Drivers and Fitness Consequences of Dispersal and Structure in Wild Sockeye Salmon Populations (*Oncorhynchus nerka*)

Samuel A. May

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Kerry Naish, Chair

Lorenz Hauser

Ray Hilborn

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

University of Washington

Abstract

Drivers and Fitness Consequences of Dispersal and Structure in Wild Sockeye Salmon Populations (*Oncorhynchus nerka*)

Samuel A. May

Chair of the Supervisory Committee:
Professor Kerry Naish
School of Aquatic and Fishery Sciences

The life histories of migratory species such as salmonids, sea turtles, and birds often involve return migrations between feeding and natal habitats. These natal homing behaviors are known to produce structured metapopulations, where geographic and demographic barriers result in non-random mating among many locally adapted subpopulations. The resulting spatial and temporal diversity across heterogeneous landscapes can buffer metapopulations against disruptive events that influence any one subunit. Dispersal and gene flow within and between subpopulations can reduce fitness losses due to inbreeding depression, influence rates of adaptation, and facilitate colonization or recolonization of newly available habitat. However, an understudied aspect of metapopulation biology is the influence of biotic and abiotic factors that lead to genetic structure

within and between subpopulations, and the effects of this structure on fitness. Therefore, the overall goal of this thesis was to investigate how environmental, behavioral, and life-history variation might influence dispersal, population structure, and fitness within and between subpopulations. To accomplish this goal, dispersal within and between two proximate stream-spawning populations of Sockeye salmon (*Oncorhynchus nerka*): A and C Creeks on the Wood River System, Bristol Bay, AK was studied over two complete generations of returning adults. First, a panel of 172 SNP loci was developed (genotyping-in-thousands by sequencing; Chapter One) and used to reconstruct a pedigree from fish returning over a 14-year period, and to identify dispersers between the two populations. Second, we investigated the drivers and fitness consequences of dispersal between A and C Creeks and found that return timing to spawning grounds and within-season variation in predation and population density influenced dispersal between the two populations (Chapter Two). Fitness consequences of dispersal depended on the direction dispersers moved; moving from A to C increased absolute fitness of dispersers (compared to individuals in their natal population) but decreased their relative fitness (compared to individuals in their new spawning population), while moving from C to A decreased absolute fitness but increased relative fitness. From these results, we concluded that dispersal was an active process in response to environmental cues and that gene flow was affected by habitat differences and within-season variation in ecological processes. Third, we aimed to examine the extent, drivers, and fitness consequences of population structure within the two streams. To achieve this aim, we quantified the scale of structure, the effect of natal homing on structure, and the fitness outcomes of homing to, and dispersing from natal sites (Chapter Three). Both spatial and temporal genetic structure was evident within both streams, and this structure was partly explained by adults returning to the same place and at the same time as they were fertilized as eggs. In addition,

phenotypes of body size and return timing were spatially segregated within the creeks. In one of the two creeks, adults returning to spawn near natal sites had greater fitness. Taken together, we concluded that these findings provided empirical evidence for how natal homing and heterogeneous habitat may lead to assortative mating systems and possible microgeographic adaptation on very small spatial and temporal scales. In other words, natal homing and dispersal within populations may result in genetic or phenotypic neighborhoods and affect fitness. Finally, we discuss the utility of these findings for predicting responses of natural populations to future environmental and anthropogenic changes such as harvest, climate change, and supportive breeding.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SUPPLEMENTAL FIGURES

LIST OF SUPPLEMENTAL TABLES

ACKNOWLEDGEMENTS

This thesis is a culmination of work, not just by one Ph.D. student, but by dozens of remarkable people working over the past two decades. Here I will mention only a handful, although there are many more.

First, I would like to thank my PhD advisor, Professor Kerry Naish, for her unwavering support of my academic and personal journey over the past six years. Kerry's calm and steady guidance allowed me room to creatively explore my own interests, without pushing me in any one direction. I am grateful for her thorough editorial work on funding applications, presentations, and publications, which has helped me immensely to develop as a scientific writer. Kerry deeply cares for her students' academic and personal growth, and it has been a privilege to learn from her.

I acknowledge my supervisory committee, who taught me to take pride in my science. I am thankful for the hard work of Professor Lorenz Hauser, for his brilliant mind in countless long meetings on data exploration methods and population genetic theory. Lorenz has been like a second advisor to me, and I am grateful for his time and creativity in helping me solve difficult data problems and become a better writer. I acknowledge Professor Ray Hilborn, whose strong desire to enjoy lunch on the beaches of Little Togiak Lake led to the development of an unparalleled longitudinal pedigree dataset of two wild populations. For two decades, Ray, alongside his family, has dedicated his summers to intensively sampling the A and C Creek populations – without which none of this research would be possible. Ray was also an unwavering source of support and inspiration for this thesis work, always willing to contribute thoughtful guidance, seek funding, and offer kind words of support. I am eternally grateful to Dr. David Paez for helping me develop my skillset in all things data exploration, statistics, and creative problem

solving. David met with me at least once a week in the year leading up to my defense, always energized to think about my data in new and creative ways. Lastly, I acknowledge Professor Jennifer Ruesink, who went above and beyond her responsibilities as my GSR by taking a genuine interest in me and my work, always with a positive and excited outlook.

I thank staff and students of the Alaska Salmon Program at the University of Washington for decades of fieldwork, especially Professor Ray Hilborn, Professor Daniel Schindler, Jackie Carter, and Chris Boatright. I acknowledge the foundational work of previous Alaska Salmon Program students who worked on A and C Creeks, particularly Drs. Jocelyn Drugan and Daniel Peterson, whose work I garnered much inspiration from.

I thank additional collaborators and mentors: Drs. Robin Waples, Jeffery Hard, Eric Ward, Mike Ford, Kim Parsons, and Garrett McKinney for help with securing funding for this work, providing additional research opportunities outside the scope of this thesis, and a steadfast willingness to discuss population and quantitative genetic theory and methodologies. I acknowledge Lisa and Jim Seeb for guidance in the early stages of this project and for use of their laboratory equipment and personnel. I thank Carita Pascal for extensive bioinformatic and laboratory technical help. I also acknowledge the hard work of undergraduate research assistants Andrea Odell, Emily Bowers, Ryan Brooks, Sarah Panciroli, and Emily Schwabe for laboratory assistance, and specifically to Andrea Odell, Emily Bowers, and Emily Schwabe for help with preliminary data exploration and analyses; it was a pleasure to learn from you all.

I am indebted to the SAFS community that supported me throughout graduate school. Thank you to my MERlab-mates who were a key support system in so many ways. In particular, thank you to Charlie Waters, Eleni Petrou, Natalie Lowell, Daniel Drinan, and Mary Fisher, whose doors were always open to an unrelenting stream of questions about coding, bioinformatics, and

lab work in my first few years of graduate school. I am grateful for my Seattle family, who took me on countless adventures in Washington and who supported me through good and bad. Specifically, I acknowledge my caring partner, Marie Zahn, and my amazing friends: Elizabeth Elmstrom, Michaela Lowe, Katie McElroy, Colin Nicol, Eileen Bates, Grant Adams, Lindsay Alma, Martini Arostegui, and John Wittouck.

I acknowledge the loving support of my family and friends outside of graduate school. Particularly I thank my parents, Earl and Denise May, for nurturing my early scientific curiosity and listening all along the way. I am grateful for the friendship of Amanda Mikalian, Trenton Hawkins, Harrison Cann, Abraham Alishaev, Dylan Bialek, Ryan Pollowitz, and Candice Johnson Pollowitz, for reminding me that there is more to life. Their love and support were persistent and instrumental to my success in graduate school.

This thesis is based upon work supported by a National Science Foundation Graduate Research Fellowship (Grant No. DGE-1762114). This work was further supported by a North Pacific Research Board Graduate Student Fellowship, a NMFS/Sea Grant Population and Ecosystem Dynamics Fellow, and a Hall Conservation Genetics Fellowship through the College of the Environment at the University of Washington. I am grateful for the continued support of the Bristol Bay Processors. Finally, I thank the School of Aquatic and Fishery Sciences at the University of Washington for fostering an outstanding learning environment and providing several quarters of funding in the form of TA, RA, and fellowship funds.

Chapter 1. POWER OF A DUAL-USE SNP PANEL FOR PEDIGREE RECONSTRUCTION AND POPULATION ASSIGNMENT

Authors: Samuel A. May, Garrett J. McKinney, Ray Hilborn, Lorenz Hauser, and Kerry A. Naish

1.1 ABSTRACT

The use of high-throughput, low-density sequencing approaches has dramatically increased in recent years in studies of eco-evolutionary processes in wild populations and domestication in commercial aquaculture. Most of these studies focus on identifying panels of SNP loci for a single downstream application, whereas there have been few studies examining the trade-offs for selecting panels of markers for use in multiple applications. Here we detail the use of a bioinformatic workflow for the development of a dual-purpose SNP panel for parentage and population assignment, which included identifying putative SNP loci, filtering for the most informative loci for the two tasks, designing effective multiplex PCR primers, optimizing the SNP panel for performance, and performing quality control steps for downstream applications. We applied this workflow to two adjacent Alaskan Sockeye Salmon populations and identified a GTseq panel of 142 SNP loci for parentage and 35 SNP loci for population assignment. Only 50-75 panel loci were necessary for >95% accurate parentage, whereas population assignment success, with all 172 panel loci, ranged from 93.9% to 96.2%. Finally, we discuss the trade-offs and complexities of the decision-making process that drives SNP panel development, optimization, and testing.

1.2 INTRODUCTION

Single nucleotide polymorphisms (SNPs) have become commonplace in ecological and evolutionary studies and are particularly useful for pedigree reconstruction and individual assignment to populations (Seeb *et al.* 2011a). The use of pedigrees in evolutionary studies has rapidly expanded, especially in investigations related to fitness and genetic diversity of wild systems. For example, pedigrees may be used to estimate inbreeding depression, relatedness, and individual reproductive success, and they are used in parentage-based population assignment (Pemberton 2008; Richards-Zawacki *et al.* 2012; Hipperson *et al.* 2017) and close-kin mark recapture studies (Bravington *et al.* 2016). While useful, the reconstruction of wild pedigrees can be constrained by processing time and the expense of obtaining molecular data for thousands of individuals. There may also be limitations associated with sampling all individuals in a population over multiple generations because of logistic constraints. Individuals might disperse between interconnected populations, which can confound data analyses but may also provide key information on the contribution of immigrants to population fitness (Peterson *et al.* 2014). As such, coupling pedigree reconstruction with approaches that assign individuals to their population of origin (population assignment) can aid in describing broader evolutionary processes across interconnected populations.

Despite the benefits of combining pedigrees with population assignments, the two objectives require different population allele frequencies for optimal performance. Pedigree reconstruction benefits from loci with high minor allele frequencies (Anderson & Garza 2005; Holman *et al*. 2017), whereas population assignment benefits from highly differentiated loci between populations (Anderson 2010). This ascertainment bias (Bradbury *et al.* 2011, Seeb *et al.*

2011b) introduced by selecting loci for a specific purpose may limit their applications in other analyses and produce important trade-offs in the design of multi-use SNP panels.

Several high-throughput sequencing technologies have been developed in recent years that allow thousands of individuals to be pooled in a single lane of sequencing and genotyped at hundreds to thousands of SNP loci (e.g., GTseq, Campbell *et al.* 2015; Rapture, Ali *et al.* 2015; and MTAseq, Onda *et al.* 2018). These approaches require the selection of specific subsets of loci and optimization of primers prior to sequencing but substantially reduce the cost of genotyping per individual, because the number of loci that can be reasonably included is limited to a few hundred (Meek & Larson 2019) by problems arising from non-specific heterodimer formation (Aykanat *et al.* 2016), unequal amplification (McKinney *et al.* 2020), and minimum required coverage.

Several studies have outlined best practices for the selection and optimization of marker panels (Liu *et al.* 2016; Holman *et al.* 2017; McKinney *et al.* 2020), including for multiple downstream applications (Aykanat *et al.* 2016). The effect of different locus numbers and population sizes on parentage (Holman *et al.* 2017; Harney *et al.* 2018) and population assignment (McKinney *et al.* 2017a; Baetscher *et al.* 2018) is also well known. However, panels are usually either assembled for a single purpose or as single-purpose, bioinformatically-separated modules in combined panels (Aykanat *et al.* 2016). To our knowledge the combined power of entire multi-purpose panels for single objectives has not been tested.

SNP panel-based approaches have been increasingly used in salmonids (Liu *et al.* 2016; Holman *et al.* 2017; Steele *et al.* 2017; Janowitz-Koch *et al.* 2019; McKinney *et al.* 2019b), a group of fishes that has been the subject of evolutionary ecology studies in natural systems, not only because of their commercial, ecological and cultural significance, but also because well-

defined spawning populations in freshwater habitats and relatively easy sampling greatly facilitate such studies. Sockeye Salmon (*Oncorhynchus nerka*) in particular have been the focus of many behavioral, ecological, and evolutionary studies that use both population assignment and pedigree analyses (i.e. Peterson *et al.* 2014, Lin *et al.* 2016, 2017). So far, these studies have relied primarily on microsatellite markers in small populations over one or two generations. However, extended eco-evolutionary questions over many generations necessitates the efficient genotyping of thousands of individuals at loci capable of both accurate pedigree reconstruction and individual assignment to population of origin. If the design of such panels is sufficiently successful and genotyping costs are sufficiently low, similar approaches may be possible for species where large-scale pedigrees and population assignment were hitherto deemed impossible; for example, rockfishes (*Sebastes* sp, Baetscher *et al.* 2019) or tuna (T*hunnus maccoyii*, Bravington *et al.* 2016).

Here we describe the development of a GTseq amplicon primer panel for rapid and effective genotyping of Sockeye Salmon to address two distinct objectives: pedigree reconstruction and determination of population of origin. We demonstrate the use of a streamlined workflow (McKinney *et al.* 2020) and detail the processes of marker selection, optimization, and testing. Furthermore, we investigate the effects of ascertainment bias on downstream applications, in an effort to determine whether dual-use SNP panels benefit from the combination of markers or effectively represent independent sets of loci that do not add much power to the alternative application.

1.3 METHODS

1.3.1 *Study system*

Tissue samples were collected from Sockeye Salmon spawning in two small creeks (A and C Creeks; 250 m and 350 m long, respectively), approximately 1 km apart on Little Togiak Lake,

Alaska (Peterson *et al.* 2014). We utilized dorsal fin tissue samples from 2010 for panel development and 2009 for panel optimization and power verification. Tissue samples were collected from nearly every fish returning to the creeks throughout the spawning season and stored in 100% ethanol (Peterson *et al.* 2014). F_{ST} values, estimated by Lin *et al.* (2008) using microsatellite markers, range from 0.02-0.04 (Lin *et al.* 2008).

1.3.2 *Ascertainment of Reference Sequences for Panel Development: RAD Sequencing and Genotyping*

DNA was extracted from 144 individuals representing both A and C creeks (Figure 1.1, Step 1) using Qiagen DNeasy Blood and Tissue kits (Qiagen, Valencia CA). Two RAD-seq libraries (Baird *et al.* 2008) were prepared following the bestRAD protocol: the first step of the RAPTURE protocol (Ali *et al.* 2015) using the *Sbf1* restriction enzyme. Paired-end 2x100-base pair sequencing was performed on an Illumina HiSeq4000. Raw RAD sequencing data were processed using *Stacks* (v.1.42, Catchen *et al.* 2013) following the bioinformatics pipeline described in Waters *et al.* (2018). Briefly, forward reads were demultiplexed and trimmed to 95bp using *process_radtags*. Individuals with fewer than 200,000 total reads were excluded from further analysis (Figure 1.1, step 1a). No assembled genome is yet available for Sockeye Salmon. Instead, reads were aligned to both the Rainbow Trout reference genome (*O. mykiss*, GenBank assembly Accession GCA_002163495) and to a Sockeye Salmon linkage map (Larson *et al.* 2015, based on haploid data) using *Bowtie*2 (v.2.3.3.1, Langmead and Salzberg 2012), allowing up to three base pair mismatches. These two alignments were treated as separate pipelines during the remaining panel development steps (Figure 1.1). Loci for each individual were identified using *pstacks* with the bounded-error SNP calling model, a minimum read depth of 10, and the default error rates. Individuals with less than 19,000 loci in the *O. mykiss* pipeline or 4,500 loci in the *O. nerka* linkage

map pipeline were excluded from further analysis, representing breaks in the distribution of loci (Figure 1.1, step 1c). The 10 individuals from each population with the highest read depth and coverage were used to construct a catalog of loci in *cstacks*; this subset was used to reduce the risk of including false polymorphisms in the catalog (as in Waters *et al.* 2018). Loci from individuals aligned to either the *O. mykiss* genome or Sockeye linkage map were matched to their respective catalogs using *sstacks*. Finally, genotypes were assigned using *populations*, with a minimum read depth of 10.

1.3.3 *Locus Filtering and SNP Selection*

Following *Stacks* genotyping, all biallelic loci were re-genotyped using a custom Python script to verify and correct genotypes, therefore minimizing potential bias in maximum likelihood genotype calls in *Stacks* due to differences in read depth between two alleles at a locus (Brieuc *et al.* 2014, but also Waters *et al.* 2018). Loci were excluded if they had a minor allele frequency of less than 0.05 in either creek population (Figure 1.1, step 2a). Loci genotyped in ≤80% of individuals, and individuals with ≥50% missing data were excluded from further analyses (Figure 1.1, step 2b).

We followed the filtering and SNP selection methods in McKinney *et al.* (2020) to reduce the effects of unequal locus amplification, cross-amplification of primer pairs, and off-target amplification (Figure 1.1, step 2). Loci were filtered using *HDplot* (McKinney *et al.* 2017b); https://github.com/gjmckinney/HDplot) to ensure no undifferentiated paralogous loci were included in the panel. To permit sufficient sequence length for primer development, polymorphic SNP loci were further filtered to exclude loci within 15 base pairs of the start or end of the 95bp RAD sequence. The program *RepeatMasker* (Smit *et al*. unpublished v.4.0.6) was used to identify and remove loci in low complexity regions or transposable elements, using the default parameters.

To reduce risk of off-target amplification, loci aligned to the *O. mykiss* genome were excluded if they matched to more than 6 unique regions. This represented a break in the distribution of matches per locus and was used as a qualitative threshold.

Loci were included in the panel to serve two primary purposes: pedigree reconstruction and population assignment. SNPs for parentage were selected from loci identified through alignment to a Sockeye Salmon linkage map (Larson *et al.* 2015), as no genome has yet been assembled for Sockeye. Loci were excluded if they had minor allele frequencies less than 0.35 in either population, as an allele frequency close to 0.5 maximizes the power for parentage (Anderson & Garza 2005). Conversely, the power to assign population of origin increases when loci are highly differentiated between populations (Anderson 2010). We anticipated that alignment to the Sockeye Salmon linkage map may have caused population-specific ascertainment bias, as the linkage map comprised loci that were polymorphic in few families and from different populations than the present study, resulting in little differentiation between our study populations (Larson *et al.* 2015). Therefore, discovery of SNPs for population assignment were instead derived from loci aligned to the *O. mykiss* genome, and SNPs that had the greatest difference in minor allele frequency between the two creek populations were selected. Beach and creek populations of Sockeye Salmon represent different morphotypes, but they often intermingle close to the spawning streams (Lin *et al.* 2008; Peterson *et al.* 2014). Therefore, an additional set of outlier loci differentiating creek and beach populations (Larson *et al.* 2017) were included to differentiate these highly divergent populations. These outlier loci were included for use in future studies, but their utility in differentiating beach and creek populations is not addressed further here.

1.3.4 *Primer Design*

Primers for all putative loci were designed in *BatchPrimer3* (v.1.0, You *et al.* 2008), with no optimum fragment or primer size. We followed the protocol of (McKinney *et al.* 2020) to optimize primers prior to sequencing. NCBI's BLAST (Altschul *et al.* 1997) was used to align all putative primer sequences to all putative amplicon sequences, to prevent potential interactions between the two. The putative primer sequences were also aligned to the *O. mykiss* genome to identify and exclude primers that may bind to more than one region of the genome.

1.3.5 *Optimization*

The final panel of primers was used to construct GTseq libraries for 24 individuals, following the methods in Campbell *et al.* (2015). Paired-end 2x75-base pair sequencing was performed on an Illumina MiSeq. We used scripts from Campbell *et al.* (2015) to genotype individuals and to identify and exclude primers which over amplified loci compared to other primer pairs. These scripts use the forward primer and an in-silico probe to count amplicon-specific sequences for each allele and assign genotypes based on observed allele ratios. This step helped to reduce off-target sequencing and increase read depth for the remaining loci by redistributing reads. We used custom scripts from McKinney *et al.* (2020) to identify primer interactions (Figure 1.1, step 4) that were not identified by BLAST (Figure 1.1, step 3b). In instances where different sets of primers interacted to amplify off-target regions, all but one of the interacting primer sets were excluded. To further reduce off-target amplification and amplification of duplicated loci, allele ratios at each locus were analyzed following the allele ratio plotting methods in McKinney *et al.* (2020); this step ensured that allele ratios conformed to those expected at single loci. In cases where loci did not fit expected allele ratios, the *in-silico* bioinformatic-probe in the genotyping script of Campbell *et al.* (2015) was extended from 15 to 30-base pairs on either side to better

exclude off-target sequence. The genotyping script was rerun, and individuals were re-genotyped with these new parameters. If this *in-silico* probe extension did not result in expected allele ratios, the locus was removed from the panel. The panel of primers was tested a second time with a set of 96 individuals using single-end 100-base pair sequencing on an Illumina MiSeq. Filtering was repeated for off-target and over-amplified loci, as above. A final set of primers was assembled.

1.3.6 *Quality Control*

To test the efficacy of the panel on a high-throughput dataset, DNA was extracted from 618 A Creek and 422 C Creek individuals from 2009, using Nexttec 1-Step DNA Extraction Kits (Nexttec Biotechnologie, Leverkusen, Germany), and GTseq libraries were prepared following the methods in Campbell *et al.* (2015). Single-end 100-base pair sequencing was performed on an Illumina HiSeq4000 and individuals were genotyped as above, using custom scripts from Campbell *et al.* (2015). A final round of locus filtering excluded loci from downstream analyses with unusual allele frequency ratios, following McKinney *et al.* (2020), as above. Pairwise linkage disequilibrium (R^2) was calculated between all SNP loci using the method from Hilland Robertson (1968) in the *LDcorSV* package (Desrousseaux *et al.* 2013) in R (R-Core-Team 2013). One locus per linked pair was excluded if pairwise R^2 values were greater than 0.25. F_{IS} values were calculated at each locus for each creek population using the *Hierfstat* package (Goudet 2005) in R, and loci were excluded if F_{IS} values were consistently greater than 0.1 or less than -0.1 in both populations.

1.3.7 *Power Analyses*

Power analyses were run to test the performance of loci for each of our two objectives: to identify parent-offspring pairs and assign individuals to their population of origin. We used the

training and hold-out method of Anderson (2010) for both parentage and population assignment to minimize high-grading bias. This method divides the dataset of empirical genotypes into two subsets: one to calculate allele frequencies for locus ranking (the training dataset) and another to calculate allele frequencies from which populations are simulated (the hold-out dataset). We evenly divided the 1,040 sequenced individuals into hold-out and training datasets by population, which were used in all subsequent analyses. The training dataset was used to rank loci by power for parentage and population assignment. The hold-out dataset was used to calculate population allele frequencies for use in simulating populations. As there is an approximately 3-12% dispersal rate between these two creek populations (Peterson *et al.* 2014, 2015), individuals in the training and hold-out datasets were cross-referenced with dispersal data from Peterson *et al.* (2014) to remove immigrants. We included only philopatric individuals, which returned to the same creek where all identified parents were born.

We tested the power of different numbers of panel loci to assign parent-offspring pairs using the unknown-sexes simulation module in the program Cervus3 (v. 3.0.7., Marshall *et al.* 1998; Kalinowski *et al.* 2007) with 100 model repetitions and 8 subsets of panel loci ranked in decreasing order of minor allele frequency in the training dataset. Populations of sizes of N = 100, and N = 500 parents were simulated to generate equivalent population sizes (100, 500) of offspring using allele frequencies in the hold-out dataset. To simulate a realistic population, training and hold-out datasets for this parentage analysis included only the C Creek population, assuming that 70% of parents were sampled, a mistyping rate of 0.005 (Campbell *et al*. 2015), and strict assignment (>95% probability). Using empirical pedigree data from Peterson *et al.* (2014), we estimated input parameters for the proportion of related individuals (0.0143), average relatedess among relatives (0.447), and rate of inbreeding (0.00). We estimated LOD scores of both single-

parent (parent-offspring pair) and parent-pair (trio of two parents, one offspring) assignments. Additionally, we examined the effect of ranking panel loci on locus-specific non-exclusion probabilities (NEP, Kalinowski *et al.* 2007). NEP measures the probability of not excluding a false parent at a given locus and can be used to measure locus specific power for parentage. We compared per-locus first-parent NEP ranked in order of decreasing power for parentage (decreasing minor allele frequency) with NEP in order of decreasing power for population assignment (decreasing difference in MAF between populations).

The accuracy of the population-specific loci in assigning individuals to their population of origin (A Creek or C Creek) compared to the broader panel was tested using the R package *rubias* (Moran & Anderson 2018). Population assignment was performed using the leave-one-out approach of Anderson *et al.* (2008), with 500 simulated individuals based on the hold-out data set. Allele frequencies were calculated from 100 individuals, randomly resampled 100 times from each population in the training dataset. Loci were ranked by allele frequency for each of these 100 draws. For each unique locus ranking order, the leave-one out assignment procedure was repeated 10 times. Therefore, there were 1,000 model repetitions (100x10) for each of 172 different panel subsets. Each subset added one additional locus to the panel, and they were used to examine the number of loci needed for accurate population assignment. Mean population assignment accuracy and 95% confidence intervals were calculated from the 1,000 repetitions of each panel subset. We then tested whether our ranking system during panel development reduced the number of panel loci necessary for accurate population assignment using the same simulation conditions. In this case, we used loci ranked in order of decreasing power for parentage in each population instead of power for population assignment.

1.4 RESULTS

1.4.1 *Marker Selection and Optimization*

No individuals had fewer than 200,000 reads; thus, 72 individuals from C Creek and 72 from A Creek were used for further panel development, with an average read depth of 3.8 million reads per individual. Four individuals were removed prior to genotyping in the *Stacks* module *populations* due to a small number of identified loci (Figure 1.1, step 1c). Alignment to the *O. mykiss* genome, followed by subsequent genotyping, yielded 4,222 bi-allelic RAD loci. Alignment to the *O. nerka* linkage map (Larson *et al.* 2015) yielded 3,864 bi-allelic loci. The numbers of loci remaining after each screening step are given in Figure 1.1. After completing the initial locus filtering steps (Figure 1.1, step 2, following McKinney *et al.* 2017b), 294 loci with minor allele frequencies greater than 0.35 across both populations were selected for parentage, whereas 60 loci with maximum differences in allele frequencies between the two creek populations (0.19-0.30) were selected for population assignment (Figure 1.2).

The final panel comprised 142 loci for parentage and 35 loci for population assignment after the final quality control step (Figure 1.1, step 5). Five of these loci were shared between both panel subsets, therefore the final panel comprised 172 primer pairs (Supplemental Table 1.1). Primer lengths ranged from 16-26 base pairs. A total of 1,175 individuals from 2009 were sequenced with the final panel of 172 loci and of these, 1,165 individuals (>99%) were successfully genotyped at more than 155 loci (90% of loci).

1.4.2 *Panel Accuracy*

In parentage simulations, the use of the full panel of 172 loci resulted in correct assignment greater than 95% in all cases (Figure 1.3). A minimum of 50 loci were needed to achieve >95%

parentage accuracy in parent-pair analyses with a population size of 100; in contrast, 75 loci were needed for a population size of 500. Assignment of offspring to parent-pairs required fewer loci than single-parent assignments (Figure 1.3) because trios (2 parents/1 offspring) are easier to identify than parent-offspring pairs. In the large population, single parent assignment required all 172 loci with an almost linear increase in assignment accuracy as loci were added. Misassignment (assignment to a false parent) slightly increased with number of panel loci but never exceeded 5.0%. Ranking loci by decreasing minor allele frequencies resulted in an increase in NEP (Figure 1.4). However, when loci were ranked by decreasing allele frequency differences between populations, NEP was randomly distributed.

Population assignment success was high in both creek populations when the full panel was used (Figure 1.5). In A Creek, 96.2% ± 0.65% of individuals and in C Creek, 93.9% ± 0.69% of individuals were correctly assigned to their population of origin. Not all panel loci were needed to achieve high population assignment success rates, as assignment success reached an asymptote after approximately 80 loci in both creeks, and only 43 loci were necessary to achieve population assignment of >90% in both creeks. However, use of the 35 loci specifically selected for population assignment was insufficient to achieve maximum assignment accuracy. When loci were ranked by decreasing power for parentage in each of the creek populations, more loci were needed to obtain high assignment accuracy: approximately 100 loci were needed to achieve >90% accuracy, compared to 40 when ranked by power to assign population (Figure 1.5). For all panel loci, mean locus-specific Fst was 0.016 ± 0.021 (Supplemental Table 1.1).

1.5 DISCUSSION

Here we developed an amplicon panel of SNP loci in Sockeye Salmon for use in pedigree reconstruction and in assignment of individuals to population of origin. These two analyses can be

combined to identify both natal and immigrant individuals in a population, ultimately supporting the study of processes such as gene flow, dispersal, and inbreeding. We described the detailed application of a bioinformatics workflow (McKinney *et al.* 2020) in panel optimization and quality control in downstream applications. We identified a final panel of 142 loci for use in parentage and 35 loci for use in individual assignment to population of origin. Genotyping success rate was high (>99% of individuals were genotyped at >90% of loci), highlighting the efficacy of the panel. We found that only 50 loci were necessary to achieve >95% parent-pair accuracy in populations of size N = 100, and 75 loci when N = 500. Population assignment success ranged from 93.9% to 96.2% in simulated populations when using all 172 panel loci.

Simulations on population sizes of up to 500 parents and 500 offspring were within the approximate range of the empirical test populations (Lin *et al.* 2016). The finding that population size had no effect on parent-pair assignment accuracy after 75 loci across these populations reveals the likely range of markers needed for similar panels. Parentage assignment in larger populations requires more loci for accuracy. The findings reported here generally agree with both theoretical (Anderson & Garza 2005) and empirical (Liu *et al.* 2016; Holman *et al.* 2017) studies that show between 50-100 loci are sufficient to achieve high parentage success. In a previous study on A Creek Sockeye Salmon, approximately 80 randomly selected polymorphic SNP loci were sufficient to achieve high parentage assignment success (Hauser *et al.* 2011). In contrast, we deliberately selected loci with high MAF, which reduced the number of loci necessary for parentage to between 50 and 75. Our finding that single-parent assignments require more loci than parent-pairs is consistent with past findings, as larger family groups are easier to identify than single related pairs (Wang 2007, Baruch and Weller 2008, Hauser *et al*. 2011). Although we simulated related individuals based on empirical estimates, more loci may be required in smaller

populations comprising a higher number of relatives. However, the linear increase of parentage assignment success across the entire range of loci suggests only a minor effect of locus ranking. Interestingly, there is an optimum number of loci because too many loci increased misassignment rate without substantially improving correct assignment. As expected, first-parent non exclusion probabilities (NEP) increased with decreasing minor allele frequency. However, NEP values showed no trend when loci were ranked by power for population assignment, demonstrating that the trade-off between these two applications was not as straightforward as one might expect and that loci selected for population assignment may still be useful for parentage.

Population assignment tests revealed that as few as 40 loci obtained population assignment accuracies >90%. Analyses based on ranking revealed that locus selection based on maximum divergence was a viable approach to improve assignment success. Given low F_{ST} values between the A and C creek populations (approximately 0.02-0.04, Lin *et al.* 2008), more loci were needed than between more divergent populations (Morin *et al.* 2009; Helyar *et al.* 2011). Therefore, we would recommend following well established methods to estimate the number of putative SNPs necessary to successfully assign individuals to populations *a priori*, based on known population F_{ST} estimates (e.g., Sylvester *et al.* 2018). We broadly recommend including approximately twice the number of SNPs necessary to account for locus dropout through optimization and quality control. Lower assignment success with loci selected for the parentage panel demonstrated the effects of ascertainment bias, but also showed that such loci are still useful for the population assignment.

While our bioinformatics protocol was successful and efficient, there are several changes and tradeoffs we would recommend considering for the development of multi-use GTseq panels. To avoid loci which may confound parentage and population assignment, we added a step to the

protocol of McKinney *et al.* (2020) which excluded putative panel loci with F_{IS} values > 0.1 or < -0.1 and were in linkage disequilibrium (LD) at the final quality control step (Figure 1.1, Step 5). We excluded loci in LD because they can confound parentage and population assignment analyses (Helyar *et al.* 2011); however, loci in LD can be informative for analyses other than parentage and population assignment – for example, they can be diagnostic for genomic features (i.e., inversions) – thus they may be considered for inclusion when developing similar panels.

We again recommend including at least twice as many loci in the primer design and optimization steps (Figure 1.1, Step 3) than would be needed for downstream applications, to buffer against locus drop out from many filtering steps. However, there are tradeoffs to consider between number of putative primers to include, cost of development, and panel efficacy. Primers currently cost approximately $20 per pair (Integrated DNA Technologies, Inc., Coralville, Iowa). The fewer loci included in a multiplex PCR reaction, the lower the likelihood of primer-primer interactions. Therefore, it is important to evaluate the relative cost of including additional primers in a test panel versus additional locus discovery and testing. For example, we included only the 60 top ranked loci in the test population assignment panel (Figure 1.1 Step 2e). However, we retained only 35 of these loci, which were insufficient to achieve maximum population assignment accuracy on their own. In retrospect, the screening of at least 100-150 loci would have cost an additional $800-1800. However, this oversight did not affect downstream applications in the final panel because the inclusion of parentage loci still resulted in greater than 95% population assignment success. A panel developed exclusively for population assignment, would have cost approximately $4,000 in additional optimization. This redundancy may be considered an additional benefit of multi-use SNP panels.

Although we used exclusively biallelic single-SNP loci, we recommend evaluating the use of microhaplotypes – multiple SNPs within an amplicon whose individual alleles can be combined into haplotype alleles. These markers are multiallelic and thus substantially increase the power for relationship inference (Baetscher *et al.* 2018) and genetic stock identification (McKinney *et al.* 2017a). Thus, microhaplotyes may also add increased power to studies where fewer amplicons are available, and because GTseq amplicons are very short (40-100 base pairs), they may be suitable for analysis of degraded DNA (Schmidt *et al.* 2019). In small populations, however, this advantage may vary as genetic drift can act to reduce the number of SNPs and resulting microhaplotype alleles per locus relative to large populations. Very few loci had two or more SNPs in the small populations we screened (15 loci with two SNPs and one locus with three SNPs). In addition, genotyping multiple SNPs per locus can be challenging. That said, Baetscher *et al.* (2018) and McKinney *et al.* (2020) have developed bioinformatic pipelines for this task (*Microhaplot* and *GTscore*, respectively). Our panel yielded high accuracy in both parentage and population assignment analyses using only single-SNP loci. Systems with larger population sizes, systems with less genetic variation, or different downstream applications may require additional power and may benefit from the inclusion of microhaplotypes.

