SPATIO-TEMPORAL GENETIC STRUCTURE, EFFECTIVE POPULATION SIZE, AND PARENTAGE SIMULATIONS FROM CONTEMPORARY GENETIC SAMPLES AND HISTORIC DEMOGRAPHIC DATA OF SOCKEYE SALMON (*ONCORHYNCHUS NERKA*) IN AUKE LAKE, ALASKA

By

Patrick D. Barry

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Fisheries

University of Alaska Fairbanks

August 2021

APPROVED:

Dr. Anthony Gharrett, Committee Chair
Dr. Megan McPhee, Committee Co-Chair
Dr. Eric Anderson, Committee Member
Dr. David Tallmon, Committee Member
Dr. Andrew Seitz, Chair
Department of Fisheries
Dr. S. Bradley Moran, Dean
College of Fisheries and Ocean Science
Dr. Richard Collins,
Director of the Graduate School

Abstract

Pacific salmon (*Oncorhynchus* spp.) have great ecological, economic, and cultural importance. Accordingly, understanding the genetic diversity of Pacific salmon populations is critical for their effective management and conservation. Spatial and temporal homing fidelity, a central life-history characteristic of Pacific salmon, generates genetic structure through reproductive isolation. Within and among populations, heterogeneity in the freshwater environment should lead to selection for traits that maximize fitness resulting in local adaptation. This adaptation increases productivity of individual populations while diversity among populations can promote long-term stability. Additionally, the demographic properties (age structure, generation length, size) of a population will affect genetic structure by regulating its response to the evolutionary forces of selection, migration, and genetic drift. The scale and extent to which reproductive isolation can produce genetic structure is incompletely understood. In this dissertation, I investigated spatial and temporal trends in population genetic structure and estimated the effective population size (N_e) of Sockeye Salmon from Auke Lake in Southeast Alaska from contemporary genetic samples (2008, 2009, 2011) and historic demographic data (1980–2017). A simulation library in the R statistical environment was developed to assess the accuracy of parentage and sibship inference from genetic markers. This library proved useful in evaluating the sibship method for estimating N_e from genetic data and evaluating genetic markers for a large-scale parentage project. I detected substantial genetic differentiation between Auke Lake and other Southeast Alaska populations (average F_{ST} = 0.1137) and an isolation-by-time pattern within the Auke Lake population. A genetically distinct cluster was identified in the late portion of the 2008 return. This group may represent a spatially segregated spawning aggregation previously described in tagging studies; however, because fish were sampled as they passed through the weir, spatial structure within Auke Lake could not be evaluated. Genetic tests for demographic change within the population indicated that the Auke Lake Sockeye Salmon population underwent a historical bottleneck event but has since increased in size. Demographic estimates of N_e from a long-term dataset from the Auke Creek weir revealed that the effective population size was low in the early 1980s and has since increased. Over the six generations evaluated, the major demographic factors that

determined N_e were variance in family size, variable contribution to the next generation by brood years within a generation, and fluctuations in population size. Contemporary estimates of N_e from genetic methods were smaller than those from demographic methods and indicated that N_e may be roughly the size of an individual return year. Genetic estimates of the ratio of the effective population size to the census size (N_e/N_c = 0.21*)* were consistent with values previously reported for other salmonids. Collectively, these chapters contribute to an improved understanding of Sockeye Salmon population genetics and provide a useful tool to assess the power of genetic markers for parentage and sibship inference.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my committee members for their assistance at every stage of this work. I am deeply grateful to my advisor Dr. Tony Gharrett for his support, patience, and dedication to my education. I lack the words to adequately describe the profound respect and admiration I have for him. Thank you to Dr. Megan McPhee for her constant encouragement and guidance. Megan's expansive knowledge of salmon ecology and genetics never ceased to amaze me. She often introduced me to new research and graciously included me on projects stringing together enough research assistantships for me to finish. I am grateful for Dr. David Tallmon's guidance throughout my graduate studies. He has been an incredible mentor and role model. Lastly, thank you to Dr. Eric Anderson at the Southwest Fisheries Science Center for his thoughtful discussions on coalescent theory and mixture models. I thoroughly enjoyed and benefited greatly working with him on our evaluation of an obscure clustering program.

