mechanisms responsible for repeated evolution of phenotypic traits for which no prior genetic information is available only became possible with the advent of genome-wide sequencing (Fraser & Whiting, 2020; Stern, 2013). Over the past decade, genomic scans have been commonly employed to identify candidate genes responsible

for the evolution of the same phenotypes in replicated populations (Deagle et al., 2012; Nichols et al., 2016; Veale & Russello, 2017b; Westram et al., 2014).

Salmonids, including sockeye salmon (*Oncorhynchus nerka*), have become a classic example of parallel evolution of divergent ecotypes in response to ecological pressures, as demonstrated by their unusually diverse life history variation. For example, multiple independent lineages of freshwater resident *O. nerk*a (known as kokanee) have evolved from anadromous sockeye salmon across the entire range of the species' pan Pacific distribution, spanning from western North America to east Asia from Russia (Kamchatka) to Japan (Kaeriyama, Urawa, & Fukuwaka, 1995; Taylor et al., 1996). Kokanee populations are genetically distinct from sympatric populations of sockeye salmon and can be further distinguished based on their habitat preferences, physiological traits, morphology, life history and behaviour (Foote et al., 1989, Taylor et al., 1996, Lemay & Russello, 2015, Veale and Russello 2016; Veale & Russello, 2017b). Moreover, these differences appear to be adaptive and have evolved due to differential selection pressures attributed to variation in their associated environments and life cycles (Foote et al., 1999). The two migratory forms can be further subdivided into reproductive ecotypes, depending on spawning location (shore-, stream- and deep-spawning), which have repeatedly evolved in different parts of the world (Moreira & Taylor, 2015; Taylor et al., 1996; Veale & Russello, 2017b).

Among these ecotypes, deep-spawning kokanee (also known as black kokanee or kunimasu) can be distinguished from the more common shore- and stream-spawning kokanee based on its morphology, unusual spawning depth and pigmentation at maturity. While stream-spawners and, to a lesser extent, shore-spawners transition to bright or brown-red coloration upon maturation, deep-spawning kokanee remain olive-black, even during the spawning season

(Moreira & Taylor, 2015; Nakabo et al., 2011, 2014). In this ecotype, spawning occurs more than 50 m below the lake surface in the lower part of the profundal zone and continues throughout the winter months (Moreira & Taylor 2015; Nakabo et al., 2014). Moreover, deep-spawning kokanee can be differentiated from stream-spawning kokanee based on the number of pyloric caeca, fin rays and gill rakers (Nakabo et al., 2014).

Geographically isolated populations of deep-spawning kokanee that share similar behavioural and morphological traits occur in Saiko Lake, Yamanashi Prefecture, Japan (Nakabo et al., 2011), as well as Anderson, Seton (Moreira & Taylor, 2015), and East Barrière Lakes (Andrusak & Morris, 2004) in British Columbia, Canada, leading to questions surrounding the origin and genomic basis of this reproductive form. To date, a handful of studies have investigated the morphology, genetics and behaviour of deep-spawning kokanee (Muto, Nakayama, & Nakabo, 2013; Nakabo et al., 2011, 2014; Nakayama, Tohkairin, Yoshikawa, & Nakabo, 2018; Veale & Russello, 2017b), only one of which combined genetic data for both Japanese and Canadian populations (Moreira & Taylor, 2015). In Saiko Lake, genetic distinctiveness of sympatric deep-spawning and stream-spawning kokanee has been previously demonstrated based on five microsatellite markers (Muto et al., 2013). Likewise, deep-spawning kokanee in Anderson and Seton Lakes were found to be genetically distinct from anadromous sockeye salmon in the connecting Portage Creek, based on both microsatellites (Moreira & Taylor, 2015) and genome-wide analyses of single nucleotide polymorphisms (SNPs) (Veale & Russello, 2017b). The only known incidence of hybridization between deep-spawning and stream-spawning kokanee has been documented in Motosu Lake (Nakayama et al., 2018). On a broader scale, deep-spawning kokanee populations in Japan and Canada are not monophyletic,

suggesting that this ecotype evolved as a result of multiple independent divergence events (Moreira & Taylor, 2015).

Here, we employed genotyping-by-sequencing to reconstruct the history and genetic basis of deep-spawning kokanee, including previously studied populations in Japan (Saiko Lake) and Canada (Anderson and Seton Lakes) as well as a newly discovered population in British Columbia (East Barrière Lake) (Figure 3.1). Using genome-wide SNP data, we conducted genomic scans to identify outlier loci between paired deep-spawning and stream-spawning populations in Canada and Japan to investigate putative genetic signatures of parallel evolution and ecotype divergence. Moreover, we characterized genome-wide structure of the four deep-spawning populations relative to each other as well as to stream-spawning kokanee populations in their respective basins to explicitly test the multiple independent origin hypothesis. Lastly, we investigated whether there was evidence for hybridization between deep- and stream-spawning kokanee in Saiko Lake, one of only two locations in the world where they have been documented to co-occur.

3.2 Methods

3.2.1 Study design and sampling

Our study included deep-spawning kokanee samples (n = 20) and stream-spawning kokanee samples (n = 17) previously collected in 2010-2011 from Saiko Lake, Yamanashi Prefecture, Japan, as well as stream-spawning kokanee samples (n = 21) from Akan Lake, Hokkaido Prefecture, Japan (Nakabo et al., 2011). The samples were originally collected by a combination of gill netting and angling and were comprised of muscle tissue preserved in 100% ethanol. For deep-spawning and stream-spawning kokanee in British Columbia, we used previously collected genotypic data from East Barrière and Nicola Lakes (Samad-zada et al., in

prep), as well as Anderson, Seton and Kootenay Lakes (Veale & Russello, 2017b). Information regarding sample collection, DNA extraction, and library construction can be found in original papers; complete distribution of kokanee populations used in this study, along with their respective ecotypes, is provided in Table 3.1.

3.2.2 Library preparation

Genomic DNA was extracted from muscle tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol with the addition of 4 µl of 100 mg/ml 7000U RNase (Qiagen) prior to ethanol precipitation. To construct genomic libraries, we employed restriction site-associated DNA sequencing (RADseq) (Baird et al., 2008) with modifications as specified in Lemay & Russello (2015). This protocol was used to ensure connectivity with previously collected data from East Barrière and Nicola Lakes (Samad-zada et al., in prep), as well as Anderson, Seton and Kootenay Lakes (Veale & Russello, 2017b). We constructed two RADseq libraries that in total contained 58 unique individuals, four of which were included as replicates (two per library) to estimate genotyping error within each library (Tintle et al., 2009). The constructed libraries were sequenced at the McGill University and Génome Québec Innovation Centre, using one lane each of Illumina 2500 PE125 (two lanes in total).

3.2.3 Genotyping and filtering

We combined the newly sequenced reads from Akan and Saiko Lakes with those previously generated for samples from Anderson, Seton and Kootenay Lakes (Veale & Russello, 2017b), and East Barrière and Nicola Lakes (Samad-zada et al., in prep) (Table 3.1). To ensure connectivity with previously sequenced samples, raw paired-end reads were demultiplexed and trimmed to 100 bp via the *process radtags* command in STACKS v2.41 (Catchen et al., 2011).

After demultiplexing, samples containing less than 2 million reads were not used in the construction of the main SNP dataset, however, they were included in the Saiko-specific pipeline (see below). Identical reads generated due to PCR amplification were removed using the *clone filter* command in STACKS v2.41 (Catchen et al., 2011). Demultiplexed, trimmed and filtered reads were aligned against the *O. nerka* reference genome (*Oner_1*, GenBank Assembly Accession ID: GCA_006149115.1, Christensen et al., 2020) using the *bwa mem* algorithm in BWA (Li & Durbin, 2009) and sorted using *samtools* (Li et al., 2009). Resulting bam files together with those generated earlier for the samples from Anderson, Seton, East Barrière, Nicola, and Kootenay Lakes were used to generate loci and call SNPs via the *gstacks* command. Next, we processed the resulting loci through the *populations* module in STACKS v2.41 (Catchen et al., 2011). We performed a sensitivity analysis (Table B.2) on the retained individuals by running the *populations* module in STACKS v2.41 (Catchen et al., 2011) with a varying set of parameters to determine the optimal set for SNP ascertainment. Based on the outcome of the sensitivity test, we only retained loci observed in 70% (r70) or more of the individuals present in all eight populations, with a minimum allele frequency of 0.03 and a maximum observed heterozygosity of 0.50. To decrease the effects of linkage disequilibrium, --*write-single-snp* flag was used to only retain one SNP per locus. Next, we calculated mean coverage data per individual using VCFtools v0.1.15 (Danecek et al., 2011) and removed individuals with mean coverage lower than 6x or mean missing data percentage more than 30%, and re-ran populations using the optimal parameters (Table B.1). We further processed this dataset through VCFtools v0.1.15 (Danecek et al., 2011) to only include sites with a minimum mean depth of 10x and a maximum mean depth of 100x and exclude sites with more than 10% missing data (--*max-missing 0.9*). Additionally, we excluded loci identified as putative paralogs

as described in McKinney et al. (2017) using HDplot function available from https://github.com/gjmckinney/HDplot. Lastly, we estimated genotyping error rate using a custom python script and removed replicate samples.

3.2.4 Outlier detection

To minimize the high incidence of false positives typically present in outlier detection methods, we used both *Fst*-based approaches, as implemented in Arlequin v3.5 (Excoffier & Lischer, 2010) and BayeScan v.2.0 (Foll & Gaggiotti, 2008), as well as a principal component analysis (PCA)-based outlier detection procedure, as implemented in *pcadapt* (Luu et al., 2017). For Arlequin, we used the hierarchical island model (Slatkin & Voelm, 1991), that allows for higher migration rates within the group than between the groups. We performed 20,000 simulations, with the number of simulated demes set to 100, and the number of simulated groups set to 10, and simulated derived allele frequency set to 0.05. We then corrected p-values of each SNP, and loci in the first quantile with FDR < 0.05 were identified as outliers. For BayeScan v2.0 (Foll & Gaggiotti, 2008), we used a pairwise approach comparing allele frequencies between all possible pairs of populations, resulting in 28 pairwise comparisons. The analyses were run for 100,000 iterations with 50,000 burn-in period with Prior Odds set to 10, and loci with q-value less than 0.05 were marked as outliers. Lastly, we inferred the number of genetic clusters through analyses of principal components (PCs), as implemented in R package *pcadapt* (Luu et al., 2017), and Cuttel's rule to infer the most likely number of PCs that explain the genetic structure within the dataset. The resulting *p*-values were corrected for multiple comparisons using the method of Benjamini-Yekutieli (2001), and loci with adjusted *p*-values of less than 0.05 were considered outliers. We performed *pcadapt* analysis three times: including all

samples in our dataset, including British Columbia samples only, and including Japanese samples only.

We also identified loci divergent between stream-spawning and deep-spawning kokanee populations, by conducting outlier scans on "paired" populations from each region. The British Columbia-specific analysis was conducted on the following pairs of populations: Nicola-East Barrière, Nicola-Anderson, and Nicola-Seton, where Nicola population is stream-spawning, and the rest of the populations are deep-spawning kokanee. Japan-specific outlier scan compared Saiko deep-spawning to Saiko stream-spawning; and Saiko deep-spawning to Akan stream-spawning kokanee population. We employed the same three outlier detection approaches as above (BayeScan, *pcadapt*, Arlequin), using identical parameters. Outliers detected in British Columbia-specific outlier scan were compared to outliers detected in the Japan-specific outlier scan, and all outliers were functionally annotated using BLAST (Altschul et al., 1990) with the e-value threshold of 1e-28, and -entrez_query set to "*Oncorhynchus*".

Lastly, we specifically examined genotypes for all Japanese samples at a SNP in the leucine-rich-repeat-containing 9 (LRRC9) gene (Veale & Russello, 2017a) that has previously been demonstrated as a candidate locus under divergent selection associated with reproductive ecotype (shore- or stream-spawning). Divergence at LRRC9 gene has previously been reported for British Columbia populations examined in this study (Samad-zada et al., in prep; Veale & Russello, 2017b).

3.2.5 Population genetics analyses

To construct a putatively neutral dataset, we removed any locus identified as an outlier in any comparison. Following outlier removal, we removed loci that significantly (-h 0.05) deviated

from Hardy-Weinberg Equilibrium in 50% or more of the populations using the *filter_hwe_by_pop.pl* script from https://github.com/jpuritz/dDocent/tree/master/scripts.

Using the resulting putatively neutral SNP dataset, we calculated locus-specific observed (H_o) and expected (H_e) heterozygosities following Nei (1987), using the *basic.stats* function in R package *hierfstat* (Goudet et al., 2015), and averaged across loci for each population. Additionally, we computed inbreeding coefficients (F_{IS}) for every population using Genetix v.4.05.2 (Belkhir et al., 2004) and 1000 permutations to determine levels of significance. We estimated effective population sizes (*Ne*) employing the Linkage Disequilibrium method (Waples & Do, 2008), as implemented in NeEstimator v.2 (Do et al., 2014), with the lowest allele frequency set to 0.05. To examine the levels of pairwise population differentiation, we calculated Weir & Cockerham's (1984) θ between all pairs of populations, performing 1000 permutations in Genetix v.4.05.2 (Belkhir et al., 2004).

To assess the number of genetic clusters representing the populations in our dataset, we employed several approaches. First, we used the Bayesian method of Pritchard et al. (2000), as implemented in STRUCTURE v2.3.4. We set the run length to 100,000 Markov Chain Monte Carlo (MCMC) iterations after the burn-in period of 100,000 MCMC iterations, using the correlated allele frequencies under the admixture model, enabling the LOCPRIOR option. We varied the number of clusters (K) was varied from one to ten, with ten replicate runs for each. We then evaluated the output of the STRUCTURE analysis in STRUCTURE Harvester (Earl & von Holdt, 2012), where we inferred the optimal K value by employing a combination of the ΔK method (Evanno et al., 2005) and plotting of the log probability of the data (Pritchard et al., 2000, to estimate where ln Pr (X|K) reached a plateau. The results of the STRUCTURE output were visualized in CLUMPAK (Kopelman et al., 2015). Second, we conducted a PCA analysis

in order to visualize the relationships among populations using *SNPRelate* v1.14.0 (Zheng et al., 2012).

In order to examine the relationship among the populations in our dataset, we used TreeMix (Pickrell & Pritchard, 2012) that implements a model accounting for both population splitting and gene flow, and hence is a more appropriate method for estimating relationships among multiple populations of the same species. Given that TreeMix requires a pruned set of markers without any missing data, this analysis was performed on a reduced dataset, where we removed SNPs which did not map to a Linkage Group (i.e. those mapped to the unplaced scaffolds on the reference genome), and only retained those that did not contain any missing data for any of the individuals *(--max-missing 1* in VCFtools v0.1.15 (Danecek et al., 2011). To maximize the number of retained SNPs for this analysis, we removed four individuals that had more than 20% missing data. Lastly, SNPs were pruned using PLINK, with the window size set to 50 kb, step size set to 10 kb, and r^2 threshold of 0.2 (Purcell et al., 2007).