In conclusion, development of multi-purpose panels should consider trade-offs in locus selection for different applications. However, these trade-offs may be less severe than expected. As such, loci developed for one purpose may still add power to other applications, such that the total number of loci needed may be substantially less than the sum of loci needed for each application. Inclusion of such multi-use loci may reduce the number of primers to test, increase the number of samples per sequencing lane, and ultimately reduce the cost of large-scale applications. In the case that some loci may decrease power for certain applications, it is simple to

bioinformatically subset the panel after sequencing (i.e., Aykanat *et al.* 2016, McKinney *et al.* 2020). Therefore, multi-purpose panels benefit from inclusion of a wide range of loci. We believe the information presented here on the processes involved in SNP selection, optimization, and quality control will be of use to others in designing successful multi-use panels for other species or populations.

1.6 FIGURES

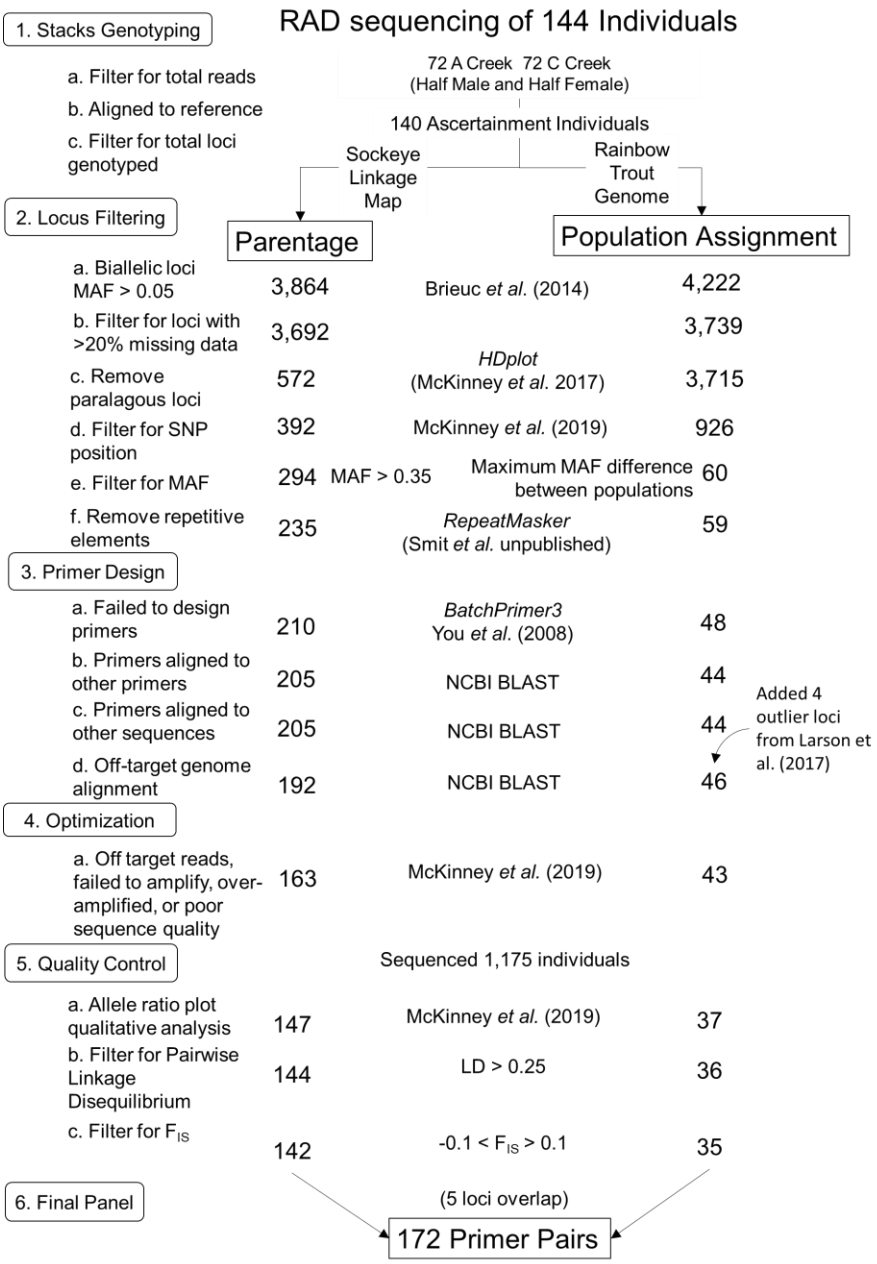


Figure 1.1. Schematic of the workflow and the results for SNP discovery, filtering, primer design, optimization, and quality control for a 172 SNP GTseq primer panel for use in individual assignment to population of origin (population assignment) and assignment of parent-offspring pairs (parentage). Numbers give the number of loci in the panel after each filtering step for use in parentage (left) or population assignment (right). Methods used for each step are given in the middle column

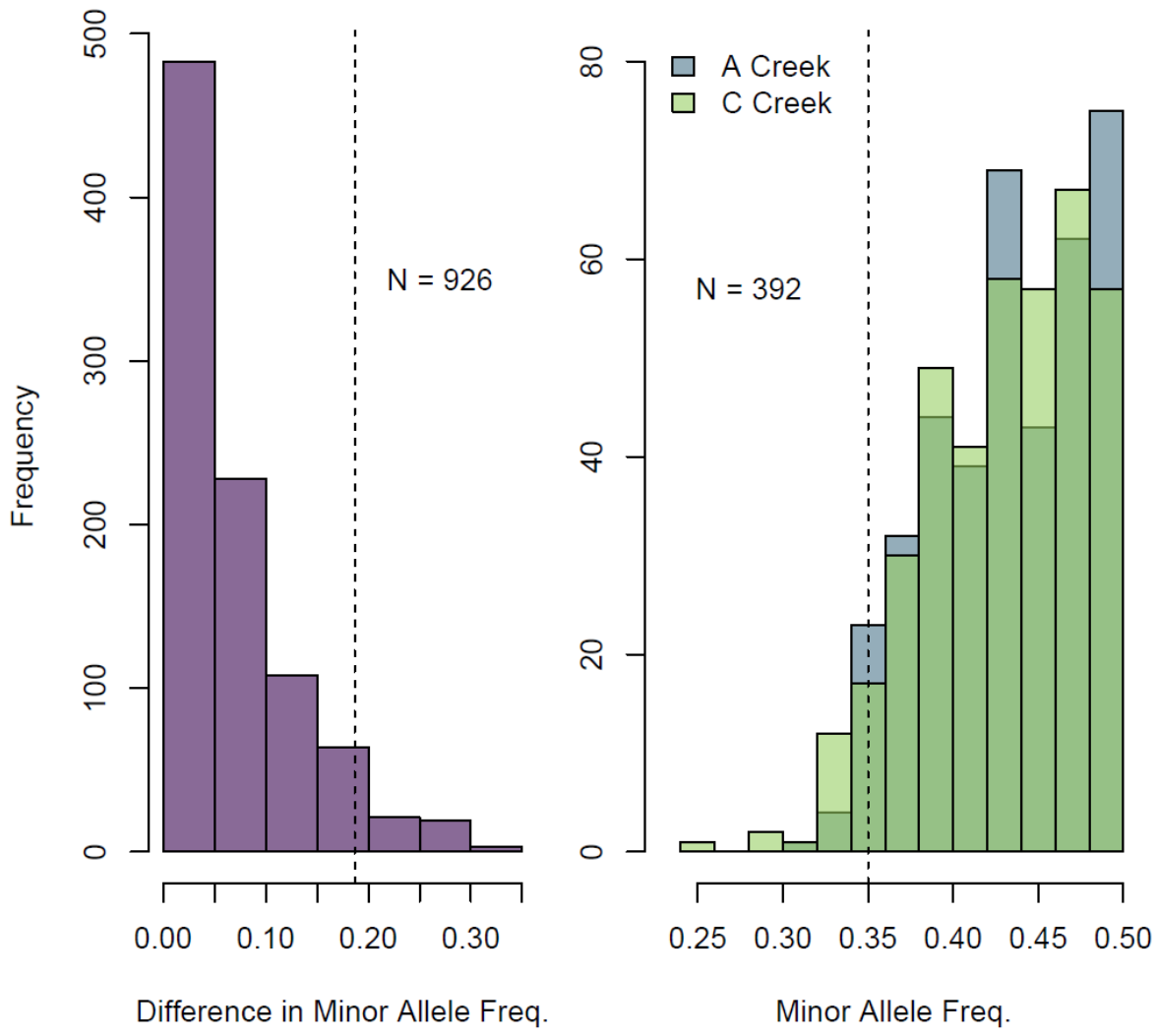


Figure 1.2. A. Distributions of the difference in minor allele frequency between the A Creek and C Creek populations of Sockeye Salmon at 926 bi-allelic loci screened for use in population assignment (left), and B. distributions of the minor allele frequencies at 392 bi-allelic loci screened for use in parentage (right). Loci were included following initial screening steps (step 2d in Figure 1.1). Loci to the right of the dashed lines were included in downstream analyses (step 2e in Figure 1.1).

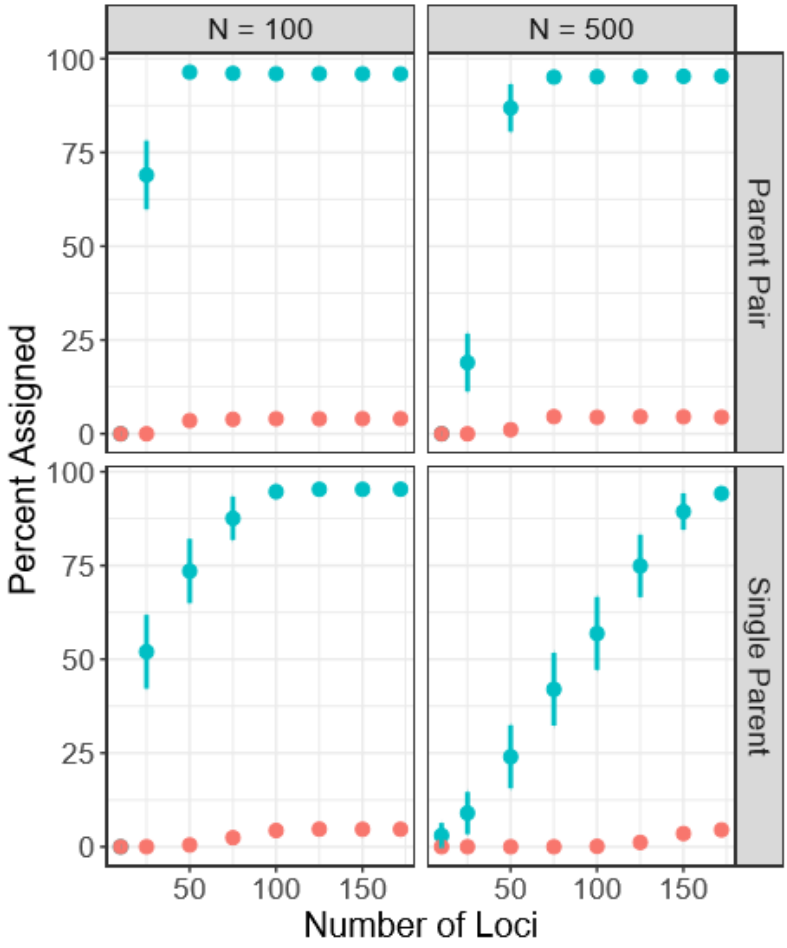


Figure 1.3. Percentage correct parentage assignment across variable numbers of SNP loci ranked from high to low minor allele frequencies in training data set. The mean percent of offspring correctly assigned (blue) or misassigned (red) to parents (y axis) is shown for 100 model repetitions. Error bars indicate 95% confidence intervals. Proportion of parents sampled was set to 0.7, and the percentage correct assignment is given as a proportion of sampled parents. The percentage misassignment was always <5% and represents the proportion of cases where a false parent was assigned, whether the true parent was sampled or not.

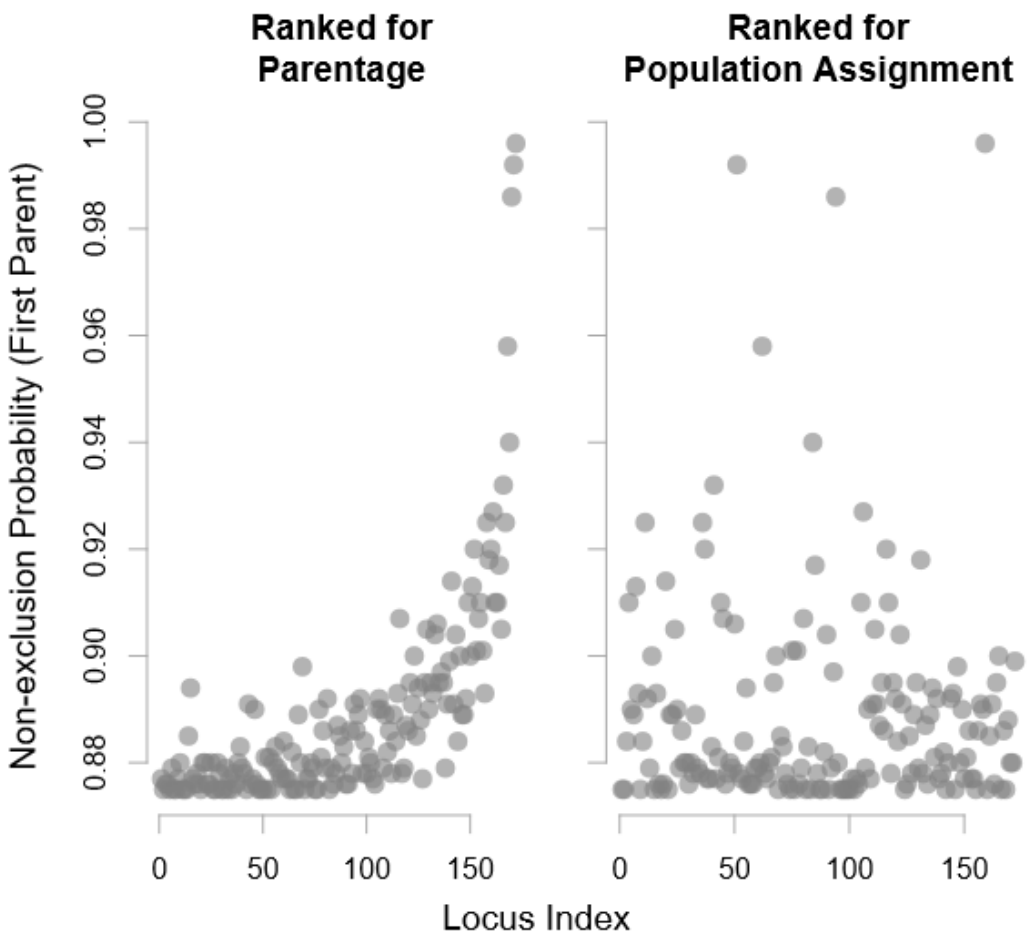


Figure 1.4. First-parent non-exclusion probabilities (NEPs, the probability of not excluding a false parent) in parentage analyses. Comparisons between parentage (left) and population assignment (right) panels NEPs (y axis) for each ranked panel locus (x axis) were calculated in Cervus3 (Kalinowski *et al.* 2007). Loci are ranked according to empirical minor allele frequencies high to low in the C Creek population (parentage markers, left) and allele frequency difference between the A and C Creek populations, high to low (population assignment markers, right).

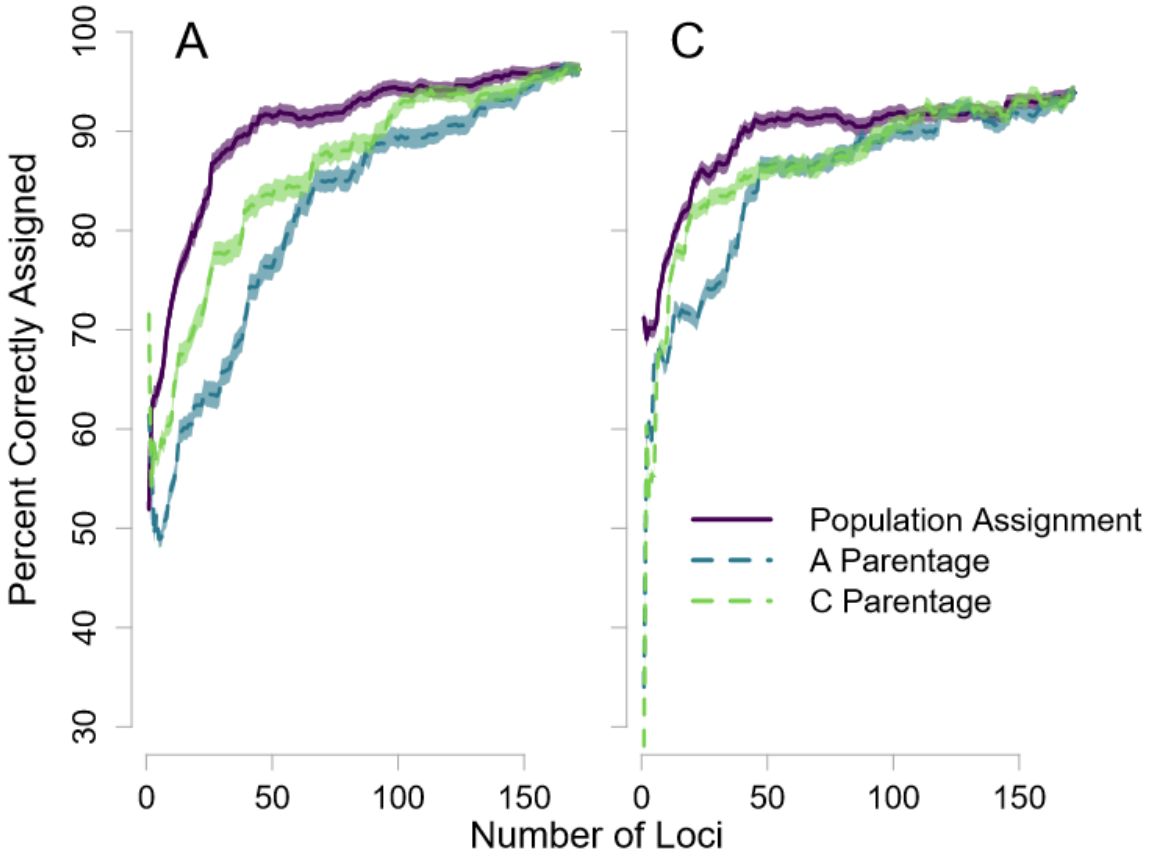


Figure 1.5. Analysis of population assignment accuracy (y axis) between marker panels developed for population assignment and parentage. Number of markers used in the analysis (x axis) are ranked by empirical allele frequency differences between populations (high to low, population assignment markers) and by within-population allele frequency (high to low, parentage panels) for each population (A and C Creeks). Shaded regions represent 95% confidence intervals of 1,000 model repetitions of 500 simulated individuals.

1.7 REFERENCES

Ackerman MW, Hand BK, Waples RK, Luikart G, Waples RS, Steele CA, Garner BA, McCane J, Campbell MR (2017) Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. *Evolutionary Applications* **10**, 146-160.

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR (2015) RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics* **202**, 389-400.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389-3402.

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

Anderson EC, Garza JC (2005) A description of full parental genotyping. *Unpublished report on file with the National Oceanic and Atmospheric Administration. Seattle, WA.*

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

Aykanat T, Lindqvist M, Pritchard V, Primmer C (2016) From population genomics to conservation and management: A workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon *Salmo salar*. *Journal of Fish Biology* **89**, 2658-2679.

Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC (2018) Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular ecology resources* **18**, 296-305.

Baetscher DS, Anderson EC, Gilbert-Horvath EA, Malone DP, Saarman ET, Carr MH, Garza JC (2019) Dispersal of a nearshore marine fish connects marine reserves and adjacent fished areas along an open coast. *Molecular Ecology* **28**, 1611-1623.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376.

Baruch E and Weller J (2008) Estimation of the number of SNP genetic markers required for parentage verification. *Animal genetics* **39**, 474-479.

Bohling JH, Starcevich SJ, Von Bargen J, Bailey EJ (2019) Describing Fine-Scale Patterns of Genetic Structure and Introgression of Redband Trout in a Complex River System. *North American Journal of Fisheries Management* **39,** 509-523.

Bradbury IR, Hubert S, Higgins B, Bowman S, Paterson IG, Snelgrove PV, Morris CJ, Gregory RS, Hardie DC, Borza T (2011) Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. *Molecular ecology resources* **11**, 218-225.

Bravington MV, Grewe PM, Davies CR (2016). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications* **7**, 13162.

Brieuc MS, Waters CD, Seeb JE, Naish KA (2014) A dense linkage map for Chinook salmon (*Oncorhynchus tshawytscha*) reveals variable chromosomal divergence after an ancestral whole genome duplication event. *G3: Genes, Genomes, Genetics* **4**, 447-460.

Campbell NR, Harmon SA, Narum SR (2015) Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular ecology resources* **15**, 855-867.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular ecology* **22**, 3124-3140.

Desrousseaux D, Sandron F, Siberchicot A, Cierco-Ayrolles C, Mangin B (2013) LDcorSV: Linkage disequilibrium corrected by the structure and the relatedness. *R package.*

Elliott L, Russello MA (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.

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5**, 184-186.

Harney E, Lachambre S, Roussel S, Huchette S, Enez F, Morvezen R, Haffray P, Boudry P (2018) Transcriptome based SNP discovery and validation for parentage assignment in

hatchery progeny of the European abalone *Haliotis tuberculata*. *Aquaculture* **491**, 105-113.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular ecology resources* **11**, 150-161.

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

Hill W, Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-231.

Hipperson RT, White SJ, Wright LI, Godley BJ (2017) A wild marine turtle pedigree reveals unequal effective contribution to the next generation. *Investigating variation in the life-history strategy of marine turtles*, 163.

Holman LE, Onoufriou A, Hillestad B, Johnston IA (2017) A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture* **476**, 59-64.

Janowitz-Koch I, Rabe C, Kinzer R, Nelson D, Hess MA, Narum SR (2019) Long-term evaluation of fitness and demographic effects of a Chinook Salmon supplementation program. *Evolutionary Applications* **12**, 456-469.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology* **16**, 1099-1106.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357.

Larson WA, McKinney GJ, Limborg MT, Everett MV, Seeb LW, Seeb JE (2015) Identification of multiple QTL hotspots in Sockeye Salmon (*Oncorhynchus nerka*) using genotyping-by-sequencing and a dense linkage map. *Journal of Heredity* **107**, 122-133.

Larson WA, Limborg MT, McKinney GJ, Schindler DE, Seeb JE, Seeb LW (2017) Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular ecology* **26**, 554-570.

Lescak E, Shedd K, Dann T (2019) Relative productivity of hatchery pink salmon in a natural stream. *NPRB Project* **1619**.

Lin J, Quinn T, Hilborn R, Hauser L (2008) Fine-scale differentiation between sockeye salmon ecotypes and the effect of phenotype on straying. *Heredity* **101**, 341.

Lin JE, Hard JJ, Naish KA, Peterson D, Hilborn R, Hauser L (2016) It's a bear market: evolutionary and ecological effects of predation on two wild sockeye salmon populations. *Heredity* **116**, 447.

Lin JE, Hard JJ, Hilborn R, Hauser L (2017) Modeling local adaptation and gene flow in sockeye salmon. *Ecosphere* **8**, e02039.

Liu S, Palti Y, Gao G, Rexroad CE (2016) Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture* **452**, 178-182.

Marshall T, Slate J, Kruuk L, Pemberton J (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular ecology* **7**, 639-655.

McKinney GJ, Seeb JE, Seeb LW (2017a) Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences* **74**, 429-434.

McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017b) Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular ecology resources* **17**, 656-669.

McKinney GJ, Pascal CE, Templin WD, Gilk-Baumer SE, Dann TH, Seeb LW, Seeb JE (2020) Dense SNP panels resolve closely related Chinook salmon populations. *Canadian Journal of Fisheries and Aquatic Sciences*. **77**, 451-461.

McKinney GJ, McPhee MV, Pascal C, Seeb JE, Seeb LW (2019) Patterns of linkage disequilibrium reveal genome architecture in chum salmon. *bioRxiv*, 729574.

Meek MH, Larson WA (2019) The future is now: amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular ecology resources*. **19**, 795-803.

Moran BM, Anderson EC (2018) Bayesian inference from the conditional genetic stock identification model. *Canadian Journal of Fisheries and Aquatic Sciences* **76**, 551-560.

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular ecology resources* **9**, 66-73.

Onda Y, Takahagi K, Shimizu M, Inoue K, Mochida K (2018) Multiplex PCR targeted amplicon sequencing (MTA-Seq): simple, flexible, and versatile SNP genotyping by highly multiplexed PCR amplicon sequencing. *Frontiers in plant science* **9**, 201.

Pemberton J (2008) Wild pedigrees: the way forward. *Proceedings of the Royal Society of London B: Biological Sciences* **275**, 613-621.

Peterson DA, Hilborn R, Hauser L (2014) Local adaptation limits lifetime reproductive success of dispersers in a wild salmon metapopulation. *Nature Communications* **5**, 3696.

Peterson DA, Hilborn R, Hauser L (2015) Exploratory behavior of dispersers within a metapopulation of sockeye salmon. *Behavioral Ecology* **27**, 126-133.

R-Core-Team (2013) R: A language and environment for statistical computing.

Richards-Zawacki CL, Wang IJ, Summers K (2012) Mate choice and the genetic basis for colour variation in a polymorphic dart frog: inferences from a wild pedigree. *Molecular ecology* **21**, 3879-3892.

Schmidt D, Campbell NR, Govindarajulu P, Larsen KW, Russello MA (2019) Genotyping-in-Thousands by sequencing (GT-seq) panel development and application to minimally-invasive DNA samples to support studies in molecular ecology. *Molecular ecology resources* **20**, 114-124.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011a) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular ecology resources* **11**, 1-8.

Seeb LW, Templin W, Sato S, Abe S, Warheit K, Park J, Seeb JE (2011b) Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Molecular ecology resources* **11**, 195-217.

Starks HA, Clemento AJ, Garza JC (2016) Discovery and characterization of single nucleotide polymorphisms in coho salmon, *Oncorhynchus kisutch*. *Molecular ecology resources* **16**, 277-287.

Steele CA, McCane M, Powell J, Vu N, Campbell MR, Narum SR (2017) Parentage Based Tagging of Snake River hatchery steelhead and Chinook salmon. *Bonneville Power Administration. Annual Progress Report, Project number 2010-031-00.*

Sylvester EV, Bentzen P, Bradbury IR, Clément M, Pearce J, Horne J, Beiko RG (2018) Applications of random forest feature selection for fine-scale genetic population assignment. *Evolutionary Applications* **11**, 153-165.

Templin WD, Seeb JE, Jasper JR, Barclay AW, and Seeb LW (2011) Genetic differentiation of Alaska Chinook salmon: the missing link for migratory studies. Molecular ecology resources **11**, 226-246.

Wang J (2007) Parentage and sibship exclusions: higher statistical power with more family members. *Heredity* **99**, 205-217.

Waters CD, Hard JJ, Brieuc MS, Fast DE, Warheit KI, Knudsen CM, Bosch WJ, Naish KA (2018) Genomewide association analyses of fitness traits in captive-reared Chinook salmon: Applications in evaluating conservation strategies. *Evolutionary Applications* **11**, 853-868.

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC bioinformatics* **9**, 253.

Zhao H, Li C, Hargrove JS, Bowen BR, Thongda W, Zhang D, Mohammed H, Beck BH, Austin JD, Peatman E (2018) SNP marker panels for parentage assignment and traceability in the Florida bass (*Micropterus floridanus*). *Aquaculture* **485**, 30-38.

Chapter 2. THE ECOLOGICAL DRIVERS AND FITNESS CONSEQUENCES OF DISPERSAL

Authors: Samuel A. May, David J. Paez, Lorenz Hauser, Ray Hilborn, Daniel E. Schindler, Kerry A. Naish

2.1 ABSTRACT

Dispersal among populations is a fundamental process in ecology and evolutionary biology due to its significance for population distribution, gene flow, and recruitment. However, examination of the ecological factors that influence dispersal, as well as the fitness consequences that follow dispersal, remains a key challenge in studies of wild populations, because of logistical difficulties associated with measuring ecological factors and fitness in wild populations. In addition, to fully examine the fitness consequences of dispersal, both absolute fitness (dispersers compared to their natal population) and relative fitness (dispersers compared to individuals in their new population) must be measured. Here, we examined the influence of ecological processes such as predation, population density, and reproductive timing on dispersal using 14 years of field surveys across two generations of returning adult Sockeye salmon. Full molecular-based pedigrees of 9,838 fish in two proximal stream-spawning populations in Bristol Bay, Alaska (A and C Creeks) were used to identify dispersers and quantify fitness of both dispersers and non-dispersers in each population. Dispersers were identified using genetic natal population assignment analyses, validated with the pedigrees. Fitness was measured as individual lifetime reproductive success: number of offspring identified for each pedigreed individual. We found that predation by bears, population density, and return timing to spawning sites were key factors influencing dispersal. We then tested whether individuals maximized their fitness by dispersing. We found asymmetric dispersal between the two streams, whereby individuals originating from C Creek had lower

dispersal rate than individuals from A Creek. Dispersers who moved from A Creek to C Creek increased their absolute fitness, whereas individuals moving in the opposite direction decreased their absolute fitness relative to non-dispersers in their respective natal streams. However, immigrants into A Creek had an increased fitness relative to local individuals in A Creek, whereas immigrants in C Creek had a reduced fitness relative to local individuals. We concluded that dispersers moved between populations in an attempt to increase their absolute fitness. This finding is relevant for gene flow and evolution because these processes depend on relative fitness, not absolute fitness. Differences in fitness of dispersers, as well as in dispersal rate, were likely due to a combination of differences in habitat and ecological variation within the two streams. This study is one of the first to provide empirical evidence of within-season ecological variation driving dispersal, fitness, and ultimate gene flow in natural populations. These findings have important implications for how human impacts to ecological processes may inadvertently affect dispersal and fitness.

2.2 INTRODUCTION

Dispersal of individuals among geographically or demographically isolated populations has dynamic effects on gene flow and is thus a fundamental process in ecology and evolutionary biology (Bowler and Benton 2005, Bonte and Dahirel 2017). Gene flow can influence the capacity of populations to respond to environmental changes such as habitat degradation or climate change (Garant et al. 2007, Räsänen and Hendry 2008, Travis et al. 2013). Thus, gene flow due to dispersal can indirectly affect population recruitment and persistence (Schreiber 2010, Yeakel et al. 2018). Dispersal also influences demography through its effects on population distributions, density, and competition (Bowler and Benton 2005, Cayuela et al. 2018). However, dispersal itself may be influenced by numerous environmental and biological processes (Clobert et al. 2004, Bowler and

Benton 2005). The processes affecting dispersal can therefore also indirectly affect fitness, productivity, and resilience. Thus, it is critical to understand the primary processes driving dispersal and their effects on fitness, to effectively predict how natural populations will respond to environmental changes (Bowler and Benton 2005, Doligez and Pärt 2008, Cayuela et al. 2018).

Dispersal has been increasingly recognized as a non-random process (Clobert et al. 2009, Cote et al. 2017). If dispersal were random, the proportion of dispersers in proximate populations should be equal, consistent across time periods, and uncorrelated with phenotypes or ecological processes. Yet, dispersal is rarely random in nature and is often dependent on individual phenotypes and environmental conditions (i.e., phenotype- or context-dependent dispersal; Clobert et al. 2004, Clobert et al. 2009). For example, studies have shown how dispersal may be affected by resource availability (Maag et al. 2018, Mishra et al. 2020), competition (Maag et al. 2018, Cayuela et al. 2019), predator avoidance (Poethke et al. 2010), and abiotic environmental conditions (Dittman et al. 2010, Cram et al. 2013). In addition, individual-level traits (i.e., phenotypes) can influence dispersal (Clobert et al. 2009), such as sex and morphology (Ducros et al. 2020) or physiology and behavior (Cote et al. 2010). However, a major challenge in studies of dispersal has been to accurately estimate dispersal between multiple populations, and to measure the ecological factors that influence dispersal in the source populations of immigrants (Cote et al. 2017, Cayuela et al. 2018). As such, there have been recent calls for further empirical investigations of the environmental and biological processes driving dispersal in wild populations, due to its relevance to gene flow, local adaptation, and productivity (Cayuela et al. 2018, Saastamoinen et al. 2018, Shizuka and Johnson 2020).

A key question in studies of dispersal and gene flow is whether individuals maximize their fitness by dispersing (Clobert et al. 2009, Bonte et al. 2014). The fitness of dispersers relative to

individuals in their breeding habitat (i.e., relative fitness) affects rates of gene flow among populations (Lenormand 2002, Orr 2009). Evolution within a population also depends on relative fitness (Wilson 2004). Therefore, dispersal is expected to evolve when the relative fitness of dispersers is increased compared to non-dispersers (Hendry et al. 2004, Duputié and Massol 2013). Yet, the relative fitness of dispersers may be reduced due to locally adapted difference among populations (Peterson et al. 2014, Mobley et al. 2019). In other words, under local adaptation, dispersal is expected to be a poor strategy for increasing relative fitness (Bonte et al. 2012). However, variable ecological factors such as competition, predation, or intrinsic habitat features may also affect the fitness of dispersers (Holt and McPeek 1996, Kawecki and Ebert 2004, Bowler and Benton 2005, Sæther and Engen 2015). Dispersers may increase their fitness compared to individuals in their source population (i.e., absolute fitness) by moving to a population with better quality habitat (Kawecki and Ebert 2004). Thus, investigations of whether individuals maximize their fitness by dispersing depend on whether the fitness of dispersers is compared to their source population (absolute fitness) or their new breeding population (relative fitness).

A key challenge for studies of dispersal and gene flow in natural populations has been to examine the compounding fitness effects of local adaptation and ecological variation (Kawecki and Ebert 2004, Räsänen and Hendry 2008). By measuring the fitness of dispersers and non-dispersers in both their natal and breeding populations, it is possible to comprehensively understand the fitness consequences of dispersal (Kawecki and Ebert 2004). Yet, measuring the fitness of both dispersers and non-dispersers in wild populations has also been a major challenge (Doligez and Pärt 2008). Measuring the fitness of dispersers and non-dispersers requires intensive sampling of parents and offspring to quantify fitness and distinguish dispersers from non-dispersers in multiple populations to make comparisons (Kawecki and Ebert 2004).

Salmonid fishes (especially *Oncorhynchus* and *Salmo* spp.) have often served as model species in studies of philopatry and dispersal, termed homing and straying respectively in the salmonid literature (Keefer and Caudill 2014). The anadromous lifecycle of many salmonids, associated with high site fidelity to natal rearing areas (i.e., natal homing), permits sampling of a high proportion of individuals as they return to their spawning grounds (Quinn 2018). In addition, the semelparous lifecycle of many salmon species means that reproduction in a single season and breeding habitat represents the net contribution of an individual to the next generation (i.e., lifetime reproductive success). Furthermore, proximate spawning populations are under nearly identical selection regimes throughout most of their lives, as juveniles mix after hatching (Quinn 2018). These life-history traits of anadromous salmonids allow for robust comparisons of the effects of ecological processes on evolutionary dynamics among heterogeneous spawning habitats (Carlson et al. 2011, Lin et al. 2016). Thus, salmon are an ideal system to study philopatry, dispersal, and their effects on fitness. Studies investigating these processes have direct implications for economically important salmon fisheries (Bett et al. 2017), hatchery supplementation efforts (Amoroso et al. 2017, Sturrock et al. 2019), and the management of at-risk stocks (Spromberg and Scholz 2011).

There is only limited empirical evidence to support the effect of environmental and biological processes on dispersal in salmonid populations, despite decades of research investigating homing and straying in salmon (Keefer and Caudill 2014, Bett et al. 2017). For example, evidence has shown that dispersal of returning adult salmon may result from variation in life-history strategies such as freshwater residency time or age at return (Westley et al. 2013, Westley et al. 2015). Abiotic environmental conditions (i.e., habitat quality) may also affect dispersal, such as substrate composition and stream morphology (Dittman et al. 2010, Cram et al.