I would also like to acknowledge Dr. William Smoker who, in collaboration with Dr. Gharrett, initiated and was a principal investigator of the project evaluating the genetic consequences of in-basin supplementation.

This dissertation has been made possible in part by decades of research at the Auke Creek weir and direct contributions of time and energy from staff at the National Marine Fisheries Service. I would like to express my sincere gratitude to Scott Vulstek, John Joyce, and Josh Russell for collecting sockeye samples regardless of conditions. Sampling of fish at the weir was conducted in accordance with the policies and regulations of the State of Alaska and under the protocols for the care and use of vertebrate animals approved by the University of Alaska Fairbanks Institutional Animal Care and Use committee.

I would also like to thank those who provided technical assistance in the laboratory. Sharon Hall, Rachel Riley, and Sarah Lyon at the University of Alaska Fairbanks extracted DNA and performed preliminary screening of microsatellite markers. Zac Grauvogel and Wei Cheng from the Alaska Department of Fish and Game were instrumental in single nucleotide polymorphism genotyping.

This work is the result of funding from many sources to which I am exceedingly grateful. The Pacific Salmon Commission Northern Fund provided funding for the genetic component of the project. I would like to extend my sincere thanks to organizations that provided financial support during my dissertation work: the Rasmuson Fisheries Research Center, Pollock Conservation Cooperative Research Center, and Douglas Island Pink and Chum.

Lastly, I would like to thank my family for their unwavering support and belief in me. Without their encouragement, this work would not have been possible.

Introduction

In Alaska, Sockeye Salmon (*Oncorhynchus nerka*) have cultural, economic, and ecological importance. Indeed, Sockeye Salmon occupy a central place in health and heritage of Alaska Native communities. The subsistence harvest of Pacific salmon (*Oncorhynchus* spp.) provides Alaska Natives access to healthy food, use of their indigenous languages, and an understanding of biology and the environment (Rose Fosdick, April 2021, Oral testimony to the NPFMC on D5 salmon genetics). Sockeye Salmon are the predominant species harvested in subsistence and personal use fisheries in Alaska, typically comprising ~ 40% of the Pacific salmon harvest (Brown et al. 2005, Fall et al. 2007, Fall et al. 2014, Fall et al. 2019). Since 1975, Sockeye Salmon have accounted for 55.9% (± 10% SD) of the total value and 27.5% (± 7% SD) of the harvest in state commercial fisheries (ADFG 2021). Pacific salmon are important vectors for nutrient transfer from the North Pacific Ocean to aquatic and terrestrial ecosystems (Bilby et al. 1996, Holtgrieve and Schindler 2011). Upon their return to freshwater to reproduce, Pacific salmon influence the structure and productivity of watersheds as well as their wildlife communities (Willson et al. 1998, Cederholm et al. 1999, Fremier et al. 2018). In recognition of their critical role in the economic and social fabric of Alaska, as well as their functional roles as an ecosystem engineer, substantial effort is devoted to managing their populations. Our understanding of the evolutionary and ecological processes characterizing populations is crucial for the effective conservation and management of our fishery resources.

The management of Sockeye Salmon in Alaska is governed by policy in the state constitution (Article VIII, § 4) stipulating that replenishable natural resources should be managed in accordance with the principle of sustained yield and for the maximum benefit of the people of the state. Within Southeast Alaska, management is complicated by Transboundary Rivers that originate in Northwest British Columbia and flow through Southeast Alaska before terminating at the Pacific Ocean. The Pacific Salmon Treaty (PST) is an agreement between the United States and Canada to cooperate in the management, research, and enhancement of transboundary stocks. A provision within the PST mandates the joint enhancement program of the Stikine and Taku Transboundary

Rivers with production goals of 100,000 Sockeye Salmon. Pacific salmon hatcheries present a tool that can maintain high commercial catches and encourage the recovery of depressed stocks; however, this tool can also disrupt the genetic structure and local adaptation that is characteristic of Pacific salmon populations thereby compromising their health and productivity. The benefits of hatcheries should outweigh their costs, but studies that evaluate the genetic effects of hatchery supplementation are few. This study provides an unprecedented, fine-scale evaluation of the genetic structure of a population of Sockeye Salmon while developing and evaluating methods to detect the genetic effects of hatchery supplementation on depressed wild stocks.