3.2.6 Saiko Lake hybridization

To assess the level of hybridization between deep-spawning and stream-spawning kokanee in Saiko Lake, we re-ran *gstacks* and *populations* in STACKS v2.41, using the same parameters as described above. Here, we used only Saiko Lake deep-spawning and stream-spawning samples (n = 41), including those individuals (n = 9) that were originally discarded from the broad-scale analyses due to a low number of reads. We removed samples with mean coverage lower than 6x and discarded loci with mean missing data > 10%, as calculated by v0.1.15 (Danecek et al., 2011), reducing the number of retained loci in order to maximize the number of samples while still maintaining a robustly filtered SNP dataset.

We used the Saiko-specific SNP dataset and the approaches implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000) and NewHybrids v 2.0 (Anderson & Thompson, 2002) to investigate evidence for admixture between the co-occurring deep- and stream-spawning kokanee in Saiko Lake. For the STRUCTURE analysis, we employed the same parameters as described above except a straight admixture model was used without LOCPRIOR and the number of clusters (*K*) was set to 2, based on the results of the broad-scale analysis (see Results below). For the NewHybrids analysis, we used a dataset of 500 SNPs that had the highest F_{ST} because NewHybrids was experiencing underflow errors with larger datasets. We defined six genotype frequency classes (deep-spawning kokanee pure stock, stream-spawning kokanee pure stock, F1, F2, B2 deep-spawning kokanee backcross; B2 stream-spawning kokanee backcross) and ran 10,000 sweeps during, and 50,000 sweeps after the burn-in period. Individuals were subsequently assigned into the genotype class that had the highest probability.

3.3 Results

3.3.1 Dataset quality

Following demultiplexing and quality filtering, newly sequenced samples (n = 107) averaged 9,014,160 retained reads. Across all 187 samples, an average of 93.95% of reads were successfully mapped to the reference genome, however, only samples (n = 173) with more than 2,000,000 reads were used to generate the main SNP dataset (see Methods). After full filtering, 9,721 SNPs were retained for 167 individuals [one sample was removed due to insufficient coverage (depth < 6x), while the five replicates were removed upon calculation of genotyping error frequency], with a mean depth of 31.42x and mean missing percentage of 2.45%. The mean within-library genotyping error rate was 1.89%.

3.3.2. Outlier detection, mapping and annotation

For genomic scans conducted across all populations, we detected the following numbers of outliers: Arlequin (n = 10); BayeScan (n = 24); and *pcadapt* (n = 444; optimal PCs retained = 5); 26 outliers were detected across two or more methods. For the *pcadapt* analyses conducted for populations independently by region, 222 and 201 outliers were detected for British Columbia (optimal PCs retained = 3) and Japan (optimal PCs retained = 2), respectively; one outlier locus was identified independently in both regions.

Genomic scans were also conducted for populations within the same basin, "paired" regionally by reproductive ecotype. For British Columbia deep-spawners (East Barrière, Anderson, Seton Lakes) versus stream-spawners (Nicola Lake), we detected the following numbers of outliers: Arlequin (n = 6); BayeScan (n = 8); and *pcadapt* (n = 148; optimal PCs retained=3); eight of these outliers were detected across two or more methods (Figure B.1). For Japanese deep-spawners (Saiko Lake) versus stream-spawners (Saiko, Akan Lakes), we detected the following numbers of outliers: Arlequin (n = 94); BayeScan (n = 0); and *pcadapt* (n = 201; optimal PCs retained = 2); 44 of these outliers were detected across two or more methods (Figure B.1). One outlier locus was detected in common between British Columbia *pcadapt* and Japan *pcadapt* analyses (SZNR01029850.1_410525), while two outliers were in common between Japan Arlequin and British Columbia *pcadapt* analysis (SZNR01029850.1_410525, SZNR01029883.1_386474). Additionally, SZNR01014727.1_80153, detected in stream versus deep-spawning comparisons in British Columbia, and SZNR01027293.1_1297100, detected in stream versus deep-spawning comparisons in Japan, annotated to the same genomic region [*Oncorhynchus mykiss* SYPG1 (SYPG1), PHF1 (PHF1), and RGL2 (RGL2) genes, complete cds; DNaseII pseudogene, complete sequence; LGN-like, PBX2 (PBX2), NOTCH-like, TAP1

(TAP1), and BRD2 (BRD2) genes, complete cds; and MHCII-alpha and Raftlin-like pseudogenes, complete sequence].

Based on the region-specific, pairwise outlier analyses, two SNPs (SZNR01019859.1_12357 and SZNR01017145.1_7778989) that were independently detected in British Columbia and Japan, respectively, annotated to the same region: PREDICTED: *Oncorhynchus kisutch* inactive phospholipid phosphatase 7 (LOC109868517), mRNA.

All Japanese kokanee samples, regardless of ecotype or location, were fixed for the "GG" genotype at SZNR01010580.1_848156 (LRRC9) that has been associated with shore-spawning behaviour in *O. nerka* (Veale & Russello, 2017a). In British Columbia, all kokanee populations were assigned to the correct reproductive ecotypes based on LRRC9 genotypes: Nicola and Kootenay Lake populations contained no individuals with the GG genotype indicative of stream-spawning, while in East Barrière, Anderson and Seton kokanee populations, the G allele frequencies were between 0.95-1.00, indicative of shore-spawning.

3.3.3 Diversity statistics and population structure estimation

We removed all SNPs that were identified as outliers (n = 700) in any of the above comparisons; none of the remaining SNPs deviated from HWE. Based on this putatively neutral dataset of 9,021 SNPs, H_o, H_e, and F_{IS} values were similar across all British Columbia populations, with both H_o and H_e, ranging from 0.17 to 0.22 and none of the F_{IS} values significantly different from zero. Levels of heterozygosity were consistently lower for the Japanese populations (0.07-0.12), all of which exhibited heterozygote deficit and positive F_{IS} values that were significantly different from zero (Table 3.1). Both deep- and stream-spawning kokanee populations in Japan demonstrated substantially lower N_e estimates [Saiko = 30.4 (deep-

spawning) and 33.2 (stream-spawning); Akan = 223.3] than those detected in British Columbia, which ranged from 814.5 (East Barrière) to 4194.2 (Nicola).

Pairwise θ estimates indicated that the Saiko Lake deep-spawning population was the most highly differentiated relative to all others in both British Columbia and Japan, with the largest values observed in comparison with stream-spawning kokanee population from Akan Lake (θ = 0.70) and Nicola Lake (θ = 0.63) (Table 3.2). The Akan Lake and Saiko Lake stream-spawning kokanee populations were not genetically distinct (θ = -0.00031). The British Columbia populations all demonstrated significant pairwise θ estimates largely associated with geographic proximity within the province, ranging from 0.00955 (Anderson Lake deep-spawning versus Seton Lake deep-spawning) to 0.25524 (Kootenay Lake stream-spawning versus Nicola Lake stream-spawning) (Table 3.2).

The STRUCTURE analysis revealed evidence for six clusters that best explained the genetic variation within our dataset (Figure 3.2). Saiko Lake deep-spawning kokanee was identified as a distinct cluster, starting from K = 2, and Akan and Saiko stream-spawning kokanee formed a separate cluster starting from K = 3, separating Japanese kokanee from populations in British Columbia. East Barrière Lake deep-spawning kokanee and Nicola Lake stream-spawning kokanee each formed distinct clusters at K = 5, and Kootenay Lake stream-spawning kokanee separated from Anderson Lake and Seton Lake deep-spawning kokanee at K = 6 (Figure 3.2). This clustering persisted even at higher values of K; Anderson Lake and Seton Lake deep-spawning kokanee never separated into distinct clusters. The PCA results largely agreed with those demonstrated by STRUCTURE, revealing differentiation by geography (Figure 3.2). The first PC explained the largest percent of variation (44.4%) and separated Saiko Lake deep-spawning kokanee from all other populations. The second PC explained 13.6% of the

variation and separated the tightly clustered Akan Lake and Saiko Lake stream-spawning kokanee from the British Columbia populations. In British Columbia, all populations formed distinct clusters, with the exception of the Anderson Lake and Seton Lake deep-spawning populations.

The maximum likelihood tree generated in TreeMix, based on the reduced dataset of 901 SNPs was consistent with patterns revealed by STRUCTURE and PCA analyses (Figure 3.3). Saiko Lake deep-spawning kokanee showed the highest level of genetic drift, relative to all other populations, followed by stream-spawning kokanee from Akan and Saiko Lakes. In British Columbia, population split estimates reflected geographic proximity, with the exception of Kootenay Lake stream-spawning kokanee that was most closely related to Anderson and Seton deep-spawning kokanee populations.

3.3.4 Hybridization analysis of Saiko Lake co-occurring ecotypes

After full filtering, 2165 SNPs and 34 samples (n = 20 deep-spawning, n = 14 stream-spawning) were retained for the Saiko-specific dataset. For NewHybrids analysis, we retained only the 500 SNPs with the highest F_{ST}. All individuals possessed STRUCTURE membership coefficients (0.80-1.00) to their population of origin. Similarly, 33/34 were assigned to their pure stock of origin based on NewHybrids analysis. There was one sample (deep-spawning kokanee) that was assigned to B2-deep-spawning kokanee backcross with a probability of 0.99849.

3.4 Discussion

The evolutionary history of kokanee has been the subject of continued study, revealing polyphyletic origins (Frazer & Russello, 2013; Taylor et al., 1996), genetic distinctiveness from anadromous sockeye salmon (Beacham & Withler, 2017) and a genomic basis for reproductive ecotype variation across their entire distribution (Lemay & Russello, 2015; Veale & Russello,

2017a, 2017b). To complement this growing body of literature, our study contributes important insights into population history and divergence of the less studied deep-spawning kokanee ecotype across its pan Pacific range.

3.4.1 Parallel genetic evolution

In salmonids, diverse life-history variation results from environmental fluctuations that create conditions for local adaptation (Waples, Pess, & Beechie, 2008). When distinct populations adapt to similar habitats, selective forces might act on the same set of traits, resulting in the evolution of similar phenotypes (Østbye et al., 2006). For example, in whitefish, adaptation to benthic or pelagic niches is associated with divergence in multiple traits, including diet composition, morphological parameters (e.g. number of gill rakers), growth rate, as well as age and size at maturity (Østbye et al., 2006). For *Oncorhynchus* spp., spawning habitat preference depends on many environmental variables, such as the flow of oxygenated water and the presence of substrate necessary for redd formation (Arostegui & Quinn, 2019). Although limited information exists on spawning ground availability in lakes, all deep-spawning kokanee populations share one defining characteristic: spawning depth. For this ecotype, the loss of bright red pigmentation in mature fish has been proposed to be an adaptation to decrease visibility to predators in deep waters with diminished light penetration or might have evolved due to trade-offs in resource allocation between pigmentation and other costly metabolic processes (Moreira & Taylor, 2015). However, the genetic basis of traits associated with deep-spawning behaviour has not yet been examined.

Identification of candidate genetic regions underlying parallelisms is an approach that can enhance our understanding of the repeatability of evolution, as well as aid in predicting the adaptive response of wild populations to different environmental conditions (Erickson et al.,

2016; Stinchcombe & Hoekstra, 2008; van Boheemen & Hodgins, 2020). In the last two decades, there has been an upsurge in the number of studies that employed genomic scans to identify molecular mechanisms responsible for the evolution of similar phenotypic traits in distinct lineages (Fraser & Whiting, 2020). In some systems, genomic scans have revealed specific regions associated with ecotype divergence: for instance, in sticklebacks, outliers that were highly divergent between limnetic and benthic ecotypes have been localized to genes associated with immune response (Jones et al., 2012). Likewise, the difference in genotype frequencies at the leucine-rich repeat-containing protein 9 (LRRC9) gene has been associated with reproductive ecotype divergence in *O. nerka* (Nichols et al., 2016; Veale & Russello, 2017a). Despite these examples, in a large proportion of studies, overlapping outliers identified between pairs of ecotypes remain rare, with the majority of outliers specific to geographic differentiation or population history of the system in question (Frazer & Russello, 2013; Perrier et al., 2013; Rougemont et al., 2017; Salisbury et al., 2020).

Similarly, our study found little overlap in outliers underlying ecotype divergence in deep-spawning and stream-spawning kokanee in Japan and Canada. Of the two outliers that were detected in both Japan- and Canada-specific outlier scans by at least one method (SZNR01029850.1_410525, SZNR01029883.1_386474), the latter did not produce significant annotations, and the former mapped to an uncharacterized region within the *O. mykiss* genome. Even solely within British Columbia, no shared outliers were detected as divergent between different population pairs of deep-spawning and stream-spawning kokanee. The one instance of where outlier loci detected in Japan and British Columbia independently localized to the same region most likely lacks biological significance, as the annotation was computationally predicted

(PREDICTED: *Oncorhynchus kisutch* inactive phospholipid phosphatase 7 (LOC109868517) and of questionable function.

The low number of outliers shared between ecotype pairs in our study can have multiple explanations. First, it is possible that *O. nerka* diversification into the deep-spawning kokanee happened independently and through different molecular processes that resulted in the evolution of morphologically and behaviourally similar populations. Moreover, if many genes with a small effect, rather than a single locus with a large effect, are responsible for the evolution of a certain phenotypic trait, identification of such genes might not be feasible via genome scans (Rockman, 2012). The absence of any lakes where deep-spawning and stream-spawning kokanee naturally co-occur further complicates the situation. Outlier detection between the two ecotypes in Saiko Lake constitutes a comparison of two populations that were artificially introduced into the ecosystem (Muto et al., 2013; Nakabo et al., 2011), and hence might be confounded (see below). Lastly, a small effective population size, and consequently reduced genetic diversity of deep-spawning kokanee in Saiko Lake might have amplified the impact of genetic drift in this population, masking the presence of shared outliers (Fraser & Whiting, 2020).

In addition to the lack of common outlier loci, analyses based on the putatively neutral dataset of 9,021 SNPs demonstrate that deep-spawning kokanee populations are more closely related to stream-spawning kokanee on the same continent, than they are to deep-spawning kokanee populations across the Pacific Ocean. These findings are consistent with the multiple origin hypothesis suggested by Moreira & Taylor (2015), which is in line with the polyphyletic origin of the resident form of sockeye salmon (Taylor et al., 1996).

3.4.2 Regional genetic diversity and divergence of deep-spawning kokanee

In British Columbia, *O. nerka* population divergence primarily reflects differentiation by geography. The three known deep-spawning kokanee populations in British Columbia constitute natural stocks that have not been subject to translocations and hence represent a unique opportunity to study divergence of this ecotype in the absence of human intervention. Based on putatively neutral data alone, differentiation between deep-spawning populations in British Columbia is best explained by geography rather than life history, as evidenced by distinct STRUCTURE clustering and reconstructed relationships based on the TreeMix analysis. Geographically proximate populations are less genetically distinct (e.g. Anderson and Seton Lake kokanee). In contrast, populations that have the same reproductive ecotype but are geographically isolated demonstrate a high level of differentiation (e.g. Anderson Lake and East Barrière Lake kokanee) (Figures 3.2-3.3; Table 3.2). These results are concordant with previous population genetic and genomic studies of kokanee reproductive ecotypes on multiple scales (Beacham & Withler, 2017; Nichols et al., 2016; Veale & Russello, 2017b; Samad-zada et al., in prep).