2013). Studies investigating the fitness of immigrants may have a reduced relative fitness compared to residents in a given spawning population (Peterson et al. 2014, Mobley et al. 2019), consistent with predicted outcomes of local adaptation (Kawecki and Ebert 2004). However, these studies did not compare the absolute fitness of dispersers to residents in their natal habitats, which is important to understand how ecological variation within spawning seasons and local adaptation may interact to affect fitness and gene flow (Kawecki and Ebert 2004). Various environmental and behavioral processes may influence dispersal at numerous points during the lifecycle of salmonids, such as habitat features, predation, competition, and mate availability (Keefer and Caudill 2014). Yet, much of what is known about dispersal and its effects on fitness in salmonids has resulted from studies of hatchery-origin individuals, not natural populations, so the causes and fitness consequences of dispersal remain unclear (Keefer and Caudill 2014).

The overall aims of this study were first, to determine the ecological and phenotypic factors influencing dispersal among two proximate populations of Sockeye salmon, and second, to investigate the fitness consequences of such dispersal events. To achieve these aims, we utilized returning adults from a long-term monitoring program to reconstruct a multi-generation pedigree, identify dispersers, and estimate individual fitness. Our first aim was accomplished by examining whether predation, population density, or phenotypic characteristics such as body shape and return day explained the dispersal of individuals among populations. Our second aim was achieved by comparing the number of offspring of dispersers and non-dispersers across both natal and breeding populations. Our findings may have important implications for how anthropogenic activities influence gene flow, recruitment, and resilience in wild metapopulations.

2.3 METHODS

2.3.1 *Study System and Sample Collection*

Tissue samples were collected from returning adults as part of a long-term field program conducted on a wild Sockeye salmon metapopulation on Little Togiak Lake, Bristol Bay, Alaska (Peterson et al. 2014, Lin et al. 2016). This system comprises two small creeks (A and C Creeks; 350m and 450m long, respectively), approximately 1km apart, with 150-800 spawners returning to each creek per year (Figure 2.1). "A" Creek is characterized by shallow, uniform habitat with few undercut banks or deep pools, making it poor habitat for predator avoidance (Pess 2009, Lin et al. 2016). By contrast, "C" Creek is characterized by more heterogeneous habitat with many undercut banks and deep pools (Pess 2009). Sockeye salmon spawn in both creeks from late July through late August. A separate, genetically distinct population also spawns along the lake beaches at the mouths of these creeks from late July through late September (hereafter "Beach" spawners, Lin et al. 2008). Sockeye salmon are semelparous, that this, individuals die after spawning. Four-year-old fish dominate this system, with some age five individuals (Lin et al. 2016).

Adult fish were exhaustively sampled from both creeks since 2004. Newly arriving adults were captured daily either with beach seines at the creek mouth or with hand nets in the creeks. Fish were tagged with uniquely labeled plastic disc tags. Tag date was recorded along with fish length (mid-eye to hypural plate), body depth, and sex. Dorsal fin tissue samples were collected from each tagged fish and stored in 100% ethanol. Daily location (in meters from the mouth of the creek) of each tagged fish was recorded using visual surveys. Otoliths of fish found dead were removed for aging and cause of death was recorded (i.e., bear predation or senescence). This study utilized samples from 14 years of field collections (2004-2017), comprising two complete generations (F0, F1, and F2, Figure 2.1).

Field observations of tagged adults were used to determine the spawning populations ("A Creek" or "C Creek"). Beach spawners were inferred for fish that were sampled on the beaches and never observed in a stream. If a fish entered both streams, its spawning population was considered as the stream where it was seen most often (as in Peterson et al. 2014).

2.3.2 *Genotyping and Pedigree Reconstruction*

To identify dispersers and quantify fitness, we reconstructed a two-generation pedigree of individuals returning to the A and C Creek populations, using samples from 14 return years (2004-2017). Only individuals observed in creek habitats were used in pedigree reconstruction; however, a representative sample of individuals observed spawning in beach habitats were genotyped to obtain population allele frequencies for population assignment analyses (May et al. 2020). Peterson et al. (2014) reconstructed a one generation, two-cohort pedigree of the two creeks, spanning the F0 (2004-2005) and the F1 generation (2008-2010) using 11 microsatellite loci. We added two additional F0 cohorts (2006-2007), their offspring, and the F2 generation (2012-2017) using SNP loci. To integrate the previous, microsatellite-based pedigree with the updated SNP-based genotyping, all F1 fish previously genotyped at microsatellite loci were re-genotyped at SNP loci. All fish returning from 2006 through 2016 were genotyped at SNP loci, as well as five-year-olds returning in 2017, so that most offspring of the 2012 return year could be included. (Figure 2.1). To identify five-year-olds in 2017, the relationship between length and otolith-based age was used; individuals with body lengths >467mm and >495mm comprised 95% of all five-year-old females and males, respectively.

To genotype individuals at SNP loci, DNA was extracted from 8,685 individuals using Nexttec 1-Step DNA extraction kits (Nexttec Biotechnologie, Hilgertshausen, Germany). DNA was sequenced following the GTseq protocol (Campbell et al. 2015) at 172 target SNP loci

developed for parentage and population assignment in this system (May et al. 2020). Briefly, a two-step PCR with barcoded primers was used to produce amplicons with individual-specific barcodes, allowing ~2,000 individuals to be pooled in each sequencing lane. Single-end 100-base pair sequencing was performed on pooled samples on an Illumina HiSeq4000. Individuals were demultiplexed according to individual barcodes and genotyped using custom scripts from Campbell et al. (2015). To ensure confidence in parentage and population assignments, we removed any samples that were genotyped at less than 50% of loci.

We used a combination of both *Colony* (Jones and Wang 2010) and the R package *Sequoia* (Huisman 2017) to reconstruct the pedigree across the microsatellite and SNP datasets, respectively. Both programs utilize similar maximum-likelihood algorithms which simultaneously assign sibship and parentage. However, *Sequoia* is more robust for relationship inference with multiple and overlapping generations, as it allows for *a priori* life history data such as return year, age priors, and sex. *Sequoia* also assigns more relationship types (i.e., third- and fourth-order relationships), whereas *Colony* only assigns first- and second-order relationships. However, *Sequoia* is incompatible with microsatellite loci.

A microsatellite-based pedigree was first reconstructed using *Colony* (Jones and Wang 2010) with microsatellite genotypes reported by Peterson et al. (2014). Briefly, we used the full-likelihood model with unknown sexes and polygamy based on the default settings of medium run length, medium likelihood precision, and weak sibship prior (Jones and Wang 2010). A conservative value of 0.7 was used as the probability of sampling a parent, although a sensitivity analysis reported by Peterson et al. (2014) indicated that this parameter had a negligible effect on parentage assignments. To account for dispersal among creeks, as well as differences in age-at-return, all adults from both creek populations were considered as potential parents. To verify

parentage assignments, *post hoc*, we later confirmed assignments with field observations of sex to ensure parent-pairs consisted of one male and one female. Only parentage assignments with a likelihood score ≥95% were included in subsequent analyses.

The second, SNP-based pedigree comprised several overlapping generations. This pedigree was reconstructed using *Sequoia* (Huisman 2017). To ensure accurate assignment with multiple, overlapping generations, informative "age priors" for reproductive events were used. Most individuals in our study system return to breed at four-years-old, some at five-years-old, and very few at other age classes (Figure 2.1). Therefore, the difference in age of returning sibling pairs is typically zero and rarely greater than one year. When overlapping generations are present, *Sequoia* uses these age priors as probabilities to assign related pairs. We derived age priors from the microsatellite-based pedigree using the *MakeAgePriors* function within *Sequoia* (procedure described in Supplemental Figure 2.1). A low genotyping error rate of 0.001, associated with GTseq (Campbell et al. 2015) was used, and one individual from each of any putatively duplicated individuals identified by *Sequoia* was removed. We limited the maximum number of sibship clustering iterations per individual to 10, and all other settings were set to default.

The two partial pedigrees reconstructed with microsatellites and SNPs were merged using custom R scripts. If different parent-pairs were identified in the two partial pedigrees for a given F1 offspring (6/2,649 F1 offspring), we prioritized the parent-pair from four years prior to the return of the offspring, as most individuals in this system are four years old. We verified that sexes of parent-pairs comprised one male and one female and that parent pairs were sampled in the same year. Where logical inconsistencies occurred, we attempted to identify data entry errors before removing one parent randomly.

2.3.3 *Population Assignment and Identification of Dispersers*

To determine the natal population of returning adults, each individual was assigned to its most likely population of origin (either A, C, or Beach) using genetic mixture analyses implemented in *rubias* (Moran and Anderson 2018). All genotyped individuals were compared to reference populations (reporting units in *rubias*). These reference populations for the two creeks consisted of 200 randomly sampled offspring from each creek that were genotyped at >80% of loci, where the parent-offspring triads (dam-sire-offspring) were observed spawning in the same creek. The reference population for beach spawners consisted of individuals visually observed spawning at the beach, and who were also genotyped at >80% loci. To account for genetic drift, fish return year and spawning creek were used as reporting units. Separate population assignment analyses were performed using the SNP dataset (2006-2017) and microsatellite dataset (2004-2005; 2008-2010). To ensure high accuracy in assigned natal locations, only population assignments with a probability score ≥0.95 (summed across collections within reporting units) were used in further analyses.

To assess the power of the SNP and microsatellite markers to assign population of origin between the two creek and beach populations, we simulated populations of 500 individuals using allele frequencies from the reference datasets detailed above, following approaches implemented in May et al. (2020). We used the self-assign, leave-one-out method of Anderson et al. (2008), implemented in *rubias* with 1,000 model repetitions. The proportion of individuals correctly assigned to each population with a probability ≥0.95 was quantified (summed across collections within reporting units). Additional information for power analyses is detailed in the supplemental materials (Supplemental Figure 2.2).

2.3.4 *Relationship Between Ecological and Phenotypic Factors and Dispersal.*

We used generalized mixed effects models (GLMMs) to determine whether ecological mechanisms (population density and predation) and phenotypic characteristics (body shape and return day) were correlated with the propensity of individuals to disperse among populations. Population density was measured as fish/m^2 on a specific return day within a given spawning season, and in an individual's natal population. Predation was measured as the proportion of individuals killed by bears on a specific return day within a given spawning season, and in an individual's natal population. Body shape was quantified as the ratio of body length to body depth (Quinn and Foote 1994). Body depth was measured at the deepest vertical point, and body length was measured from mid-eye to hypural plate (in millimeters). Return day was defined as the Julian entry date into the spawning stream (hereafter "return day"). Only fish assigned natal populations in the creeks were included in these models, since density or predation was not measured in lake habitats.

We first built a full model where individual disperser phenotype (a binary trait; disperser vs. non-disperser) was assumed to be binomially distributed, and a function of return day, density, predation, sex, and natal population, which were all fixed effects. Sampling year was included as a random effect to account for interannual variation in dispersal rates.

(Model 2.1) $$dispersal \sim return\ day + return\ day^2 + density + density^2 + predation + natal\ population + sex + natal\ population + [YEAR]$$

We tested for possible variation in dispersal among the two streams and sexes by including these terms as fixed effects. We included quadratic effects of return day and natal population density on dispersal, following evidence of non-linear relationships between these variables

observed in preliminary analyses. Return day, density, and predation were scaled and standardized among all cohorts to a mean of zero and standard deviation of one to improve model convergence.

Next, we examined the effects of body shape on dispersal by constructing an additional model where dispersal was again assumed binomially distributed, and a function of individual body shape, sex, and natal population, which were fixed effects. This model was constructed separately from Model 2.1 to avoid type II error associated with including many variables in a single model.

(Model 2.2) $dispersal \sim body\ shape + \ body\ shape^2 + sex + natal\ population + [YEAR]$

As above, we accounted for variation among sexes and streams by including these terms as fixed effects, and interannual variation in dispersal by including a random effect of return year. A quadratic effect of body shape on dispersal was included, again following evidence of a non-linear relationship in this variable observed in preliminary analyses.

To ascertain which explanatory variables affected the response variable, a best-fit model for each GLMM (Models 2.1 and 2.2, separately) was selected from simpler models that omitted different fixed effects. Inference for this suite of models was accomplished using weighted Akaike Information Criteria values, a commonly used measure of the proportion of the total predictive power provided by the full suite of models (*w*AIC; Symonds and Moussalli 2011). The effects of explanatory variables on response variables were interpreted in the best-fit models.

2.3.5 *Fitness Consequences of Dispersal*

GLMMs were also used to test for fitness differences between non-dispersers and dispersers in each spawning population. Fitness was quantified using pedigree-based measures of individual lifetime reproductive success, defined as the total number of returning adult offspring

per adult individual (hereafter "reproductive success"). Individuals were categorized into three groups (i.e., 'dispersal category'): non-disperser, creek-to-creek disperser, or beach-to-creek disperser. In addition, fish 'dispersal direction' was defined by inferred natal and spawning populations. Thus, dispersal direction could occur from A Creek to A Creek or from C to C, to describe fish that spawned in the same creek where they incubated as eggs; or from A Creek to C Creek or C Creek to A Creek, to describe fish that dispersed. These models included adults returning from 2004-2012 (F0 and F1 generations), as reproductive success values for adults returning in subsequent years would have required an additional generation of sampling.

First, a full model with a negative binomial error distribution was constructed where reproductive success (zero and positive integer values) was a function of fixed effects for dispersal category, sex, and spawning population.

(Model 2.3**)** $$reproductive\ success \sim dispersal\ category + sex + spawning\ population + dispersal\ category{:}\ sex + dispersal\ category{:}\ spawning\ population + sex{:}\ spawning\ population + [YEAR]$$

To account for variation in reproductive success among sexes and spawning populations, both terms were included as fixed effects. Pairwise interaction terms between sex and dispersal category as well as spawning population and dispersal category were included to account for possible fitness differences in dispersal between the sexes and populations. In addition, a pairwise interaction term between sex and spawning population was included to account for possible fitness differences between males and females spawning in the two populations. Sampling year was treated as a random effect to account for interannual variation in fitness. Our preliminary investigation of interannual variation in reproductive success is detailed in the supplemental materials (Supplemental Figure 2.2).

Individuals dispersing into A Creek were shown to have reduced fitness compared to those dispersing into C Creek (Model 2.3, results section), suggesting an interaction between natal and spawning habitats. Thus, we also investigated fitness differences between dispersers from the two stream populations by constructing an additional model where reproductive success was a function of the direction of dispersal and sex. This model included creek-origin individuals, only, as we could not measure reproductive success of non-dispersers in the beach-spawning population.

(Model 2.4) $reproductive\ success \sim dispersal\ direction + sex + [YEAR]$

As above, inference for these models was performed using model selection through *w*AIC on a suite of simpler models. The effects of the explanatory variables on reproductive success were interpreted in the best fit models.

2.4 RESULTS

2.4.1 *Study System and Sample Collection*

A total of 4,095 individuals sampled in 2004-2005 and 2008-2010 were genotyped using microsatellites; of these, 3,926 (95.9%) were observed in the creeks and utilized in pedigree reconstruction. An additional 166 (4.1%) fish were sampled after being observed spawning on the beaches (Lin et al. 2008, Peterson et al. 2014) and used as baseline for population assignment of the beach fish. An additional 8,685 samples collected during 2006-2017 were genotyped with SNPs; of these, 8,391 (96.6%) were observed in the creeks and utilized for pedigree reconstruction, and an additional 294 (3.4%) were observed spawning on the beaches and utilized for population assignment. Only 104 of 6,123 individuals in the F0 and F1 generations were observed in both streams and were subsequently attributed to the spawning population where they were seen most often.

2.4.2 *Genotyping and Pedigree Reconstruction*

We used genotypes of microsatellite markers for individuals returning from 2004 through 2005 and 2008 through 2010 reported by Peterson et al. (2014; sourced from www.datadryad.org). Briefly, DNA was extracted from 4,377 fin-clipped individuals. After testing for evidence of null alleles and linkage disequilibrium, 27 individuals genotyped at less than eight of 11 microsatellite loci were removed. We used genotypes from the remaining 4,350 individuals for pedigree reconstruction and population assignment in this study.

Fish returning from 2006 through 2017 were genotyped at SNP loci. A total of 48 out of the 8,391 fish sampled in the creeks and genotyped at SNP loci were removed due to missing tissue samples or degraded DNA; 322 were removed due to poor sequencing quality; 18 were duplicated samples, and 38 duplicated genotypes identified by *Sequoia* were removed. Therefore, 7,965 individuals were included in pedigree reconstruction (95% of 8,391). Of the 294 representative beach spawners selected for sequencing, 168 were genotyped at >80% of loci and used for population assignment analyses. This low proportion of genotyped individuals was likely due to poor DNA quality in the older beach samples (i.e., from 2004-2005).

The full pedigree comprised 9,838 individuals after merging the SNP- and microsatellite-pedigrees (Figure 2.1). Of the 7,146 offspring returning as adults in the F1 and F2, only 15 were assigned parents who returned in different years; in these cases, one parent from each pair was removed, while retaining the parent from four years prior to the return of the offspring since this was the most frequent age class. Both parents were assigned to 5,031 individuals (70.4% of F1-F2), and a further 1,494 individuals were assigned one parent (20.8% of F1-F2); therefore, at least one parent was assigned to 6,525 individuals (91.2% of F1-F2). Lifetime reproductive success ranged from 0 to 33 offspring (mean = 1.78, σ^2 = 12.53. No offspring were assigned to 68.8% of

female and 57.7% of male parents in the F0 and F1. The overall sex ratio of individuals in the full pedigree was 61.7% females to 38.3% males, ranging from 48% to 73% females across populations and years (counts of individuals by sex and year provided in Figure 2.1).

2.4.3 *Population Assignment and Identification of Dispersers*

Spawning populations were identified for 4,243 individuals spawning in A Creek and 5,014 individuals spawning in C Creek through direct observation (total 9,257). Spawning populations were not assigned to 443 individuals observed only at the mouths of the creeks (4.6% of 9,838 individuals in pedigree) and 7 individuals which were observed in both creeks equally.

Individuals were assigned to natal populations using genetic population assignments (95% cut off criterion) for 3,847 fish spawning in A creek (91% of 4,243 observed spawners) and 4,270 fish spawning in C creek (85% of 5,014 observed spawners). Therefore, 8,117 individuals were assigned to both natal and spawning populations. Of these, 3,634 and 3,751 (total 7,385) were non-dispersers, 171 and 320 (total 491) were beach-origin immigrants and 42 and 199 were creek-origin immigrants (total 241) in A and C Creek, respectively. Beach-origin immigrants were more common than creek-origin immigrants in most, but not all, years (Figure 2.2). The proportion of immigrants in the stream populations ranged from zero to 0.25 and zero to 0.17 for creek- and beach-origin individuals, respectively. The proportion of all creek immigrants from A to C was greater than immigrants of the same sex from C to A (Figure 2.2); 7.2 times greater for females and 4.3 times greater for males across all years.

A power analysis suggested that population assignment analyses accurately assigned natal populations to individuals. A 95% probability cutoff indicated that the proportion of simulated individuals that were correctly assigned to their natal population (assignment accuracy) for SNPs

and microsatellites, respectively, was 0.79 and 0.91 to A Creek, 0.95 and 0.95 to C Creek, and 0.94 and 0.88 to the beach population, respectively (Supplemental Figure 2.3). This proportion was similar to a power analysis reported in a previous study which examined assignment accuracy in exclusively creek-origin individuals (May et al. 2020). Individual population assignments were compared to pedigree assignments; there was 98.6% agreement between individual population assignment and pedigree assignment in cases where the spawning population of both parents were known (4,369 out of 4,430 F1-F2 offspring). Due to the high concordance among genetic population assignments and the pedigree, we concluded that genetic population assignments accurately assigned natal populations and were suitable for identifying dispersers between populations. In the 61 cases where assigned natal population disagreed with pedigree-inferred natal populations, the population assignment method was used for consistency.

2.4.4 *Relationship Between Ecological and Phenotypic Factors and Dispersal.*

We investigated ecological factors influencing dispersal of individuals between creeks using measures of the proportion of individuals killed by bears in the natal population, density in the natal population (fish/m^2), and return day (Model 2.1). We found that dispersal was correlated with predation, population density, return day, and natal population. Specifically, the best fit model for these explanatory variables included terms for predation, a quadratic effect of density, a quadratic effect of return day, and natal population (AIC = 1628.99, *w*AIC = 0.71, Supplemental Table 2.2; $\beta_{predation} = 0.35$, $\beta_{density} = -0.89$, $\beta_{density}^2 = 0.49$, $\beta_{return\ day} = 0.50$, $\beta_{return\ day}^2 = -0.39$, $\beta_{natal\ population\ C} = -1.40$; Supplemental Figure 2.4).

The quadratic effect of return day on dispersal indicated that dispersal was more likely for individuals returning in the middle of the spawning season, although this effect was greater in A Creek than in C (Figure 2.3, left panels; Supplemental Figure 2.4). The peak dispersal day,

estimated by the model (day 223) was several days after the mean return day in both creeks (return day A: $\mu = 219$, sd = 6.75; return day C $\mu = 217$, sd = 6.26). Second, the quadratic relationship observed between density (measured in the natal population on a given return day) and dispersal indicated that dispersal was most likely during periods of either high (i.e., >0.6 fish/m^2) or low (i.e., <0.1 fish/m^2) density in the natal stream and least likely during periods of intermediate density (i.e., 0.3 fish/m^2; Figure 2.3, middle panels; Supplemental Figure 2.4). However, a positive relationship between density and dispersal was only observed during periods of high densities in A Creek, not in C Creek where density was generally lower (density A $\mu = 0.27$, sd = 0.20; C $\mu =$ 0.19, sd = 0.12). Third, individuals were most likely to disperse when bear predation (measured in the natal population on a given return day) was highest (Figure 2.3, right panels; Supplemental Figure 2.4). However, the effect of predation on dispersal was greater in A Creek where predation was greater (proportion killed by bears A $\mu = 0.05$, sd = 0.14; C $\mu = 0.02$, sd = 0.05). Sex was not correlated with dispersal.

The effects of body shape on dispersal were investigated using the ratio of body depth to body length (Model 2.2). The best fit model included only the term natal population; therefore, no relationship was detected between body shape and dispersal (AIC = 1775.11, *w*AIC = 0.35, $\beta_{\text{natal population C}} = -1.73$; Supplemental Table 2.3).

2.4.5 *Fitness Consequences of Dispersal*

Investigation of the relationship between dispersal and fitness found that fitness varied among non-dispersers, creek-to-creek, and beach-to creek individuals (Model 2.3). The best fit model contained dispersal category, sex, spawning population, and no interaction terms (AIC = 15331.37, *w*AIC = 0.25, Supplemental Table 2.4; $\beta_{\text{dispersal category creek}} = -0.114$, $\beta_{\text{dispersal category beach}} = -1.02$, $\beta_{\text{sex males}} = 0.58$, $\beta_{\text{spawning population C}} = 0.44$, Supplemental Figure 2.5, left panel). This result indicated that

creek-to-creek dispersers, with both creeks combined, had a reduced fitness compared to non-dispersers; although, the difference in fitness was small (Non-dispersers $\mu = 1.83$, $\sigma^2 = 13.10$; Creek Dispersers $\mu = 2.07$, $\sigma^2 = 14.3$). In addition, beach immigrants had a greatly reduced fitness compared to creek-origin individuals (Beach $\mu = 0.66$, $\sigma^2 = 3.04$; Creek-origin, combined $\mu = 1.83$, $\sigma^2 = 13.03$; Figure 2.4). The positive model coefficient for spawning population indicated that individuals spawning in C Creek had higher mean fitness than those spawning in A Creek (A Creek $\mu = 2.21$, $\sigma^2 = 18.0$; C Creek $\mu = 1.38$, $\sigma^2 = 7.57$). Lastly, the positive model coefficient for sex indicated that males had a higher mean fitness than females in both creeks (creeks combined: Males $\mu = 2.47$, $\sigma^2 = 19.36$; Females $\mu = 1.36$, $\sigma^2 = 8.24$).

We further examined the effect of migration direction on fitness and found that the absolute and relative fitness dispersers depended on whether dispersers moved from A to C or from C to A (Model 2.4). The best fit model contained terms for dispersal direction and sex (AIC = 14689.50, *w*AIC = 1, Supplemental Table 2.5; $\beta_{C\text{ to }C} = 0.45$, $\beta_{C\text{ to }A} = -0.23$, $\beta_{A\text{ to }C} = 0.35$, $\beta_{sex\ males} = 0.60$; Supplemental Figure 2.5). Individuals dispersing from A to C ($n = 137$, $\mu = 2.04$, $\sigma^2 = 15.60$) had increased absolute fitness compared to non-dispersers in A ($n = 2{,}599$, $\mu = 1.41$, $\sigma^2 = 7.73$, Figure 2.4). However, individuals dispersing from C to A ($n = 23$, $\mu = 2.22$, $\sigma^2 = 6.54$) had decreased absolute fitness compared to non-dispersers in C ($n = 1{,}942$, $\mu = 2.39$, $\sigma^2 = 19.60$). Yet, relative fitness of dispersers in C (immigrants) compared to non-dispersers in C was decreased, while relative fitness of immigrants in A was increased (Figure 2.4).

2.5 DISCUSSION

We examined the behavioral and ecological drivers and fitness consequences of dispersal among two stream spawning populations of Sockeye salmon (A and C Creeks), studied over two generations (14 years). We found that high predation rates in natal populations resulted in greater

dispersal rates. High and low population densities in natal populations also resulted in greater dispersal rates, while intermediate densities resulted in lower dispersal. Finally, returning slightly after the middle of the spawning season resulted in the highest likelihood of dispersing among streams. There was no support for an effect of sex or body shape on the probability of dispersing. Investigation of the fitness consequences of dispersal revealed several interesting results. First, there was higher mean reproductive success for individuals in C Creek compared to A Creek. Second, more fish moved from A Creek to C Creek than in the opposite direction, by a factor of 4.7. Third, dispersal behavior also had asymmetric fitness consequences. Fish that moved from A Creek to C Creek increased their absolute reproductive success compared to individuals in their natal streams, whereas this value decreased for individuals that moved from C Creek to A Creek. Importantly, the opposite was found when relative fitness was considered. Individuals that moved to C Creek from A creek increased their relative reproductive success compared to individuals in their new spawning streams, whereas this value decreased for fish that moved to A Creek from C Creek. These results provide empirical evidence of ecological and behavioral processes influencing dispersal and highlight mechanisms generating fitness variation in natural populations. From these results we concluded that dispersal was not a random process, and was driven by ecological dynamics such as predation, population density, and spawn timing. We also concluded that dispersers attempted to maximize their absolute fitness by moving to better quality habitat, comprising a lower predation risk and population density. Yet, relative fitness was likely affected by within-season variation in predation and density, which may have important implications for how ecological processes affect gene flow between wild populations.

2.5.1 *Relationship Between Ecological and Phenotypic Factors and Dispersal*

Increased predation by brown bears on specific days of the spawning season resulted in increased dispersal among streams (i.e., predator-induced dispersal; Poethke et al. 2010). Increased dispersal as a result of predation by brown bears demonstrates how predator-prey dynamics may affect rates of gene flow among populations within a metapopulation. One explanation for observed predator-induced dispersal may be the "scent of death" hypothesis, whereby individuals sense the presence of predators through olfaction of predator kairomones or of conspecific stress hormones (Kats and Dill 1998, Sih et al. 2010, Clinchy et al. 2013, Takahashi 2014). In this study, few dispersers were observed in their natal streams, indicating no evidence of physical interaction among dispersers and predators. Alternatively, dispersers may have entered their natal streams and were not sampled, as sampling took place only once per day during the spawning season, and a previous study showed exploratory behavior of dispersers in this system (Peterson et al. 2015).

The different effects of predation on dispersal in the two streams provides insight into how differences in habitat features might affect dispersal rates. Although predator-induced dispersal was found in both stream-spawning populations, the effect of predation on dispersal was greater in A Creek than in C Creek, consistent with previous findings of increased predation in A Creek (Lin et al. 2016). Differences in habitat likely explain these population differences in both predation rate and subsequent predator-induced dispersal. A Creek is characterized by shallower, more exposed habitat than C Creek, which has many undercut stream banks, deep pools, and more woody debris (Pess 2009, Bentley et al. 2014, Lin et al. 2016). These habitat differences likely reduced perceived or actual individual predation risk (Cunningham et al. 2013). Thus, there likely exists an interaction between habitat features, predation, and dispersal, which may affect connectivity among nearby populations.

Here, we provide the first empirical evidence that population density on a specific day of the spawning season can influence dispersal. Specifically, dispersal was greatest at either low or high densities (i.e., negative and positive density-dependent dispersal, respectively; Rodrigues and Johnstone 2014). At low densities, selectivity for body size by brown bears, and thus individual mortality risk increases (Cunningham et al. 2013). In addition, fewer potential mates are available at low densities (i.e., low resource availability). These factors likely led to increased dispersal. On the other hand, high densities likely resulted in increased competition for quality spawning habitat and mates, which drove dispersal at high densities. However, these density-dependent effects on dispersal varied between the creeks; in A Creek, dispersal increased at both high and low densities, while in C Creek dispersal was only observed to increase at low densities. As mean population density was approximately 30% lower in C Creek than in A, competition for mates and spawning sites was also likely reduced in C Creek. In addition, the proportion of individuals killed by bears was 60% lower in C Creek, likely due to habitat differences among the two streams. Thus, observed patterns of density-dependent dispersal likely resulted from a cost-benefit relationship between mate availability, competition, and predation risk, resulting in an optimal density where dispersal was least likely at intermediate density values. These patterns empirically validate recent theoretical work (e.g., Yeakel et al. 2018) demonstrating the importance of density-dependent dispersal for metapopulation connectivity and resilience to natural and anthropogenic disturbances such as habitat degradation, fragmentation, and climate change.

Our finding of a polynomial relationship between dispersal and spawning season highlights the importance of temporal dynamics to gene flow and connectivity (Fraser et al. 2007). Yet, we also observed that the peak likelihood of dispersal occurred 2-4 days after the mean return day of the populations. This finding might be explained by the fact that individuals returning slightly after

the mean return day may experience the highest population densities or predation. In salmon populations, selection coefficients for spawn timing are large and vary according to local optima (Tillotson and Quinn 2018). These optima in turn are influenced by competition for breeding habitat, egg incubation temperatures, and food availability after juvenile emergence (Webb and McLay 1996, Morbey and Ydenberg 2003, Gharrett et al. 2013). Therefore, when density or predation risk is high, a higher proportion of individuals may seek alternative spawning grounds to increase their chances of reproducing during the optimum return time. Yet, we also found that the effect of return time on dispersal was greatly reduced in C Creek compared to A Creek, likely due to the much lower density and predation risk in C Creek. Therefore, we concluded that dispersal was likely influenced by dynamic interactions among habitat features, predation intensity, competition, and mate availability, while individual body shape and sex had no effect.

2.5.2 *Fitness Consequences of Dispersal*

A key question in dispersal studies is whether individuals maximize their fitness by dispersing, as expected by theory (Hamilton and May 1977, Ronce 2007, Clobert et al. 2009, Cote et al. 2010). Yet empirical validations of this theory are rare, and most dispersal studies demonstrate a decreased relative fitness of dispersers in their new habitat attributed to local adaptation (e.g., Peterson et al. 2014, Mobley et al. 2019, but see Bonte et al. 2014). Here, we were able to examine the symmetry of dispersal and both the relative fitness (i.e., local vs. foreign) and the absolute fitness (i.e., home vs. away) of dispersers (Kawecki and Ebert 2004). Our finding that dispersal was asymmetric between the two streams (i.e., more dispersal from A to C Creek) indicated that dispersal was not random and was likely an informed process (Clobert et al. 2009). Furthermore, our finding that both dispersing and non-dispersing individuals spawning in C Creek had more offspring on average than those returning to A Creek indicated that individuals likely

dispersed in attempt to increase their absolute fitness. Combined, both the increased frequency of dispersers and increased fitness of dispersers moving to C Creek provide empirical validation of the theory that dispersal is expected to increase mean fitness.

The sum of the fitness benefits and costs to individuals reproducing in a given habitat comprise its overall intrinsic habitat quality, which is often quantified using demographic measures such as survival and reproduction (Johnson 2007, Bieri et al. 2018). The frequency of dispersal and the absolute fitness of dispersers is expected to increase from low quality to high quality habitats (Lin and Batzli 2001, Baguette et al. 2011). Here, asymmetric dispersal, in combination with differences in habitat features, predation risk, population density, and fitness are consistent with the expected outcomes of greater habitat quality in C Creek during the course of this study (Kawecki and Ebert 2004, Johnson 2007, Bieri et al. 2018). Indeed, dispersers from A Creek to C Creek had 79% more offspring on average, compared to non-dispersers in their natal stream. In contrast, few individuals dispersed in the opposite direction from high to low quality habitat, although the 22 individuals moving in this direction had 7% fewer offspring on average than non-dispersers in their natal habitat. (Hamilton and May 1977, Ronce 2007, Clobert et al. 2009, Cote et al. 2010, Bonte et al. 2014). From this evidence, we concluded that individuals dispersed in an attempt to maximize their absolute fitness by moving from low quality (A Creek) to high quality (C Creek) habitats (Cram et al. 2013, Bonte et al. 2014). These results provide empirical evidence for the theory that dispersal and fitness should increase from low to high quality habitats.

The expected fitness outcomes of local adaptation between two populations predict that relative fitness of maladapted, foreign individuals should be lower than local individuals in both populations (Kawecki and Ebert 2004, Blanquart et al. 2013). Here, fitness differences among creek immigrants and non-dispersers reproducing in the same habitat (i.e., local vs foreign) were

not consistent with these expected outcomes. Here, relative fitness was decreased for immigrants in C Creek but increased for immigrants in A Creek. From this result, we concluded that A and C Creeks were not adaptively differentiated, despite previous evidence demonstrating significant genetic differentiation (Lin et al. 2008).

There are two likely explanations that could explain the absence of local adaptation between these two populations. First, gene flow, weak selection, or genetic drift may have been sufficient to prevent adaptive divergence (Garant et al. 2007). If this were the only explanation for the absence of local adaptation, we would expect no difference in the relative fitness of local and foreign individuals reproducing in a given habitat (Kawecki and Ebert 2004). Yet, we did observe differences in relative fitness. A second explanation is that within-season temporal variation in ecological conditions are expected to result in differences in relative fitness of local and foreign individuals (Kawecki and Ebert 2004). Temporal variability is also expected to promote the evolution of dispersal (Levin et al. 1984, Holt and McPeek 1996); however, empirical validations of this theory remain a key challenge in studies of natural populations (Cayuela et al. 2018). Here, we observed greater variation in predation and density in A Creek relative to C. Therefore, increased variation in density and predation may have conveyed a fitness benefit to immigrants in A Creek during periods when conditions were favorable, despite predation and competition being greater overall (i.e., on average) in A Creek. This study did not explicitly test for the evolution of dispersal phenotypes. However, if dispersal has a genetic basis, our findings may empirically demonstrate how temporal variation in ecological processes may increase the relative fitness of dispersers and promote the evolution of dispersal phenotypes.