In Japan, on the other hand, *O. nerka* divergence patterns are more consistent with the documented history of kokanee transplantations (Kogura et al., 2011). Kokanee in Japan constitute a valuable commercial fishery and, due to their economic and recreational value, are subject to on-going translocation and supplementation efforts (Yamamoto, Kitamura, Sakano, & Morita, 2011). Here, we analyzed three Japanese kokanee populations, spanning two different lakes and representing two ecotypes (Table 3.1); all three populations were artificially introduced into the lakes they currently inhabit. All stream-spawning kokanee in Japan, including the contemporary Saiko Lake population, have their origin in Akan Lake. More specifically, Saiko

Lake stream-spawning kokanee were transplanted from Tawada or Shikotsu Lakes (Muto et al., 2013). However, they were initially introduced to these lakes from Akan Lake (Kogura et al., 2011), explaining the absence of genetic differentiation between Saiko Lake and Akan Lake stream-spawning kokanee. Similarly, deep-spawning kokanee in Saiko Lake are the progeny of transplants that were relocated from Tazawa Lake in 1935 (Nakabo et al., 2011). Fine-scale population structure analyses of the two ecotypes in Saiko Lake demonstrated high differentiation between stream-spawning and deep-spawning kokanee (Figures 3.2-3.3), concordant with previous findings in this system based on microsatellite analyses (Muto et al., 2013). Furthermore, 97% of Saiko Lake kokanee were assigned to their correct respective genotype frequency classes, suggesting that deep- and stream-spawning kokanee in Saiko Lake likely represent reproductively isolated populations. Previous research has attributed this divergence to the temporal and spatial distribution of the two ecotypes (Muto et al., 2013), which is a known driver of reproductive isolation in *O. nerka* (Arostegui & Quinn, 2019). However, the presence of one individual that was assigned to the B2 deep-spawning kokanee backcross demonstrates that there exists a potential for hybridization. These results should be considered in the context of used dataset because SNP panels vary in their assignment accuracy and power (Elliott & Russello, 2018). Previous research in a different system demonstrated that the 500 highest F_{ST} SNP dataset is sufficient to correctly assign backcrosses (Elliott and Russello, 2018) but the applicability of this to Saiko Lake kokanee requires additional testing. Hybridization was previously reported in Motosu Lake (Nakayama et al., 2018), the only other known location in the world where deep- and stream-spawning kokanee co-occur. These results need to be considered for effective conservation strategies of the unique deep-spawning kokanee ecotype.

In terms of genetic diversity, all three Japanese kokanee populations exhibited unusually low estimates of N_e and $H_{o,}$ compared to populations in British Columbia. Moreover, all Japanese populations demonstrated evidence for inbreeding, as revealed by positive F_{IS} coefficients that were significantly different from 0 (Table 3.1). These values are indicative of reduced genetic diversity that is often observed in transplanted populations, even in cases when a viable population is successfully established. Loss of diversity is characteristic of a founder effect, as the subset chosen for transplantation rarely represents the diversity of an entire donor stock and not all transplants survive in the new locality (Quinn, Graynoth, Wood, & Foote, 1998).

Lastly, our results showed that all Japanese populations were shore-spawners based on genotypes at the LRRC9 locus. Limited information exists regarding the original reproductive ecotype of kokanee in Japan (Nakayama, pers. comm.), which is attributed to repeated transplantations events. Still, several factors might explain the shore-spawning genotype of populations examined in this study, even those that were originally described as stream-spawners elsewhere (Moreira & Taylor, 2015). The contemporary kokanee population in Akan Lake consists of both stream- and shore-spawning individuals but they belong to the same population and do not seem to be reproductively isolated (Nakayama, pers. comm.). It is also unclear which ecotype was transplanted to Saiko Lake (Nakayama, pers. comm.). Alternatively, some studies propose that acidification of lacustrine waters can impact the reproductive behaviour of kokanee during spawning (Ikuta, Suzuki, & Kitamura, 2003), which might introduce additional selective pressure on already genetically depauperate populations. Nevertheless, information on the spawning behaviour of Japanese kokanee remains scarce, which prevents us from making any definitive conclusions regarding the history of reproductive (shore- or stream-spawning) ecotype(s) in these systems.

3.5 Implications

To summarize, we found no evidence for parallel genetic signature underlying deep-spawning ecotype divergence in Canada and Japan, based on the low number of shared outliers. However, we recognize that the usage of reduced representation sequencing might not be suitable for identifying candidate genes under selection, as these approaches only capture a fraction of the genome. Given that the sockeye salmon genome is estimated to be 2.6 Gbp (Christensten et al., 2020), the probability that a gene or genes underlying a certain phenotypic trait will be found among the several thousand examined markers is low. Therefore, a thorough understanding of the genomic basis of deep-spawning kokanee requires a whole-genome analysis, ideally coupled with an extended genomic context, such as mutation and recombination rates. Furthermore, genomic scans alone might not be sufficient for identifying statistically significant gene candidates, if landscape and demographic-specific factors are not considered (Fraser & Whiting, 2020). Alternatively, direct investigation of the functional significance of targeted genes hypothesized to play a role in deep-spawning kokanee morphology or behaviour might shed more light on the evolution of this ecotype. For example, beta-carotene oxygenase 2-like gene is associated with red pigmentation in Chinook salmon (Lehnert et al., 2019) and thus might be a potential candidate for divergent selection in deep-spawning kokanee. Divergent selection might also act on genes associated with adaptations to deep water conditions, such as haemoglobins (oxygen-binding and transport) or rhodopsins (visual perception in low-light conditions) (Hahn, Genner, Turner, & Joyce, 2017).

In conclusion, our data demonstrate that the population structure of the eight kokanee populations examined in this study is associated with geographic isolation or translocation history rather than ecotype. Our results show a high level of divergence between kokanee populations in Canada and Japan, as well as between deep-spawning and stream-spawning

kokanee populations regionally, suggesting independent evolution of deep-spawning ecotype on the two continents. We also detected a strong heterozygote deficiency and inbreeding levels in all Japanese kokanee populations, likely associated with documented transplantation history, which should be considered within the context of on-going fisheries management in these systems.

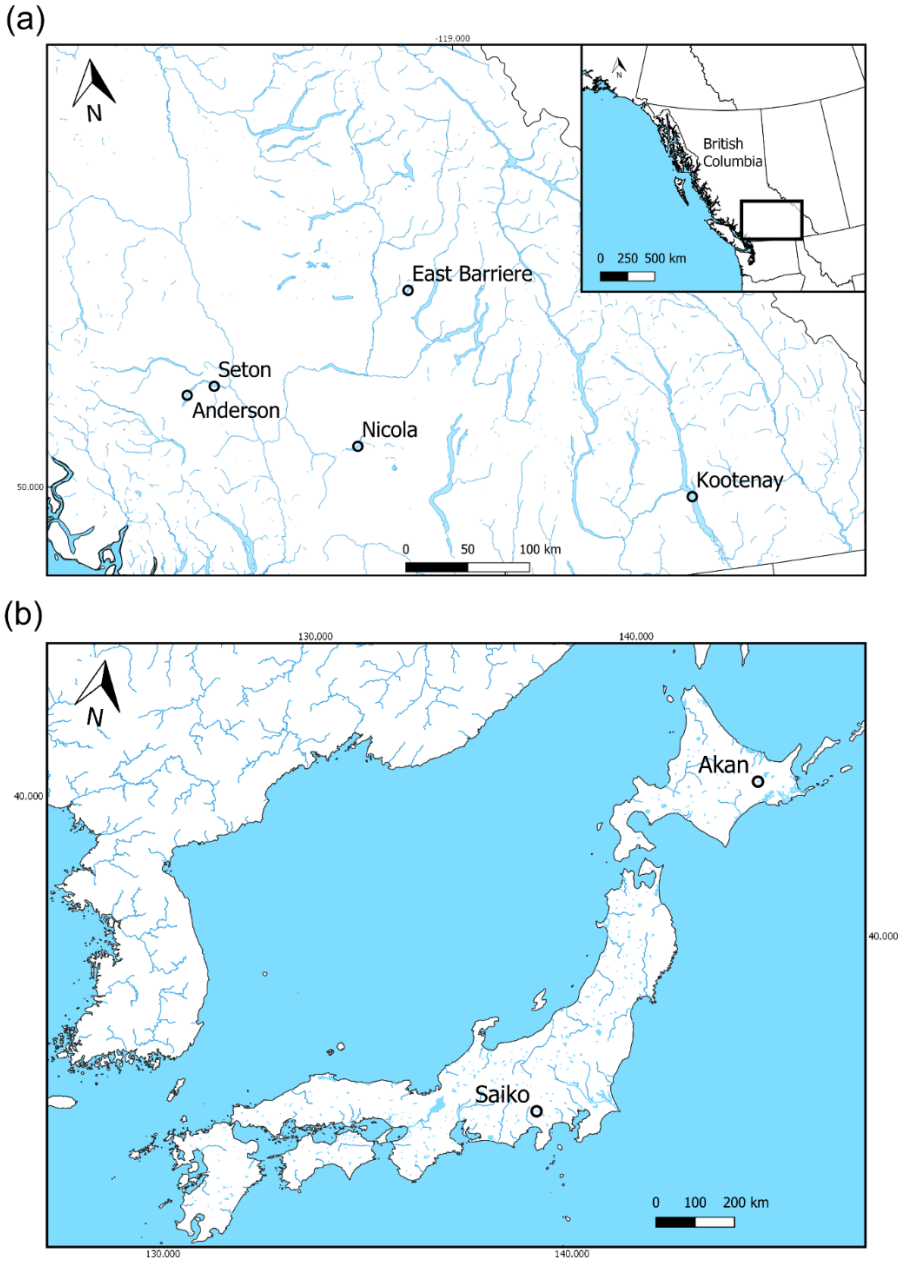

Figure 3.1. Locations of the study lakes. Maps produced using QGIS Geographic Information System, Open Source Geospatial Foundation Project (http://qgis.org). (a) Map of the Fraser River drainage, British Columbia, Canada. Shapefiles were retrieved from B.C. Data Catalogue (https://catalogue.data.gov.bc.ca/dataset). (b) Map of Japan. Shapefiles were retrieved from FAO Geonetwork (http://www.fao.org/geonetwork/srv/en/main.home).

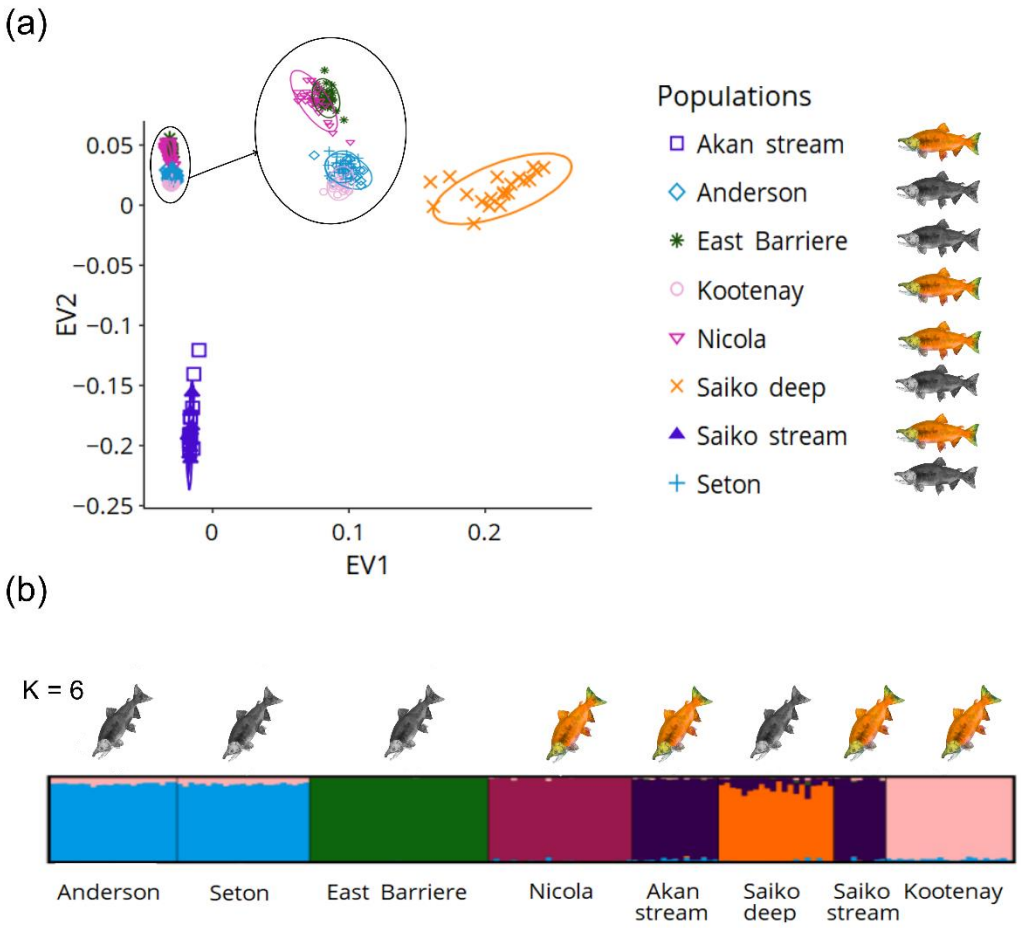

Figure 3.2. (a) PCA results, produced using R package *SNPRelate* (Zheng et al., 2012), showing population clustering based on the first two eigenvectors that explain 44.4% and 13.6% of the variation, respectively. (b) Results of Bayesian clustering method, as implemented in STRUCTURE v3.4 (Pritchard et al., 2000). Output results represent the optimal K value (K = 6), as determined by the ΔK method (Evanno et al., 2005), as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). Visualized using CLUMPAK (Kopelman et al., 2015). Orange fish represent stream-spawning kokanee, black fish represent deep-spawning kokanee. Fish illustrations are a courtesy of Eileen Klatt.