Beach-origin individuals had almost three times fewer offspring than creek-origin individuals in both stream habitats, and differences in the mean fitness of beach individuals

spawning in the two streams were small. Fitness differences between beach- and creek-origin fish are likely due to locally adapted differences among the beach and creek ecotypes, which have distinct body morphologies (Peterson et al. 2014, Larson et al. 2017). Yet reduced fitness in beach immigrants could also be due to exploratory behavior of beach individuals that subsequently returned to the beach habitats to spawn (Peterson et al. 2015). It was not possible to quantify fitness in the beach population, so we could not fully test whether the beach and creek populations exhibit locally adapted differences (Kawecki and Ebert 2004). In other words, while the beach fish had reduced reproductive success in the creek habitats, we cannot test the relative fitness of creek fish who reproduce in the beach habitats.

2.5.3 *Assumptions and Limitations*

This study made the following assumptions. First, we assumed that missing data in the pedigree were random and did not systematically bias our assessment of ecological and phenotypic characteristics or reproductive success values. We found that 70% and 20% of the 7,146 individuals in the F1 and F2 generations were assigned both parents and one parent, respectively. Thus, pedigree size and completeness were high relative to similar pedigree-based studies (e.g., Aguillon et al. 2017, Reid et al. 2021). However, we did not sample the adjacent beach spawning population, and it is likely that some individuals dispersed from the creeks to the beaches (Lin et al. 2008). A higher proportion of creek to beach dispersers may have come from A Creek, following our finding of increased dispersal out of A Creek. However, due to the close proximity of the stream habitats and low competitive ability of small-bodied stream fish relative to beach fish in the lake habitat (Lin et al. 2008), we found it unlikely that dispersal to the beach population was asymmetric between the two creeks.

Second, we assumed that population assignment analyses accurately determined individual natal populations and that subsequent identification of dispersers was unbiased regarding population. Power analyses indicated that between 79% and 95% of A and C individuals could be correctly assigned to their natal populations of origin. However, this test depended on a random draw of 200 reference individuals from each population and was lower than expected in some populations, likely due to undetected hybrids in the reference populations used. We assigned population of origin to all individuals using a repeated leave-one-out approach with a 95% confidence threshold, utilizing all individuals in the pedigree to increase power. In comparing natal population assignments to the pedigree, we found >97% agreement in cases where two parents were assigned. From this we concluded that natal populations were assigned accurately, with few misassigned individuals. The use of a 95% confidence threshold may have increased type II error associated with correctly assigned individuals being assigned no natal location; however, this would merely have decreased the number of individuals in the dataset and would not have resulted in any systematic bias.

Third, we assumed that estimation of individual return days approximated individual spawn timing. We found this likely as Sockeye salmon are known to spawn with 1-3 days of entering spawning streams (Esteve 2005, Lisi et al. 2013).

Fourth, we estimated population density and predation in the natal population on the return day of individuals to their spawning populations. Thus, we assumed that dispersing individuals explored their natal population on the same day that they entered their spawning population. While this may not have been true in all cases, we still found it likely that the emigration, transience, and immigration phases of dispersal (Cayuela et al. 2018) took place within 1-3 days, consistent with previous findings of exploratory behaviors of dispersers in this system (Peterson et al. 2015). From

assessments of these assumptions, we conclude that our study design was unbiased with regard to population, although the effect of the beach population on between-stream dispersal remains unclear.

2.5.4 *Conclusions and Impacts*

This study is one of the first to provide empirical evidence of the effects of ecological dynamics and habitat on dispersal and its fitness consequences. Specifically, we conclude that dispersal was likely an informed process, resulting from a combination of differences in habitat features, predation intensity, and population densities. These findings have implications for how anthropogenic impacts on local environmental or demographic processes may have consequences for dispersal rates, gene flow, adaptation, and ultimately population recruitment and resilience.

Anthropogenic impacts such as habitat degradation, harvest, supplementation, or climate change are of growing concern to wild salmonid populations (Reed et al. 2011, Crozier et al. 2019, Thompson et al. 2019). Such anthropogenic effects may influence ecological and environmental dynamics that are essential to wild population viability. For example, population densities in salmonids are affected by supplementation programs (Buhle et al. 2009, Tatara and Berejikian 2012, Amoroso et al. 2017) and commercial harvest (Minto et al. 2008, Kolchin et al. 2021). In addition, anthropogenic development projects and climate change can also affect the availability and quality of salmon spawning habitats at various lifecycle stages (Pess et al. 2003, Schindler et al. 2008, Pitman et al. 2021). We demonstrate how changes to wild population dynamics such as density, predation, and habitat can affect dispersal and gene flow among populations. Therefore, human activities which affect these dynamics may also indirectly impact fitness, genetic diversity, and adaptive divergence in wild metapopulations (Thompson et al. 2019). Adaptive life history diversity in heterogenous landscapes increases the stability of metapopulations (Schindler et al.

2015, Brennan et al. 2019). It follows that anthropogenic effects on dispersal and gene flow may decrease the effectiveness of ecological portfolios to buffer against future changes.

2.6 FIGURES

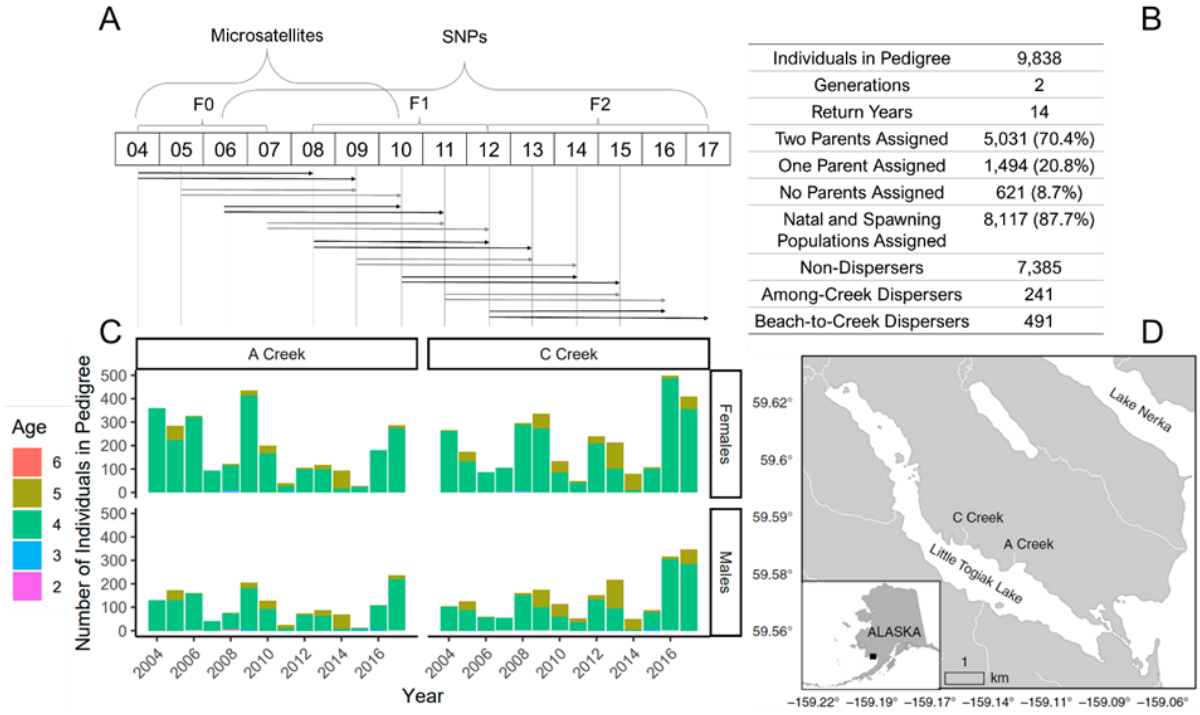


Figure 2.1. A) Schematic of pedigree structure; includes only four- and five-year-old individuals. Horizontal arrows track cohorts from brood year to return year, and alternating gray colors indicates age structure for individuals born in the same year. Brackets indicate generation (F0, F1, or F2) and marker type used for sequencing (microsatellites or SNPs). B) Summary statistics for pedigree reconstruction, population assignment, and identification of dispersers. Percentages are given in reference to the F1-F2 generations for parentage assignments and to total number of individuals in the pedigree for population assignment. C) Number of individuals included in the reconstructed pedigree (y axis) across adult return years (x axis). Data separated by stream spawning population (A Creek or C Creek), sex, and age class determined from pedigree (where known; colors). D) Map of study sites previously published by Peterson et al. (2014). Square inset map indicates sample location on the Wood River System, Bristol Bay, AK.

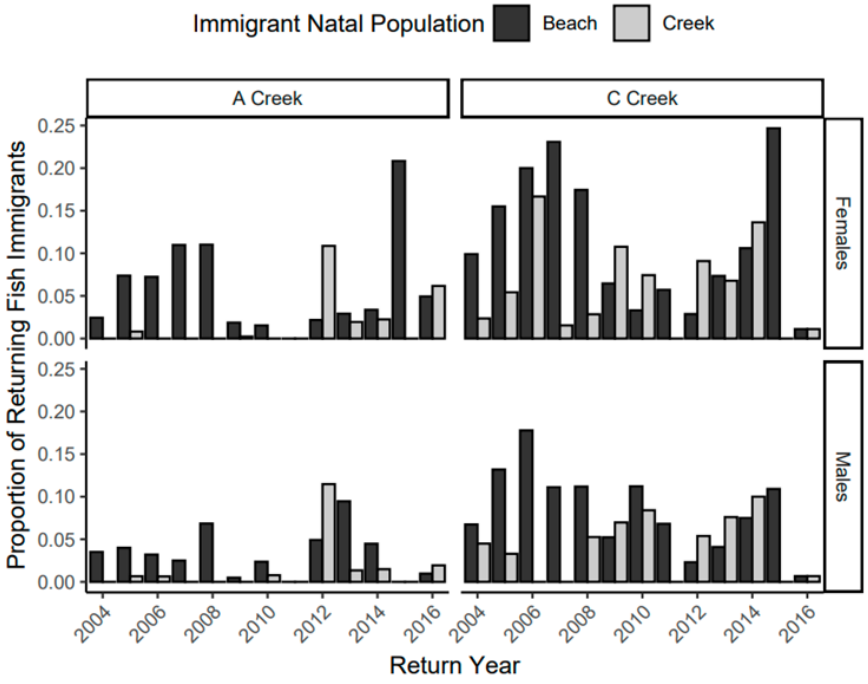


Figure 2.2. Proportion of stream-spawning individuals identified as either creek-to-creek immigrants (grey; Creek), or beach-to-creek immigrants (black; Beach) for each sex and spawning population. Proportions are given for each return year (x-axis).

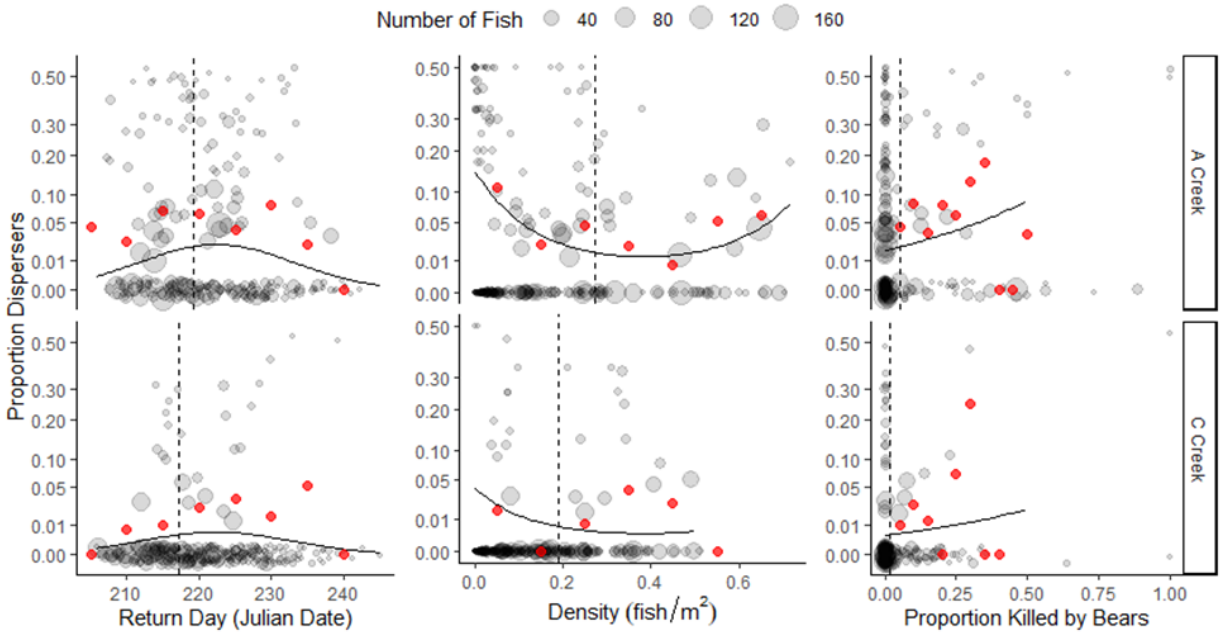


Figure 2.3. Effects of return day, population density, and predation (x-axis) on the proportion of dispersers (y –axis, square root scale) across two creeks. Return day was estimated as the Julian date each fish was first seen in the stream. Density was estimated as the density (fish/m2) in the individual's natal population on its return day. Predation was estimated as the proportion of individuals killed by bears in the individual's natal population on its return day. Dashed lines give the mean of each x-axis variable in each population. Empirical estimates for proportion of dispersers (points) are given as either the proportion of dispersers for each x-axis value within return years (gray) or the proportion of dispersers for binned x-axis values among all return years (red). Solid lines give the fitted values of the best-fit mixed effects model (Model 2.1 in text).

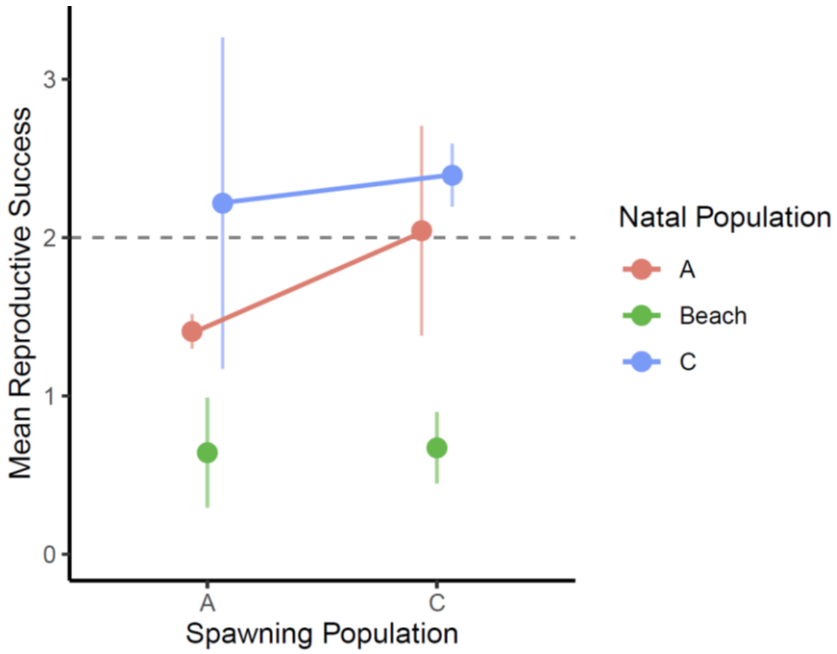
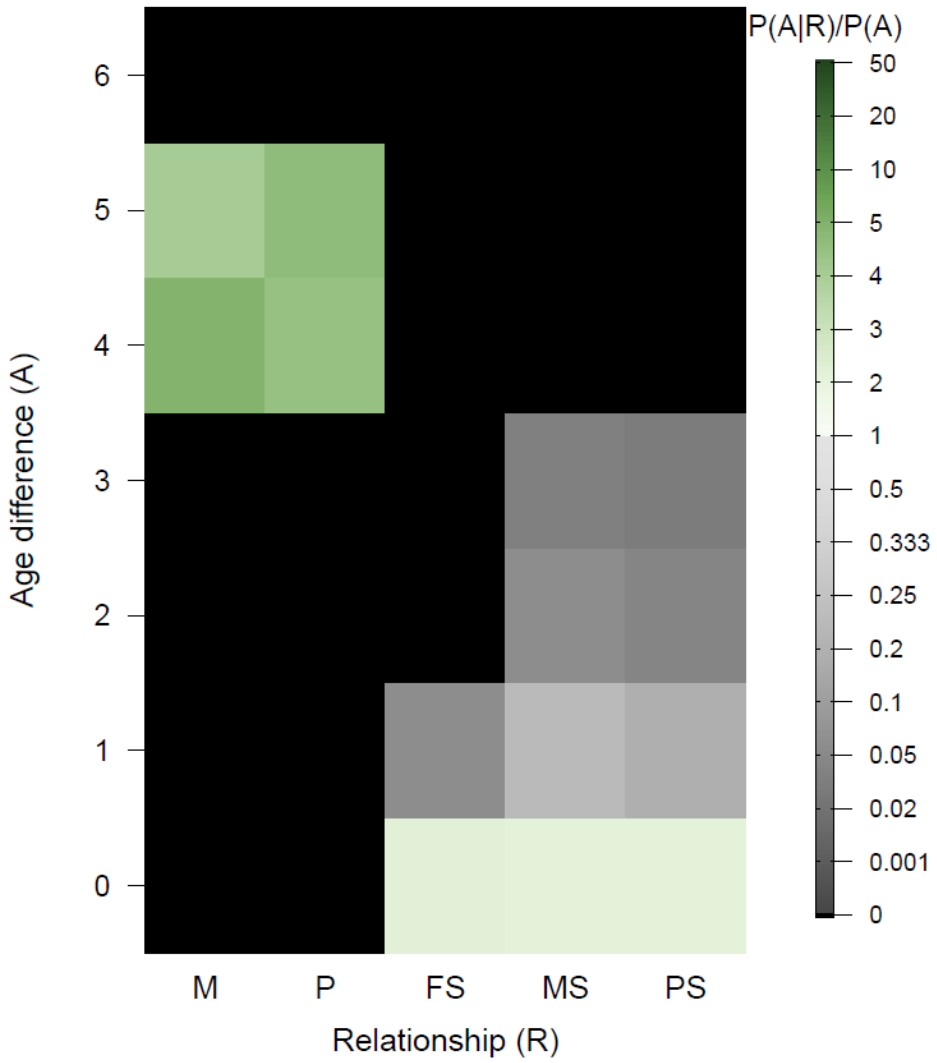


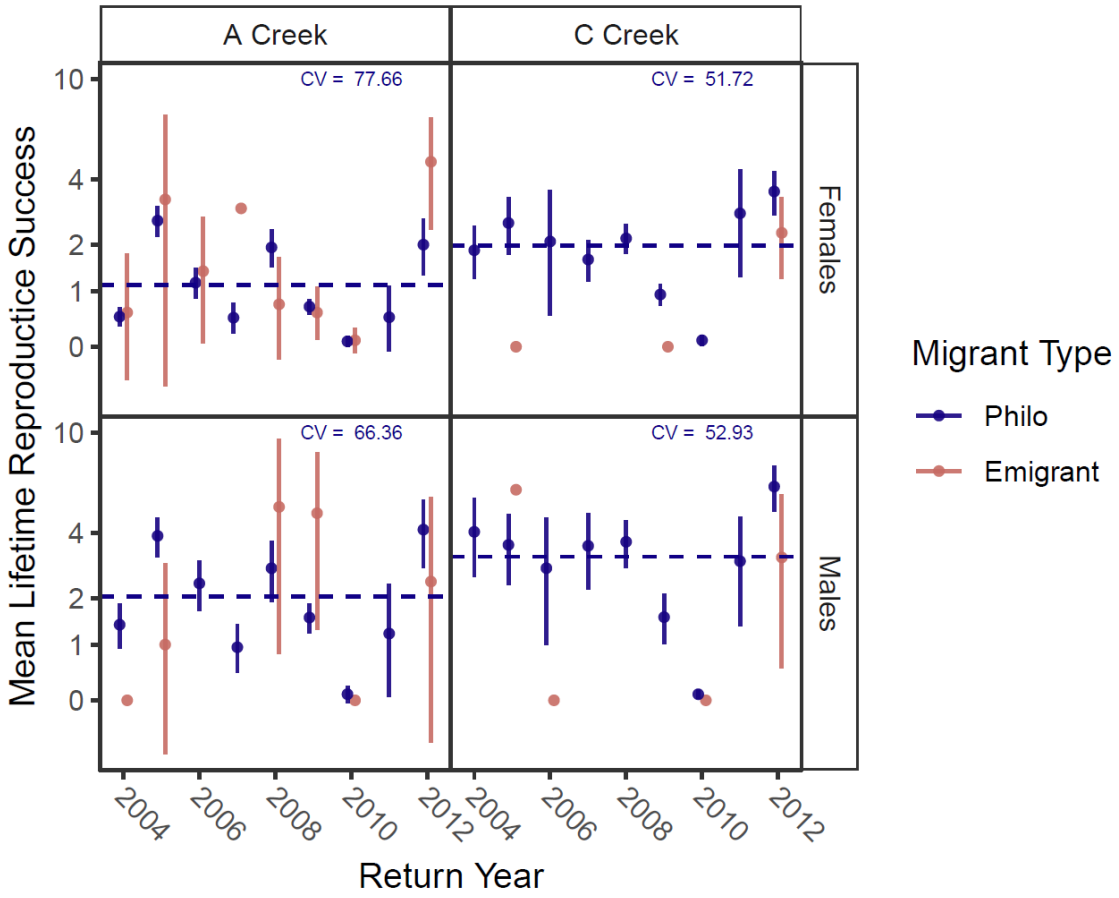
Figure 2.4. Mean reproductive success (fitness) of non-dispersing and dispersing individuals representing three natal habitats. Individuals are grouped by their natal population (colors) and spawning population (x axis). Reproductive success (y axis) is the mean number of offspring of individuals in each dispersal category, derived from the pedigree. Error bars are given as 95% confidence intervals. Solid lines compare the absolute fitness of individuals who dispersed from their natal population. Fitness of fish from the beach population, in their natal habitat was not known. Dashed line indicates the mean reproductive success value of two, necessary to maintain a stable population size over time. Inference was performed using mixed effect models (Models 3-4 in text, but see also Supplemental Figure 2.5 for model coefficients and standard errors).

2.7 SUPPLEMENTAL MATERIALS



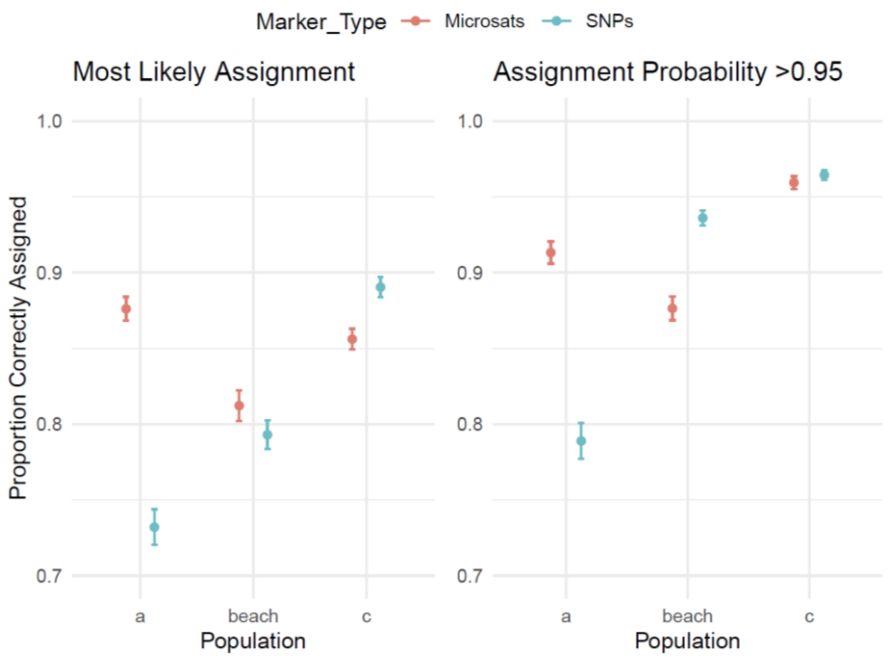
Supplemental Figure 2.1. Age priors used for SNP pedigree reconstruction in Sequoia. Priors were generated from the microsatellite-based pedigree (see text for details) using the MakeAgePrior function in Sequoia. Relationship (x-axis) is either mother (M), father (P), full sibling (FS), maternal half-sibling (MS), or paternal half-sibling (PS). Possible age differences are provided within relationship categories (y-axis). Age priors (color scale) are a frequency function of the age difference in years (A, y-axis) given each relationship (R), divided by the frequency of each age difference in the pedigree (Huisman 2017). These results show that nearly all parent-offspring relationships are either four or five years apart, reflecting the high frequency of four- and five-year-old individuals in these populations. Similarly, most full-sibships are zero

years apart, with some pairs one year apart, as >80 individuals in the system are four years old and the remainder are primarily five years old. The greatest variance in age difference is among half-sib pairs, as these may be up to two years apart, and rarely three years apart. Although, most half-sib pairs return in the same year (zero years apart).

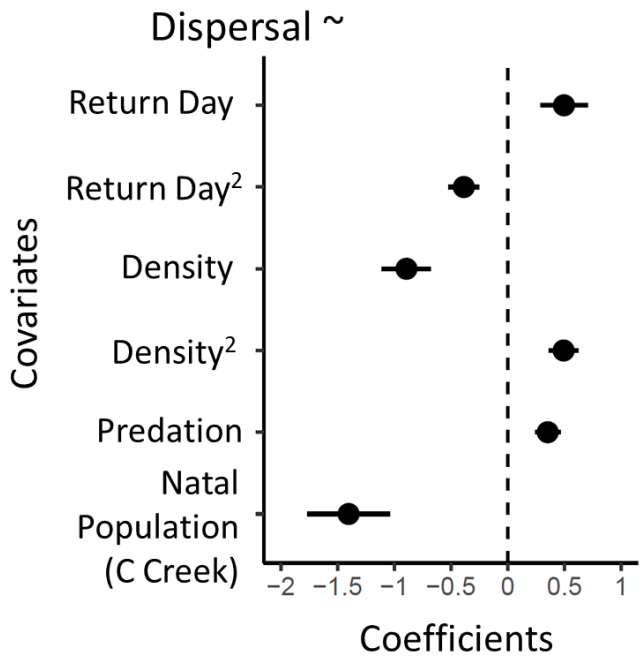


Supplemental Figure 2.2. Analyses of mean lifetime reproductive success values (y-axis) across return years (x-axis) for individuals originating in either creek population. Individuals were grouped by dispersal category (color) as either philopatric (i.e., non-dispersers, blue) or emigrants originating in one stream but spawning in the other stream (i.e., dispersers, orange). Error bars are given as 95% confidence intervals about the mean. To examine interannual variation in reproductive success in each stream habitat, the coefficient of variation among years was estimated for non-dispersing individuals, only (CV, upper right values). Non-dispersers were used to remove differences in fitness estimates due to differences among populations. These results showed that the mean fitness of dispersers was higher than that of non-dispersers for females in A Creek in five of eight years with detected dispersers. Yet male dispersers in A Creek had higher fitness than non-dispersers only in two of six years with detected dispersers. However, years where dispersers had lower fitness than non-dispersers also had a smaller sample size (i.e., one individual, shown here as no 95% confidence interval about the mean). In C Creek, these results showed that dispersers were detected in few years compared to A Creek (C: 5 years;

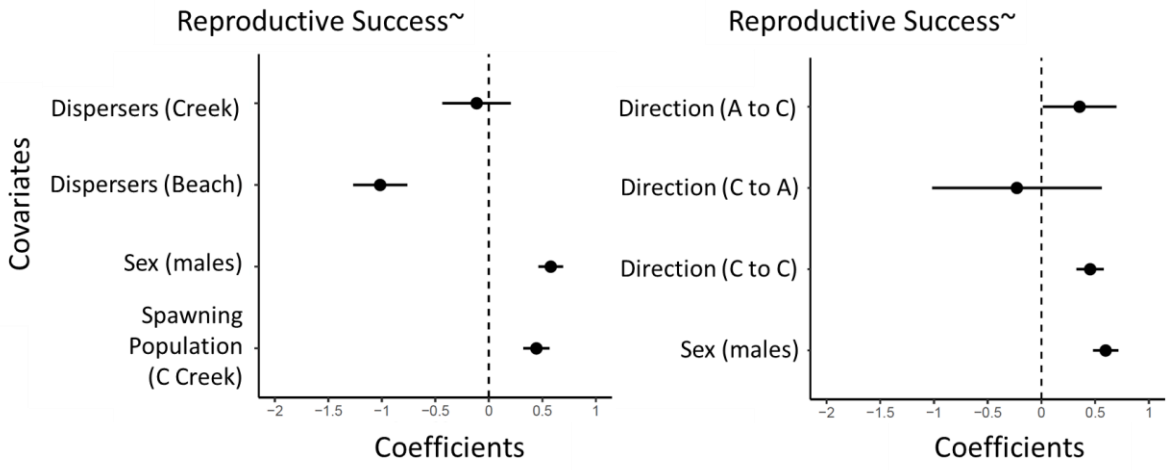
A: 8 years). In addition, dispersers in C Creek had decreased fitness compared to non-dispersers in two of three years for females and three of four years for males where dispersers were detected.



Supplemental Figure 2.3. Power analysis investigating population assignment accuracy, measured as the proportion of simulated individuals assigned to the population they were simulated from (y-axis), among 500 simulated individuals for each population (A Creek, C Creek, and Beach; x-axis). Separate analyses were performed for SNPs (blue) and microsatellites (red) for individual most-likely assignment (left, does not account for probability of assignment) and for individuals whose probability of assignment was >0.95 (right). Error bars represent 95% confidence intervals about the mean of 1,000 model repetitions. As expected, there was a higher proportion of individuals assigned correctly to their natal location when a 95% probability threshold was used. We ultimately chose to proceed with a 95% probability threshold to ensure accurate estimates of natal locations and identification of dispersers, as all further analyses in this study depended on accurate identification of dispersers. This resulted in natal assignments for 91% and 85% of individuals in A and C Creeks, respectively. Natal population assignments were later validated by comparing them the spawning location of pedigreed parents. We found that 98.6% of natal population assignments agreed with parent spawning locations in cases where both parents were identified.



Supplemental Figure 2.4. Model coefficients, (x-axis) for GLMMs investigating the relationship between ecological covariates, (fixed effects, y-axis) and dispersal (response variable). Results reported for the best-fit model (Model 2.1 in text). Point estimates are bounded by model-estimated standard error bars. Sex was the only fixed effect examined that was not included in the best fit model.



Supplemental Figure 2.5. Model coefficients (x-axes) for GLMMs investigating the relationship between dispersal (fixed effects, y-axes) and reproductive success (response variables). Results reported for the best-fit models (Model 2.3 – Left; Model 2.4 – Right). Point estimates are bounded by model-estimated standard error bars. Specifically, these models were constructed to investigate the effects of 'dispersal category' (Model 2.3 – Left; either non-dispersers, creek-to-creek dispersers, or beach-to-creek dispersers) and 'dispersal direction' (Model 2.4 – Right; either A to A, A to C, C to A, or C to C) on reproductive success. These were included as discrete fixed effects, and the first category in each variable was incorporated into model intercepts (non-dispersers – Left; A to A – Right). We interpreted the effects of all covariates in these best-fit models on reproductive success by examining empirical mean reproductive success values for each of these categories (Figure 2.4).

Supplemental Table 2.1. Summary statistics for reproductive success for each dispersal category, migration direction, and sex: sample size (n), mean (μ), standard deviation (σ), and maximum value (max).