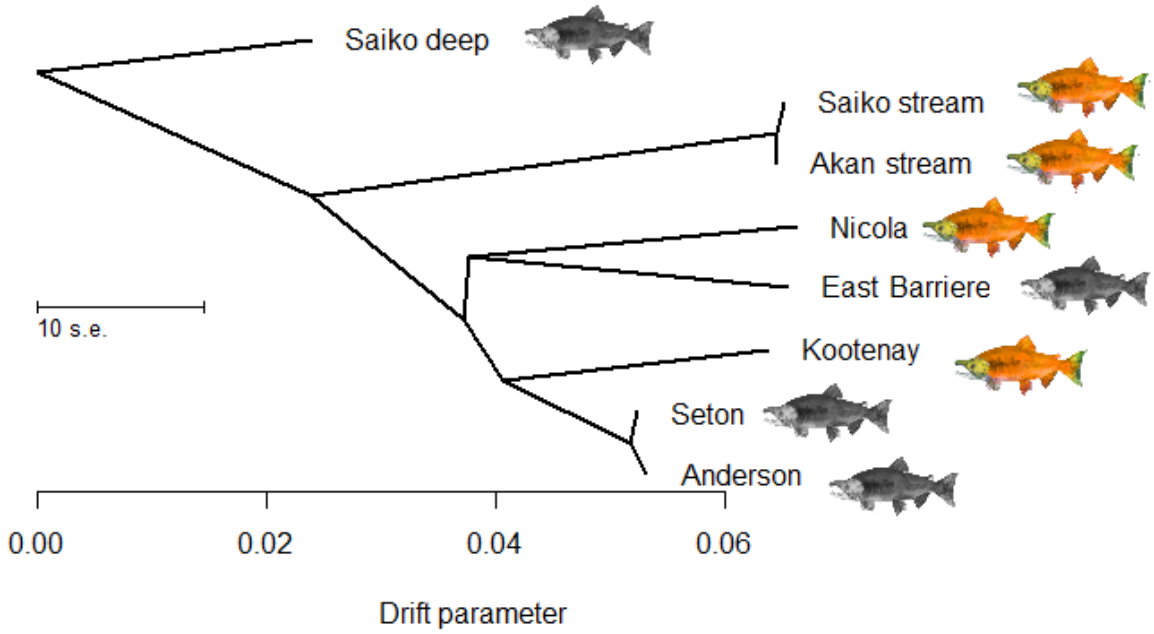

Figure 3.3. Maximum likelihood (ML) tree generated by TreeMix (Pickrell & Pritchard, 2012), based on the pruned dataset of 901 putatively neutral SNPs. ML Tree was rooted at Saiko deep-spawning population. Orange fish represent stream-spawning kokanee, black fish represent deep-spawning kokanee. Fish illustrations are a courtesy of Eileen Klatt.

Table 3.1. Sample size, spawner type, summary of the diversity statistics, and estimated Ne for the eight populations examined in this study. *FIS* values that were significantly different from 0 at $\alpha = 0.01$ are indicated by **.

Population	**Spawner type**	**N_S**	**N_R**	***Ho***	***He***	***FIS***	**Ne (95% CI)**
Anderson	deep-spawning	22	22	0.2153	0.212	-0.01573	2287.8 (1765.9-3245.4)
Seton	deep-spawning	23	23	0.2174	0.2145	-0.01314	1463.4(1247.2 - 1769.9)
East Barrière	deep-spawning	31	31	0.2017	0.1921	-0.04993	814.5(753.7-886.0)
Nicola	stream-spawning	25	25	0.1685	0.1679	-0.00369	4194.2(2594.0-10914.4)
Akan	stream-spawning	21	15	0.0769	0.09	0.14569**	223.3(196.6-258.1)
Saiko	deep-spawning	20	20	0.1126	0.1212	0.07087**	30.4 (30.1-30.6)
Saiko	stream-spawning	17	9	0.0737	0.0862	0.14437**	33.2 (32.1-34.4)
Kootenay	stream-spawning	22	22	0.1869	0.1888	0.01050	2101.7(1615.0-3006.2)

Abbreviations: number of sequenced samples (N_S), number of samples in the main SNP dataset retained after full filtering (N_R), confidence interval (CI), expected heterozygosity (H_e), observed heterozygosity (H_o), inbreeding coefficient (F_{IS}), and effective population size (N_e).

Table 3.2. Weir & Cockerham's (1984) θ estimates between each pair of populations based on 1000 permutations, as calculated by Genetix. ** indicate values that are significant at α = 0.01.

Population	Seton	East Barrière	Nicola	Akan stream-spawning	Saiko deep-spawning	Saiko stream-spawning	Kootenay
Anderson	0.00955**	0.17221**	0.21146**	0.35993**	0.56557**	0.34399**	0.13696**
Seton	-	0.17046**	0.20593**	0.35616**	0.562**	0.34044**	0.13518**
East Barrière	-	-	0.25303**	0.4079**	0.58897**	0.39593**	0.22313**
Nicola	-	-	-	0.46237**	0.62815**	0.44937**	0.25524**
Akan stream-spawning	-	-	-	-	0.7005**	-0.00031	0.40332**
Saiko deep-spawning	-	-	-	-	-	0.69688**	0.59703**
Saiko stream-spawning	-	-	-	-	-	-	0.38911**

Chapter 4: Conclusion

4.1 Overview and significance

Sockeye salmon represent a valuable model for studying ecotype divergence and the genomic basis of local adaptation due to an unusually high level of morphological and behavioural variation that is exhibited by this species. Furthermore, due to the high economic value of salmonids, research that focuses on *Oncorhynchus* can be of direct value for conservation planning purposes, especially for populations that are facing a threat of decline or extirpation. To complement the expanding body of literature on the topic of life history variation in salmonids, our study contributes important insights into population history and divergence of the rare and less studied deep-spawning ecotype of *O. nerka.* Our work offers valuable insights into the genetic basis of deep-spawning ecotype on multiple scales. First, this is the first genome-wide analysis of deep-spawning *O. nerka* in Alouette Lake, which represents a unique system where construction of a dam caused major changes in demography and life history. Second, this is the first genome-wide analysis of deep-spawning kokanee that combines populations across Canada and Japan.

4.2 Alouette Lake *O. nerka*: significance and limitations

4.2.1 Research findings and fisheries management implications

Overall, our findings demonstrate that in the Fraser Basin *O. nerka* differentiation is associated with geography rather than ecotype, with Alouette Lake *O. nerka* being highly differentiated from all other populations examined in this study. We found no evidence of differentiation between resident and migrant *O. nerka* in Alouette Lake at the neutral, genome-wide loci. High levels of differentiation relative to other examined Fraser River populations taken together with small estimated effective population size indicate the strong influence of a

genetic drift, consistent with a bottleneck scenario hypothesized by previous research in this system (Godbout et al., 2011; Samarasin et al., 2017). These findings also provide support for Alouette Lake *O. nerka* representing a landlocked sockeye population that has likely undergone significant declines after the completion of Alouette Dam. The nutrient restoration program that has been in effect since 1999 increased productivity of the lake and caused an upsurge of the *O. nerka* numbers present in the lake (Harris et al., 2011; Scott et al., 2017; van Poorten et al., 2018; Vainionpaa et al., 2020). Nonetheless, the current effective population size of Alouette Lake *O. nerka* remains low. Reduced genetic diversity is typically associated with decreased population fitness and lower survival rates under stressful conditions (Markert et al., 2010). If adaptive variation specific for a certain population is irreparably lost, this might decrease its response potential in the face of future disturbance events, such as a warming climate (Garcia de Leaniz et al., 2007). Given this information, conservation efforts that focus on Alouette Lake *O. nerka* need to consider not just the current population size but also remaining genetic diversity of this population.

Furthermore, our results demonstrate that migrant and resident *O. nerka* in Alouette Lake could be differentiated by outlier loci, all of which are located on sex chromosomes and overlap with outlier loci divergent between male and female resident *O. nerka*. Based on these findings, we suggested that alternative selective pressures may exist for male and female *O. nerka* in populations that have access to migration, in contrast to what is observed in solely resident populations. Our study suggests that a shift in selective pressures resulting in different evolutionary trajectories for male and female sockeye salmon can be an overlooked consequence of river obstruction and habitat fragmentation. Considering that the main advantage of anadromy is increased feeding opportunities, theory predicts that anadromous behaviour should occur in

cases where benefits of migration are higher than predation and disease costs encountered en route or during the at-sea rearing stage (Økland, Jonsson, Jensen, & Hansen, 1993). Moreover, the physiological transition that enables salmon to live in a saltwater environment is also costly, and mechanisms that trigger smoltification are usually the opposite of those that promote maturation (Foote, Mayer, Wood, Clarke, & Blackburn, 1994). For example, in residual sockeye salmon population in Cultus Lake, fast-growing individuals were more likely to mature in the lake and slow-growing individuals were more likely to smolt and migrate to sea (Jonsson & Jonsson, 1993; Ricker, 1938). To generalize these observations, it appears that if a residual individual can grow to a certain size and mature without the marine phase, then it will forego migration. It has also been shown that maturation rates tend to be faster for *O. mykiss* populations that are reared in environments with abundant food resources (Tipping & Byrne, 1996). Given the nutrient enrichment program that is in place in the Alouette Lake that was likely responsible for the population upsurge of the *O. nerka* population, it is likely that migration in this system is determined in part by food availability, and subsequently growth, maturation rates and sexual conflict.

To summarize, our study indicates that *O. nerka* in Alouette Lake is a landlocked sockeye population that residualized after the dam was constructed. We conclude that efforts to provide passage for migrant individuals are justified. However, conservation decisions regarding migration in this system would be incomplete if sex-specific factors are not considered. We demonstrated that at the identified migrant-resident outlier loci, migrant *O. nerka* and female resident *O. nerka* demonstrate heterozygote deficiency, compared to male resident *O. nerka*. These results indicate that when presented with passage opportunity, it is either female *O. nerka* or *O. nerka* with specific genotypes that are more likely to migrate. Deciphering between these

two possibilities would require obtaining sex information for migrant individuals (both juveniles at the time of downstream migration and adults at the time of return). It should, however, be noted that Christensen et al. (2020) also observed higher heterozygosity in male sockeye compared to female sockeye at Linkage Groups 9a and 9b. This might serve as indirect evidence that the majority of migrant *O. nerka* in this study were female. If the sex ratio of migrant *O. nerka* is indeed biased towards females, this can have direct consequences for conservation planning. For instance, a possible conservation strategy would be to decrease food availability in Alouette Lake that might "force" more males to migrate to the ocean in search of feeding opportunities.

4.2.2 Limitations and future studies

To expand on our understanding of migration in Alouette Lake *O. nerka*, future studies focusing on evolutionary dynamics within this system should further investigate the link between migration and sex. Such a study would greatly benefit from sex information of migrant individuals, and ideally extended sex data of the resident individuals spanning a more diverse temporal period. No sex data were recorded for AUM and JDM samples at the time they were collected, likely due to concerns regarding the invasiveness of the sexing procedure. An alternative approach would be to use molecular sexing methods on samples that were already collected. An example of molecular sexing method in salmon is genotyping at the master sex determining region, *sdY* (Larson, McKinney, Seeb, & Seeb, 2016), however, the currently available methodology does not assign correct phenotypic sex in 100% of the cases, and hence is not applicable for all *O. nerka* populations.

Additionally, further investigation into the genomic basis of migratory behaviour of Alouette Lake *O. nerka* might benefit from gene-environment association studies coupled with

transcriptome analyses. Subjecting *O. nerka* juveniles to conditions with varying levels of food availability might demonstrate whether migratory tendency in this system depends on environmental factors, and comparative transcriptome analysis might reveal genes showing differential expression between resident and migrant individuals. These approaches have been used in understanding the basis of migratory tactics in different systems. For instance, Wysujack, Greenberg, Bergman, & Olsson (2009) demonstrated that low food environment increases the proportion of migratory individuals in the brown trout population, and Miller et al. (2011) used gene expression profiles to predict survival and spawning success of sockeye salmon following migration.

4.3 Deep-spawning kokanee: findings and significance

4.3.1 Research findings and implications

In Chapter 3, we demonstrate that divergence patterns in deep-spawning kokanee populations examined in this study are associated with geography or translocation history rather than ecotype. These results are concordant with previous research on ecotype divergence in *O. nerka*, both at the migratory (anadromous/resident), and reproductive (shore-/stream-spawning) levels (Veale & Russello, 2017b). These findings have multiple implications. First, differentiation patterns demonstrated in our study are consistent with the hypothesis that isolated deep-spawning kokanee populations in Canada and Japan evolved independently, rather than from a common deep-spawning ancestor (Moreira & Taylor, 2015). However, the exact evolutionary trajectory of this ecotype remains unclear due to the low number of known deep-spawning kokanee populations and absence of known lakes, where deep-spawning kokanee naturally co-occur with other kokanee ecotypes. Second, our study found a very low number of common outliers between deep-spawning and stream-spawning paired kokanee populations in

Canada and Japan. This provides further support for the independent evolution hypothesis and suggests that different molecular pathways may have resulted in the evolution of a similar phenotype in different parts of the world.

Our study was not designed to address any specific management aims regarding the conservation of the deep-spawning ecotype. However, several conclusions that emerged from this analysis might be important for conservation. All three Japanese kokanee populations examined in this study demonstrated low levels of heterozygosity and positive inbreeding values, indicative of population bottleneck and reduced genetic diversity. As mentioned previously, genetic diversity has direct consequences for fitness and survival, and this might be especially critical for populations inhabiting freshwater habitats (Garcia de Leaniz et al., 2007). These findings are not surprising when the translocation history of kokanee populations is taken into account (Kogura et al., 2011).

4.3.2 Future studies

Currently, limited information exists on the subject of genetics of deep-spawning kokanee. While reduced representation markers, such as those used in this study, constitute a great tool for examining population structure and conducting exploratory genomic scans, they only capture a fraction of the genome. Given that the sockeye salmon genome is estimated to be 2.6 Gbp (Christensten et al., 2020), the probability that a gene or genes underlying a certain phenotypic trait will be found among the several thousand examined markers is low. Therefore, a thorough understanding of the genomic basis of deep-spawning kokanee would require a whole-genome analysis. Alternatively, direct investigation of the functional significance of targeted genes hypothesized to play a role in deep-spawning kokanee morphology or behaviour might shed more light on the evolution of this ecotype. For example, the beta-carotene oxygenase 2-

like gene is associated with red pigmentation in Chinook salmon (Lehnert et al., 2019) and thus might be a potential candidate for divergent selection in deep-spawning kokanee. Divergent selection might also act on genes associated with adaptations to deep water conditions, such as haemoglobins (oxygen-binding and transport), or rhodopsins (visual perception in low-light conditions) (Hahn et al., 2017). This approach might involve conducting a transcriptome analysis to quantify gene expression of the targeted genes in various tissues between deep-spawners and kokanee that spawn at a shallower depth. Comparative transcriptome analysis is a powerful technique for identifying genes under divergent selection that has been successfully implemented in other systems. For instance, analysis of different migratory ecotypes in sticklebacks has uncovered candidate genes underlying salinity tolerance (Kusakabe et al., 2017), and comparison of various tissue types in cichlid fish led to identification of genes associated with body colouration (Ahi et al., 2020).

4.4 Ecotype divergence and candidate genes: the future

Identification of candidate outlier loci that were shown to be highly divergent between sockeye and kokanee populations, deep-spawning and stream-spawning populations, as well as Alouette Lake resident and migrant individuals adds to the growing literature on the genomic basis of ecotype divergence and genetic basis of migration (Larson et al., 2017, 2019; Larson, McKinney, Seeb, & Seeb, 2016; Lemay & Russello, 2015; Nichols et al., 2016; Veale & Russello, 2017b). These loci can contribute to the design of panels that can be applied for ecotype or stock assignment and, coupled together with novel methods in SNP panel design, can be highly valuable for conservation management. Additionally, identification of outlier loci offers a list of candidate genes that can be directly targeted by studies investigating molecular processes that underlie migratory and reproductive variation in *O. nerka*. The availability of the

sockeye salmon genome (Christensen et al., 2020) will likely improve the existing functional annotation in the near future providing a better biological context for identified outliers, eventually leading to more insights regarding the complex evolution of ecotype divergence in *O. nerka*.

References

Ahi, E. P., Lecaudey, L. A., Ziegelbecker, A., Steiner, O., Glabonjat, R., Goessler, W., … Sefc, K. M. (2020). Comparative transcriptomics reveals candidate carotenoid color genes in an East African cichlid fish. *BMC Genomics*, *21*(1), 54. https://doi.org/10.1186/s12864-020-6473-8

Almeida-Gomes, M., & Rocha, C. F. D. (2015). Habitat Loss Reduces the Diversity of Frog Reproductive Modes in an Atlantic Forest Fragmented Landscape. *Biotropica*, *47*(1), 113–118. https://doi.org/10.1111/btp.12168

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, *215*(403–410), 8.

Anderson, E. C., & Thompson, E. A. (2002). A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics*, *160*, 1217–1229.

Andreev, D., Kreitman, M., Phillips, T. W., Beeman, R. W., & ffrench-Constant, R. H. (1999). Multiple Origins of Cyclodiene Insecticide Resistance in *Tribolium castaneum* (Coleoptera: Tenebrionidae). *Journal of Molecular Evolution*, *48*(5), 615–624. https://doi.org/10.1007/PL00006504

Andrusak, H., & Morris, A. (2004). *Final survey results of select lakes in the Kamloops area inhabited by kokanee—Year 4*. Nelson, BC: Redfish Consulting Ltd.

Angilletta, M. J., Steel, A. E., Bartz, K. K., Kingsolver, J. G., Scheuerell, M. D., Beckman, B. R., & Crozier, L. G. (2008). Big dams and salmon evolution: changes in thermal regimes and their potential evolutionary consequences: Thermal adaptation of salmon. *Evolutionary Applications*, *1*(2), 286–299. https://doi.org/10.1111/j.1752-4571.2008.00032.x

Arantes, C. C., Fitzgerald, D. B., Hoeinghaus, D. J., & Winemiller, K. O. (2019). Impacts of hydroelectric dams on fishes and fisheries in tropical rivers through the lens of functional traits.

Current Opinion in Environmental Sustainability, *37*, 28–40. https://doi.org/10.1016/j.cosust.2019.04.009

Arostegui, M. C., & Quinn, T. P. (2019). Reliance on lakes by salmon, trout and charr (*Oncorhynchus*, *Salmo* and *Salvelinus*): An evaluation of spawning habitats, rearing strategies and trophic polymorphisms. *Fish and Fisheries*, *20*, 775–794. https://doi.org/10.1111/faf.12377

Arthington, A. H., Dulvy, N. K., Gladstone, W., & Winfield, I. J. (2016). Fish conservation in freshwater and marine realms: Status, threats and management: Fish Conservation in Freshwater and Marine Realms. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *26*(5), 838–857. https://doi.org/10.1002/aqc.2712

Aykanat, T., Rasmussen, M., Ozerov, M., Niemelä, E., Paulin, L., Vähä, J., … Primmer, C. R. (2020). Life-history genomic regions explain differences in Atlantic salmon marine diet specialization. *Journal of Animal Ecology*, *89*(11), 2677–2691. https://doi.org/10.1111/1365-2656.13324

Aykanat, T., Thrower, F. P., & Heath, D. D. (2011). Rapid evolution of osmoregulatory function by modification of gene transcription in steelhead trout. *Genetica*, *139*(2), 233–242. https://doi.org/10.1007/s10709-010-9540-2

Baggio, R. A., Araujo, S. B. L., Ayllón, D., & Boeger, W. A. (2018). Dams cause genetic homogenization in populations of fish that present homing behavior: Evidence from a demogenetic individual-based model. *Ecological Modelling*, *384*, 209–220. https://doi.org/10.1016/j.ecolmodel.2018.06.019

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Balcke, A. (2009). *Alouette River Sockeye Trap and Truck Maple Ridge, BC. 2008-2009.* Maple Ridge, BC: BC Hydro Bridge Coastal Fish and Wildlife Restoration Program. Retrieved from: http://a100.gov.bc.ca/appsdata/acat/documents/r40710/08.ALU.04_truck_trap_1388677798432_8673845708.pdf (Accessed 12/11/2020)

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., … Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*(7582), 405–408. https://doi.org/10.1038/nature16062

Baxter, B. E., & Bocking, R. C. (2006). *Field trials to assess Coho smolt migration success through the Alouette reservoir, 2005*. Burnaby, BC: BC Hydro Bridge Coastal Fish and Wildlife Restoration Program. Retrieved from: http://a100.gov.bc.ca/appsdata/acat/documents/r40680/06.ALU.02_Steelhead_smolt_1388416674665_8414187995.pdf (Accessed 11/13/2020)

Beacham, T. D., & Withler, R. E. (2017). Population structure of sea-type and lake-type sockeye salmon and kokanee in the Fraser River and Columbia River drainages. *PLOS ONE*, *12*(9), e0183713. https://doi.org/10.1371/journal.pone.0183713

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). *GENETIX 4.05, Windows TM software for population genetics.* Laboratoire génome, populations, interactions, CNRS UMR, 5000.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188.

Berejikian, B. A., Bush, R. A., & Campbell, L. A. (2014). Maternal Control over Offspring Life History in a Partially Anadromous Species, *Oncorhynchus mykiss*. *Transactions of the American Fisheries Society*, *143*(2), 369–379. https://doi.org/10.1080/00028487.2013.862181

Borick-Cunningham, G. (2018). *Alouette Watershed Sockeye-Fish Passage Feasibility Project Year 1. Prepared for: Fish and Wildlife Compensation Program*. Maple Ridge, BC: Alouette River Management Society.

Boyle, S. A., Zartman, C. E., Spironello, W. R., & Smith, A. T. (2012). Implications of habitat fragmentation on the diet of bearded saki monkeys in central Amazonian forest. *Journal of Mammalogy*, *93*(4), 959–976. https://doi.org/10.1644/11-MAMM-A-286.1

Bryant D., & Moulton V. (2003). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, *21*(2), 255–265. https://doi.org/10.1093/molbev/msh018

Burkhead, N. M. (2012). Extinction Rates in North American Freshwater Fishes, 1900–2010. *BioScience*, *62*(9), 798–808. https://doi.org/10.1525/bio.2012.62.9.5

Castro, L. F. C., Wilson, J. M., Gonçalves, O., Galante-Oliveira, S., Rocha, E., & Cunha, I. (2011). The evolutionary history of the stearoyl-CoA desaturase gene family in vertebrates. *BMC Evolutionary Biology*, *11*(1), 132. https://doi.org/10.1186/1471-2148-11-132

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). *Stacks*: Building and Genotyping Loci *De Novo* From Short-Read Sequences. *Genes|Genomes|Genetics*, *1*(3), 171–182. https://doi.org/10.1534/g3.111.000240

Chen, G., Selbie, D. T., Finney, B. P., Schindler, D. E., Bunting, L., Leavitt, P. R., & Gregory-Eaves, I. (2011). Long-term zooplankton responses to nutrient and consumer subsidies arising from migratory sockeye salmon (*Oncorhynchus nerka*). *Oikos*, *120*(9), 1317–1326. https://doi.org/10.1111/j.1600-0706.2011.19042.x

Chen, K.-Y., Euclide, P. T., Ludsin, S. A., Larson, W. A., Sovic, M. G., & Gibbs, H., Lisle. (2020). RAD-Seq Refines Previous Estimates of Genetic Structure in Lake Erie Walleye. *Transactions of the American Fisheries Society*, *149*, 159–173. https://doi.org/10.1002/tafs.10215

Cheptou, P.-O., Hargreaves, A. L., Bonte, D., & Jacquemyn, H. (2017). Adaptation to fragmentation: Evolutionary dynamics driven by human influences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1712), 20160037. https://doi.org/10.1098/rstb.2016.0037

Christensen, K. A., Rondeau, E. B., Minkley, D. R., Sakhrani, D., Biagi, C. A., Flores, A.-M., … Koop, B. F. (2020). The sockeye salmon genome, transcriptome, and analyses identifying population defining regions of the genome. *PLOS ONE*, *15*(10), e0240935. https://doi.org/10.1371/journal.pone.0240935

Coleman, R. A., Gauffre, B., Pavlova, A., Beheregaray, L. B., Kearns, J., Lyon, J., … Sunnucks, P. (2018). Artificial barriers prevent genetic recovery of small isolated populations of a low-mobility freshwater fish. *Heredity*, *120*(6), 515–532. https://doi.org/10.1038/s41437-017-0008-3

Connallon, T., & Chenoweth, S. F. (2019). Dominance reversals and the maintenance of genetic variation for fitness. *PLOS Biology*, *17*(1), e3000118. https://doi.org/10.1371/journal.pbio.3000118

Cooke, S., Paukert, C., & Hogan, Z. (2012). Endangered river fish: Factors hindering conservation and restoration. *Endangered Species Research*, *17*(2), 179–191. https://doi.org/10.3354/esr00426

Crossin, G. T., Hinch, S. G., Cooke, S. J., Welch, D. W., Patterson, D. A., Jones, S. R. M., … Farrell, A. P. (2008). Exposure to high temperature influences the behaviour, physiology, and survival of sockeye salmon during spawning migration. *Canadian Journal of Zoology*, *86*(2), 127–140. https://doi.org/10.1139/Z07-122

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davis, C. D., Garza, J. C., & Banks, M. A. (2017). Identification of multiple genetically distinct populations of Chinook salmon (*Oncorhynchus tshawytscha*) in a small coastal watershed. *Environmental Biology of Fishes*, *100*(8), 923–933. https://doi.org/10.1007/s10641-017-0616-z

Deagle, B. E., Jones, F. C., Chan, Y. F., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2012). Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1732), 1277–1286. https://doi.org/10.1098/rspb.2011.1552

Díaz-Arce, N., Arrizabalaga, H., Murua, H., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. *Molecular Phylogenetics and Evolution*, *102*, 202–207. https://doi.org/10.1016/j.ympev.2016.06.002

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214. https://doi.org/10.1111/1755-0998.12157

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., … Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, *81*(02), 163. https://doi.org/10.1017/S1464793105006950

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Erickson, P. A., Glazer, A. M., Killingbeck, E. E., Agoglia, R. M., Baek, J., Carsanaro, S. M., … Miller, C. T. (2016). Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks: REPEATABLE EVOLUTION IN BENTHIC STICKLEBACKS. *Evolution*, *70*(4), 887–902. https://doi.org/10.1111/evo.12897

Elliott, L., & Russello, M. A. (2018). SNP panels for differentiating advanced-generation hybrid classes in recently diverged stocks: A sensitivity analysis to inform monitoring of sockeye salmon re-stocking programs. *Fisheries Research*, *208*, 339–345. https://doi.org/10.1016/j.fishres.2018.09.001

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Ferguson, A., Reed, T., McGinnity, P., & Prodöhl, P. (2017). Anadromy in brown trout (*Salmo trutta*): A review of the relative roles of genes and environmental factors and the implications for management and conservation. In: G. Harris (Ed.), Sea trout: Management & science (pp. 1– 40). Leicestershire, England: Matador Publishing Ltd.

Foerster, R. E. (1930). The Reactions of Some Sockeye Salmon to a Power Dam. *Transactions of the American Fisheries Society*, *60*(1), 244–246.

Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Folmar, L., C., & Dickhoff, W., W. (1980). The parr-smolt transformation (smoltification) and seawater adaptation in salmonids. A review of selected literature. *Aquaculture*, *21*, 1–37.

Foote, C. J., Brown, G. S., & Wood, C. C. (1997). Spawning success of males using alternative mating tactics in sockeye salmon, *Oncorhynchus nerka*. *Canadian Journal of Fisheries and Aquatic Sciences*, *54*, 1785–1795.

Foote, C. J., Mayer, I., Wood, C. C., Clarke, W. C., & Blackburn, J. (1994). On the developmental pathway to nonanadromy in sockeye salmon, *Oncorhynchus nerka*. *Canadian Journal of Zoology*, *72*(3), 397–405. https://doi.org/10.1139/z94-056

Foote, C. J., Moore, K., Stenberg, K., Craig, K., J., Wenburg, J., K., & Wood, C. C. (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, C. J., Wood, C. C., & Withler, R. E. (1989). Biochemical Genetic Comparison of Sockeye Salmon and Kokanee, the Anadromous and Nonanadromous forms of *Oncorhynchus nerka*. *Canadian Journal of Fisheries and Aquatic Sciences*, *46*(1), 149–158. https://doi.org/10.1139/f89-020

Fraser, B. A., & Whiting, J. R. (2020). What can be learned by scanning the genome for molecular convergence in wild populations? *Annals of the New York Academy of Sciences*, *1476*(1), 23–42. https://doi.org/10.1111/nyas.14177

Fraser, D. J., Debes, P. V., Bernatchez, L., & Hutchings, J. A. (2014). Population size, habitat fragmentation, and the nature of adaptive variation in a stream fish. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1790), 20140370. https://doi.org/10.1098/rspb.2014.0370

Frazer, K. K., & Russello, M. A. (2013). Lack of parallel genetic patterns underlying the repeated ecological divergence of beach and stream-spawning kokanee salmon. *Journal of Evolutionary Biology*, *26*(12), 2606–2621. https://doi.org/10.1111/jeb.12250

Fry, J. D. (2010). The genomic location of sexually antagonistic variation: Some cautionary comments. *Evolution*. https://doi.org/10.1111/j.1558-5646.2009.00898.x

Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., … Quinn, T. P. (2007). A critical review of adaptive genetic variation in Atlantic salmon: Implications for conservation. *Biological Reviews*, *82*(2), 173–211. https://doi.org/10.1111/j.1469-185X.2006.00004.x

Garner, K., Parfitt, B., & Pacific Fisheries Resource Conservation Council. (2006). *First Nations, salmon fisheries and the rising importance of conservation*. Vancouver, B.C.: Pacific Fisheries Resource Conservation Council. Retrieved from http://www.deslibris.ca/ID/222930 (Accessed 11/13/2020)

Gende, S. M., Edwards, R. T., Willson, M. F., & Wipfli, M. S. (2002). Pacific Salmon in Aquatic and Terrestrial Ecosystems. *BioScience*, *52*(10), 917. https://doi.org/10.1641/0006-3568(2002)052[0917:PSIAAT]2.0.CO;2

Gislason, G., Lam, E., Knapp, G., & Guettabi, M. (2017). *Economic Impacts of Pacific Salmon Fisheries. Prepared for: Pacific Salmon Commission, Vancouver, Canada*. Anchorage, AK: Institute of Social & Economic Research University of Alaska Anchorage. Retrieved from https://www.psc.org/download/333/specialreports/9337/economic-impacts-of-pacific-salmonfisheries.pdf (Accessed 1/25/2021)

Godbout, L., Wood, C. C., Withler, R. E., Latham, S., Nelson, R. J., Wetzel, L., … McKeegan, K. D. (2011). 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*(9), 1590–1602. https://doi.org/10.1139/f2011-089

Goudet, J., & Jombart, T. (2015). *hierfstat*: Estimation and tests of hierarchical F-statistics. R package version 0.04-22. Retrieved from https:// CRAN.R-project.org/package=hierfstat (Accessed 1/25/2021)

Gross, J. B., Borowsky, R., & Tabin, C. J. (2009). A Novel Role for Mc1r in the Parallel Evolution of Depigmentation in Independent Populations of the Cavefish *Astyanax mexicanus*. *PLoS Genetics*, *5*(1), e1000326. https://doi.org/10.1371/journal.pgen.1000326

Gustafson, R. G., Waples, R. S., Myers, J. M., Weitkamp, L. A., Bryant, G. J., Johnson, O. W., & Hard, J. J. (2007). Pacific Salmon Extinctions: Quantifying Lost and Remaining Diversity. *Conservation Biology*, *21*(4), 1009–1020. https://doi.org/10.1111/j.1523-1739.2007.00693.x

Haag, T., Santos, A. S., Sana, D. A., Morato, R. G., Cullen Jr, L., Crawshaw Jr, P. G., … Eizirik, E. (2010). The effect of habitat fragmentation on the genetic structure of a top predator: Loss of diversity and high differentiation among remnant populations of Atlantic Forest jaguars (*Panthera onca*): Jaguar Conservation Genetics. *Molecular Ecology*, *19*(22), 4906–4921. https://doi.org/10.1111/j.1365-294X.2010.04856.x

Hahn, C., Genner, M. J., Turner, G. F., & Joyce, D. A. (2017). The genomic basis of cichlid fish adaptation within the deepwater “twilight zone” of Lake Malawi: Genomic basis of deep-water cichlid fish adaptation. *Evolution Letters*, *1*(4), 184–198. https://doi.org/10.1002/evl3.20

Haldane, J. B. S. (1932). *The Causes of Evolution*. London: Longmans Green.