Categories			Reproductive Success			
Disp.	Migrant Type	Sex	n	μ	σ^2	max
Non-Disperser	A to A	F	1690	1.06	4.97	16
		M	909	2.05	12.25	25
	C to C	F	1193	1.90	13.18	31
		M	749	3.18	28.94	33
Creek	C to A	F	13	1.77	3.20	4
		M	10	2.80	11.09	8
	A to C	F	94	1.71	11.59	16
		M	43	2.77	24.30	20
Beach	Beach to A	F	81	0.57	3.06	9
		M	28	0.86	4.41	7
	Beach to C	F	149	0.65	3.03	14
		M	74	0.72	2.79	8

Supplemental Table 2.2. Results of stepwise generalized linear mixed-effect model selection procedures to explain dispersal as response variables (Model 2.1 in text). Fixed effects included return day (day of first entry), density (fish/m^2), predation (proportion killed by bears) sex (Male or Female), natal population (A Creek or C Creek), or null (1). All models included a random effect of year. Models were evaluated with the Akaike Information Criterion (AIC). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (wAIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Fixed Effects	**K**	**LL**	**AIC**	**ΔAIC**	***w*AIC**
return day + return day^2 + density + density^2 + predation + natal population	8	-806.50	1628.99	0.00	0.71
return day + return day^2 + density + density^2 + predation + sex + natal population	9	-806.38	1630.76	1.77	0.29
return day + return day^2 + density + density^2 + natal population	7	-823.43	1660.85	31.86	0.00
density + density^2 + predation + natal population	6	-826.85	1665.70	36.71	0.00
return day + density + density^2 + predation + natal population	7	-825.97	1665.94	36.95	0.00
density + density^2 + predation + sex + natal population	7	-826.22	1666.43	37.44	0.00
return day + density + density^2 + predation + sex + natal population	8	-825.54	1667.08	38.09	0.00
return day + return day^2 + density + predation + natal population	7	-834.67	1683.33	54.34	0.00
return day + return day^2 + density + predation + sex + natal population	8	-834.55	1685.10	56.11	0.00
return day + density + density^2 + natal population	6	-838.11	1688.22	59.23	0.00
return day + density + density^2 + sex + natal population	7	-837.73	1689.46	60.47	0.00
density + density^2 + natal population	5	-840.29	1690.58	61.59	0.00
density + density^2 + sex + natal population	6	-839.58	1691.16	62.17	0.00
return day + return day^2 + density + density^2 + predation	7	-839.03	1692.06	63.07	0.00
return day + return day^2 + density + density^2 + predation + sex	8	-838.95	1693.89	64.90	0.00

return day + return day^2 + predation + natal population	6	-844.92	1701.85	72.86	0.00
return day + return day^2 + predation + sex + natal population	7	-844.86	1703.72	74.73	0.00
return day + return day^2 + density + natal population	6	-848.61	1709.22	80.22	0.00
return day + density + predation + natal population	6	-856.36	1724.72	95.72	0.00
return day + density + predation + sex + natal population	7	-855.99	1725.97	96.98	0.00
return day + predation + sex + natal population	6	-858.02	1728.04	99.05	0.00
return day + return day^2 + natal population	5	-859.04	1728.09	99.10	0.00
return day + return day^2 + sex + natal population	6	-858.95	1729.91	100.92	0.00
density + predation + natal population	5	-860.85	1731.70	102.71	0.00
density + predation + sex + natal population	6	-859.94	1731.88	102.89	0.00
return day + return day^2 + density + density^2	6	-860.92	1733.83	104.84	0.00
return day + density + density^2 + predation	6	-861.24	1734.49	105.49	0.00
return day + return day^2 + density + density^2 + sex	7	-860.81	1735.62	106.63	0.00
return day + density + density^2 + predation + sex	7	-860.85	1735.70	106.71	0.00
predation + natal population	4	-864.44	1736.87	107.88	0.00
predation + sex + natal population	5	-863.63	1737.26	108.27	0.00
return day + density + natal population	5	-865.39	1740.78	111.79	0.00
density + density^2 + predation + sex	6	-864.49	1740.99	111.99	0.00
density + density^2 + predation	5	-865.56	1741.11	112.12	0.00
return day + density + sex + natal population	6	-865.04	1742.08	113.09	0.00
return day + natal population	4	-867.75	1743.51	114.52	0.00
return day + sex + natal population	5	-867.50	1745.01	116.02	0.00
density + sex + natal population	5	-870.61	1751.22	122.23	0.00
density + natal population	4	-871.63	1751.26	122.27	0.00
natal population	3	-875.61	1757.22	128.22	0.00
sex + natal population	4	-874.74	1757.49	128.50	0.00
return day + return day^2 + density + predation	6	-875.45	1762.91	133.92	0.00
return day + density + density^2	5	-876.57	1763.14	134.15	0.00
return day + density + density^2 + sex	6	-876.20	1764.39	135.40	0.00

return day + return day^2 + density + predation + sex	7	-875.44	1764.87	135.88	0.00
return day + return day^2 + predation	5	-878.71	1767.41	138.42	0.00
return day + return day^2 + predation + sex	6	-878.70	1769.40	140.41	0.00
density + density^2 + sex	5	-883.41	1776.82	147.83	0.00
density + density^2	4	-884.76	1777.52	148.52	0.00
return day + return day^2 + density	5	-895.01	1800.01	171.02	0.00
return day + predation	4	-896.12	1800.23	171.24	0.00
return day + predation + sex	5	-895.94	1801.89	172.89	0.00
return day + return day^2 + density + sex	6	-894.99	1801.97	172.98	0.00
return day + density + predation	5	-896.09	1802.18	173.19	0.00
return day + density + predation + sex	6	-895.91	1803.81	174.82	0.00
return day + return day^2	4	-898.43	1804.86	175.87	0.00
return day	3	-908.93	1823.86	194.87	0.00
predation + sex	4	-908.13	1824.26	195.27	0.00
predation	3	-909.28	1824.56	195.57	0.00
density + predation + sex	5	-907.59	1825.19	196.19	0.00
return day + density	4	-908.78	1825.56	196.57	0.00
density + predation	4	-908.84	1825.69	196.70	0.00
return day + density + sex	5	-908.63	1827.27	198.28	0.00
sex	3	-925.83	1857.67	228.68	0.00
density + sex	4	-925.09	1858.18	229.19	0.00
1	2	-927.11	1858.22	229.23	0.00
density	3	-926.53	1859.06	230.07	0.00

Supplemental Table 2.3. Results of stepwise generalized linear mixed-effect model selection procedures to explain dispersal as response variables (Model 2.2 in text). Fixed effects included body shape (depth/length), sex (Male or Female), natal population (A Creek or C Creek), or null (1). All models included a random effect of year. Models were evaluated with the Akaike Information Criterion (AIC). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (*w*AIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Fixed Effects	**K**	**LL**	**AIC**	**ΔAIC**	***w*AIC**
natal population	3	-884.56	1775.11	0	0.39
sex + natal population	4	-883.79	1775.57	0.46	0.31
shape + natal population	4	-884.36	1776.71	1.60	0.17
shape + sex + natal population	5	-883.68	1777.35	2.24	0.13
sex	3	-940.09	1886.19	111.07	0.00
1	2	-941.24	1886.49	111.37	0.00
shape	3	-940.38	1886.75	111.64	0.00
shape + sex	4	-940.06	1888.12	113.01	0.00

Supplemental Table 2.4. Results of stepwise generalized linear mixed-effect model selection procedures to explain fitness as response variables (Model 2.3 in text). Fixed effects included dispersal category (Non-disperser, Creek-to-Creek, or Beach-to-Creek), sex (Male or Female), spawning population (A Creek or C Creek), or null (1). All models included a random effect of year. Models were evaluated with the Akaike Information Criterion (AIC). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (*w*AIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Fixed Effects	**K**	**LL**	**AIC**	**ΔAIC**	***w*AIC**
dispersal category + sex + spawning population	7	-806.50	1628.99	0.00	0.71
dispersal category + spawning population + spawning population:sex	8	-806.38	1630.76	1.77	0.29
dispersal category + sex + spawning population + sex:spawning population	8	-823.42	1660.85	31.86	0.00
dispersal category + sex + spawning population + dispersal category:sex	9	-826.85	1665.70	36.71	0.00
dispersal category + sex + dispersal category:sex + sex:spawning population	10	-825.97	1665.94	36.95	0.00
dispersal category + sex + spawning population + dispersal category:sex + sex:spawning population	10	-826.22	1666.43	37.44	0.00
dispersal category + sex + spawning population + dispersal category:spawning population	9	-825.54	1667.08	38.09	0.00
dispersal category + sex + spawning population + dispersal category:spawning population + sex:spawning population	10	-834.66	1683.33	54.34	0.00
dispersal category + sex + spawning population + dispersal category:sex + dispersal category:spawning population	11	-834.55	1685.10	56.11	0.00
dispersal category + sex + spawning population + dispersal category:sex + dispersal category:spawning population + sex:spawning population	12	-838.11	1688.22	59.23	0.00

dispersal category + sex + dispersal category:sex	8	-837.73	1689.46	60.47	0.00
sex + spawning population	5	-840.29	1690.58	61.59	0.00
sex + spawning population + sex:spawning population	6	-839.58	1691.16	62.17	0.00
sex	4	-839.03	1692.06	63.07	0.00
dispersal category + spawning population	6	-838.95	1693.89	64.90	0.00
dispersal category	5	-844.92	1701.85	72.86	0.00
spawning population	4	-844.86	1703.72	74.73	0.00
1	3	-848.61	1709.22	80.22	0.00

Supplemental Table 2.5. Results of stepwise generalized linear mixed-effect model selection procedures to explain fitness as response variables (Model 2.4 in text). Fixed effects included dispersal direction (A to A, C to C, A to C, or C to A), sex (Male or Female), or null (1). All models included a random effect of year. Models were evaluated with the Akaike Information Criterion (AIC). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (*w*AIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Fixed Effects	**K**	**LL**	**AIC**	**ΔAIC**	***w*AIC**
Stray direction + sex	7	-7337.75	14689.50	0	1.00
Sex	4	-7363.88	14735.76	46.26	0.00
Stray direction	6	-7388.53	14789.06	99.56	0.00
1	3	-7414.18	14834.35	144.85	0.00

2.8 REFERENCES

Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, G. Coop, and N. Chen. 2017. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLoS genetics **13**:e1006911.

Amoroso, R. O., M. D. Tillotson, and R. Hilborn. 2017. Measuring the net biological impact of fisheries enhancement: Pink salmon hatcheries can increase yield, but with apparent costs to wild populations. Canadian Journal of Fisheries and Aquatic Sciences **74**:1233-1242.

Anderson, E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences **65**:1475-1486.

Baguette, M., J. Clobert, and N. Schtickzelle. 2011. Metapopulation dynamics of the bog fritillary butterfly: experimental changes in habitat quality induced negative density-dependent dispersal. Ecography **34**:170-176.

Bentley, K. T., D. E. Schindler, T. J. Cline, J. B. Armstrong, D. Macias, L. R. Ciepiela, and R. Hilborn. 2014. Predator avoidance during reproduction: diel movements by spawning sockeye salmon between stream and lake habitats. Journal of Animal Ecology **83**:1478-1489.

Bett, N. N., S. G. Hinch, N. J. Burnett, M. R. Donaldson, and S. M. Naman. 2017. Causes and consequences of straying into small populations of Pacific salmon. Fisheries **42**:220-230.

Bieri, J. A., C. Sample, W. E. Thogmartin, J. E. Diffendorfer, J. E. Earl, R. A. Erickson, P. Federico, D. T. Flockhart, S. Nicol, and D. Semmens. 2018. A guide to calculating habitat-quality metrics to inform conservation of highly mobile species. Natural Resource Modeling **31**:e12156.

Blanquart, F., O. Kaltz, S. L. Nuismer, and S. Gandon. 2013. A practical guide to measuring local adaptation. Ecology letters **16**:1195-1205.

Bonte, D., and M. Dahirel. 2017. Dispersal: a central and independent trait in life history. Oikos **126**:472-479.

Bonte, D., A. De Roissart, N. Wybouw, and T. Van Leeuwen. 2014. Fitness maximization by dispersal: evidence from an invasion experiment. Ecology **95**:3104-3111.

Bonte, D., H. Van Dyck, J. M. Bullock, A. Coulon, M. Delgado, M. Gibbs, V. Lehouck, E. Matthysen, K. Mustin, and M. Saastamoinen. 2012. Costs of dispersal. Biological reviews **87**:290-312.

Bowler, D. E., and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biological reviews **80**:205-225.

Brennan, S. R., D. E. Schindler, T. J. Cline, T. E. Walsworth, G. Buck, and D. P. Fernandez. 2019. Shifting habitat mosaics and fish production across river basins. Science **364**:783-786.

Buhle, E. R., K. K. Holsman, M. D. Scheuerell, and A. Albaugh. 2009. Using an unplanned experiment to evaluate the effects of hatcheries and environmental variation on threatened populations of wild salmon. Biological Conservation **142**:2449-2455.

Carlson, S. M., T. P. Quinn, and A. P. Hendry. 2011. Eco-evolutionary dynamics in Pacific salmon. Heredity **106**:438-447.

Cayuela, H., Q. Rougemont, J. G. Prunier, J. S. Moore, J. Clobert, A. Besnard, and L. Bernatchez. 2018. Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. Molecular ecology **27**:3976-4010.

Cayuela, H., B. R. Schmidt, A. Weinbach, A. Besnard, and P. Joly. 2019. Multiple density-dependent processes shape the dynamics of a spatially structured amphibian population. Journal of Animal Ecology **88**:164-177.

Clinchy, M., M. J. Sheriff, and L. Y. Zanette. 2013. Predator-induced stress and the ecology of fear. Functional Ecology **27**:56-65.

Clobert, J., R. A. Ims, and F. Rousset. 2004. Causes, mechanisms and consequences of dispersal. Pages 307-335 Ecology, genetics and evolution of metapopulations. Elsevier.

Clobert, J., J. F. Le Galliard, J. Cote, S. Meylan, and M. Massot. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecology letters **12**:197-209.

Cote, J., E. Bestion, S. Jacob, J. Travis, D. Legrand, and M. Baguette. 2017. Evolution of dispersal strategies and dispersal syndromes in fragmented landscapes. Ecography **40**:56-73.

Cote, J., J. Clobert, T. Brodin, S. Fogarty, and A. Sih. 2010. Personality-dependent dispersal: characterization, ontogeny and consequences for spatially structured populations. Philosophical Transactions of the Royal Society B: Biological Sciences **365**:4065-4076.

Cram, J. M., C. E. Torgersen, R. S. Klett, G. R. Pess, D. May, T. N. Pearsons, and A. H. Dittman. 2013. Tradeoffs between homing and habitat quality for spawning site selection by hatchery-origin Chinook salmon. Environmental Biology of Fishes **96**:109-122.

Crozier, L. G., M. M. McClure, T. Beechie, S. J. Bograd, D. A. Boughton, M. Carr, T. D. Cooney, J. B. Dunham, C. M. Greene, and M. A. Haltuch. 2019. Climate vulnerability assessment for Pacific salmon and steelhead in the California Current Large Marine Ecosystem. PLoS One **14**:e0217711.

Cunningham, C. J., G. T. Ruggerone, and T. P. Quinn. 2013. Size selectivity of predation by brown bears depends on the density of their sockeye salmon prey. The American Naturalist **181**:663-673.

Dittman, A. H., D. May, D. A. Larsen, M. L. Moser, M. Johnston, and D. Fast. 2010. Homing and spawning site selection by supplemented hatchery-and natural-origin Yakima River spring Chinook salmon. Transactions of the American Fisheries Society **139**:1014-1028.

Doligez, B., and T. Pärt. 2008. Estimating fitness consequences of dispersal: a road to 'know-where'? Non-random dispersal and the underestimation of dispersers' fitness. Journal of Animal Ecology **77**:1199-1211.

Ducros, D., N. Morellet, R. Patin, K. Atmeh, L. Debeffe, B. Cargnelutti, Y. Chaval, B. Lourtet, A. Coulon, and A. M. Hewison. 2020. Beyond dispersal versus philopatry? Alternative behavioural tactics of juvenile roe deer in a heterogeneous landscape. Oikos **129**:81-92.

Duputié, A., and F. Massol. 2013. An empiricist's guide to theoretical predictions on the evolution of dispersal. Interface focus **3**:20130028.

Esteve, M. 2005. Observations of spawning behaviour in Salmoninae: Salmo, Oncorhynchus and Salvelinus. Reviews in fish biology and fisheries **15**:1-21.

Fraser, D. J., M. M. Hansen, S. ØSTERGAARD, N. Tessier, M. Legault, and L. Bernatchez. 2007. Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. Molecular ecology **16**:3866-3889.

Garant, D., S. E. Forde, and A. P. Hendry. 2007. The multifarious effects of dispersal and gene flow on contemporary adaptation. Functional Ecology **21**:434-443.

Gharrett, A. J., J. Joyce, and W. W. Smoker. 2013. Fine-scale temporal adaptation within a salmonid population: mechanism and consequences. Molecular ecology **22**:4457-4469.

Hamilton, W. D., and R. M. May. 1977. Dispersal in stable habitats. Nature **269**:578-581.

Hendry, A. P., V. Castric, M. T. Kinnison, T. P. Quinn, A. Hendry, and S. Stearns. 2004. The evolution of philopatry and dispersal. Evolution illuminated: salmon and their relatives:52-91.

Holt, R. D., and M. A. McPeek. 1996. Chaotic population dynamics favors the evolution of dispersal. The American Naturalist **148**:709-718.

Johnson, M. D. 2007. Measuring habitat quality: a review. The Condor **109**:489-504.

Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular ecology resources **10**:551-555.

Kats, L. B., and L. M. Dill. 1998. The scent of death: chemosensory assessment of predation risk by prey animals. Ecoscience **5**:361-394.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecology letters **7**:1225-1241.

Keefer, M. L., and C. C. Caudill. 2014. Homing and straying by anadromous salmonids: a review of mechanisms and rates. Reviews in fish biology and fisheries **24**:333-368.

Kolchin, S. A., E. V. Volkova, L. V. Pokrovskaya, and A. V. Zavadskaya. 2021. Consequences of a Sockeye Salmon shortage for the brown bear in the basin of Lake Kurilskoe, Southern Kamchatka. Nature Conservation Research. Заповедная наука **6**:53-65.

Larson, W. A., M. T. Limborg, G. J. McKinney, D. E. Schindler, J. E. Seeb, and L. W. Seeb. 2017. Genomic islands of divergence linked to ecotypic variation in sockeye salmon. Molecular ecology **26**:554-570.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends in Ecology & Evolution **17**:183-189.

Levin, S. A., D. Cohen, and A. Hastings. 1984. Dispersal strategies in patchy environments. Theoretical population biology **26**:165-191.

Lin, J., T. Quinn, R. Hilborn, and L. Hauser. 2008. Fine-scale differentiation between sockeye salmon ecotypes and the effect of phenotype on straying. Heredity **101**:341-350.

Lin, J. E., J. J. Hard, K. A. Naish, D. Peterson, R. Hilborn, and L. Hauser. 2016. It's a bear market: evolutionary and ecological effects of predation on two wild sockeye salmon populations. Heredity.

Lin, Y.-T. K., and G. O. Batzli. 2001. The influence of habitat quality on dispersal, demography, and population dynamics of voles. Ecological Monographs **71**:245-275.

Lisi, P. J., D. E. Schindler, K. T. Bentley, and G. R. Pess. 2013. Association between geomorphic attributes of watersheds, water temperature, and salmon spawn timing in Alaskan streams. Geomorphology **185**:78-86.

Maag, N., G. Cozzi, T. Clutton-Brock, and A. Ozgul. 2018. Density-dependent dispersal strategies in a cooperative breeder. Ecology **99**:1932-1941.

Minto, C., R. A. Myers, and W. Blanchard. 2008. Survival variability and population density in fish populations. Nature **452**:344-347.

Mishra, A., S. Tung, V. Shree Sruti, S. Srivathsa, and S. Dey. 2020. Mate-finding dispersal reduces local mate limitation and sex bias in dispersal. Journal of Animal Ecology **89**:2089-2098.

Mobley, K. B., H. Granroth-Wilding, M. Ellmen, J.-P. Vähä, T. Aykanat, S. E. Johnston, P. Orell, J. Erkinaro, and C. R. Primmer. 2019. Home ground advantage: Local Atlantic salmon have higher reproductive fitness than dispersers in the wild. Science advances **5**:eaav1112.

Morbey, Y. E., and R. C. Ydenberg. 2003. Timing games in the reproductive phenology of female Pacific salmon (*Oncorhynchus* spp.). The American Naturalist **161**:284-298.

Orr, H. A. 2009. Fitness and its role in evolutionary genetics. Nature Reviews Genetics **10**:531-539.

Pess, G., D. R. Montgomery, T. J. Beechie, and L. Holsinger. 2003. Anthropogenic alterations to the biogeography of Puget Sound salmon. Restoration of Puget Sound rivers. University of Washington Press, Seattle, Washington:129-154.

Pess, G. R. 2009. Patterns and processes of salmon colonization.

Peterson, D. A., R. Hilborn, and L. Hauser. 2014. Local adaptation limits lifetime reproductive success of dispersers in a wild salmon metapopulation. Nature communications **5**.

Peterson, D. A., R. Hilborn, and L. Hauser. 2015. Exploratory behavior of dispersers within a metapopulation of sockeye salmon. Behavioral Ecology:arv129.

Pitman, K. J., J. W. Moore, M. Huss, M. R. Sloat, D. C. Whited, T. J. Beechie, R. Brenner, E. W. Hood, A. M. Milner, and G. R. Pess. 2021. Glacier retreat creating new Pacific salmon habitat in western North America. Nature communications **12**:1-10.

Poethke, H. J., W. W. Weisser, and T. Hovestadt. 2010. Predator-induced dispersal and the evolution of conditional dispersal in correlated environments. The American Naturalist **175**:577-586.

Quinn, T. P. 2018. The behavior and ecology of Pacific salmon and trout. University of Washington press.

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

Räsänen, K., and A. P. Hendry. 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. Ecology letters **11**:624-636.

Reed, T. E., D. E. Schindler, M. J. Hague, D. A. Patterson, E. Meir, R. S. Waples, and S. G. Hinch. 2011. Time to evolve? Potential evolutionary responses of Fraser River sockeye salmon to climate change and effects on persistence. PLoS One **6**:e20380.

Reid, J. M., P. Arcese, P. Nietlisbach, M. E. Wolak, S. Muff, L. Dickel, and L. F. Keller. 2021. Immigration counter-acts local micro-evolution of a major fitness component: Migration-selection balance in free-living song sparrows. Evolution letters **5**:48-60.

Roberts, L. J., and C. G. de Leaniz. 2011. Something smells fishy: predator-naïve salmon use diet cues, not kairomones, to recognize a sympatric mammalian predator. Animal Behaviour **82**:619-625.

Rodrigues, A. M., and R. A. Johnstone. 2014. Evolution of positive and negative density-dependent dispersal. Proceedings of the Royal Society B: Biological Sciences **281**:20141226.

Ronce, O. 2007. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. Annu. Rev. Ecol. Evol. Syst. **38**:231-253.

Saastamoinen, M., G. Bocedi, J. Cote, D. Legrand, F. Guillaume, C. W. Wheat, E. A. Fronhofer, C. Garcia, R. Henry, and A. Husby. 2018. Genetics of dispersal. Biological reviews **93**:574-599.

Sæther, B.-E., and S. Engen. 2015. The concept of fitness in fluctuating environments. Trends in Ecology & Evolution **30**:273-281.

Schindler, D. E., J. B. Armstrong, and T. E. Reed. 2015. The portfolio concept in ecology and evolution. Frontiers in Ecology and the Environment **13**:257-263.

Schindler, D. E., X. Augerot, E. Fleishman, N. J. Mantua, B. Riddell, M. Ruckelshaus, J. Seeb, and M. Webster. 2008. Climate change, ecosystem impacts, and management for Pacific salmon. Fisheries **33**:502-506.

Schreiber, S. J. 2010. Interactive effects of temporal correlations, spatial heterogeneity and dispersal on population persistence. Proceedings of the Royal Society B: Biological Sciences **277**:1907-1914.

Shizuka, D., and A. E. Johnson. 2020. How demographic processes shape animal social networks. Behavioral Ecology **31**:1-11.

Sih, A., D. I. Bolnick, B. Luttbeg, J. L. Orrock, S. D. Peacor, L. M. Pintor, E. Preisser, J. S. Rehage, and J. R. Vonesh. 2010. Predator–prey naïveté, antipredator behavior, and the ecology of predator invasions. Oikos **119**:610-621.

Spromberg, J. A., and N. L. Scholz. 2011. Estimating the future decline of wild coho salmon populations resulting from early spawner die-offs in urbanizing watersheds of the Pacific Northwest, USA. Integrated Environmental Assessment and Management **7**:648-656.

Sturrock, A. M., W. H. Satterthwaite, K. M. Cervantes-Yoshida, E. R. Huber, H. J. Sturrock, S. Nusslé, and S. M. Carlson. 2019. Eight decades of hatchery salmon releases in the California Central Valley: factors influencing straying and resilience. Fisheries **44**:433-444.

Takahashi, L. K. 2014. Olfactory systems and neural circuits that modulate predator odor fear. Frontiers in behavioral neuroscience **8**:72.

Tatara, C. P., and B. A. Berejikian. 2012. Mechanisms influencing competition between hatchery and wild juvenile anadromous Pacific salmonids in fresh water and their relative competitive abilities. Environmental Biology of Fishes **94**:7-19.

Thompson, T. Q., M. R. Bellinger, S. M. O'Rourke, D. J. Prince, A. E. Stevenson, A. T. Rodrigues, M. R. Sloat, C. F. Speller, D. Y. Yang, and V. L. Butler. 2019. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. Proceedings of the National Academy of Sciences **116**:177-186.

Tillotson, M. D., and T. P. Quinn. 2018. Selection on the timing of migration and breeding: A neglected aspect of fishing-induced evolution and trait change. Fish and Fisheries **19**:170-181.

Travis, J. M., M. Delgado, G. Bocedi, M. Baguette, K. Bartoń, D. Bonte, I. Boulangeat, J. A. Hodgson, A. Kubisch, and V. Penteriani. 2013. Dispersal and species' responses to climate change. Oikos **122**:1532-1540.

Webb, J. H., and H. A. McLay. 1996. Variation in the time of spawning of Atlantic salmon (*Salmo salar*) and its relationship to temperature in the Aberdeenshire Dee, Scotland. Canadian Journal of Fisheries and Aquatic Sciences **53**:2739-2744.

Westley, P. A., A. H. Dittman, E. J. Ward, and T. P. Quinn. 2015. Signals of climate, conspecific density, and watershed features in patterns of homing and dispersal by Pacific salmon. Ecology **96**:2823-2833.

Westley, P. A., T. P. Quinn, and A. H. Dittman. 2013. Rates of straying by hatchery-produced Pacific salmon (*Oncorhynchus spp.)* and steelhead (*Oncorhynchus mykiss*) differ among species, life history types, and populations. Canadian Journal of Fisheries and Aquatic Sciences **70**:735-746.

Wilson, D. S. 2004. What is wrong with absolute individual fitness? Trends in Ecology & Evolution **19**:245-248.

Yeakel, J. D., J. P. Gibert, T. Gross, P. A. Westley, and J. W. Moore. 2018. Eco-evolutionary dynamics, density-dependent dispersal and collective behaviour: implications for salmon metapopulation robustness. Philosophical Transactions of the Royal Society B: Biological Sciences **373**:20170018.

Chapter 3. HOMING IN SPACE AND TIME DRIVES FINE-SCALE POPULATION STRUCTURE AND INFLUENCES FITNESS

Authors: Samuel A. May, Emily K. Bowers, Andrea N. Odell, David J. Paez, Ray Hilborn, Lorenz Hauser, and Kerry A. Naish

3.1 ABSTRACT

Most eco-evolutionary models currently used in management to predict the impacts of anthropogenic activities on wild population dynamics assume random mating within populations. Yet, environmental, behavioral, and life-history variation may increase the likelihood of mating between individuals who share phenotypes or ancestry (i.e., assortative mating). In salmonid systems, natal homing behaviors may result in assortative mating systems if individuals return to natal sites within continuous breeding populations. However, the fine-scale resolution at which homing occurs in both space and time and the effects of homing on genetic structure and fitness variation within populations remains largely unknown. This study aimed to examine the extent, drivers, and fitness consequences of natal homing and population structure within Sockeye salmon populations. To achieve this aim, we used 14 years of field surveys and full pedigrees of two stream-spawning populations of Sockeye salmon in Bristol Bay, Alaska (A and C Creeks). First, to test whether related individuals clustered together in space and time, we compared pairwise pedigree-based relatedness values to pairwise distance values using a series of autocorrelation analyses. Findings indicated evidence of both spatial and temporal genetic structure within both streams; individuals returning within 50 meters and within five days of each other were more related than expected by chance. Second, to determine if genetic structure resulted from natal

homing behaviors, we examined the correlation between individual natal and spawning locations within the streams. Results showed that spatial and temporal structure within populations could be partly explained by fine-scale homing; individuals returned to the same place and at the same time as they were fertilized as eggs. Third, to quantify the fitness consequences associated with straying away from natal sites, stray distances were measured as the distance between individual natal and spawning locations in the streams. We examined the effect of these stray distance values on individual lifetime reproductive success and found that, in one of the two creeks, individuals returning to spawn near natal sites had greater reproductive success. This result may provide evidence for microgeographic adaptation if individuals were adapted to specific natal locations within the streams. However, there was no evidence of a fitness benefit to individuals returning at the same time as they were fertilized as eggs. Spawning habitat also influenced fitness, as reproductive success was correlated with spawning location within both streams. Results were consistent with the expected outcomes of an assortative mating system, as phenotypes of body shape, return timing, and sex were spatially segregated within the streams. Taken together, the clustering of both relatives and phenotypes within the streams provided strong evidence for an assortative mating system within these small populations. The results of this study provide empirical evidence demonstrating how natal homing, reproductive timing, and habitat characteristics can generate population structure on very small spatial and temporal scales. While eco-evolutionary models currently used in management assume random mating within populations, we show how processes influencing non-random mating may better represent natural population dynamics.

3.2 INTRODUCTION

The life histories of aggregating species such as salmonids, sea turtles, and birds often involve round-trip movement among different habitats for reproduction, refuge, and feeding (Milner-Gulland et al. 2011). In many species, homing behaviors to natal habitats produce genetically structured metapopulations, where non-random mating, combined with geographic and demographic barriers, may result in locally adapted populations (Hanski 1999, Xu 2018). This population diversity can buffer metapopulations against disruptive events that influence any one subunit through portfolio effects (Schindler et al. 2010, Brennan et al. 2019). Yet, the scale and drivers of population structure and its effects on fitness are poorly understood in many wild populations. As a result, current models used in management to predict the impact of anthropogenic activities often assume random mating (i.e., panmixia) within populations (e.g., Araki et al. 2008, Willoughby and Christie 2019). However, processes that influence non-random mating may more accurately represent wild population dynamics (Edelaar and Bolnick 2012). Thus, by examining the scale and fitness consequences of structure within populations, it may be possible to better predict how wild populations may respond to environmental and anthropogenic changes.

The scale and extent of population structure is controlled by the rate of gene flow within and among populations (Waples and Gaggiotti 2006). Genetic structure within populations might result from a combination of precise natal homing and ecological processes which influence movements within habitat patches (Neville et al. 2006). Within population structure may increase the likelihood of mating among relatives (i.e., inbreeding) and among individuals who share phenotypes (i.e., assortative mating), both of which can affect fitness and rates of gene flow (Jiang et al. 2013, Kardos et al. 2016). Recent studies have examined how assortative mating, strong

selection, and limited gene flow may result in local adaptation within populations, leading to microgeographic adaptation; (Torres Dowdall et al. 2012, Gharrett et al. 2013, Richardson et al. 2014, Reid et al. 2021). Thus, homing to a specific place or time within a continuous breeding population (hereafter, 'fine-scale homing') may convey fitness benefits if individuals are better suited to conditions in their natal environments (Richardson et al. 2014).

Salmonid fishes (especially *Oncorhynchus* and *Salmo* spp.) have served as model species in studies of population structure and gene flow (Hendry et al. 2004). The majority of adults return to their natal habitats, although some fraction disperse (Keefer and Caudill 2014). These dispersers are referred to as termed 'strays' in the salmonid literature (Hendry et al. 2004). The anadromous lifecycle of many salmonids permits sampling of returning populations on or near their spawning habitats (Lin et al. 2016, Barnett et al. 2019, Link and English 2020).

There is evidence for departures from panmixia within continuous breeding populations of salmonids. Fine-scale spatial structure in salmonids suggests that individuals may home to within kilometers of their natal site in a particular river-reach (Bentzen et al. 2001, Stewart et al. 2003, Neville et al. 2006, Stelkens et al. 2012, Barnett et al. 2019). Experimental evidence using heat-tagged otoliths has shown how Sockeye salmon (*Oncorhynchus nerka*) may return within meters of their natal site of incubation, suggesting homing as a mechanism for widespread structure within geographically continuous populations (Quinn et al. 2006). In addition, a recent study by Camacho and Hendry (2020) found a positive correlation between body shape, stream depth in spawning locations, and reproductive lifespan in Sockeye salmon, indicating how phenotypic clustering may convey fitness benefits if increased reproductive lifespan results in more offspring.

Furthermore, salmon home not just in space, but also in time; high heritability values have been estimated for return timing (Hendry and Day 2005, Carlson and Seamons 2008). Individuals

may also return to spawn at the same time as others with similar phenotypes (Doctor and Quinn 2009). Such temporal structuring may indicate possible adaptation-by-time if certain phenotypes have greater fitness at particular times during the spawning season (Hendry et al. 1999). These findings of fine-scale spatial and temporal structure suggest that non-random mating may play a larger role in controlling gene flow and adaptation within populations than previously thought. Yet, no study has quantified both the scale of population structure and the fitness consequences of this structure within salmonid populations. By examining structure and its fitness effects within salmonid populations, we may better understand the relative influence of natal homing behaviors and phenotypic clustering on assortative mating regimes and fitness.

In natural environments, the detection of fine-scale genetic structure and its fitness consequences requires the quantification of relatedness between breeding individuals and their fitness (Edelaar and Bolnick 2012). Pedigree-based studies of wild populations have the ability to provide this information (Pemberton 2008, Peterson et al. 2014). Thus, an ideal study system to examine fine-scale genetic structure and its effects on fitness would be in a species that homes to natal breeding grounds, has a pedigree, and has individual-level data for breeding locations. Coupled with ecological and individual-level data from long-term monitoring programs, such a pedigree would allow for powerful inference into the effects of ecology and behavior on gene flow and fitness within populations.

Comparisons of ecological and evolutionary processes among multiple populations and habitats provide valuable experimental replication and insights into how processes might differ among habitats (Fraser et al. 2020). Here, we used pedigrees of two intensively sampled breeding populations of Sockeye salmon in Bristol Bay, Alaska (May et al. in prep, Ch2). This program has been conducted over 14 years and across two complete generations of returning adults (F0 to F2).

Morphological and behavioral data, including arrival timing and movement on the spawning grounds has been collected for individual returning adults, permitting both within and between population analyses of fine-scale processes. For example, our previous study investigating A and C Creek provided insights into how ecological mechanisms influenced dispersal, as well as how fitness outcomes of dispersal varied among the two habitats (May et al. in prep, Ch2). The use of two wild populations allows for replication and validation of empirical findings and comparisons of how behavior and patterns of genetic structure may vary in different habitats.

The present study seeks to quantify the scale and fitness consequences of spatial and temporal structure within continuous breeding populations of Sockeye salmon. This work will inform our understanding of the mechanisms that control assortative mating regimes and fitness variation within populations. Our objectives were: 1) to quantify the scale of spatial and temporal structure by estimating autocorrelation in pairwise relatedness values among individuals within cohorts, across two generations of returning adults; 2) to test whether any spatial and temporal structure within a returning cohort is a result of homing to natal incubation sites; and 3) to determine whether homing and straying behaviors in space and time are correlated with the reproductive success of returning adults. We discuss the ecological and behavioral drivers of genetic structure within populations. This information in turn may aid efforts to predict the effects of ecological and anthropogenic changes on fitness and resilience in these culturally and economically significant species.

3.3 METHODS

3.3.1 *Sample Collection and Pedigree Reconstruction*

Tissue samples were collected as part of a long-term field program from a wild Sockeye salmon metapopulation on Little Togiak Lake, Bristol Bay, Alaska (Peterson et al. 2014, Lin et al. 2016) May et al. in prep, Ch2). The system comprises two distinct creeks: known as "A" and "C". These creeks are approximately 1km apart, 350m and 450m long, respectively, and each host 150-800 spawners per year. "A" Creek is shallow with little refuge, making it poor habitat for predator avoidance (Pess 2009, Lin et al. 2016). In contrast, "C" Creek is more heterogeneous habitat with many undercut banks and deep pools, particularly in the upper reaches of the creek (Pess 2009), resulting in a longer average in-creek lifespan and reduced predation risk (Lin et al. 2016). Sockeye salmon spawn in the streams from late July through late August. A separate, morphologically and genetically distinct population also spawns along the lake beaches at the mouths of these streams from late July through late September (Lin et al. 2008). Sockeye salmon senesce after spawning and do not return to breed in subsequent years (they are semelparous).

Adults were tagged daily upon entrance into the creeks with uniquely labeled plastic disc tags, although some were captured using beach seines and tagged at the mouths of the creeks before entry. At tagging, fish length (mid-eye to hypural plate), body depth, and sex were recorded, and dorsal fin tissue samples were collected and stored in 100% ethanol. Daily location measured in 10m meter sections from the mouth of the creek of each tagged fish was recorded using visual surveys. Upon death, otoliths were removed for fish aging where possible. Four-year-old fish dominate this system, with some age five individuals (May et al. in prep, Ch2).

A two-generation pedigree of adults returning to the A and C Creek populations, originally reported in a previous study (May et al. in prep, Ch2), was used to estimate relationships between individuals and to estimate the number of offspring from each returning adult (i.e., lifetime reproductive success). Briefly, the pedigree included F0-F2 adults returning between years 2004 and 2017 (Figure 3.1). A combination of 11 microsatellite loci (Peterson et al. 2014) and 172 single nucleotide polymorphism (SNP; May et al. 2020) loci were used to genotype individuals observed in the stream habitats. In addition, a representative sample of individuals spawning on the beach were included for later use in identifying beach immigrants using assignment tests (May et al. 2020). The microsatellite-based pedigree for years 2004 through 2005 and 2008 through 2010 (Peterson et al. 2014) was constructed using *Colony2* (Jones and Wang 2010), and additional cohorts returning in 2006 through 2007 and 2011 through 2017 were added using SNP genotypes (May et al. in prep, Ch2) and *Sequoia* (Huisman 2017).

3.3.2 *Estimation of Spatial, Temporal, and Relatedness Parameters*

The first objective was to investigate the scale of spatial and temporal genetic structure within the two creeks. Spatial genetic structure was defined as related individuals spawning closer together in space than expected in a randomly distributed population. Similarly, temporal genetic structure was defined as related individuals spawning closer together in time than expected in a randomly mixed population. To test whether closely related individuals clustered in space or time compared to randomly mixed (i.e., panmictic) populations, we examined the extent of autocorrelation among pairwise relatedness values across spatial (river meter) and temporal (return day) measures (Smouse and Peakall 1999).

Pedigree-based relatedness values (e.g., full-siblings = 0.5, half siblings = 0.25) were estimated among all individuals within cohorts using *Kinship2* (Sinnwell et al. 2014). Pairwise

relationships among individuals genotyped in 2017 were excluded, as only a portion of the population were included in the pedigree in that year (May et al. in prep, Ch2).

Pairwise spatial distances were defined as the difference, in meters, between the spawning locations of returning adults (i.e., redd locations). Spawning locations were not recorded in the field and were inferred using an average of observed in-stream daily locations for individual fish, until they died post-spawning. Sockeye salmon are known to explore the stream for suitable spawning habitat when they first enter (Rich Jr et al. 2006) and thus, the initial tagging location was excluded. Following death, carcasses often drifted downstream (in the case of senescence) or were transported to the stream banks (in the case of predation by bears), and thus location of carcass recovery was similarly excluded. Finally, any outlier observations that significantly affected the mean were also removed. A statistic dfBeta (Belsley et al. 2005) was calculated for each observation, where dfBeta is the difference between the i^{th} observation and the mean excluding the i^{th} observation. dfBeta was scaled by the standard error calculated with the i^{th} observation omitted and dfBeta values greater than two were used to identify outliers (Belsley et al. 2005). The value for dfBeta was iteratively recalculated until no outliers remained or until only two observations remained, as dfBeta requires at least three observations. Each individual's estimated spawning location was calculated as the mean of remaining observed locations. To maximize accuracy in spawning location estimation, individuals with only two remaining observations greater than 100 meters apart were not assigned estimated spawning locations. To maximize sample size, individuals with only one remaining daily location after removing the day of tagging and death were assigned estimated spawning locations at the location of the single observation. We explored the effects of screening criteria on sample size, as well as estimation accuracy. The latter was measured as the coefficient of determination (R^2) of a regression of

estimated spawning locations of female and male parents (detailed in the supplemental materials; Supplemental Figure 3.1).

To quantify pairwise temporal distance among individuals within cohorts, the difference in return day between individuals was used. Return day was designated as the date each individual was first recorded in the stream habitats. To account for potential annual variation in population return timing, all return days were standardized to the Julian date that the first returning fish of the season entered each creek, following similar approaches in salmonid studies (e.g., Poirier et al. 2012, Naish et al. 2013). The date that the first returning fish of the season entered each creek varied by a maximum of five calendar days, and therefore standardization of return days was assumed to have a negligible effect on further analyses. Standardized return days (hereafter, 'return days') were used in all further analyses to estimate pairwise temporal distance in return timing between individuals. Return days were assumed to be a reasonable proxy for spawn timing, as Sockeye salmon are known to spawn within three days of entering stream habitats (Esteve 2005, Lisi et al. 2013).

To concentrate on processes that affect structure within populations, immigrants from other populations were removed from further analyses. These immigrants were identified by assigning a natal and spawning population to pedigreed individuals from genetic population assignment analyses and visual observations respectively, previously described in May et al. (in prep, Ch2). Briefly, genetic population assignments were performed using the leave-one-out approach of (Anderson et al. 2008) in *rubias*, using a 95% confidence threshold for individual assignments (Moran and Anderson 2018). Individuals with different natal and spawning locations were designated as immigrants and removed from all further analyses.

3.3.3 *Fine-Scale Spatial and Temporal Genetic Structure*

Following preliminary analyses showing variation in relatedness throughout the streams and spawning season (Supplemental Figures 3.2 and 3.3), we performed a series of genetic autocorrelation analyses (Smouse and Peakall 1999). The autocorrelation coefficient (*r*) was estimated using GENALEX (version 6.51b2; Peakall and Smouse 2006), where the value of *r* ranges from -1 to 1. A positive (or negative) *r* indicates that individuals within a certain distance class are more (or less) related than expected by chance alone (Peakall et al. 2003). A pedigree-based relatedness matrix among all pairs of individuals within populations and years was used as a measure of genetic distance. As relatedness values measure genetic similarity, as opposed to genetic distance, all values output by GENALEX were multiplied by negative one to ensure that results were comparable to similar studies using genetic autocorrelation analyses with other genetic distance measures (e.g., Neville et al. 2006). To allow for comparative analyses in population structure, creek populations were analyzed separately.

The following subsets of pedigreed individuals were used in autocorrelation analyses. First, to avoid imprecise estimates of autocorrelation due to a small number of comparisons, only cohorts with more than 100 pairwise comparisons were included. Second, environmental conditions experienced by individuals from different age classes may vary, therefore, only four-year-old individuals were included. Age has been previously reported (May et al; Ch2, *in prep*), and was determined using a hierarchical approach: for individuals assigned at least one parent in the pedigree, the difference between the return year of parent-offspring pairs was used (pedigree age). When pedigree age was unavailable, otolith readings were used to determine age (Geffen 1992). If no otoliths were sampled, length was used as a proxy for age (May et al.; Ch2, in prep). If no size information was available, as in the case of bear-consumed carcasses, individuals were

removed from further analysis. As most individuals in this system are four years old (approximately 82%), removing five-year-olds likely did not substantially reduce power to infer autocorrelation. Finally, individuals were included in spatial autocorrelation analyses if they were assigned an estimated spawning location.

Autocorrelation analyses require binned distance classes. Spatial distance classes of 50 meters were used, following an assumption that Sockeye rarely move more than 50m away from their redds and courted mates (McPhee and Quinn 1998, Rich Jr et al. 2006). In addition, habitat availability changed during the years; marginal habitat near the headwaters of both streams was only available in some years. Therefore, to standardize distance comparisons among years, individuals estimated to spawn within 300m and 400m of the mouths of A and C Creek, respectively, were included. Individuals were separated into populations based on return year, and a single autocorrelation coefficient was quantified for each distance category using the multi-population model in GENALEX (Peakall and Smouse 2006).

Individuals were included in temporal autocorrelation analyses if they had a recorded return day into the stream habitats. Pairwise temporal distance values were binned into temporal distance classes of five days, following the assumption that individuals typically spawn within five days following stream entry (Pess 2009, Lin et al. 2016).

Two recommended methods for testing significance of autocorrelation results (Peakall et al. 2003) were used. Permutations were used to estimate the upper and lower 95% confidence intervals about $r = 0$, which represented the null hypothesis of no structure. A total of 999 random permutations were performed by randomly assigning individuals to distance classes. Bootstrapping was also used to generate confidence limits ($\alpha = 0.05$) about r estimates for each

distance class; 999 bootstraps were used to randomly resample individuals within each distance class. Values of *r* were considered significant if both of these criteria were met.

3.3.4 *Fine-Scale Homing Behaviors in Space and Time*

Following evidence of both spatial and temporal genetic structure within populations (results section), the second objective of this study was to investigate the mechanisms producing structure. Structure may result from individuals returning to the same place ('natal location') or at the same time ('fertilization day') as they were fertilized as eggs. Structure may also result from a clustering of relatives in space or time from kin-recognition or phenotypic habitat matching (Edelaar et al. 2008). We hypothesized that observed patterns of spatial structure might result from natal homing behaviors if individual spawning and natal locations were positively correlated. We inferred natal locations from the estimated spawning locations of pedigreed parents. Specifically, mother spawning locations were used, as female Sockeye salmon are known to dig a single redd and remain nearby until senescence, whereas males may mate with multiple females and move more within spawning habitats (Rich Jr et al. 2006). Therefore, we expected higher precision in estimates of mother spawning location.

Generalized linear mixed models (GLMMs) were used to determine whether natal location was a significant predictor of spawning location. A full model with spawn location was a function of fixed effects for natal location, return day, body shape, and sex was constructed with a Gaussian error distribution.

(Model 3.1) $$spawn\ location \sim natal\ location + return\ day + return\ day^2 + body\ shape + \ sex + body\ shape{:}\ sex + \ [YEAR]$$

The effect of body shape on spawn location was tested, following suggestion of a relationship in preliminary analyses. Body shape was quantified as the ratio of body length to body depth

(Quinn and Foote 1994). Body depth was measured at the deepest vertical point, and body length was measured from mid-eye to hypural plate (in millimeters). An interaction term between body shape and sex was included to account for known differences in the body shapes of males and females. Return day was included as a quadratic term following evidence in previous studies that return day may affect spawn location within stream habitats (Doctor and Quinn 2009, Lin 2012). Return year was included as a random effect to account for interannual variation in spawning locations. To examine possible differences in the processes influencing spawning location among the two populations, previously shown differ in population density and predation intensity (May et al. in prep, Ch2) and have habitat features (Pess 2009), the populations were modeled separately. Natal location, return day, and body shape were scaled and standardized among all cohorts to a mean of zero and standard deviation of one to improve model convergence.

'Temporal homing' was defined as the spawning of individuals (return day) near the same day during the season when they were fertilized as eggs (fertilization day). Temporal autocorrelation might be explained by a positive correlation between individual return days and fertilization days. However, it was not possible to record individual fertilization day in the field. This parameter was inferred from mother return day, since female Sockeye salmon generally only spawn once (Rich Jr et al. 2006). We assumed that return day served as an appropriate proxy for spawn day; the same measure was used for all individuals, so we predicted no systematic bias.

A full model with a Gaussian error distribution was constructed, where return day was a function of fixed effects for fertilization day, body shape, and sex.

(Model 3.2) $$return\ day \sim fertilization\ day + body\ shape + sex + body\ shape{:}\,sex + [YEAR]$$

The model accounted for differences in return day due to body shape and sex with an interaction term for these variables (Doctor and Quinn 2009, Lin et al. 2016). All models included

a random effect of return year to account for interannual variation in return timing. To highlight differences in the processes influencing return day among the two populations, the populations were modeled separately. Fertilization day and body shape were scaled and standardized among all cohorts to a mean of zero and standard deviation of one to improve model convergence.

To determine which explanatory variables affected the response variables of spawn location and return day, a best-fit model for each GLMM was selected from simpler models that omitted different fixed effects. Inference for this suite of models was performed using weighted Akaike Information Criteria values, a measure of the proportion of the total predictive power provided by the full set of models (*w*AIC; Symonds and Moussalli 2011). The effects of the explanatory variables on the response variables were interpreted in the best fit models.

3.3.5 *Fitness Consequences of Fine-Scale Spatial and Temporal Structure*

The third objective was to determine whether homing and straying within populations influenced individual fitness. We hypothesized that individuals spawning closer to when or where they were born would have increased fitness compared to individuals who spawned further away. To test this hypothesis, individual measures of 'stray distance' were quantified, defined as the absolute geographic distance, in meters, between estimated natal and spawning locations. In addition, individual measures of 'temporal stray distance' were quantified, defined as the absolute temporal distance, in days, between estimated fertilization and return days. Fitness was quantified using pedigree-based measures of individual lifetime reproductive success, defined as the total number of returning adult offspring per adult individual (hereafter, "reproductive success"). To facilitate comparative analyses, the fitness consequences of homing and straying were modeled separately in the two creek populations. Generally, these models included individuals from the F1

generation, only, as natal locations and reproductive success of individuals was only known for this generation (Figure 3.1).

To determine the effect of spatial stray distance on fitness, a full GLM model was built where reproductive success (zero and positive integer values) was assumed to follow a negative binomial error distribution and was a function of stray distance, spawn location, body shape, and sex.

(Model 3.3) $$reproductive\ success \sim stray\ distance + spawn\ location + stray\ distance{:}\ spawn\ location + body\ shape + sex + [YEAR]$$

To test for variation in reproductive success due to differences in body shape or spawn location within the streams, these terms were included as fixed effects. To account for variation in reproductive success due to known differences in fitness among males and females (May et al. in prep, Ch2), we included a fixed effect of sex. Sampling year was treated as a random effect in this model, following earlier results in May et al. (in prep, Ch2) where substantial interannual variation in fitness was reported in these populations. To test for possible interactions among spawning location and stray distance, we included an interaction term among these variables. Stray distance, body shape, and spawn location were scaled and standardized among all cohorts to a mean of zero and standard deviation of one to improve model convergence.

To determine the effects of temporal stray distance on fitness, a full model was developed where reproductive success was a function of temporal stray distance, return day, body shape, and sex.

(Model 3.4) $$reproductive\ success \sim temporal\ stray\ distance + return\ day + temporal\ stray\ distance{:}\ return\ day + body\ size + sex + [YEAR]$$

To test for variation in reproductive success due to differences in body shape or return day during the spawning season, these terms were included as fixed effects. We also included a fixed

effect of sex and a random effect of sampling year. To test for a possible interaction among return day and temporal stray distance, we included an interaction term among these variables. Temporal stray distance, return day, and body shape were scaled and standardized among all cohorts to a mean of zero and a standard deviation of one to improve model convergence.

As above, inference for these models was performed using model selection through *w*AIC on a suite of simpler models, and the effects of the explanatory variables on reproductive success were interpreted in the best fit models.

3.4 RESULTS

3.4.1 *Spatial, Temporal, and Relatedness Parameters*

We have previously reported the results of the pedigree reconstruction of Sockeye salmon in the A and C Creek system (May et al. in prep, Ch2). Briefly, the pedigree comprised 4,243 spawners in A Creek, 5,012 spawners in C Creek, and 582 individuals observed on the beach habitat (total 9,838). Of 7,146 fish in the F1-F2 generations, both parents were assigned to 5,031 individuals (70.4%), and a further 1,494 individuals were assigned one parent (20.8%); therefore, at least one parent was assigned to 6,525 individuals (91.2%). Number of returning adult offspring per adult individual (i.e., reproductive success) ranged from 0 to 33 (mean=1.78, $\sigma^2 = 12.53$), and no offspring were assigned to 68.8% of female and 57.7% of male parents in the F0 and F1.

Genetic population assignment analyses with 95% confidence thresholds (natal population) and visual observations (spawning population), resulted in both natal and spawning populations assigned to 8,117 fish (full results reported in May et al. in prep). Of these, 732 immigrants were excluded from this study. Thus, 7,385 fish were identified as residents (i.e., philopatric), comprising 3,634 and 3,751 fish spawning in A and C Creek, respectively. We further removed

1,005 individuals sampled in 2017 due to incomplete sampling in that year, although these fish were used as offspring to estimate individual reproductive success values for their assigned parents. The final dataset comprised 3,207 and 3,173 spawners in A and C Creeks, respectively (total 6,308).

The number of individuals included in the spatial autocorrelation analyses depended on identification of spawn location for each individual, while the analyses on spatial homing depended on both natal and spawn location. Spawn location was estimated for 2,309 and 2,748 fish in A and C Creeks, respectively (in meters from creek mouth; A: $\mu = 165.0$, sd = 80.3; C: $\mu = 244.0$, sd = 104.0; Figure 3.2). Natal locations were estimated for 1,514 and 2,200 fish in A and C, respectively (in meters from creek mouth; A: $\mu = 162.0$, sd = 83.0; C: $\mu = 296.0$, sd = 77.7). Stray distance was estimated for 1,010 and 1,864 individuals that were attributed both natal and spawning locations in A and C Creek, respectively (in meters from natal location; A: $\mu = 92.3$, sd = 68.3; C: $\mu = 99.0$, sd = 80.1).

Similarly, the number of individuals included in the temporal autocorrelation analyses depended on identification of return day for each individual, while the analyses on temporal homing depended on both return day and fertilization day estimates. Return day was estimated for all individuals (in days from standardized first day of season; A: $\mu = 9.42$, sd = 6.05; C: $\mu = 7.54$, sd = 5.59; Figure 3.2). Fertilization day was estimated for 1,629 and 2,297 individuals in A and C Creek, respectively (in days from standardized first day of season; A: $\mu = 10.4$, sd = 5.92; C: $\mu = 8.0$, sd = 5.68), Temporal stray distance was estimated for all individuals that were attributed fertilization days (days between fertilization and spawn days; A: $\mu = 5.50$, sd = 4.46; C: $\mu = 5.84$, sd = 4.96).

When all cohorts from all generations were combined into a single two-dimensional density plot, higher numbers of individuals were observed in the first half of the spawning season and in the upper regions of both creeks (Figure 3.2, sections A-B). However, inter-annual variation was observed in both space and time (Figure 3.2, sections C-D).

3.4.2 *Fine-Scale Spatial and Temporal Genetic Structure*

Significant spatial autocorrelation values were identified across both creeks, suggesting that individuals returning to the same creek sections were more related than expected by chance alone, and individuals returning far apart from one another were less related than expected by chance alone. Spatial autocorrelation analyses included 1,801 four-year-old individuals from A Creek and 2,117 from C Creek (total 3,918), after cohorts with less than 100 pairwise comparisons were removed (2011, 2014, 2015 in A Creek; 2014 in C Creek). Spatial autocorrelation analyses (Figure 3.3) showed small but significant positive *r* values for the pairwise distance categories between zero and 50 meters in both creeks, and between 50 and 100 meters in C Creek (A $r_{0\text{-}50} = 0.03 \pm 0.02$; C $r_{0\text{-}50} = 0.05 \pm 0.04$, $r_{50\text{-}100} = 0.022 \pm 0.020$). In both creeks, all other significant *r* values were negative across distance classes (Figure 3.3). Positive *r* values in the 0-50m distance class were larger in C Creek by a factor of 1.7, indicating stronger spatial structure.

The results of the temporal autocorrelation analyses indicated that individuals returning to the creeks within five days were more related to each other, and individuals returning at greater times apart were less related to each other than expected by chance alone. Temporal autocorrelation analyses included 2,583 and 2,422 four-year-old individuals from A and C Creek, respectively, after excluding the 2014 and 2015 cohorts from A Creek and the 2014 cohort from C Creek due to less than 100 pairwise comparisons. Individuals returning between zero and five days from each other had significant and positive *r* values; values for this distance category were greater in A

Creek by a factor of 3.2 (A $r_{0\text{-}5}$ = 0.16±0.02; C $r_{0\text{-}5}$ = 0.05±0.01). All other distance classes had significant and negative *r* values. Negative autocorrelation values were greater in magnitude in A Creek than in C Creek in the 25-30 day pairwise distance category by a factor of 2.3 (A $r_{25\text{-}30}$ = -0.76±0.08; C $r_{25\text{-}30}$ =-0.33±0.10). In addition, temporally distant individuals in A Creek were less related than those in C Creek. Overall, these results indicated strong within-population temporal structure in both populations.

3.4.3 *Fine-Scale Homing Behaviors in Space and Time*

Investigation of the drivers of spawning location within the streams indicated that observed spatial structure was likely driven by natal homing behaviors within continuous breeding habitats, within C Creek, but this result was not supported in A Creek (Figure 3.4). The best fit model for C Creek included natal location, in addition to parameters for body shape and sex (AIC = 17581.32, *w*AIC = 0.17, $\beta_{\text{natal location}}$ = 19.83, β_{shape} = 4.51, $\beta_{\text{sex males}}$ = 17.9, Supplemental Table 3.1). In contrast, the best fit model for A Creek did not include natal location but did include a linear effect of return day on spawning location (AIC = 9750.70, *w*AIC = 0.23, $\beta_{\text{return day}}$ = -10.44, Supplemental Table 3.1). The model for C Creek showed that large individuals of both sexes, in addition to a higher proportion of males than females, spawned in the upper reaches of C Creek (Figure 3.4). In addition, individuals returning earlier in the season spawned closer to the mouth of the stream in A Creek (Figure 3.4). These results indicated that individuals were spatially segregated by phenotypes of body shape and sex in C Creek and by return timing in A Creek.

Investigation of the drivers of spawn timing strongly indicated that observed patterns of temporal genetic structure resulted from individuals returning at the same time during the spawning season as they were fertilized as eggs (Figure 3.5). The effects of temporal homing on within-population temporal genetic structure were investigated by examining the relationship

between fertilization day of individuals as eggs and return day of returning adults. The best fit models for both creeks contained an effect of fertilization day and sex (A Creek: $\beta_{\text{fertilization day}}$ = 3.20, $\beta_{\text{sex males}}$ = -2.38; C Creek: $\beta_{\text{fertilization day}}$ = 2.26, $\beta_{\text{sex males}}$ = -4.13, Supplemental Table 3.2). These model results also indicated that males returned earlier in the season than females in both creeks (Figure 3.5).

3.4.4 *Fitness Consequences of Fine-Scale Spatial and Temporal Structure*

Examination of the fitness effects of spatial homing indicated that individuals returning closer to their natal location had increased fitness in C Creek, but this result was not supported in A Creek (Figure 3.6). Yet also, individuals spawning in the lower reaches of A Creek and in the upper reaches of C Creek had increased fitness (Supplemental Figure 3.4). The best fit model for C Creek included stray distance, as well as the parameters spawn location, sex, and an interaction term between stray distance and spawn location (AIC = 3517.20, *w*AIC =0.44, $\beta_{\text{stray distance}}$ = -0.11, $\beta_{\text{spawn location}}$ = 0.56, $\beta_{\text{sex males}}$ = 0.47, $\beta_{\text{stray distance:spawn location}}$ = -0.11, Supplemental Table 3.3). In contrast, the best fit model for A Creek did not include stray distance but did include the parameters spawn location, body shape, and sex (AIC = 2118.32, *w*AIC = 0.19, $\beta_{\text{spawn location}}$ = -0.19, $\beta_{\text{body shape}}$ = -0.58, $\beta_{\text{sex males}}$ = 0.37, Supplemental Table 3.3). The interaction term between stray distance and spawn location in the model for C creek indicated that fish born in the upper reaches but returning to the lower reaches of the stream had decreased fitness compared to individuals who were born in and returned to the lower reaches (Supplemental Figure 3.5). In contrast, individuals who were born in and returned to the upper reaches of C Creek had greater fitness than individuals who were born in the lower reaches but spawned in the upper reaches. The two models also showed that males had higher fitness than females in both streams (Figure 3.6). Finally, individuals with a

lower body length to depth ratio had increased fitness in A Creek, but there was no effect of body shape on fitness in C Creek (Figure 3.6).

Investigation of the effects of temporal structure on fitness indicated that temporal stray distance (the absolute difference between estimated fertilization day and return day) had no effect on fitness (Figure 3.7). The best fit models for both creeks did not include the parameter for temporal stray distance. The model for A Creek included the parameters for return day, body shape, and sex (AIC = 2708.84, *w*AIC = 0.34, $\beta_{\text{return day}}$ = -0.26, $\beta_{\text{body shape}}$ = -0.39, $\beta_{\text{sex males}}$ = 0.44, Supplemental Table 3.4), while the only well-supported parameter in C Creek was sex (AIC = 4080.47, *w*AIC = 0.32, $\beta_{\text{sex males}}$ = 0.42, Supplemental Table 3.4). Model results indicated that individuals returning earlier in the season had greater reproductive success in A Creek compared to individuals returning later (Figure 3.7).

3.5 DISCUSSION

This study examined the scale, drivers, and fitness consequences of spatial and temporal structure within two stream spawning populations of Sockeye salmon (A and C Creeks in Bristol Bay, Alaska) over 14 years. We found evidence of both spatial and temporal genetic structure within both streams; individuals returning within 50 meters and within five days of one another were more related than expected by chance. This spatial and temporal structure within populations could be partly explained by fine-scale homing both spatially and temporally. In one of the two creeks, individuals returning to spawn at natal sites within the streams had greater reproductive success. Spawning habitat also influenced fitness, as reproductive success was correlated with spawning location within the streams. Despite observations of strong temporal homing, we did not find evidence of a fitness effect associated with fine-scale temporal homing in either creek.

The results of this study provide empirical evidence demonstrating how natal homing, reproductive timing, and habitat characteristics can generate population structure and generate fitness variation on very small spatial and temporal scales. This empirical evidence may have implications for our understanding of gene flow and adaptive processes within philopatric populations: natal homing may result in assortative mating systems which can limit gene flow and promote microgeographic adaptation. While eco-evolutionary models currently used in management assume random mating within populations, we show how processes influencing non-random mating may be important for fitness and adaptation.

3.5.1 *Fine-Scale Spatial and Temporal Genetic Structure*

Sockeye salmon were more likely to spawn near relatives in both space and time than expected in a randomly mixed population. This finding demonstrated how aggregative breeding populations may exhibit genetic structuring, even when occupying small, continuous habitats that are often assumed to be randomly mixed. While many studies of salmonids have examined genetic structure among populations (e.g., Beacham et al. 2006, Lin et al. 2008, Larson et al. 2014), few have examined it within populations. The few studies that have demonstrated genetic structure within salmonid populations have done so on the scale of tens to hundreds of kilometers (Bentzen et al. 2001, Neville et al. 2006, Kitanishi et al. 2009, Stelkens et al. 2012). To our knowledge, our findings represent the finest-scale example of within-population genetic structuring in any salmonid species.

We observed differences in the magnitude of structure among the two creek populations. Spatial structure was greater in C Creek, but temporal structure was greater in A Creek. Differences in the magnitude of spatial structure may be explained by the more heterogenous habitat in C Creek, such as undercut banks and deep pools, particularly between 200 and 400 meters from the

stream mouth (Pess 2009). Such heterogenous habitat may result in more precise imprinting in juveniles rearing in C Creek and could result in more accurate olfaction to natal sites in returning adults (Dittman and Quinn 1996, Stewart et al. 2003, Rich Jr et al. 2006). Spatial or temporal differences in structure among streams may also be due to the more chaotic population dynamics and homogenous habitat in A Creek, which are thought to facilitate the evolution of dispersive behaviors (Holt and McPeek 1996, Hendry et al. 2004). For example, competition (due to greater population density) and predation (due to a greater proportion of individuals killed by bears) is greater in A Creek than in C Creek (May et al. in prep, Ch2). Thus, if structure is important for fitness or population resilience, it may be more easily maintained spatially in the heterogeneous habitat of C Creek but temporally in the homogenous and chaotic habitat of A Creek.

3.5.2 *Natal Homing Drives Population Structure; Possible Evidence for Assortative Mating*

While other studies have demonstrated within-population homing in anadromous salmonids on the scale of kilometers (i.e., in rivers or lakes; Stewart et al. 2003, Barnett et al. 2018), this study is one of the first to demonstrate that Sockeye salmon can home within meters of their natal sites of incubation (Quinn et al. 2006) and within days of when they were fertilized as eggs. Thus, we concluded that natal homing behaviors contributed significantly to structure in both space and time, in addition to other likely drivers such as habitat features and social dynamics (Rich Jr et al. 2006, Camacho and Hendry 2020). These results confirm previous findings which suggested that natal homing behaviors may drive within-population genetic structuring in small Sockeye salmon stream habitats (Quinn et al. 2006, Barnett et al. 2019). Our finding that Sockeye salmon homed to their natal site on the scale of meters also supports a long-held hypothesis that Sockeye salmon may home the most accurately of all salmonid species (Keefer and Caudill 2014).

We present possible evidence for an assortative mating system (Jiang et al. 2013) within small, continuous stream habitats. Patterns of fine-scale structure and natal homing in both space and time may increase the likelihood of mating between relatives or individuals who share phenotypes (Jiang et al. 2013). One expected outcome of assortative mating systems is a segregation of phenotypes in time or space (Jiang et al. 2013, Richardson et al. 2014). Indeed, phenotypes were segregated in the creeks; larger individuals and a higher proportion of males spawned in the upper reaches of C Creek, and earlier returning individuals spawned closer to the mouth of A Creek. Our results corroborate findings from previous studies showing that phenotypes (i.e., body size, return timing) may be segregated in Sockeye salmon stream spawning populations (Doctor and Quinn 2009, Camacho and Hendry 2020). In addition, the number of days individuals survive once entering these streams has been shown to range from approximately three to ten days (Lin et al. 2016), and Sockeye salmon are known to reproduce within just three days of entering stream habitats (Esteve 2005, Lisi et al. 2013). The short reproductive lifespan of Sockeye salmon in the streams lends additional support to our hypothesis that these small populations exhibit an assortative mating system. This assortative mating of clustered groups of phenotypes could lead to microgeographic adaptation if fitness is increased for individuals who reproduce in their natal habitat patches (Richardson et al. 2014).

3.5.3 *Fitness Effects of Fine-Scale Spatial and Temporal Structure*

We demonstrate, for the first time in a vertebrate species, a fitness benefit associated with natal homing within a continuous breeding population. This finding in C Creek was consistent with the expected outcomes of microgeographic adaptation: individuals should have increased fitness if, in the context of a continuous and variable landscape, they reproduce in habitat patches to which they are adapted (Richardson et al. 2014). Yet, the fitness benefits of homing were also compounded

by fitness variation along the length of the stream, which may indicate variation in intrinsic habitat qualities (Johnson 2007, Richardson et al. 2014, Bieri et al. 2018). Individuals spawning in the upper reaches of the stream had increased fitness compared to the lower reaches, indicating better quality habitat in the upper reaches. Indeed, the upper sections on C Creek comprise more woody debris, undercut banks, and deep pools than the lower half, which is dominated by shallow, uniform habitat with little refuge (Pess 2009). A major challenge for studies investigating the drivers of fitness variation has been to separate the fitness effects of local adaptation and habitat variation (Kawecki and Ebert 2004, Snowberg and Bolnick 2012, Savolainen et al. 2013, Richardson et al. 2014). Our results empirically demonstrate that both microgeographic adaptation and habitat variation can contribute to fitness on very small spatial scales.

An alternative explanation for the result that fitness varied throughout the stream may be that larger bodied individuals were more competitive and therefore outcompeted smaller individuals for higher quality sites in the upper reaches of C creek (Foote 1990, Quinn and Foote 1994). It is also possible that individuals may have chosen habitats that best match their phenotypes (Camacho and Hendry 2020). Still, if competition or habitat choice were the only factors driving population structure and fitness, we would not have found an effect of stray distance on reproductive success. Therefore, we find it likely that a combination of differences in competitive ability, habitat characteristics, and microgeographic adaptation contributed to fitness outcomes in C Creek.

We found no evidence supporting a pattern of adaptation-by-time in these populations (Hendry and Day 2005, Doctor and Quinn 2009); individuals returning closer in time to when they were fertilized as eggs had no difference in reproductive success relative to individuals who returned farther away in time. Observed patterns of fine-scale temporal structure may have resulted from strong stabilizing selection on return timing, shown in some studies to occur due to local optima

for egg incubation temperatures, fry emergence timing, and available spawning habitat (Einum and Fleming 2000, Anderson et al. 2010, Crozier et al. 2019). In other words, as long as individuals arrived during the spawning season, they may have had offspring, while individuals who return outside of that reproductive period produced no offspring. High heritability values have also been attributed to phenological traits in salmonids, including return timing (Hendry and Day 2005, Carlson and Seamons 2008). In fact, recent studies have identified specific loci or genetic architecture that explain significant variation in phenological life-history traits (Barson et al. 2015, Hess et al. 2016, Prince et al. 2017, Pearse et al. 2019, Thompson et al. 2019). Broadly, it appears that observed patterns of temporal structure were a byproduct of strong stabilizing selection for return timing. Thus, local spatial, demographic, and ecological processes likely played a larger role in fitness outcomes than temporal homing within a given spawning season (May et al. in prep, Ch2).

3.5.4 *Assumptions and Limitations*

This study made several assumptions that may have influenced interpretations. First, we assumed that spawning locations, estimated from individual daily position observations, were representative of true locations of redds. The estimated spawning locations of parent pairs were highly correlated, suggesting that estimated spawning locations were a reasonable approximation of true redd locations. It is likely that the estimated spawning locations of females were more precise than males, because males traverse greater distances on spawning grounds (Rich Jr et al. 2006). Thus, the homing-based analyses used mother location as a proxy for natal locations. This approach in turn reduced the sample size; a higher sample size may have improved power of the study. Second, we assumed that individual return day represented spawn timing, based on the finding that Sockeye salmon spawn within several days of entering stream habitats (Esteve 2005,

Lisi et al. 2013). Because the same return day measure was used for all individuals, we expected no systemic bias. Third, we used pedigree-based relatedness values for spatial autocorrelation while other authors used genetic distance and thus coalescence of an extended time. Yet pedigree-based relatedness values provided better estimates for assortative mating for the aims of this study. Finally, we excluded immigrants, even though they constituted 20-30% of all fish in some years (May et al. in prep, Ch2). Immigrants were removed from all analyses, because these individuals had disproportionately large stray distances and would have confounded detections of within-population structure. However, it is likely that immigrants reduced the magnitude of structure and assortative mating than reported here. Furthermore, immigrants may have reduced the risk of inbreeding or kin-competition, which was not examined here.

3.5.5 *Conclusions and Future Directions*

This study is one of the first to empirically demonstrate how natal homing behaviors might lead to assortative mating systems within local population subunits. Consistent with the expected outcomes of microgeographic local adaptation, we also demonstrate how this population structure can lead to a segregation of phenotypes and fitness variation within populations (Richardson et al. 2014). Local adaptation is typically considered a population-level process, but here we show how it may occur on finer spatial scales than previously thought. As a result, we conclude that these populations may harbor increased genetic diversity compared to a randomly mixed population (i.e., panmixia). Yet further study is needed to examine whether assortative mating systems increase genetic diversity in wild populations, particularly when population sizes are small and genetic drift may play a larger role in genetic diversity (Adkison 1995, Consuegra et al. 2005). If assortative mating does result in increased diversity, then we posit that it may contribute to a fine-scale portfolio effect (Schindler et al. 2010), buffering discrete populations against perturbations

in any one population segment more than expected in a panmictic population. As a result, selective harvest by predators or fisheries may also reduce genetic diversity more than expected in a panmictic population. For example, brown bears are known to kill a high proportion of fish in these streams on a given day (May et al. in prep, Ch2). Given our findings of fine-scale temporal structure, these predation events may selectively remove family groups or 'genetic neighborhoods' from the population gene pool (Richardson et al. 2014). Likewise, commercial salmon fisheries operating on openings and closures might reduce genetic diversity more than expected by chance if relatives collectively migrate upriver (Berdahl et al. 2016). However, current models used in management to predict effects of anthropogenic perturbations often assume random mating within populations (e.g., Araki et al. 2008, Reed et al. 2011, Willoughby and Christie 2019). Here we demonstrate how processes that influence non-random mating may more accurately represent wild population dynamics.

3.6 FIGURES

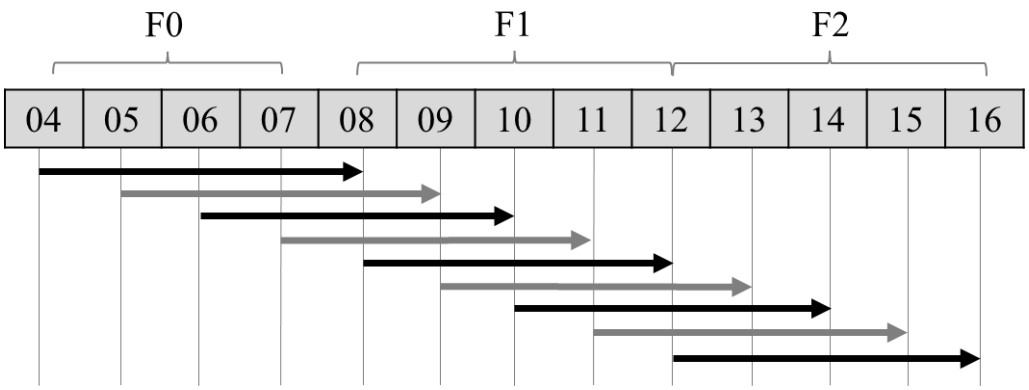


Figure 3.1. A schematic of a two-generation pedigree (F0 to F2) of two adjacent wild Sockeye salmon populations (A and C Creeks). This pedigree spans 14 years, from 2004 through 2017 (redrawn from May et al. in prep, Ch2). Arrows trace cohorts from brood year to return year for four-year-old individuals (81.7% of individuals, 2004 through 2016); therefore, the 2017 return year is not shown, as five-year-olds were sampled in that year, only.