Harris, S. L., Andrusak, H., Andrusak, G., Sebastian, D., Scholten, G., Weir, T., … Down, N. E. (2010). *The Alouette Reservoir nutrient restoration program, 2003-2008* (Fisheries project

Report No. RD 129). Province of British Columbia, Ministry of Environment, Aquatic Ecosystem Protection Branch.

Hébert, A. S. (2019). *Reproductive ecology and habitats of deep spawning Kokanee and Sockeye Salmon Oncorhynchus nerka* (University of British Columbia). University of British Columbia. Retrieved from https://open.library.ubc.ca/collections/ubctheses/24/items/1.0381033 (Accessed 12/19/2020)

Hecht, B. C., Campbell, N. R., Holecek, D. E., & Narum, S. R. (2013). Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, *22*(11), 3061–3076. https://doi.org/10.1111/mec.12082

Hirose, S., Narita, K., Asano, K., & Nakane, A. (2018). Salmon cartilage proteoglycan promotes the healing process of *Staphylococcus aureus*-infected wound. *Heliyon*, *4*(3), e00587. https://doi.org/10.1016/j.heliyon.2018.e00587

Hirst, S. M. (1991). Impacts of the operation of existing hydroelectric developments on fishery resources in British Columbia. Inland fisheries news. Volume 1. *Canadian Manuscript Report of Fisheries and Aquatic Sciences 2093*.

Holmlund, C. M., & Hammer, M. (1999). Ecosystem services generated by fish populations. *Ecological Economics*, *29*(2), 253–268. https://doi.org/10.1016/S0921-8009(99)00015-4

Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Ikuta, K., Suzuki, Y., & Kitamura, S. (2003). Effects of low pH on the reproductive behavior of salmonid fishes. *Fish Physiology and Biochemistry*, *28*, 407–410.

Jacob, C., McDaniels, T., & Hinch, S. (2010). Indigenous culture and adaptation to climate change: Sockeye salmon and the St'át'imc people. *Mitigation and Adaptation Strategies for Global Change*, *15*(8), 859–876. https://doi.org/10.1007/s11027-010-9244-z

Jenkins, M. (2003). Prospects for Biodiversity. *Science*, *302*(5648), 1175–1177. https://doi.org/10.1126/science.1088666

Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., … Kingsley, D. M. (2012). A Genome-wide SNP Genotyping Array Reveals Patterns of Global and Repeated Species-Pair Divergence in Sticklebacks. *Current Biology*, *22*(1), 83–90. https://doi.org/10.1016/j.cub.2011.11.045

Jonsson, B., & Jonsson, N. (1993). Partial migration: Niche shift versus sexual maturation in fishes. *Reviews in Fish Biology and Fisheries*, *3*(4), 348–365. https://doi.org/10.1007/BF00043384

Junge, C., Museth, J., Hindar, K., Kraabøl, M., & Vøllestad, L. A. (2014). Assessing the consequences of habitat fragmentation for two migratory salmonid fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *24*(3), 297–311. https://doi.org/10.1002/aqc.2391

Kaeriyama, M. (1996). Effects of Population Density and Habitat Environment on Life History Strategy and Migration of Juvenile Sockeye (*Oncorhynchus nerka*) and Chum Salmon (*O.keta*). *Scientific Reports of the Hokkaido Salmon Hatchery*, *50*, 101–111.

Kaeriyama, M., Urawa, S., & Fukuwaka, M. (1995). Variation in Body Size, Fecundity, and Egg Size of Sockeye and Kokanee Salmon, *Oncorhynchus nerka*, Released from Hatchery. *Scientific Reports of the Hokkaido Salmon Hatchery*, *49*, 1–9.

Kendall, N. W., McMillan, J. R., Sloat, M. R., Buehrens, T. W., Quinn, T. P., Pess, G. R., … Zabel, R. W. (2015). Anadromy and residency in steelhead and rainbow trout (*Oncorhynchus mykiss*):

A review of the processes and patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(3), 319–342. https://doi.org/10.1139/cjfas-2014-0192

Kim, N. N., Shin, H. S., Choi, Y. J., Yamamoto, Y., Fukaya, K., Ueda, H., & Choi, C. Y. (2013). Effect of hypo-osmotic environmental changes on the expression of gonadotropin-releasing hormone, its receptor, and gonadotropin hormone subunit mRNA in adult chum salmon (*Oncorhynchus keta*). *Marine and Freshwater Behaviour and Physiology*, *45*(6), 397–410. https://doi.org/10.1080/10236244.2013.777215

Kogura, Y., Seeb, J. E., Azuma, N., Kudo, H., Abe, S., & Kaeriyama, M. (2011). The genetic population structure of lacustrine sockeye salmon, *Oncorhynchus nerka*, in Japan as the endangered species. *Environmental Biology of Fishes*, *92*(4), 539–550. https://doi.org/10.1007/s10641-011-9876-1

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kusakabe, M., Ishikawa, A., Ravinet, M., Yoshida, K., Makino, T., Toyoda, A., … Kitano, J. (2017). Genetic basis for variation in salinity tolerance between stickleback ecotypes. *Molecular Ecology*, *26*(1), 304–319. https://doi.org/10.1111/mec.13875

Larson, W. A., Dann, T. H., Limborg, M. T., McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2019). Parallel signatures of selection at genomic islands of divergence and the major histocompatibility complex in ecotypes of sockeye salmon across Alaska. *Molecular Ecology*, *28*(9), 2254–2271. https://doi.org/10.1111/mec.15082

Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E., & Seeb, L. W. (2017). Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular Ecology*, *26*(2), 554–570. https://doi.org/10.1111/mec.13933

Larson, W. A., McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2016). Identification and Characterization of Sex-Associated Loci in Sockeye Salmon Using Genotyping-by-Sequencing and Comparison with a Sex-Determining Assay Based on the *sdY* Gene. *Journal of Heredity*, *107*(6), 559–566. https://doi.org/10.1093/jhered/esw043

Larson, W. A., Seeb, J. E., Dann, T. H., Schindler, D. E., & Seeb, L. W. (2014). Signals of heterogeneous selection at an MHC locus in geographically proximate ecotypes of sockeye salmon. *Molecular Ecology*, *23*(22), 5448–5461. https://doi.org/10.1111/mec.12949

Leaver, M. J., Villeneuve, L. A., Obach, A., Jensen, L., Bron, J. E., Tocher, D. R., & Taggart, J. B. (2008). Functional genomics reveals increases in cholesterol biosynthetic genes and highly unsaturated fatty acid biosynthesis after dietary substitution of fish oil with vegetable oils in Atlantic salmon (*Salmo salar*). *BMC Genomics*, *9*(1), 299. https://doi.org/10.1186/1471-2164-9-299

Lee, Y. W., Gould, B. A., & Stinchcombe, J. R. (2014). Identifying the genes underlying quantitative traits: A rationale for the QTN programme. *AoB PLANTS*, *6*, plu004. https://doi.org/10.1093/aobpla/plu004

Lehnert, S. J., Christensen, K. A., Vandersteen, W. E., Sakhrani, D., Pitcher, T. E., Heath, J. W., … Devlin, R. H. (2019). Carotenoid pigmentation in salmon: Variation in expression at *BCO2-l* locus controls a key fitness trait affecting red coloration. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1913), 20191588. https://doi.org/10.1098/rspb.2019.1588

Lemay, M. A., & Russello, M. A. (2015). Genetic evidence for ecological divergence in kokanee salmon. *Molecular Ecology*, *24*(4), 798–811. https://doi.org/10.1111/mec.13066

Lessard, J. L., & Hayes, D. B. (2003). Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. *River Research and Applications*, *19*(7), 721–732. https://doi.org/10.1002/rra.713

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liermann, C. R., Nilsson, C., Robertson, J., & Ng, R. Y. (2012). Implications of Dam Obstruction for Global Freshwater Fish Diversity. *BioScience*, *62*(6), 539–548. https://doi.org/10.1525/bio.2012.62.6.5

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., … Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Luu, K., Bazin, E., & Blum, M. G. B. (2017). *pcadapt*: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592

Macqueen, D. J., Kristjánsson, B. K., Paxton, C. G. M., Vieira, V. L. A., & Johnston, I. A. (2011). The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression: Gene expression evolution in fish populations. *Molecular Ecology*, *20*(15), 3167–3184. https://doi.org/10.1111/j.1365-294X.2011.05172.x

Maekawa, K., & Koseki, Y. (2001). Skewed reproductive success among male white-spotted charr land-locked by an erosion control dam: Implications for effective population size. *Ecological Research*, *16*, 727–735.

Mank, J. E. (2017). Population genetics of sexual conflict in the genomic era. *Nature Reviews Genetics*, *18*(12), 721–730. https://doi.org/10.1038/nrg.2017.83

Markert, J. A., Champlin, D. M., Gutjahr-Gobell, R., Grear, J. S., Kuhn, A., McGreevy, T. J., … Nacci, D. E. (2010). Population genetic diversity and fitness in multiple environments. *BMC Evolutionary Biology*, *10*(205), 1–13. https://doi.org/10.1186/1471-2148-10-205

Mattocks, S., Hall, C. J., & Jordaan, A. (2017). Damming, Lost Connectivity, and the Historical Role of Anadromous Fish in Freshwater Ecosystem Dynamics. *BioScience*, *67*(8), 713–728. https://doi.org/10.1093/biosci/bix069

McKinney, G. J., Hale, M. C., Goetz, G., Gribskov, M., Thrower, F. P., & Nichols, K. M. (2015). Ontogenetic changes in embryonic and brain gene expression in progeny produced from migratory and resident *Oncorhynchus mykiss*. *Molecular Ecology*, *24*(8), 1792–1809. https://doi.org/10.1111/mec.13143

McKinney, G. J., Pascal, C. E., Templin, W. D., Gilk-Baumer, S. E., Dann, T. H., Seeb, L. W., & Seeb, J. E. (2020). Dense SNP panels resolve closely related Chinook salmon populations. *Canadian Journal of Fisheries and Aquatic Sciences*, *77*(3), 451–461. https://doi.org/10.1139/cjfas-2019-0067

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669. https://doi.org/10.1111/1755-0998.12613

Meybeck, M. (2003). Global analysis of river systems: From Earth system controls to Anthropocene syndromes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1440), 1935–1955. https://doi.org/10.1098/rstb.2003.1379

Miller, K. M., Li, S., Kaukinen, K. H., Ginther, N., Hammill, E., Curtis, J. M. R., … Farrell, A. P. (2011). Genomic Signatures Predict Migration and Spawning Failure in Wild Canadian Salmon. *Science*, *331*(6014), 214–217. https://doi.org/10.1126/science.1196901

Miller, K. M., Schulze, A. D., Ginther, N., Li, S., Patterson, D. A., Farrell, A. P., & Hinch, S. G. (2009). Salmon spawning migration: Metabolic shifts and environmental triggers. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *4*(2), 75–89. https://doi.org/10.1016/j.cbd.2008.11.002

Miller, K. M., Teffer, A., Tucker, S., Li, S., Schulze, A. D., Trudel, M., … Hinch, S. G. (2014). Infectious disease, shifting climates, and opportunistic predators: Cumulative factors potentially impacting wild salmon declines. *Evolutionary Applications*, *7*(7), 812–855. https://doi.org/10.1111/eva.12164

Moore, J. W., Hayes, S. A., Duffy, W., Gallagher, S., Michel, C. J., & Wright, D. (2011). Nutrient fluxes and the recent collapse of coastal California salmon populations. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*(7), 1161–1170. https://doi.org/10.1139/f2011-054

Moreira, A. L., & Taylor, E. B. (2015). The origin and genetic divergence of “black” kokanee, a novel reproductive ecotype of *Oncorhynchus nerka*. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(10), 1584–1595. https://doi.org/10.1139/cjfas-2015-0145

Morita, K., Sahashi, G., Miya, M., Kamada, S., Kanbe, T., & Araki, H. (2019). Ongoing localized extinctions of stream-dwelling white-spotted charr populations in small dammed-off habitats of

Hokkaido Island, Japan. *Hydrobiologia*, *840*(1), 207–213. https://doi.org/10.1007/s10750-019-3891-1

Morita, K., Yamamoto, S., & Hoshino, N. (2000). Extreme life history change of white-spotted char (*Salvelinus leucomaenis*) after damming. *Canadian Journal of Fisheries and Aquatic Sciences*, *57*, 1300–1306.

Muto, N., Nakayama, K., & Nakabo, T. (2013). Distinct genetic isolation between "Kunimasu" (*Oncorhynchus kawamurae*) and "Himemasu" (*O. nerka*) in Lake Saiko, Yamanashi Prefecture, Japan, inferred from microsatellite analysis. *Ichthyological Research*, *60*(2), 188–194. https://doi.org/10.1007/s10228-012-0323-x

Nakabo, T., Nakayama, K., Muto, N., & Miyazawa, M. (2011). (*Oncorhynchus kawamurae(* "Kunimasu," a deepwater trout, discovered in Lake Saiko, 70 years after extinction in the original habitat, Lake Tazawa, Japan. *Ichthyological Research*, *58*(2), 180–183. https://doi.org/10.1007/s10228-011-0204-8

Nakabo, T., Tohkairin, A., Muto, N., Watanabe, Y., Miura, Y., Miura, H., … Kai, Y. (2014). Growth-related morphology of "Kunimasu" (*Oncorhynchus kawamurae*: Family Salmonidae) from Lake Saiko, Yamanashi Prefecture, Japan. *Ichthyological Research*, *61*(2), 115–130. https://doi.org/10.1007/s10228-013-0382-7

Nakayama, K., Tohkairin, A., Yoshikawa, A., & Nakabo, T. (2018). Detection and morphological characteristics of "Kunimasu" (*Oncorhynchus kawamurae*)/"Himemasu" (*O. nerka*) hybrids in Lake Motosu, Yamanashi Prefecture, Japan. *Ichthyological Research*, *65*(2), 270–275. https://doi.org/10.1007/s10228-018-0613-z

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841–2847. https://doi.org/10.1111/mec.12350

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press. New York, NY.