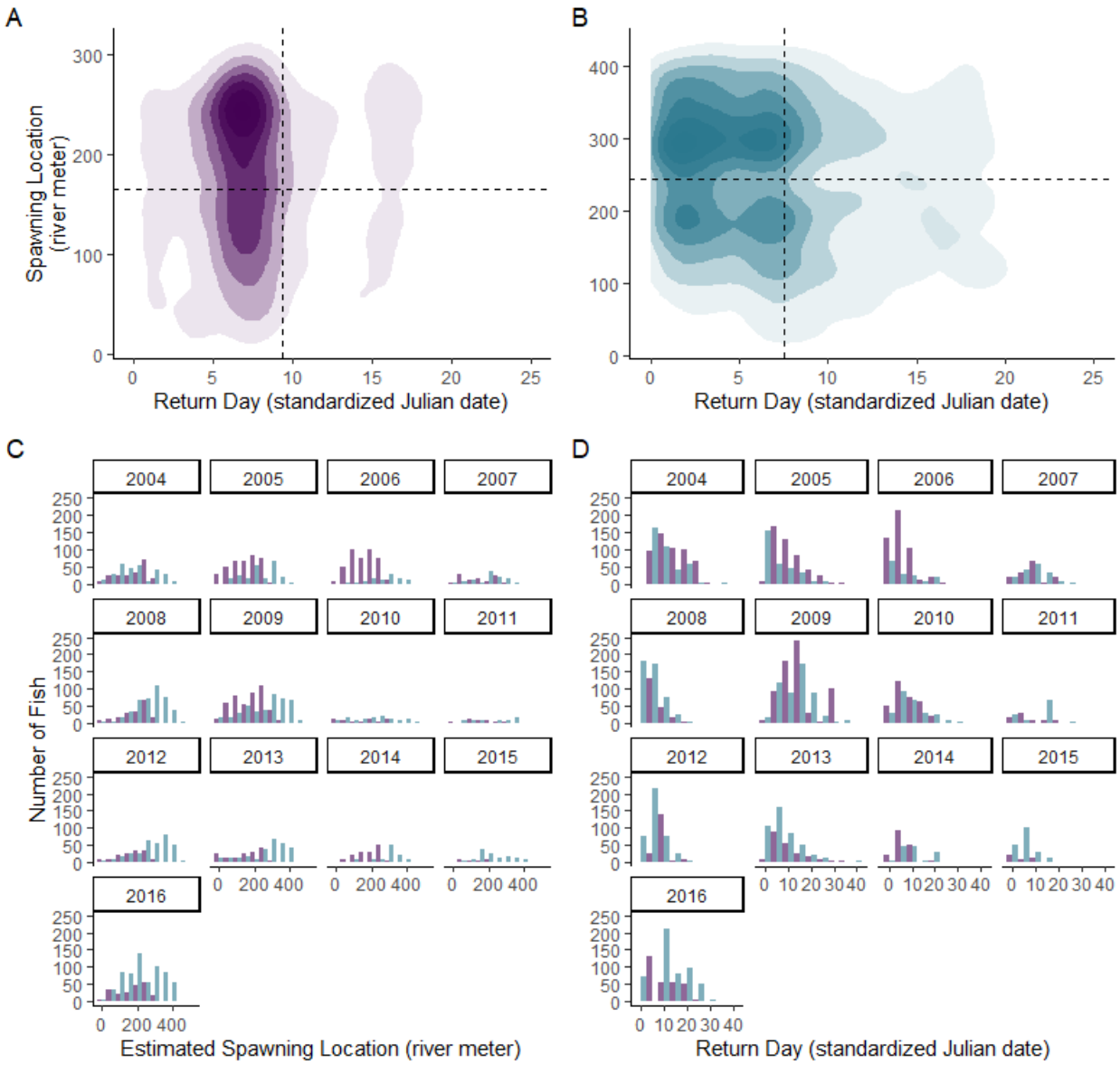


Figure 3.2. Distributions of standardized return days and estimated spawning locations (see text for details) for three generations of Sockeye salmon in two proximate populations (A – purple, C – blue). Panels A-B are kernel density plots using data from all cohorts, combined, which include only individuals attributed values for both estimated return day and spawning location. Contours show density kernels (ggplot2; Hadley 2016), which sum to a value of one. Mean values are given for spawning location (y-axis) and return day (x-axis, dashed lines). Sections C-D are histograms of number of fish returning by return year and include all individuals in the final dataset used in this study (see text for details).

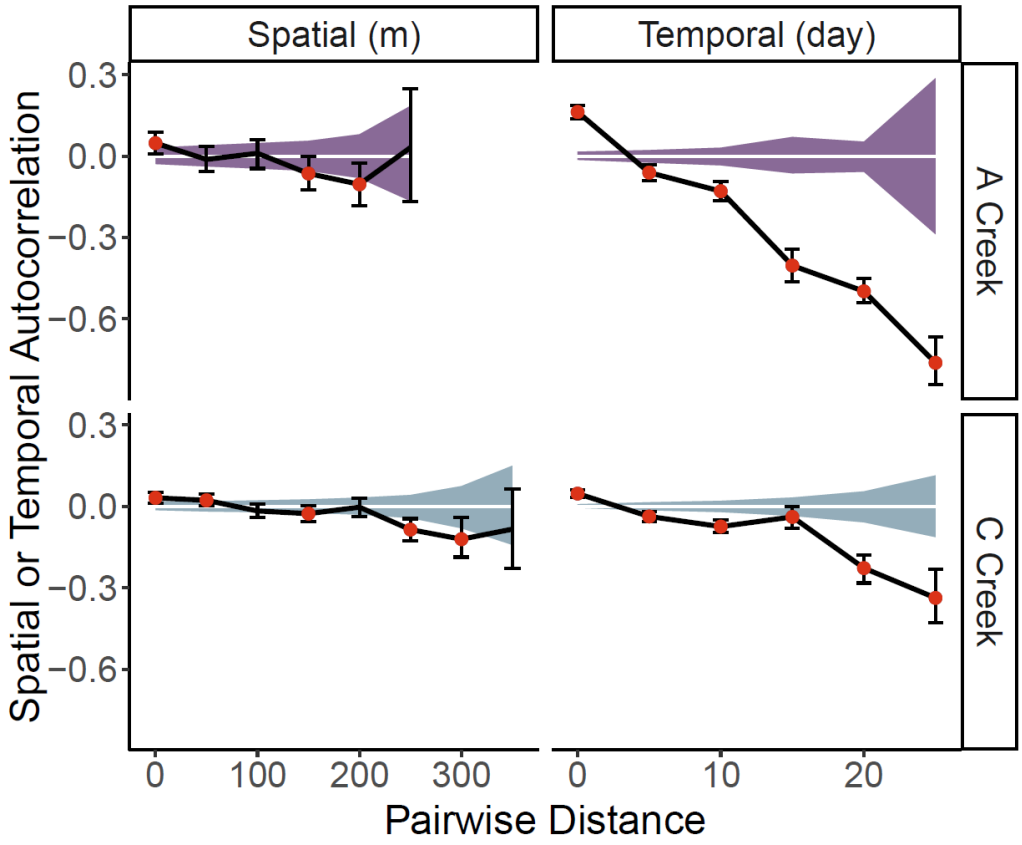


Figure 3.3. Results of spatial (left) and temporal (right) autocorrelation analyses on two adjacent populations of wild Sockeye salmon (A Creek – purple, C Creek – blue). The autocorrelation coefficient (r, y-axis) is given for each distance class (x-axis) in river meters (spatial, left) and standardized return days (temporal, right). Return years were aggregated using the multi-population model in GENALEX (Peakall and Smouse 2006). Shaded regions bound the 95% confidence intervals about r = 0, which represents the null hypothesis of no spatial or temporal structure, obtained through permutations. Additionally, error bars indicate 95% confidence limits for each bootstrapped r-value. Significant points (red) indicate r-values for which the bootstrapped error bars do not include zero and which fall outside the shaded confidence intervals (Smouse and Peakall 1999).

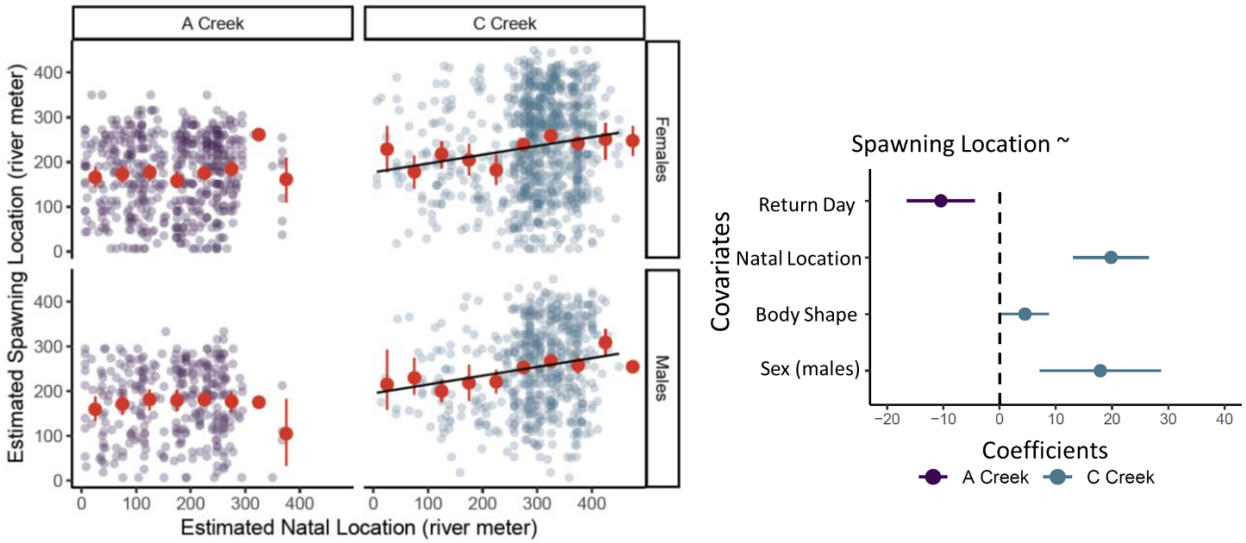


Figure 3.4. Left) Relationship between estimated natal locations (x-axis) and spawning locations (y-axis) in two adjacent Sockeye salmon populations (A – purple, C – blue). Empirical estimates of locations (see text for detail) are given as either individual values (small points) or mean spawning locations (red), binned in 50m x-axis intervals of natal locations among all return years and bounded by bootstrapped 95% confidence intervals. Solid lines show the fitted values of the best-fit mixed effects model (Model 3.1 in text). Right) Model coefficients (x-axis) for the effects of fixed effects (covariates, y-axis) on spawning location (response variable) in best-fit models (Model 3.1 in text; A – purple, C – blue). Point estimates are bounded by model-estimated standard error bars.

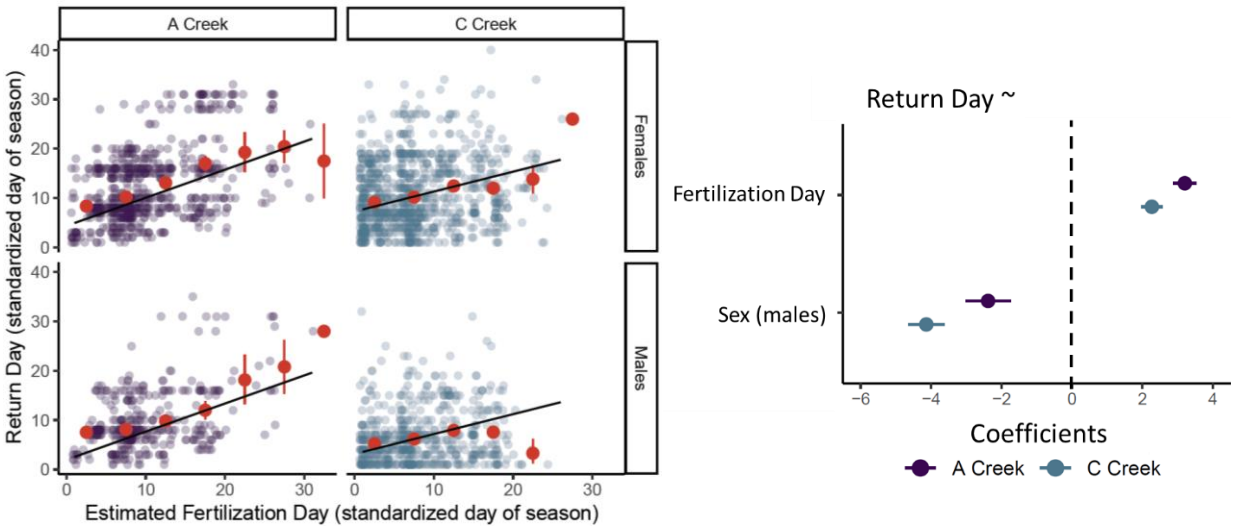


Figure 3.5. Left) Relationship between estimated fertilization day (x-axis) and return day (y-axis) in two adjacent Sockeye salmon populations (A – purple, C – blue). Empirical estimates of return day and fertilization day are given as either individual values (small points) or mean y-axis values (red), binned in five-day x-axis intervals among all return years and bounded by bootstrapped 95% confidence intervals. Solid lines give the fitted values of the best-fit mixed effects models examining the relationship between these parameters (Model 3.2 in text). Right) Model coefficients (x-axis) for the effects of fixed effects (covariates, y-axis) on return day (response variable) in the best-fit models (Model 3.2 in text; A – purple, C – blue). Point estimates are bounded by model-estimated standard error bars.

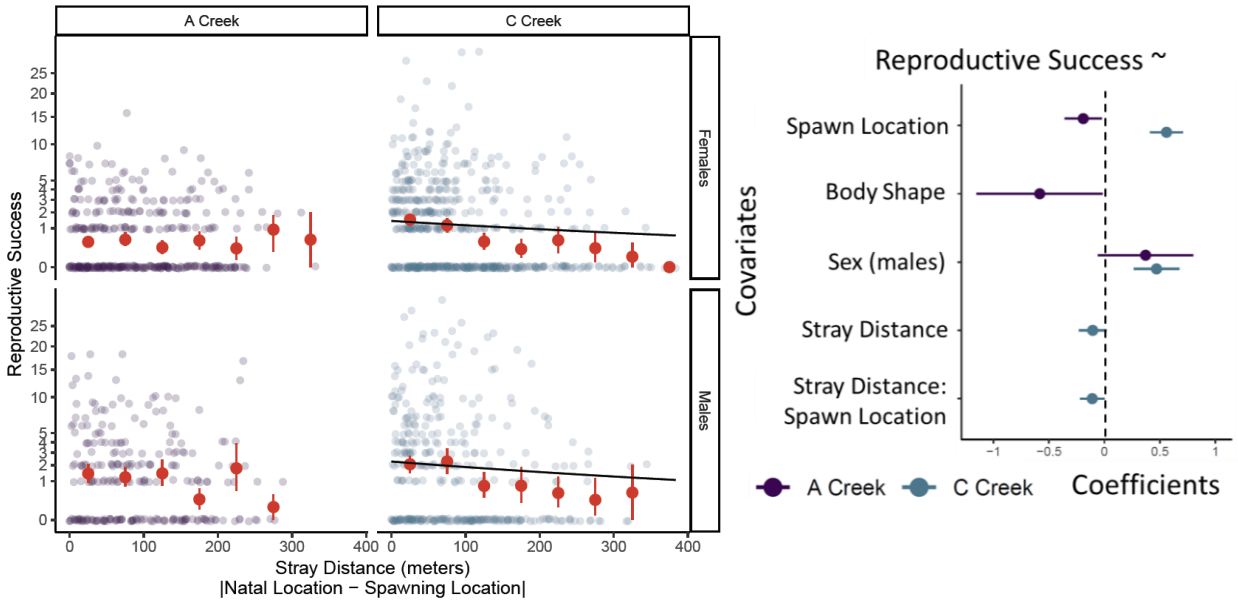


Figure 3.6. Left) Relationship between reproductive success (fitness) and stray distance across two spawning populations of Sockeye salmon (A Creek – purple, C Creek – blue). Stray distance was estimated as the absolute distance between estimated natal and spawning locations for each individual (x-axis). The lifetime reproductive success, measured as number of returning adult offspring was derived from the pedigree (y-axis, square-root scale). Empirical estimates of reproductive success (points) are given as either individual values (small points) or mean values (red), binned in 50m intervals among all return years and bounded by bootstrapped 95% confidence intervals. Solid lines give the fitted response values of the best-fit mixed effects models (Model 3.3 in text). Right) Model coefficients (x-axis) for the effects of fixed effects (covariates, y-axis) on reproductive success (response variable) in the best-fit models (Model 3.2 in text; A – purple, C – blue). Point estimates are bounded by model-estimated standard error bars. Additional effects are shown in the supplemental materials (Supplemental Figure 3.4).

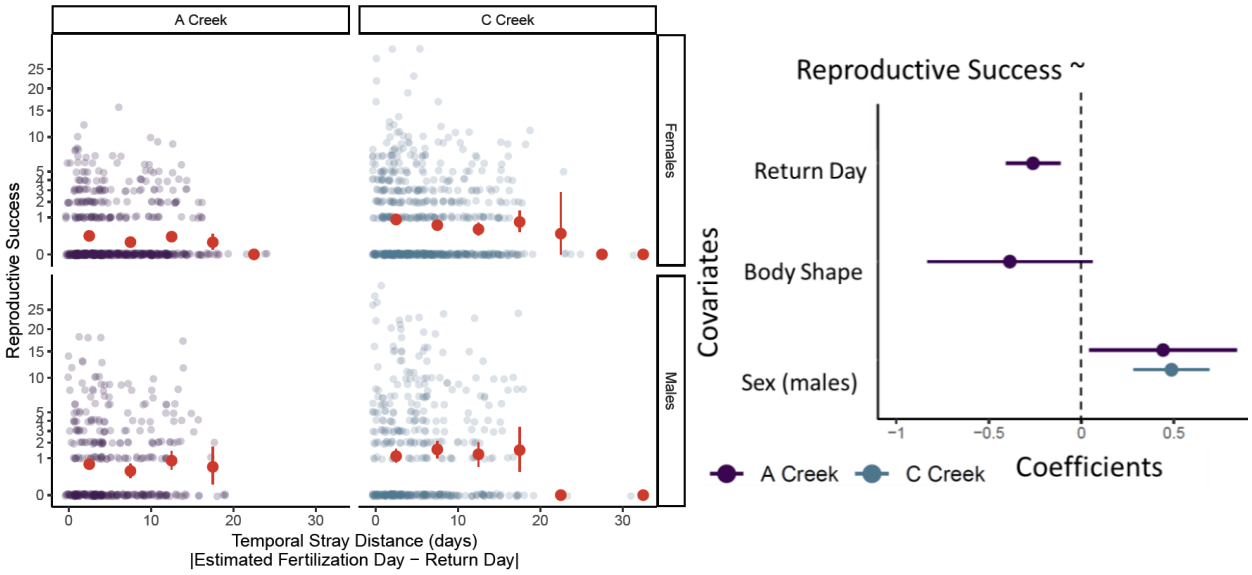
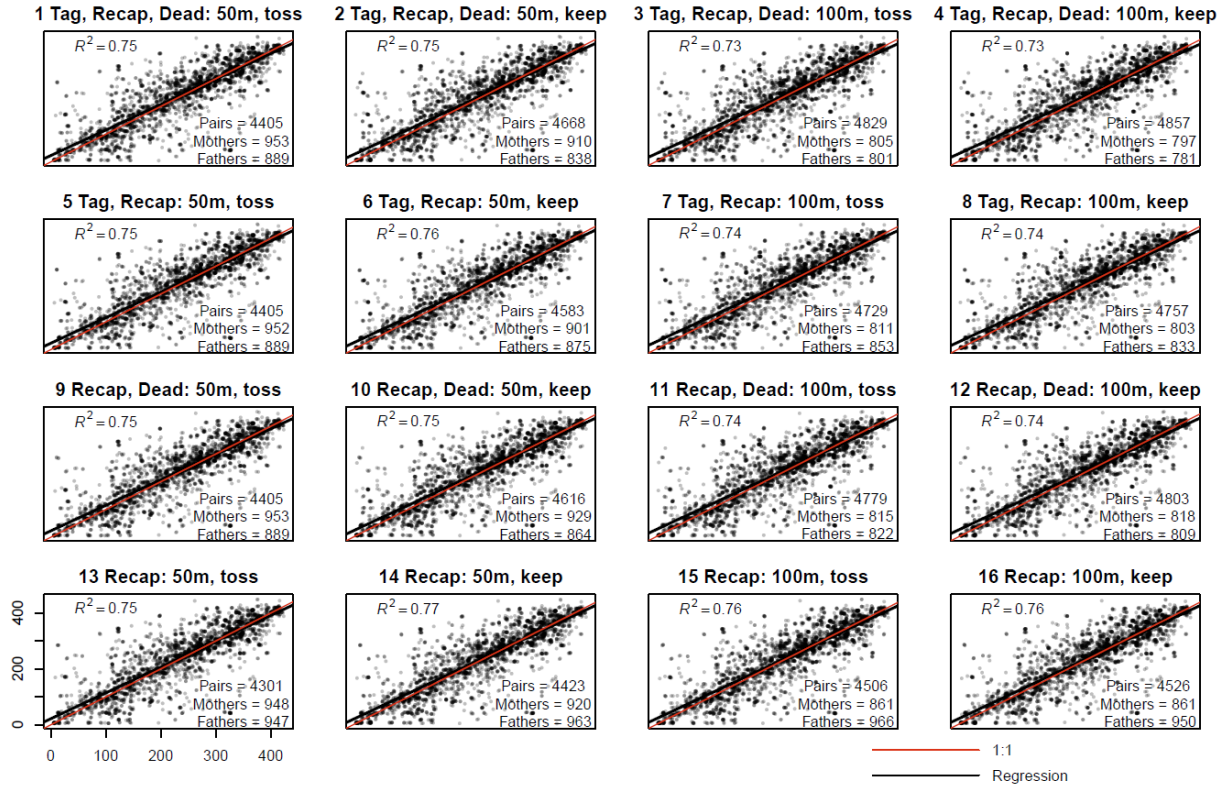


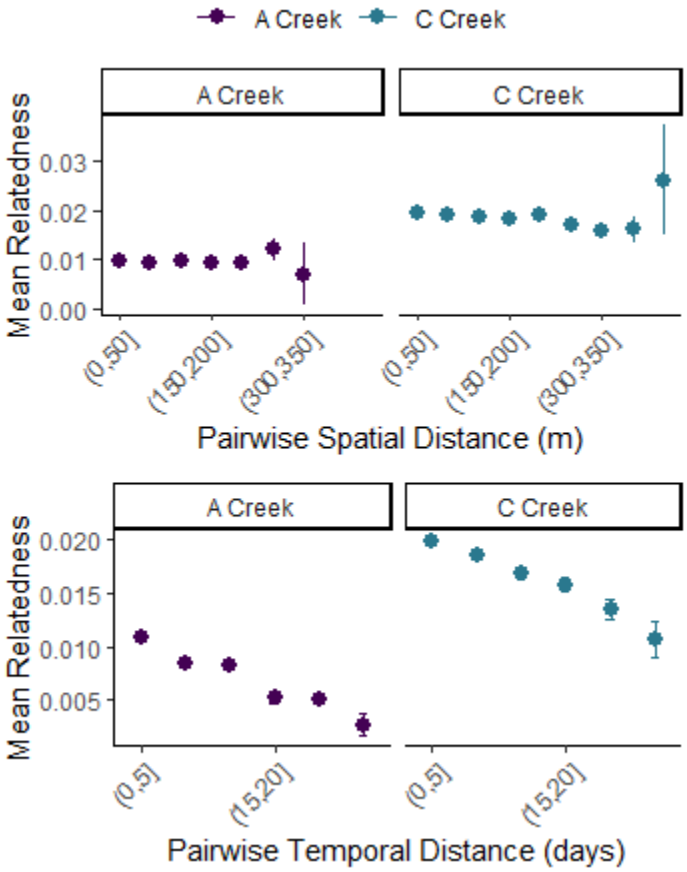
Figure 3.7. Left) Relationship between reproductive success and temporal stray distance across two spawning populations of Sockeye salmon (A Creek – purple, C Creek – blue). Temporal stray distance was estimated as the absolute distance (in days) between estimated individual day of fertilization and return day (x-axis). Empirical estimates of reproductive success (points) are given as either individual values (small points) or mean values (red), binned in five-day intervals among all return years and bounded by bootstrapped 95% confidence intervals. The best-fit mixed effects model exploring this relationship did not include temporal stray distance (Model 3.4 in text). Right) Model coefficients (x-axis) for the effects of fixed effects (covariates, y-axis) on reproductive success (response variable) in the best-fit models (Model 3.2 in text; A – purple, C – blue). Point estimates are bounded by model-estimated standard error bars. Additional effects are shown in the supplemental materials (Supplemental Figure 3.6).

3.7 SUPPLEMENTAL MATERIALS



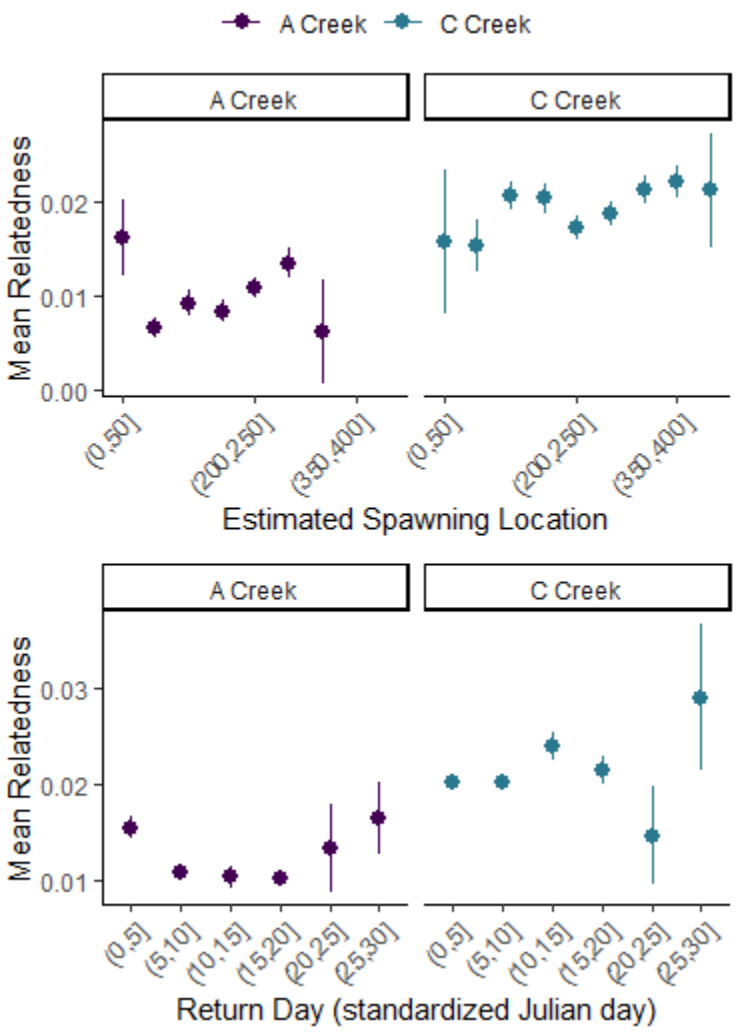
Supplemental Figure 3.1. Relationship between estimated spawning location (meters) for mothers (y-axes) and fathers (x-axes) for 16 methods of estimating spawning location from daily location observations using dfBeta values to remove outliers and quantify mean location (use of dfBeta detailed in text). Here we used all genotyped individuals with known pedigree-based parent-offspring relationships (2004-2017, May et al. in prep, Ch2). We excluded individuals observed only on the beaches or at the creek mouths. Regression lines of best fit (black) conformed closely to 1:1 lines (red), indicating a high likelihood that individuals were assigned to correct spawning locations using all methods. R^2 values provided additional information about the model fit and ranged from 0.73-0.77. Plot titles indicate 1) which observations were included (tag date, death date, and all other observations: recaptures); 2) if only two observations remained, they were included only if the two observations were less than either 50 or 100 meters apart; and 3) if only one observation remained, it was either retained (kept) or removed (tossed). Methods were qualitatively compared to optimize for the largest possible dataset and smallest possible R^2 value, while removing individuals with very few or inaccurate observations. For

example, the first observation of the locations where carcasses were recovered may not accurately reflect spawning location within these short streams. We ultimately proceeded with method 16, in which we excluded observations taken on the first day of sampling or where the fish was found dead. If only two observations remained, we estimated a spawning location only if they were <100m apart, under the assumption that greater distances would not provide accurate estimates of spawning location. Finally, we retained singleton observations, under the assumption that, in most cases, this was 1 of 3 total observations (tag location, recapture location, and carcass location), and was likely the best indicator of the three of spawning location. This method provided confident assignments for 4,526 parent-pairs, 861 dam-offspring pairs, and 950 sire-offspring pairs (no dam). We ultimately used only dam-offspring pairs to estimate individual natal locations (total 5,387), under the assumption that daily observations of females were closer to true redd locations and therefore were better predictors of individual natal location.



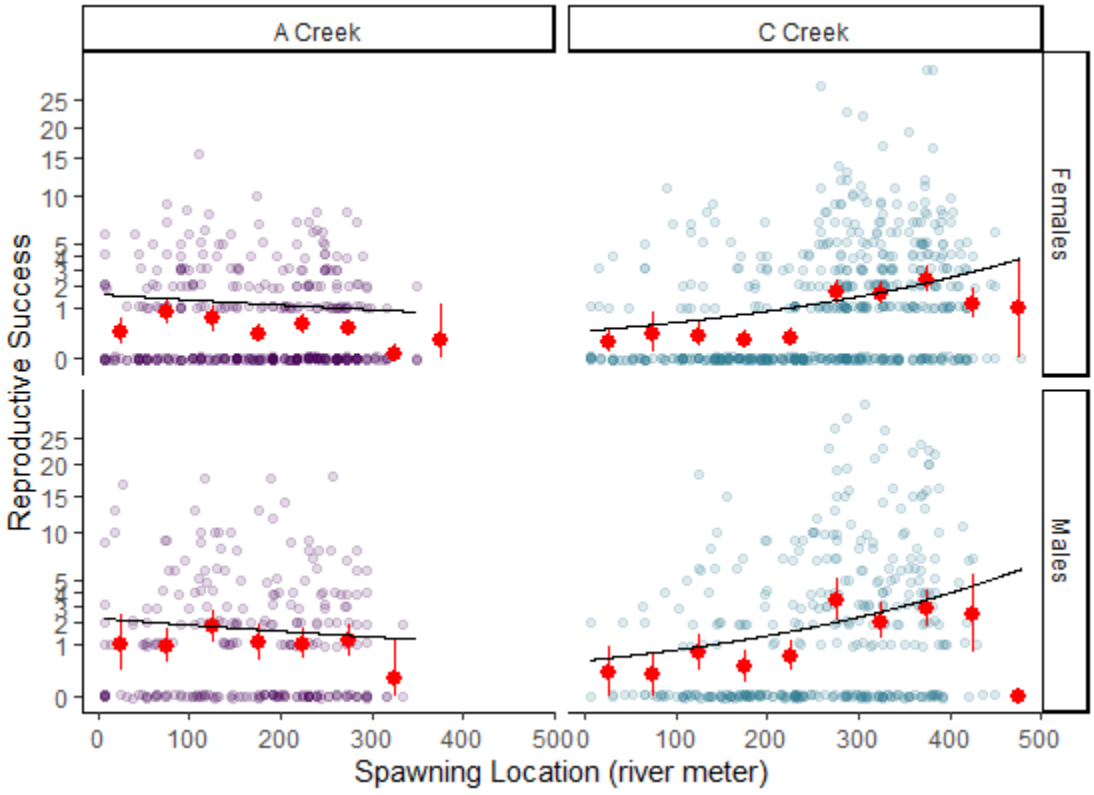
Supplemental Figure 3.2 Preliminary analysis examining the relationship between pairwise geographic distance (x-axis, top) or pairwise temporal distance (x-axis, bottom) and mean pairwise relatedness (y-axis) for two proximate populations of wild Sockeye salmon (A Creek – purple, C Creek – blue). These analyses are similar to autocorrelation analyses used in the main text (Figure 3.3); however, mean relatedness was used here instead of the autocorrelation coefficient, r. Mean relatedness was estimated from pedigree-based pairwise relatedness values of all individuals within cohorts and populations for each distance category in 50m increments (spatial, top) and temporal distance classes in five-day increments (temporal, bottom). Error bars indicate 95% confidence limits about mean relatedness values. As in autocorrelation analyses in the text (Figure 3.3), to eliminate the effects of different ecological conditions experienced by individuals in different brood years, only four-year-olds were included in these analyses. In addition, to examine the two populations separately, immigrants were removed from these analyses. Spatial analyses (top) indicated a slight negative relationship between pairwise distance class and mean relatedness among individuals in C Creek, but no evident trend in A Creek. This result was similar to the spatial autocorrelation results reported in the main text; however, by

using spatial autocorrelation analyses, significant genetic structure was detected in A Creek. This preliminary analysis also highlighted the small number of comparisons in the greatest distance classes between individuals spawning in the marginal habitat at the very top of the creeks (only available in some years) and the very bottom of the creeks (near the mouth). Based on this finding, these distance classes were excluded from spatial autocorrelation analyses. Temporal analyses (bottom) indicated a strong negative relationship between relatedness and pairwise temporal distance in both populations. In addition, mean relatedness appears to be greater overall in C Creek. This result is surprising given the increased proportion of immigrants in C Creek relative to A Creek (May et al. in prep, Ch2). Ultimately, we chose to use autocorrelation analyses to investigate structure, as these analyses allow for the detection of structure within specific distance classes. Autocorrelation analyses are more sensitive to detect fine-scale structure than Mantel tests (Mantel 1967), which typically detect genetic structure only if a strong signal is present in the entire dataset (Peakall et al. 1995, Peakall et al. 2003). Similarly, the autocorrelation methods of Peakall and Smouse (2006) – which use all available pairwise comparisons of individuals – are more sensitive to detecting autocorrelation under restricted gene flow than Moran's I (Cliff and Ord 1981), which compares allele frequencies between groups of genotypes (Epperson 1995, Smouse and Peakall 1999). Lastly, the methods of Peakall and Smouse (2006) also allow for an overall estimate of r to be calculated for each distance class, across multiple populations – substantially increasing power to detect autocorrelation by increasing the total number of comparisons (Peakall et al. 2003).