Nichols, K. M., Kozfkay, C. C., & Narum, S. R. (2016). Genomic signatures among *Oncorhynchus nerka* ecotypes to inform conservation and management of endangered Sockeye Salmon. *Evolutionary Applications*, *9*(10), 1285–1300. https://doi.org/10.1111/eva.12412

Økland, F., Jonsson, B., Jensen, A., J., & Hansen, L., P. (1993). Is there a threshold size regulating seaward migration of brown trout and Atlantic salmon? *Journal of Fish Biology*, *42*, 541–550. https://doi.org/10.1111/j.1095-8649.1993.tb00358.x

Østbye, K., Amundsen, P.-A., Bernatchez, L., Klemetsen, A., Knudsen, R., Kristoffersen, R., … Hindar, K. (2006). Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times: Parallel evolution in european whitefish. *Molecular Ecology*, *15*(13), 3983–4001. https://doi.org/10.1111/j.1365-294X.2006.03062.x

Palkovacs, E. P., Kinnison, M. T., Correa, C., Dalton, C. M., & Hendry, A. P. (2012). Fates beyond traits: Ecological consequences of human-induced trait change: Fates beyond traits. *Evolutionary Applications*, *5*(2), 183–191. https://doi.org/10.1111/j.1752-4571.2011.00212.x

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., … Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, *3*(12), 1731–1742. https://doi.org/10.1038/s41559-019-1044-6

Perrier, C., Bourret, V., Kent, M. P., & Bernatchez, L. (2013). Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Molecular Ecology*, *22*(22), 5577–5593. https://doi.org/10.1111/mec.12500

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Plate, E. M., Mathews, M. A., & Bocking, R. C. (2014). *Technical feasibility and recommendations for Alouette Lake sockeye salmon re-establishment above the Alouette dam*. BC Hydro Fish and Wildlife Compensation Program. Retrieved from: http://a100.gov.bc.ca/appsdata/acat/documents/r49028/13.ALU.02_1443026016059_3025327969.pdf (Accessed 11/13/2020)

Prince, D. J., O'Rourke, S., M., Thompson, T., Q., Ali, O., A., Lyman, H., S., Saglam, I. K., … Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, *3*, e1603198.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*, 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Quesada, H., Posada, D., Caballero, A., Morán, P., & Rolán-Alvarez, E. (2007). Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution*, *61*(7), 1600–1612. https://doi.org/10.1111/j.1558-5646.2007.00135.x

Quinn, T. P. (2005). *The Behavior and Ecology of Pacific Salmon and Trout*. Vancouver, B.C.: UBC Press.

Quinn, T. P., & Foote, C. J. (1994). The effects of body size and sexual dimorphism on the reproductive behaviour of sockeye salmon, *Oncorhynchus nerka*. *Animal Behaviour*, *48*, 751–761.

Quinn, T.P., Graynoth, E., Wood, C. C., & Foote, C. J. (1998). Genotypic and Phenotypic Divergence of Sockeye Salmon in New Zealand from Their Ancestral British Columbia Populations. *Transactions of the American Fisheries Society*, *127*, 517–534.

Quinn, T. P., Helfield, J. M., Austin, C. S., Hovel, R. A., & Bunn, A. G. (2018). A multidecade experiment shows that fertilization by salmon carcasses enhanced tree growth in the riparian zone. *Ecology*, *99*(11), 2433–2441. https://doi.org/10.1002/ecy.2453

Quinn, T. P., & Myers, K. W. (2004). Anadromy and the marine migrations of Pacific salmon and trout: Rounsefell revisited. *Reviews in Fish Biology and Fisheries*, *14*(4), 421–442. https://doi.org/10.1007/s11160-005-0802-5

R Core Team (2018). *R: A language and environment for statistical computing*.

Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/ (Accessed 2/1/2019)

Rand, P. S., Goslin, M., Gross, M. R., Irvine, J. R., Augerot, X., McHugh, P. A., & Bugaev, V. F. (2012). Global Assessment of Extinction Risk to Populations of Sockeye Salmon *Oncorhynchus nerka*. *PLoS ONE*, *7*(4), e34065. https://doi.org/10.1371/journal.pone.0034065

Ricker, W. E. (1938). “Residual” and kokanee salmon in Cultus lake. *Journal of the Fisheries Research Board of Canada*, *4*(3), 192–218.

Roberts, J. H., Angermeier, P. L., & Hallerman, E. M. (2013). Distance, dams and drift: What structures populations of an endangered, benthic stream fish? *Freshwater Biology*, *58*(10), 2050–2064. https://doi.org/10.1111/fwb.12190

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, *66*(1), 1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x

Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: Effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources*, *16*(4), 991–1001. https://doi.org/10.1111/1755-0998.12518

Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S., & Evanno, G. (2017). Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, *26*(1), 142–162. https://doi.org/10.1111/mec.13664

Russello, M. A., Kirk, S. L., Frazer, K. K., & Askey, P. J. (2012). Detection of outlier loci and their utility for fisheries management: Outlier loci for fisheries management. *Evolutionary Applications*, *5*(1), 39–52. https://doi.org/10.1111/j.1752-4571.2011.00206.x

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews Molecular Cell Biology*, *14*(10), 630–642. https://doi.org/10.1038/nrm3658

Salisbury, S. J., McCracken, G. R., Perry, R., Keefe, D., Layton, K. K. S., Kess, T., … Ruzzante, D. E. (2020). Limited genetic parallelism underlies recent, repeated incipient speciation in geographically proximate populations of an Arctic fish (*Salvelinus alpinus*). *Molecular Ecology*, mec.15634. https://doi.org/10.1111/mec.15634

Samarasin, P., Shuter, B. J., & Rodd, F. H. (2017). After 100 years: Hydroelectric dam-induced life-history divergence and population genetic changes in sockeye salmon (*Oncorhynchus nerka*). *Conservation Genetics*, *18*(6), 1449–1462. https://doi.org/10.1007/s10592-017-0992-0

Scott, D. C., Harris, S. L., Hébert, A. S., & van Poorten, B. T. (2017). Nutrient dynamics in a highly managed reservoir system: Considering anadromous sockeye salmon (*Oncorhynchus nerka*) and nutrient restoration. *Lake and Reservoir Management*, *33*(1), 14–22. https://doi.org/10.1080/10402381.2016.1247391

Scribner, K. T., Lowe, W. H., Landguth, E., Luikart, G., Infante, D. M., Whelan, G. E., & Muhlfeld, C. C. (2016). Applications of Genetic Data to Improve Management and Conservation of River Fishes and Their Habitats. *Fisheries*, *41*(4), 174–188. https://doi.org/10.1080/03632415.2016.1150838

Sheer, M. B., & Steel, E. A. (2006). Lost Watersheds: Barriers, Aquatic Habitat Connectivity, and Salmon Persistence in the Willamette and Lower Columbia River Basins. *Transactions of the American Fisheries Society*, *135*(6), 1654–1669. https://doi.org/10.1577/T05-221.1

Slatkin, M., & Voelm, L. (1991). FST in a Hierarchical Island Model. *Genetics*, *127*, 627–629.

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, *14*(11), 751–764. https://doi.org/10.1038/nrg3483

Stinchcombe, J. R., & Hoekstra, H. E. (2008). Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity*, *100*(2), 158–170. https://doi.org/10.1038/sj.hdy.6800937

Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: Recent progress and future challenges. *Journal of the North American Benthological Society*, *29*(1), 344–358. https://doi.org/10.1899/08-171.1

Suzuki, T., Suzuki, N., & Tojo, K. (2019). Parallel evolution of an alpine type ecomorph in a scorpionfly: Independent adaptation to high-altitude environments in multiple mountain locations. *Molecular Ecology*, mec.15119. https://doi.org/10.1111/mec.15119

Sveen, L. R., Timmerhaus, G., Krasnov, A., Takle, H., Handeland, S., & Ytteborg, E. (2019). Wound healing in post-smolt Atlantic salmon (*Salmo salar L.*). *Scientific Reports*, *9*(1), 3565. https://doi.org/10.1038/s41598-019-39080-x

Taylor, E. B., & Foote, C. J. (1991). Critical swimming velocities of juvenile sockeye salmon and kokanee, the anadromous and non-anadromous forms of Oncorhynchus nerka (Walbaum). *Journal of Fish Biology*, *38*(3), 407–419. https://doi.org/10.1111/j.1095-8649.1991.tb03130.x

Taylor, E. B., Harvey, S., Pollard, S., & Volpe, J. (1997). Postglacial genetic differentiation of reproductive ecotypes of kokanee *Oncorhynchus nerka* in Okanagan Lake, British Columbia. *Molecular Ecology*, *6*(6), 503–517. https://doi.org/10.1046/j.1365-294X.1997.00213.x

Taylor, E. B., Foote, C. J., & Wood, C. C. (1996). Molecular genetic evidence for parallel life-history evolution within a Pacific salmon (sockeye salmon and kokanee, *Oncorhynchus nerka*). *Evolution*, *50*(1), 401–416. https://doi.org/10.1111/j.1558-5646.1996.tb04502.x

Tintle, N., Gordon, D., Van Bruggen, D., & Finch, S. (2009). The Cost Effectiveness of Duplicate Genotyping for Testing Genetic Association. *Annals of Human Genetics*, *73*(3), 370–378. https://doi.org/10.1111/j.1469-1809.2009.00516.x

Tipping, J. M., & Byrne, J. B. (1996). Reducing Feed Levels during the Last Month of Rearing Enhances Emigration Rates of Hatchery-Reared Steelhead Smolts. *The Progressive Fish-Culturist*, *58*, 128–130.

Tomalty, K. M. H., Meek, M. H., Stephens, M. R., Rincón, G., Fangue, N. A., May, B. P., & Baerwald, M. R. (2015). Transcriptional Response to Acute Thermal Exposure in Juvenile

Chinook Salmon Determined by RNAseq, *Genes|Genomes|Genetics*, *5*(7), 1335–1349. https://doi.org/10.1534/g3.115.017699

Vainionpaa, H. E., Sarchuk, J. A., Andrusak, H., & Harris, S. L. (2020). *Alouette Reservoir nutrient restoration project 5-year review, 2014-2018* (No. RD168). Province of British Columbia Ministry of Environment and Climate Change Strategy Ecosystems Branch.

Valenzuela-Quiñonez, F. (2016). How fisheries management can benefit from genomics? *Briefings in Functional Genomics*, *15*(5), 352–357. https://doi.org/10.1093/bfgp/elw006

van Poorten, B. T., Harris, S., & Hébert, A. (2018). Evaluating benefits of stocking on sockeye recovery projections in a nutrient-enhanced mixed life history population. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(12), 2280–2290. https://doi.org/10.1139/cjfas-2017-0438

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Veale, A. J., & Russello, M. A. (2017a). An ancient selective sweep linked to reproductive life history evolution in sockeye salmon. *Scientific Reports*, *7*(1), 1747. https://doi.org/10.1038/s41598-017-01890-2

Veale, A. J., & Russello, M. A. (2017b). Genomic Changes Associated with Reproductive and Migratory Ecotypes in Sockeye Salmon (*Oncorhynchus nerka*). *Genome Biology and Evolution*, *9*(10), 2921–2939. https://doi.org/10.1093/gbe/evx215

Waples, R. S., & Do, C. (2008). LDNE: A program for estimating effective population size from data on linkage disequilibrium: Computer programs. *Molecular Ecology Resources*, *8*(4), 753–756. https://doi.org/10.1111/j.1755-0998.2007.02061.x

Waples, R. S., Pess, G. R., & Beechie, T. (2008). SYNTHESIS: Evolutionary history of Pacific salmon in dynamic environments: Evolutionary history of Pacific salmon in dynamic

environments. *Evolutionary Applications*, *1*(2), 189–206. https://doi.org/10.1111/j.1752-4571.2008.00023.x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370.

Westram, A. M., Galindo, J., Rosenblad, A. M., Grahame, J. W., Panova, M., & Butlin, R. K. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology*, *23*(18), 4603–4616. https://doi.org/10.1111/mec.12883

Whiteley, A. R., Coombs, J. A., Hudy, M., Robinson, Z., Colton, A. R., Nislow, K. H., & Letcher, B. H. (2013). Fragmentation and patch size shape genetic structure of brook trout populations. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(5), 678–688. https://doi.org/10.1139/cjfas-2012-0493

Willis, S. C., Hess, J. E., Fryer, J. K., Whiteaker, J. M., Brun, C., Gerstenberger, R., & Narum, S. R. (2020). Steelhead (*Oncorhynchus mykiss*) lineages and sexes show variable patterns of association of adult migration timing and age-at-maturity traits with two genomic regions. *Evolutionary Applications*, eva.13088. https://doi.org/10.1111/eva.13088

Willson, M. F., & Halupka, K. C. (1995). Anadromous Fish as Keystone Species in Vertebrate Communities. *Conservation Biology*, *9*(3), 489–497. https://doi.org/10.1046/j.1523-1739.1995.09030489.x

Withler, R. E., Le, K. D., Nelson, R. J., Miller, K. M., & Beacham, T. D. (2000). Intact genetic structure and high levels of genetic diversity in bottlenecked sockeye salmon (*Oncorhynchus nerka*) populations of the Fraser River, British Columbia, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, *57*, 1985–1998.

Wood, C. C., Bickham, J. W., Nelson, J.R., Foote, C. J., & Patton, J. C. (2008). Recurrent evolution of life history ecotypes in sockeye salmon: Implications for conservation and future evolution: Recurrent evolution of ecotypes in sockeye salmon. *Evolutionary Applications*, *1*(2), 207–221. https://doi.org/10.1111/j.1752-4571.2008.00028.x

Wood, T. E., Burke, J. M., & Rieseberg, L. H. (2005). Parallel genotypic adaptation: When evolution repeats itself. *Genetica*, *123*, 157–170.

Wysujack, K., Greenberg, L. A., Bergman, E., & Olsson, I. C. (2009). The role of the environment in partial migration: Food availability affects the adoption of a migratory tactic in brown trout *Salmo trutta*. *Ecology of Freshwater Fish*, *18*(1), 52–59. https://doi.org/10.1111/j.1600-0633.2008.00322.x

Yamamoto, S., Kitamura, S., Sakano, H., & Morita, K. (2011). Genetic structure and diversity of Japanese kokanee *Oncorhynchus nerka* stocks as revealed by microsatellite and mitochondrial DNA markers. *Journal of Fish Biology*, *79*(5), 1340–1349. https://doi.org/10.1111/j.1095-8649.2011.03098.x

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Appendices

Appendix A: Chapter 2 Supplementary materials

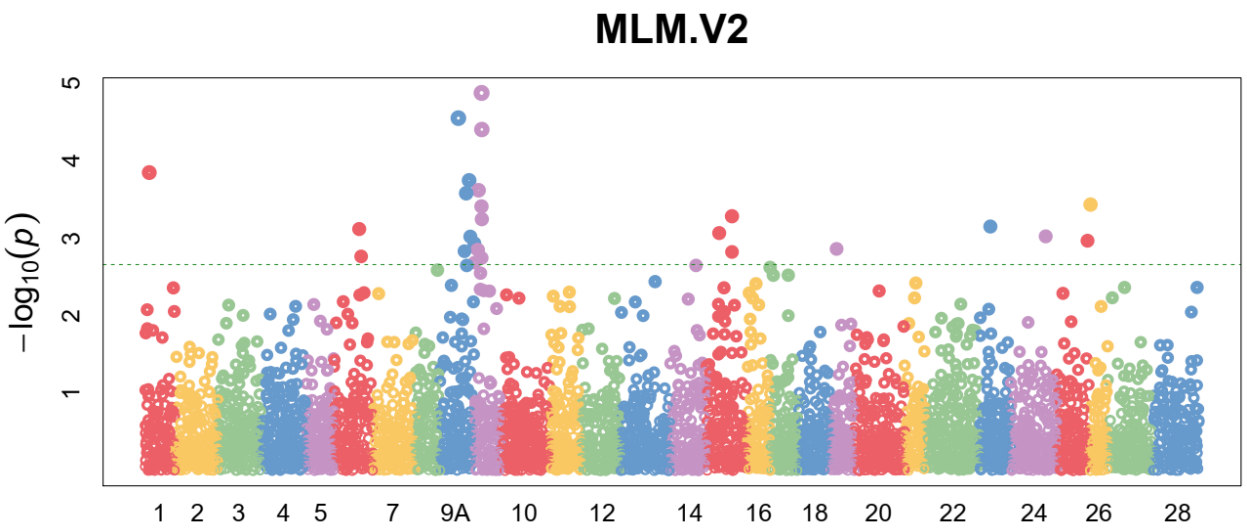

Figure A.1. Manhattan plot generated using GWAS Mixed Linear Model (MLM) analysis, using GAPIT v3.0 (Lipka et al., 2012). Numbers on x-axis correspond to Linkage Groups. V2 corresponds to phenotype (migratory or resident).

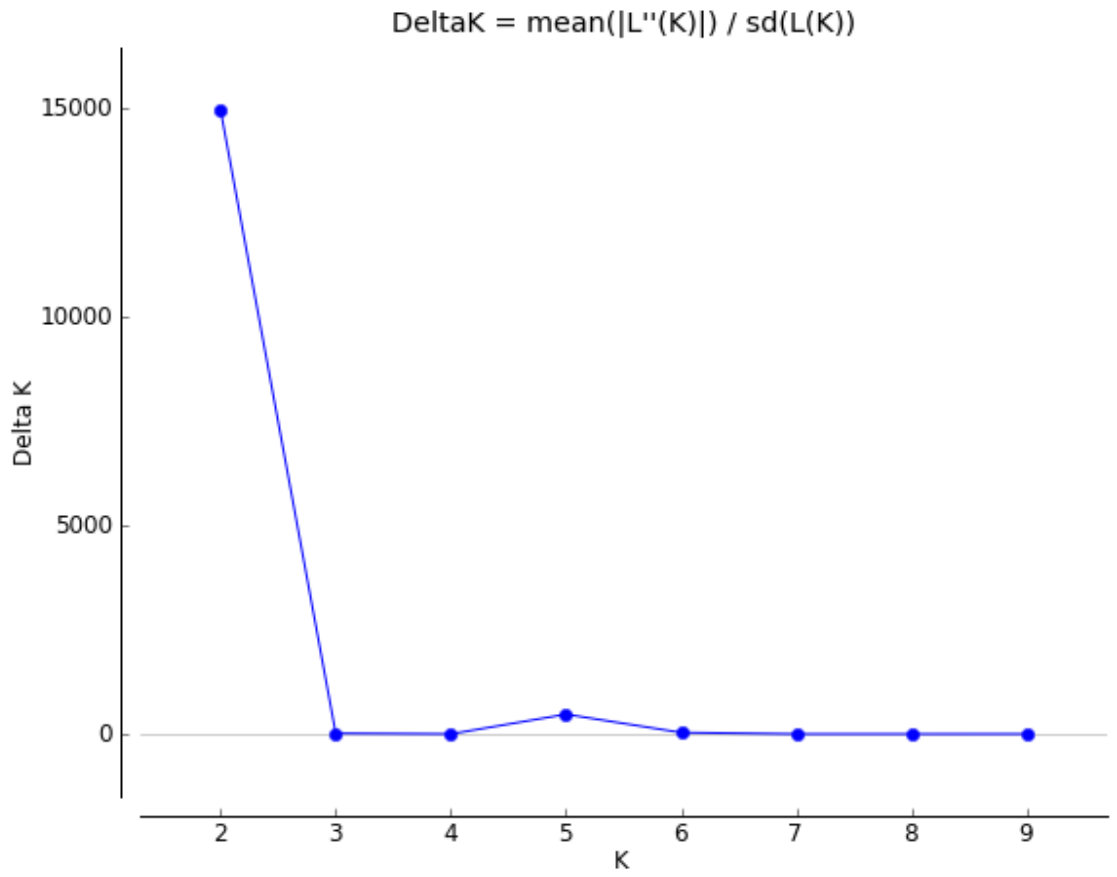

Figure A.2. Delta K values for the 10 iterations of STRUCTURE with the number of clusters varying from 1 to 10. Generated using STRUCTURE Harvester.

Table A.1. Populations, morphs and spawner types of *O. nerka* populations examined in this study.

Population	Morphs	Spawner type	NS	NR	NF	Source
Alouette Lake	resident	deep-spawner	68	8	61	New data
	adult migrant (AUM)	deep-spawner[±]	85	7	78	New data
	juvenile migrant (JDM)	deep-spawner[±]	26	0	24	
Scotch Creek	sockeye	stream-spawner	25	2	25	New data
Portage Creek	sockeye	stream-spawner	NA	NA	23	Veale & Russello (2017b)
Anderson Lake	kokanee	deep-spawner	NA	NA	22	Veale & Russello (2017b)
Seton Lake	kokanee	deep-spawner	NA	NA	23	Veale & Russello (2017b)
East Barrière	kokanee	deep-spawner	31	2	31	New data
Nicola Lake	kokanee	stream-spawner	25	0	25	New data
Total			260		312	

Abbreviations: number of unique samples sequenced (NS), number of replicates (NR), number of unique samples retained after filtering (NF).

[±] See Chapter 2 Introduction for spawning information of Alouette Lake *O. nerka* migrants.

Table A.2. Sensitivity analysis to identify the optimal set of parameters for the *populations* run (Stacks v2.41; Catchen et al., 2011).

r	min_maf	N (Depth>6x)	SNPs	Mean missing %	Mean depth
0.7	0.01	330	14948	4.1	24.58
	0.03	330	11313	4.5	24.75
	0.05	330	9700	4.75	24.75
0.75	0.01	330	14257	3.74	24.76
	0.03	330	10752	4.11	24.91
	0.05	330	9220	4.37	24.92
0.8	0.01	330	13213	3.26	25.06
	0.03	330	9899	3.61	25.24
	0.05	330	8458	3.84	25.27
0.85	0.01	330	11805	2.74	25.48
	0.03	330	8723	3.05	25.72
	0.05	330	7390	3.25	25.79
0.9	0.01	330	9153	1.96	26.23
	0.03	330	6516	2.18	26.59
	0.05	330	5355	2.35	26.8
0.95	0.01	330	5233	1.08	27.52
	0.03	330	3310	1.22	28.03
	0.05	330	2471	1.30	28.47

Table A.3. Evanno Table generated using STRUCTURE Harvester, showing Ln'(K) and Delta K values, based on 10 iterations of STRUCTURE output, with the number of clusters (K) varying from 1 to 10.

# K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-1963925	10.6608	NA	NA	NA
2	10	-1782303	9.3031	181621.3	139275.4	14970.9
3	10	-1739957	334.5973	42345.87	5070.43	15.15383
4	10	-1702682	12741.36	37275.44	8329.22	0.653715
5	10	-1673736	59.529	28946.22	28137.88	472.6752
6	10	-1672927	948.4455	808.34	32638.89	34.41304
7	10	-1704758	93309.67	-31830.6	13395.52	0.14356
8	10	-1749984	225893.5	-45226.1	33303.29	0.147429
9	10	-1761907	270804.9	-11922.8	217664	0.803767
10	10	-1991494	527410.6	-229587	NA	NA

Appendix B: Chapter 3 Supplementary materials

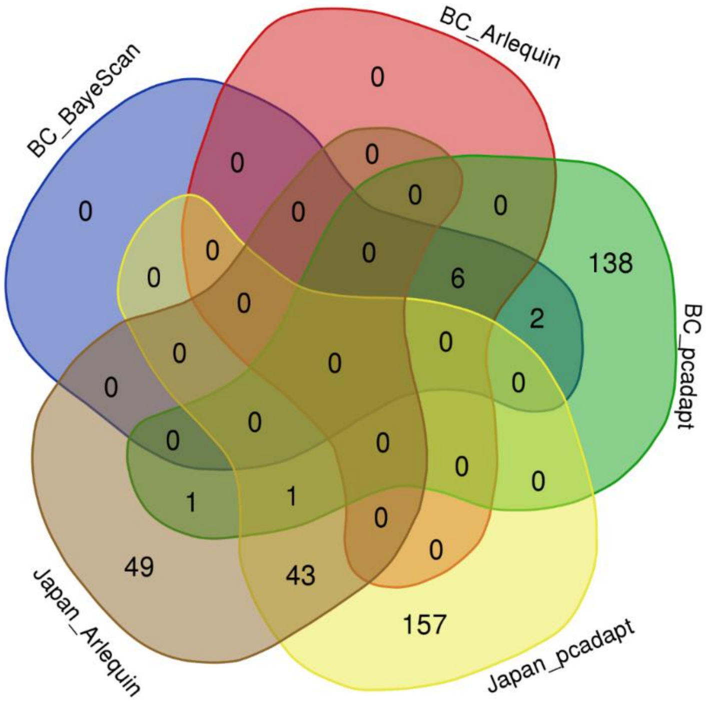

Figure B.1. Venn Diagram demonstrating the number of outliers detected by three different approaches (BayeScan, Arlequin, *pcadapt)* in British Columbia- and Japan-specific outlier scans. Outlier scans were conducted between "paired" populations of stream-spawning and deep-spawning kokanee. Venn Diagram produced using http://bioinformatics.psb.ugent.be/webtools/Venn/; results verified using R (R Core Team, 2018).

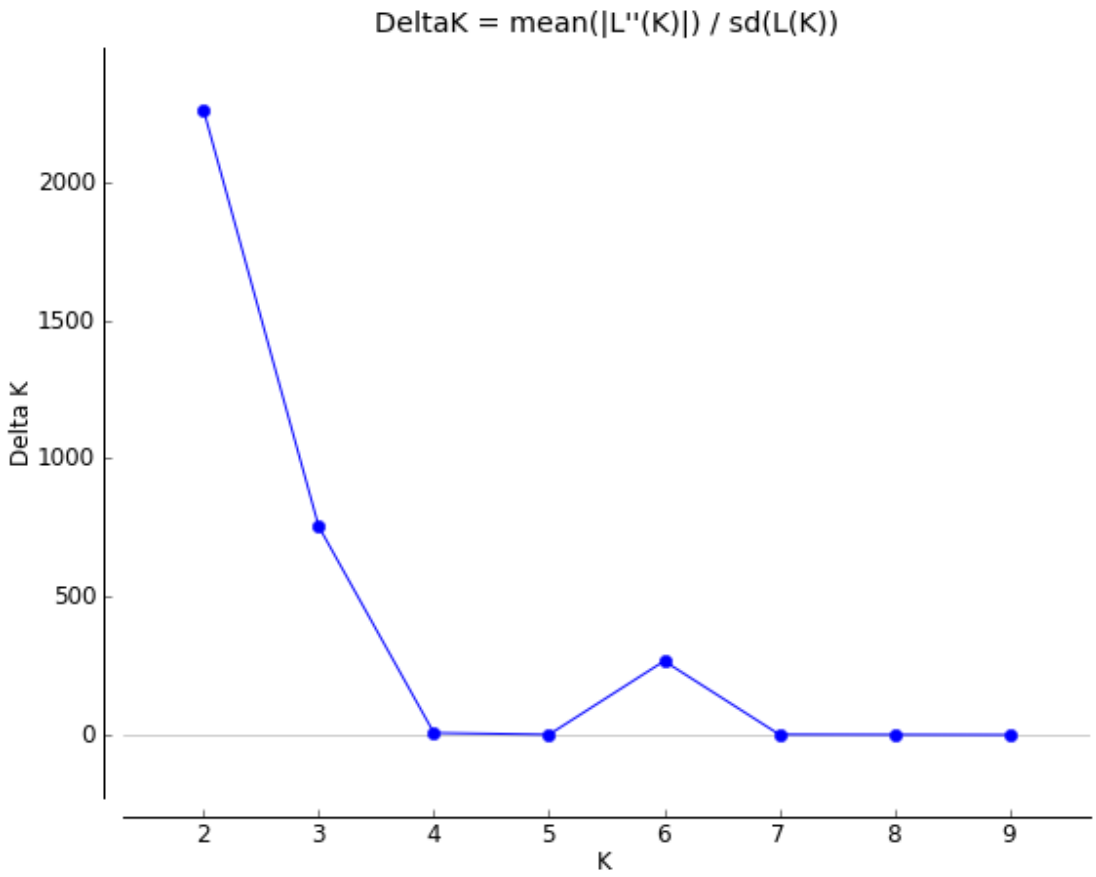

Figure B.2. Delta K values for the 10 iterations of STRUCTURE with the number of clusters varying from 1 to 10. Generated using STRUCTURE Harvester.

Table B.1. Sensitivity analysis to identify the optimal set of parameters for the *populations* run (Stacks v2.41; Catchen et al., 2011).

r	min_maf	N (Depth>6x)	SNPs	Mean missing %	Mean depth
0.7	0.03	173	10704	2.90	32.58
	0.05	173	9372	2.98	32.68
0.75	0.03	173	9495	2.43	33.27
	0.05	173	8286	2.49	33.38
0.8	0.03	173	9179	2.27	33.32
	0.05	173	8023	2.33	33.43
0.85	0.03	173	7119	1.69	34.4
	0.05	173	6189	1.75	34.52
0.9	0.03	173	5646	1.33	35.24
	0.05	173	4866	1.38	35.38

Table B.2. Evanno Table generated using STRUCTURE Harvester, showing Ln'(K) and Delta K values, based on 10 iterations of STRUCTURE output, with the number of clusters (K) varying from 1 to 10.

# K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-1213392	13.91	NA	NA	NA
2	10	-1006343	50.8609	207049.2	114979.5	2260.666
3	10	-914273	58.3206	92069.72	43972.67	753.9817
4	10	-866176	1158.917	48097.05	8418.89	7.264448
5	10	-826498	10080.49	39678.16	9279.1	0.920501
6	10	-796099	153.95	30399.06	41126.63	267.1428
7	10	-806827	9882.559	-10727.6	12732.33	1.288364
8	10	-804822	9223.82	2004.76	6396.1	0.693433
9	10	-809213	22698.17	-4391.34	1254.45	0.055267
10	10	-814859	15889.94	-5645.79	NA	NA