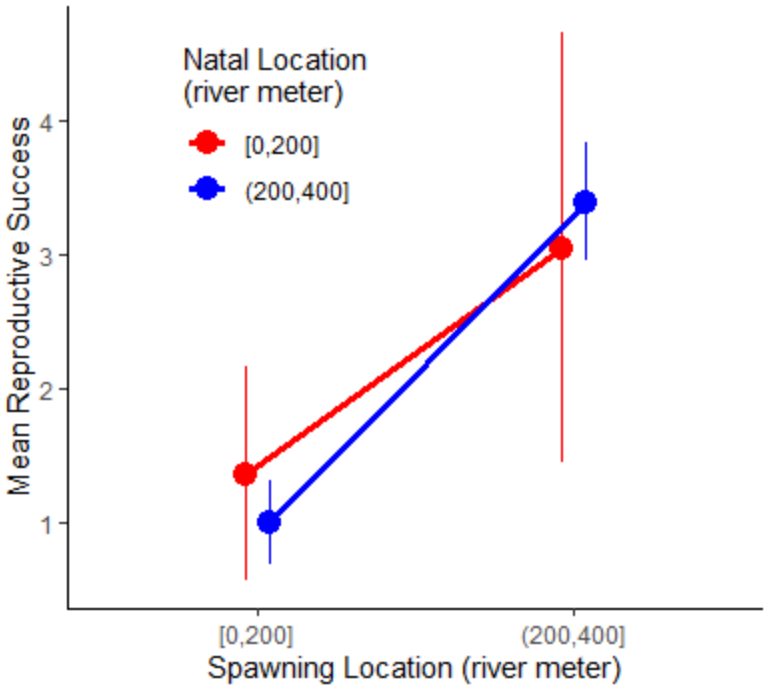


Supplemental Figure 3.3. Preliminary analysis examining the relationship between estimated spawning location (x-axis, top) or return day (x-axis, bottom) and average pairwise relatedness (y-axis) for two proximate populations of wild Sockeye salmon (A Creek – purple, C Creek – blue). Mean relatedness was estimated from pedigree-based pairwise relatedness values of all individuals within cohorts and populations for each distance bin (x axis) in 50m increments (spatial, top) and standardized return days (temporal, bottom). Error bars indicate 95% confidence limits for relatedness values. As in autocorrelation analyses in the text (Figure 3.3), to eliminate the effects of different ecological conditions experienced by individuals in different brood years, only four-year-olds were included in these analyses. In addition, to examine the two populations separately, immigrants were removed from these analyses. Spatial analyses (top) indicate a positive relationship between distance from the mouth of the stream and relatedness among individuals, although confidence in this result decreases at the extreme distance bins at

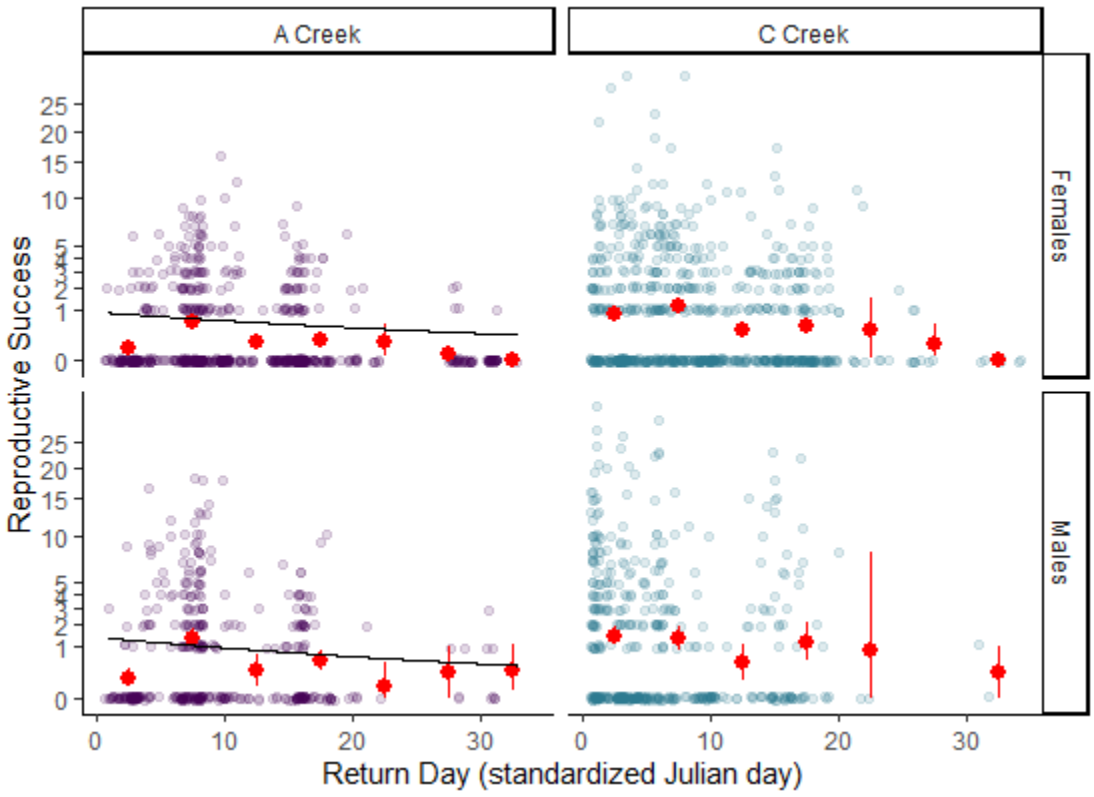
the top and bottom of both creeks. Temporal analyses (bottom) indicate a possible quadratic relationship between relatedness and return day in A Creek, although no trend is evident in C Creek. This quadratic relationship in A Creek supports the hypothesis that individuals attempt to return during the local optima for return day, as it appears temporal structure is weakest during the middle of the season (presumed ecological optimum).



Supplemental Figure 3.4. Fitness effects of spawning location across two spawning populations of Sockeye salmon (A Creek – purple, C Creek – blue). Spawning locations were estimated from individual daily position observations (x-axis, see text for detail). Fitness was measured as the lifetime reproductive success of individuals, derived from the pedigree (y-axis, square-root scale). Empirical estimates of reproductive success (points) are given as either individual values (small points) or mean values, binned in 50m intervals among all return years and bounded by bootstrapped 95% confidence intervals (red). Solid lines give the fitted values of the best-fit mixed effects models (Model 3.3 in text).



Supplemental Figure 3.5. Preliminary analysis of mean reproductive success (y-axis) for individuals born in the lower (red) and upper (blue) half of C Creek. Points give the spawning location (x-axis) for adults returning to either the lower or upper half of the creek. Points are bounded by 95% confidence intervals about the mean reproductive success. This analysis was used to interpret the interaction term between spawning location and stray distance in Model 3.3 in the text. These results indicate that individuals who spawn in the upper regions of C Creek have greater fitness than those in the lower half of the creek. However, of the individuals spawning in the upper regions of the stream, those who were born there have increased fitness compared to immigrants. Similarly, individuals born in the lower half of the stream have greater fitness compared to immigrants. These results are consistent with the expected outcomes of local adaptation (Kawecki and Ebert 2004), but on a microgeographic scale (Richardson et al. 2014).



Supplemental Figure 3.6. Fitness effects of return day across two spawning populations of Sockeye salmon (A Creek – purple, C Creek – blue). Return day was estimated as the day each individual was first observed in the stream, standardized to the Julian date the first individual was observed in each stream, each year (x-axis). Fitness was measured as the lifetime reproductive success of individuals, derived from the pedigree (y-axis, square-root scale). Empirical estimates of reproductive success (points) are given as either individual values (small points) or mean values, binned in five-day intervals among all return years and bounded by bootstrapped 95% confidence intervals (red). Solid lines give the fitted values of the best-fit mixed effects models (Model 3.4 in text).

Supplemental Table 3.1. Results of stepwise generalized linear mixed-effect model selection procedures to explain estimated offspring spawn location (river meter; Model 3.1 in text). Fixed effects included natal location (meters from stream mouth), a quadratic effect of return day (standardized Julian day), body size (length/depth), sex (male or female), or null (1). Separate models were constructed for each spawning population (A Creek or C Creek). Models were evaluated with the Akaike Information Criterion (AIC). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (wAIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Creek	**Fixed Effects**	**K**	**LL**	**AIC**	**ΔAIC**	***w*AIC**
A	return day	4	-4871.35	9750.70	0.00	0.23
A	natal location + return day	5	-4871.04	9752.09	1.39	0.12
A	return day + shape	5	-4871.08	9752.17	1.47	0.11
A	return day + sex	5	-4871.34	9752.67	1.98	0.09
A	return day + return day^2	5	-4871.34	9752.69	1.99	0.09
A	natal location + return day + shape	6	-4870.77	9753.54	2.84	0.06
A	return day + shape + sex	6	-4870.90	9753.80	3.10	0.05
A	natal location + return day + sex	6	-4871.03	9754.06	3.37	0.04
A	natal location + return day + return day^2	6	-4871.04	9754.07	3.37	0.04
A	return day + return day^2 + shape	6	-4871.08	9754.16	3.46	0.04
A	return day + return day^2 + sex	6	-4871.33	9754.66	3.96	0.03
A	natal location + return day + return day^2 + shape	7	-4870.76	9755.52	4.83	0.02
A	return day + return day^2 + shape + sex	7	-4870.90	9755.79	5.10	0.02
A	return day + shape + sex + shape:sex	7	-4870.90	9755.80	5.10	0.02
A	natal location + return day + return day^2 + sex	7	-4871.02	9756.04	5.35	0.02
A	natal location + return day + shape + sex + shape:sex	8	-4870.58	9757.15	6.46	0.01
A	return day + return day^2 + shape + sex + shape:sex	8	-4870.90	9757.79	7.09	0.01
A	natal location + return day + return day^2 + shape + sex + shape:sex	9	-4870.57	9759.14	8.45	0.00
A	1	3	-4877.13	9760.26	9.56	0.00
A	natal location	4	-4876.60	9761.19	10.50	0.00

A	sex	4	-4877.05	9762.10	11.41	0.00
A	shape	4	-4877.12	9762.24	11.54	0.00
A	natal location + sex	5	-4876.53	9763.05	12.36	0.00
A	natal location + shape	5	-4876.59	9763.17	12.48	0.00
A	shape + sex	5	-4876.84	9763.68	12.99	0.00
A	natal location + shape + sex	6	-4876.33	9764.66	13.97	0.00
A	natal location + shape + sex + shape:sex	7	-4876.33	9766.66	15.96	0.00
C	natal location + shape + sex	6	-8783.66	17579.33	0.00	0.46
C	natal location + shape + sex + shape:sex	7	-8783.66	17581.32	1.99	0.17
C	natal location + sex	5	-8785.81	17581.62	2.29	0.15
C	natal location + return day + shape + sex + shape:sex	8	-8783.50	17583.00	3.67	0.07
C	natal location + return day + sex	6	-8785.67	17583.35	4.02	0.06
C	natal location + return day + return day^2 + shape + sex + shape:sex	9	-8783.35	17584.70	5.37	0.03
C	natal location + return day + return day^2 + sex	7	-8785.47	17584.95	5.62	0.03
C	natal location + return day + shape	6	-8787.75	17587.50	8.17	0.01
C	natal location + return day	5	-8788.82	17587.63	8.30	0.01
C	natal location	4	-8789.85	17587.70	8.37	0.01
C	natal location + shape	5	-8789.00	17588.01	8.68	0.01
C	natal location + return day + return day^2 + shape	7	-8787.72	17589.43	10.11	0.00
C	natal location + return day + return day^2	6	-8788.75	17589.50	10.17	0.00
C	shape + sex	5	-8800.18	17610.36	31.03	0.00
C	return day + shape + sex	6	-8799.86	17611.71	32.38	0.00
C	sex	4	-8802.15	17612.30	32.97	0.00
C	return day + return day^2 + shape + sex	7	-8799.71	17613.42	34.09	0.00
C	return day + shape + sex + shape:sex	7	-8799.85	17613.70	34.37	0.00
C	return day + sex	5	-8801.86	17613.73	34.40	0.00
C	return day + return day^2 + sex	6	-8801.66	17615.32	36.00	0.00
C	return day + return day^2 + shape + sex + shape:sex	8	-8799.71	17615.42	36.09	0.00
C	return day	4	-8804.81	17617.63	38.30	0.00
C	return day + shape	5	-8803.83	17617.67	38.34	0.00
C	1	3	-8806.20	17618.39	39.06	0.00
C	shape	4	-8805.46	17618.91	39.59	0.00

C	return day + return day^2	5	-8804.75	17619.49	40.16	0.00
C	return day + return day^2 + shape	6	-8803.80	17619.59	40.27	0.00

Supplemental Table 3.2. Results of stepwise generalized linear mixed-effect model selection procedures to explain return day (Model 3.2 in Text). Models were evaluated with the Akaike Information Criterion (AIC). Fixed effects included fertilization day (return day of pedigreed mothers), body shape (length/depth), sex (male or female), or null (1). Separate models were constructed for each spawning population (A Creek or C Creek). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (wAIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Creek	**Fixed Effects**	**K**	**LL**	**AIC**	**ΔAIC**	***w*AIC**
A	fertilization day + sex	5	-3791.67	7593.35	0.00	0.57
A	fertilization day + shape + sex	6	-3791.55	7595.09	1.74	0.24
A	fertilization day + shape + sex + shape:sex	7	-3790.80	7595.60	2.25	0.19
A	fertilization day + shape	5	-3801.01	7612.03	18.68	0.00
A	fertilization day	4	-3817.55	7643.09	49.74	0.00
A	sex	4	-3953.39	7914.78	321.43	0.00
A	shape + sex	5	-3953.26	7916.52	323.17	0.00
A	shape + sex + shape:sex	6	-3953.25	7918.50	325.15	0.00
A	shape	4	-3960.95	7929.90	336.55	0.00
A	1	3	-3974.80	7955.60	362.25	0.00
C	fertilization day + sex	5	-5205.38	10420.77	0.00	0.66
C	fertilization day + shape + sex	6	-5205.38	10422.76	1.99	0.24
C	fertilization day + shape + sex + shape:sex	7	-5205.34	10424.69	3.92	0.09
C	sex	4	-5303.45	10614.91	194.14	0.00
C	shape + sex	5	-5303.33	10616.66	195.90	0.00
C	shape + sex + shape:sex	6	-5302.64	10617.28	196.51	0.00
C	fertilization day + shape	5	-5314.03	10638.06	217.29	0.00
C	fertilization day	4	-5320.76	10649.51	228.75	0.00
C	shape	4	-5393.85	10795.69	374.93	0.00
C	1	3	-5401.50	10809.01	388.24	0.00

Supplemental Table 3.3. Results of stepwise generalized linear mixed-effect model selection procedures to explain fitness as response variables (Model 3.3 in text). Fixed effects included stray distance (i.e., |estimated natal location – estimated spawn location|), spawn location (meters from stream mouth), sex (Male or Female), or null (1). Separate models were constructed for each spawning population (A Creek or C Creek). All models included a random effect of year. Models were evaluated with the Akaike Information Criterion (AIC). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (wAIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Creek	**Fixed Effects**	**K**	**LL**	**AIC**	**ΔAIC**	***w*AIC**
A	spawn location + shape + sex	6	-1053.16	2118.33	0.00	0.19
A	stray distance + spawn location + shape + sex	7	-1052.31	2118.62	0.30	0.17
A	spawn location + shape	5	-1054.58	2119.17	0.84	0.13
A	stray distance + spawn location + shape	6	-1053.78	2119.56	1.24	0.10
A	stray distance + spawn location + sex	6	-1054.20	2120.39	2.07	0.07
A	spawn location + sex	5	-1055.20	2120.39	2.07	0.07
A	stray distance + spawn location + shape + sex + stray distance:spawn location	8	-1052.24	2120.48	2.15	0.07
A	shape + sex	5	-1055.65	2121.29	2.97	0.04
A	stray distance + spawn location + shape + stray distance:spawn location	7	-1053.73	2121.45	3.13	0.04
A	shape	4	-1057.03	2122.07	3.74	0.03
A	stray distance + shape + sex	6	-1055.06	2122.12	3.80	0.03
A	stray distance + spawn location + sex + stray distance:spawn location	7	-1054.11	2122.22	3.89	0.03
A	stray distance + shape	5	-1056.49	2122.98	4.66	0.02
A	sex	4	-1057.77	2123.54	5.21	0.01

A	stray distance + sex	5	-1057.06	2124.13	5.80	0.01
A	stray distance + spawn location	5	-1068.74	2147.48	29.15	0.00
A	spawn location	4	-1070.22	2148.44	30.12	0.00
A	stray distance + spawn location + stray distance:spawn location	6	-1068.70	2149.40	31.07	0.00
A	stray distance	4	-1071.97	2151.94	33.61	0.00
A	1	3	-1073.07	2152.15	33.82	0.00
C	stray distance + spawn location + sex + stray distance:spawn location	7	-1751.60	3517.20	0.00	0.44
C	stray distance + spawn location + sex	6	-1753.47	3518.94	1.74	0.19
C	stray distance + spawn location + shape + sex + stray distance:spawn location	8	-1751.48	3518.96	1.77	0.18
C	spawn location + sex	5	-1755.30	3520.59	3.39	0.08
C	stray distance + spawn location + shape + sex	7	-1753.36	3520.72	3.52	0.08
C	spawn location + shape + sex	6	-1755.17	3522.34	5.15	0.03
C	stray distance + spawn location + shape + stray distance:spawn location	7	-1760.43	3534.87	17.67	0.00
C	stray distance + spawn location + stray distance:spawn location	6	-1761.74	3535.47	18.27	0.00
C	stray distance + spawn location + shape	6	-1762.10	3536.20	19.00	0.00
C	stray distance + spawn location	5	-1763.34	3536.68	19.49	0.00
C	spawn location + shape	5	-1764.22	3538.44	21.24	0.00
C	spawn location	4	-1765.54	3539.08	21.89	0.00
C	stray distance + sex	5	-1779.19	3568.37	51.17	0.00
C	stray distance + shape + sex	6	-1779.17	3570.34	53.14	0.00
C	stray distance	4	-1787.32	3582.64	65.44	0.00
C	stray distance + shape	5	-1786.59	3583.19	65.99	0.00

C	sex	4	-1796.77	3601.54	84.35	0.00
C	shape + sex	5	-1796.76	3603.52	86.32	0.00
C	1	3	-1805.61	3617.22	100.03	0.00
C	shape	4	-1804.88	3617.76	100.56	0.00

Supplemental Table 3.4. Results of stepwise generalized linear mixed-effect model selection procedures to explain fitness as response variables (Model 3.4 in text). Fixed effects included temporal stray distance (i.e., |fertilization day – return day|), spawn location (meters from stream mouth), sex (Male or Female), or null (1). Separate models were constructed for each spawning population (A Creek or C Creek). All models included a random effect of year. Models were evaluated with the Akaike Information Criterion (AIC). All models were compared to the best-fit model by comparing AIC values (ΔAIC). The weighted AIC score (wAIC) is the proportion of the total predictive power provided by the full set of models. Log-likelihood (LL) accounts for the number of model parameters (K). The null model refers to an intercept-only model with no covariates.

Creek	Fixed Effects	K	LL	AIC	ΔAIC	*w*AIC
A	return day + shape + sex	6	-1348.42	2708.84	0.00	0.34
A	return day + sex	5	-1349.87	2709.75	0.91	0.22
A	temporal stray distance + return day + shape + sex	7	-1348.40	2710.80	1.95	0.13
A	return day + shape	5	-1350.75	2711.51	2.66	0.09
A	temporal stray distance + return day + sex	6	-1349.80	2711.61	2.77	0.09
A	temporal stray distance + return day + shape + sex + temporal stray distance:return day	8	-1348.33	2712.67	3.83	0.05
A	temporal stray distance + return day + shape	6	-1350.75	2713.50	4.66	0.03
A	temporal stray distance + return day + sex + temporal stray distance:return day	7	-1349.76	2713.52	4.68	0.03
A	temporal stray distance + return day + shape + temporal stray distance:return day	7	-1350.59	2715.18	6.34	0.01
A	temporal stray distance + shape + sex	6	-1353.22	2718.44	9.59	0.00
A	shape + sex	5	-1354.38	2718.77	9.92	0.00
A	temporal stray distance + sex	5	-1354.56	2719.12	10.28	0.00
A	sex	4	-1356.01	2720.02	11.18	0.00
A	shape	4	-1357.10	2722.21	13.37	0.00
A	temporal stray distance + shape	5	-1356.14	2722.28	13.44	0.00
A	return day	4	-1367.04	2742.08	33.23	0.00

A	temporal stray distance + return day	5	-1366.98	2743.95	35.11	0.00
A	temporal stray distance + return day + temporal stray distance:return day	6	-1366.50	2745.00	36.16	0.00
A	temporal stray distance	4	-1373.87	2755.74	46.90	0.00
A	1	3	-1375.69	2757.38	48.54	0.00
C	return day + sex	5	-2035.24	4080.47	0.00	0.32
C	temporal stray distance + return day + sex	6	-2035.12	4082.25	1.77	0.13
C	return day + shape + sex	6	-2035.19	4082.37	1.90	0.12
C	temporal stray distance + return day + sex + temporal stray distance:return day	7	-2034.21	4082.41	1.94	0.12
C	sex	4	-2037.30	4082.60	2.13	0.11
C	temporal stray distance + return day + shape + sex	7	-2035.08	4084.16	3.68	0.05
C	temporal stray distance + return day + shape + sex + temporal stray distance:return day	8	-2034.15	4084.31	3.83	0.05
C	shape + sex	5	-2037.28	4084.56	4.09	0.04
C	temporal stray distance + sex	5	-2037.30	4084.60	4.13	0.04
C	temporal stray distance + shape + sex	6	-2037.28	4086.56	6.09	0.02
C	1	3	-2048.26	4102.52	22.05	0.00
C	shape	4	-2047.27	4102.53	22.06	0.00
C	return day + shape	5	-2046.83	4103.67	23.20	0.00
C	return day	4	-2047.97	4103.95	23.47	0.00
C	temporal stray distance + return day + shape + temporal stray distance:return day	7	-2045.06	4104.12	23.65	0.00
C	temporal stray distance + return day + temporal stray distance:return day	6	-2046.18	4104.36	23.89	0.00
C	temporal stray distance	4	-2048.26	4104.52	24.05	0.00
C	temporal stray distance + shape	5	-2047.27	4104.53	24.06	0.00
C	temporal stray distance + return day + shape	6	-2046.82	4105.63	25.16	0.00
C	temporal stray distance + return day	5	-2047.95	4105.90	25.43	0.00

3.8 REFERENCES

Adkison, M. D. 1995. Population differentiation in Pacific salmons: local adaptation genetic drift, or the environment? Canadian Journal of Fisheries and Aquatic Sciences **52**:2762-2777.

Anderson, E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences **65**:1475-1486.

Anderson, J., P. Faulds, W. I. Atlas, G. Pess, and T. Quinn. 2010. Selection on breeding date and body size in colonizing coho salmon, *Oncorhynchus kisutch*. Molecular ecology **19**:2562-2573.

Araki, H., B. A. Berejikian, M. J. Ford, and M. S. Blouin. 2008. Fitness of hatchery-reared salmonids in the wild. Evolutionary applications **1**:342-355.

Barnett, H., T. P. Quinn, K. L. Fresh, M. E. Koehler, K. D. Burton, M. Bhuthimethee, and N. Phil Peterson. 2018. Differential marking of embryos by location and date of release reveals within-river natal and parental influence in progeny return timing in Sockeye Salmon.

Barnett, H. K., T. P. Quinn, K. L. Fresh, M. E. Koehler, K. D. Burton, M. Bhuthimethee, and N. P. Peterson. 2019. Differential Marking of Embryos by Location and Date of Release Reveals Within-River Natal Homing and Parental Influence on Progeny Return Timing in Sockeye Salmon. Transactions of the American Fisheries Society **148**:393-405.

Barson, N. J., T. Aykanat, K. Hindar, M. Baranski, G. H. Bolstad, P. Fiske, C. Jacq, A. J. Jensen, S. E. Johnston, and S. Karlsson. 2015. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature **528**:405-408.

Beacham, T. D., B. McIntosh, C. MacConnachie, K. M. Miller, R. E. Withler, and N. Varnavskaya. 2006. Pacific Rim population structure of sockeye salmon as determined from microsatellite analysis. Transactions of the American Fisheries Society **135**:174-187.

Belsley, D. A., E. Kuh, and R. E. Welsch. 2005. Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons.

Bentzen, P., J. Olsen, J. McLean, T. Seamons, and T. Quinn. 2001. Kinship analysis of Pacific salmon: insights into mating, homing, and timing of reproduction. Journal of Heredity **92**:127-136.

Berdahl, A., P. A. Westley, S. A. Levin, I. D. Couzin, and T. P. Quinn. 2016. A collective navigation hypothesis for homeward migration in anadromous salmonids. Fish and Fisheries **17**:525-542.

Bieri, J. A., C. Sample, W. E. Thogmartin, J. E. Diffendorfer, J. E. Earl, R. A. Erickson, P. Federico, D. T. Flockhart, S. Nicol, and D. Semmens. 2018. A guide to calculating habitat-quality metrics to inform conservation of highly mobile species. Natural Resource Modeling **31**:e12156.

Camacho, C., and A. P. Hendry. 2020. Matching habitat choice: it's not for everyone. Oikos **129**:689-699.

Carlson, S. M., and T. R. Seamons. 2008. A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. Evolutionary applications **1**:222-238.

Cliff, A. D., and J. K. Ord. 1981. Spatial processes: models & applications. Taylor & Francis.

Consuegra, S., E. Verspoor, D. Knox, and C. G. De Leániz. 2005. Asymmetric gene flow and the evolutionary maintenance of genetic diversity in small, peripheral Atlantic salmon populations. Conservation Genetics **6**:823-842.

Crozier, L. G., M. M. McClure, T. Beechie, S. J. Bograd, D. A. Boughton, M. Carr, T. D. Cooney, J. B. Dunham, C. M. Greene, and M. A. Haltuch. 2019. Climate vulnerability assessment for Pacific salmon and steelhead in the California Current Large Marine Ecosystem. PLoS One **14**:e0217711.

Dittman, A., and T. Quinn. 1996. Homing in Pacific salmon: mechanisms and ecological basis. The Journal of experimental biology **199**:83-91.

Doctor, K., and T. Quinn. 2009. Potential for adaptation-by-time in sockeye salmon (*Oncorhynchus nerka*): the interactions of body size and in-stream reproductive life span with date of arrival and breeding location. Canadian Journal of Zoology **87**:708-717.

Edelaar, P., and D. I. Bolnick. 2012. Non-random gene flow: an underappreciated force in evolution and ecology. Trends in Ecology & Evolution **27**:659-665.

Edelaar, P., A. M. Siepielski, and J. Clobert. 2008. Matching habitat choice causes directed gene flow: a neglected dimension in evolution and ecology. Evolution: International Journal of Organic Evolution **62**:2462-2472.

Einum, S., and I. A. Fleming. 2000. Selection against late emergence and small offspring in Atlantic salmon (*Salmo salar*). Evolution **54**:628-639.

Epperson, B. K. 1995. Fine-scale spatial structure: correlations for individual genotypes differ from those for local gene frequencies. Evolution **49**:1022-1026.

Esteve, M. 2005. Observations of spawning behaviour in Salmoninae: *Salmo, Oncorhynchus* and *Salvelinus*. Reviews in fish biology and fisheries **15**:1-21.

Foote, C. J. 1990. An experimental comparison of male and female spawning territoriality in a Pacific salmon. Behaviour:283-314.

Fraser, H., A. Barnett, T. H. Parker, and F. Fidler. 2020. The role of replication studies in ecology. Ecology and Evolution **10**:5197-5207.

Geffen, A. 1992. Validation of otolith increment deposition rate. Pages 101-113 Otolith microstructure examination and analysis. Canada, Ottawa: Can. Spec. Publ. Fish. Aquat. Sci.

Gharrett, A. J., J. Joyce, and W. W. Smoker. 2013. Fine-scale temporal adaptation within a salmonid population: mechanism and consequences. Molecular ecology **22**:4457-4469.

Hadley, W. 2016. Ggplot2: Elegant graphics for data analysis. Springer.

Hanski, I. 1999. Metapopulation ecology. Oxford University Press.

Hendry, A. P., O. K. Berg, and T. P. Quinn. 1999. Condition dependence and adaptation-by-time: breeding date, life history, and energy allocation within a population of salmon. Oikos:499-514.

Hendry, A. P., V. Castric, M. T. Kinnison, T. P. Quinn, A. Hendry, and S. Stearns. 2004. The evolution of philopatry and dispersal. Evolution illuminated: salmon and their relatives:52-91.

Hendry, A. P., and T. Day. 2005. Population structure attributable to reproductive time: isolation by time and adaptation by time. Molecular ecology **14**:901-916.

Hess, J. E., J. S. Zendt, A. R. Matala, and S. R. Narum. 2016. Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. Proceedings of the Royal Society B: Biological Sciences **283**:20153064.

Holt, R. D., and M. A. McPeek. 1996. Chaotic population dynamics favors the evolution of dispersal. The American Naturalist **148**:709-718.

Huisman, J. 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Molecular ecology resources **17**:1009-1024.

Jiang, Y., D. I. Bolnick, and M. Kirkpatrick. 2013. Assortative mating in animals. The American Naturalist **181**:E125-E138.

Johnson, M. D. 2007. Measuring habitat quality: a review. The Condor **109**:489-504.

Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular ecology resources **10**:551-555.

Kardos, M., H. R. Taylor, H. Ellegren, G. Luikart, and F. W. Allendorf. 2016. Genomics advances the study of inbreeding depression in the wild. Evolutionary applications **9**:1205-1218.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecology letters **7**:1225-1241.

Keefer, M. L., and C. C. Caudill. 2014. Homing and straying by anadromous salmonids: a review of mechanisms and rates. Reviews in fish biology and fisheries **24**:333-368.

Kitanishi, S., T. Yamamoto, and S. Higashi. 2009. Microsatellite variation reveals fine-scale genetic structure of masu salmon, *Oncorhynchus masou*, within the Atsuta River. Ecology of Freshwater Fish **18**:65-71.

Larson, W. A., L. W. Seeb, M. V. Everett, R. K. Waples, W. D. Templin, and J. E. Seeb. 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evolutionary applications **7**:355-369.

Lin, J. 2012. Microevolution, local adaptation, and demography in wild populations of Pacific salmon. University of Washington.

Lin, J., T. Quinn, R. Hilborn, and L. Hauser. 2008. Fine-scale differentiation between sockeye salmon ecotypes and the effect of phenotype on straying. Heredity **101**:341-350.

Lin, J. E., J. J. Hard, K. A. Naish, D. Peterson, R. Hilborn, and L. Hauser. 2016. It's a bear market: evolutionary and ecological effects of predation on two wild sockeye salmon populations. Heredity **116**:447.

Link, M. R., and K. K. English. 2020. Long-term, sustainable monitoring of Pacific salmon populations using fishwheels to integrate harvesting, management, and research. Pages 667-674 Sustainable Fisheries Management. CRC Press.

Lisi, P. J., D. E. Schindler, K. T. Bentley, and G. R. Pess. 2013. Association between geomorphic attributes of watersheds, water temperature, and salmon spawn timing in Alaskan streams. Geomorphology **185**:78-86.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer research **27**:209-220.

May, S. A., G. J. McKinney, R. Hilborn, L. Hauser, and K. A. Naish. 2020. Power of a dual-use SNP panel for pedigree reconstruction and population assignment. Ecology and Evolution.

McPhee, M. V., and T. P. Quinn. 1998. Factors affecting the duration of nest defense and reproductive lifespan of female sockeye salmon, *Oncorhynchus nerka*. Environmental Biology of Fishes **51**:369-375.

Milner-Gulland, E. J., J. M. Fryxell, and A. R. Sinclair. 2011. Animal migration: a synthesis. Oxford University Press.

Moran, B. M., and E. C. Anderson. 2018. Bayesian inference from the conditional genetic stock identification model. Canadian Journal of Fisheries and Aquatic Sciences **76**:551-560.

Naish, K., T. Seamons, M. Dauer, L. Hauser, and T. Quinn. 2013. Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (*Oncorhynchus mykiss*) population released in the wild. Molecular ecology **22**:1295-1309.

Neville, H., D. Isaak, J. Dunham, R. Thurow, and B. Rieman. 2006. Fine-scale natal homing and localized movement as shaped by sex and spawning habitat in Chinook salmon: insights from spatial autocorrelation analysis of individual genotypes. Molecular ecology **15**:4589-4602.

Peakall, R., M. Ruibal, and D. B. Lindenmayer. 2003. Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. Evolution **57**:1182-1195.

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

Peakall, R., P. E. Smouse, and D. R. Huff. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss (*Buchloe dactyloides*). Molecular ecology **4**:135-148.

Pearse, D. E., N. J. Barson, T. Nome, G. Gao, M. A. Campbell, A. Abadía-Cardoso, E. C. Anderson, D. E. Rundio, T. H. Williams, and K. A. Naish. 2019. Sex-dependent dominance maintains migration supergene in rainbow trout. Nature ecology & evolution **3**:1731-1742.

Pemberton, J. 2008. Wild pedigrees: the way forward. Proceedings of the Royal Society of London B: Biological Sciences **275**:613-621.

Pess, G. R. 2009. Patterns and processes of salmon colonization.

Peterson, D. A., R. Hilborn, and L. Hauser. 2014. Local adaptation limits lifetime reproductive success of dispersers in a wild salmon metapopulation. Nature Communications **5**:3696.

Poirier, J. M., T. A. Whitesel, and J. R. Johnson. 2012. Chum salmon spawning activity in tributaries below Bonneville dam: the relationship with tailwater elevation and seasonal precipitation. River Research and Applications **28**:882-892.

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

Quinn, T. P., and C. J. Foote. 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., I. J. Stewart, and C. P. Boatright. 2006. Experimental evidence of homing to site of incubation by mature sockeye salmon, Oncorhynchus nerka. Animal Behaviour **72**:941-949.

Reed, T. E., D. E. Schindler, M. J. Hague, D. A. Patterson, E. Meir, R. S. Waples, and S. G. Hinch. 2011. Time to evolve? Potential evolutionary responses of Fraser River sockeye salmon to climate change and effects on persistence. PLoS One **6**:e20380.

Reid, J. M., P. Arcese, P. Nietlisbach, M. E. Wolak, S. Muff, L. Dickel, and L. F. Keller. 2021. Immigration counter-acts local micro-evolution of a major fitness component: Migration-selection balance in free-living song sparrows. Evolution letters **5**:48-60.

Rich Jr, H. B., S. M. Carlson, B. E. Chasco, K. C. Briggs, and T. P. Quinn. 2006. Movements of male sockeye salmon, Oncorhynchus nerka, on spawning grounds: effects of in-stream residency, density and body size. Animal Behaviour **71**:971-981.

Richardson, J. L., M. C. Urban, D. I. Bolnick, and D. K. Skelly. 2014. Microgeographic adaptation and the spatial scale of evolution. Trends in Ecology & Evolution **29**:165-176.

Savolainen, O., M. Lascoux, and J. Merilä. 2013. Ecological genomics of local adaptation. Nature Reviews Genetics **14**:807-820.

Schindler, D. E., R. Hilborn, B. Chasco, C. P. Boatright, T. P. Quinn, L. A. Rogers, and M. S. Webster. 2010. Population diversity and the portfolio effect in an exploited species. Nature **465**:609-612.

Sinnwell, J. P., T. M. Therneau, and D. J. Schaid. 2014. The kinship2 R package for pedigree data. Human heredity **78**:91-93.

Smouse, P. E., and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity **82**:561-573.

Snowberg, L., and D. Bolnick. 2012. Partitioning the effects of spatial isolation, nest habitat, and individual diet in causing assortative mating within a population of threespine stickleback. Evolution: International Journal of Organic Evolution **66**:3582-3594.

Stelkens, R. B., G. Jaffuel, M. Escher, and C. Wedekind. 2012. Genetic and phenotypic population divergence on a microgeographic scale in brown trout. Molecular ecology **21**:2896-2915.

Stewart, I. J., T. P. Quinn, and P. Bentzen. 2003. Evidence for fine-scale natal homing among island beach spawning sockeye salmon, Oncorhynchus nerka. Environmental Biology of Fishes **67**:77-85.

Thompson, T. Q., M. R. Bellinger, S. M. O'Rourke, D. J. Prince, A. E. Stevenson, A. T. Rodrigues, M. R. Sloat, C. F. Speller, D. Y. Yang, and V. L. Butler. 2019. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. Proceedings of the National Academy of Sciences **116**:177-186.

Torres Dowdall, J., C. A. Handelsman, E. W. Ruell, S. K. Auer, D. N. Reznick, and C. K. Ghalambor. 2012. Fine-scale local adaptation in life histories along a continuous environmental gradient in Trinidadian guppies. Functional Ecology **26**:616-627.

Waples, R. S., and O. Gaggiotti. 2006. Invited Review: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular ecology **15**:1419-1439.

Willoughby, J. R., and M. R. Christie. 2019. Long-term demographic and genetic effects of releasing captive-born individuals into the wild. Conservation biology **33**:377-388.

Xu, J. 2018. Ecology and Evolution of Dispersal in Metapopulations.