The research conducted makes use of a unique Sockeye Salmon population. The Auke Lake Sockeye Salmon population is relatively small, which permits the exhaustive sampling of all returning adults. Since 1980, escapement has averaged 2778 returning adults. The National Marine Fisheries Service has operated the Auke Creek station, a two-way fish counting weir and small research hatchery, at the head of the tidewater connecting Auke Bay and Auke Creek since 1980. Operation of the weir begins in late winter before ice-out of the lake. In the downstream configuration, all emigrating fry and smolt from early March until June are counted. Over multiple days, scales are sampled from a portion of the outmigration to estimate the age composition. In late June the weir is converted to its upstream configuration to count all immigrating adults. Again, a portion of the returning fish is sampled for scales to estimate the age proportion of the return. The work detailed in these three dissertation chapters would have been unfeasible without the weir infrastructure and personnel.

The first chapter examines the spatio-temporal genetic structure of Sockeye Salmon in Auke Lake, Alaska. Phenotypic diversity exists within and among populations, arising from evolutionary adaptation to variable habitats. This variability can ensure the long-term viability of populations as well as the species by acting as a buffer against both anthropogenic (exploitation and climate change) and stochastic environmental events (Waples and Teel 1990, Wang et al. 2002). Among Sockeye Salmon populations, distinct life history 'ecotypes' have been recognized. With respect to their freshwater residence three main ecotypes have been described: lake-type, sea/river-type, and kokanee (Gilbert 1913, Wood et al. 1987). The 'lake-type,' typically displays strong natal philopatry.

Adults spawn in lake tributaries or lake beaches and juveniles rear for 1–2 years before migrating to the ocean (Burgner 1991, Quinn 1993, Gustafson and Winans 1999). Theories to explain the evolutionary advantage to philopatry include certainty of mate availability, transgenerational resource transfer (nutrients from parental carcasses increasing lake productivity for future generations), and suitable spawning habitat (Hendry et al. 2004). Irrespective of the purpose, this precise homing results in reproductive isolation among populations and substantial spatial structure (Varnavskaya et al. 1994, Wood 1994, Beacham et al. 2006). In contrast, adults of the 'sea/river ecotype' spawn in tributaries of the mainstem river or side channels and their offspring rear in slough and side-channel river habitat for 1–2 years (river-type), while sea-type forego an extended freshwater rearing stage and enter the marine environment as subyearlings (Wood et al. 1987, Gilbert 1913). Sea/river-type sockeye populations are typically weakly differentiated and are theorized to represent the ancestral colonizing form of *O. nerka* (Wood et al. 1987, 2008, Wood 1995, Gustafsen and Winans 1999, McPhee et al. 2009, but see Beacham and Withler 2017). The northern southeast region of Alaska, with its large transboundary rivers where lake-rearing habitat is limited, has a large number of sea/river type populations (Wood et al. 1987, Eiler et al. 1992). This gives rise to the potential for elevated migration between sea/river and lake-type populations (Quinn et al. 2021).

Other prominent ecotypes in salmonids (e.g., Chinook [*Oncorhynchus tshawytscha*] and chum salmon [*Oncorhynchus keta*], steelhead [*Oncorhynchus mykiss*]) are defined relative to their adult migration timing (i.e., winter/spring/fall/late-fall runs). Migration timing is highly heritable in Pacific salmon (Smoker et al. 1998, Quinn et al. 2000, Carlson and Seamons 2008). This can lead to temporal structure among subpopulations (McGregor et al. 1998, Fillatre et al. 2003, Hendry et al. 2002, Hendry and Day 2005) or within populations (Gharrett et al. 2013, Manhard et al. 2017). The genetic variation within subpopulations is shaped by the selective landscape, favoring individuals that are well adapted to their natal stream leading to local adaptation (Taylor 1991). Local adaptation is prevalent in salmonid fishes (Fraser et al. 2011). While spatial heterogeneity in selection regimes leading to local adaptation has been well studied, temporal heterogeneity in selection and its scale leading to adaptation-by-time is less well

understood (Hendry and Day 2005). For adaptation by time to manifest, there must be temporal restriction of gene flow leading to isolation by time. Chapter 1 explores the spatial and temporal genetic structure of Auke Lake sockeye. Drawing on an exhaustive sample of the entire return over three years, the chapter evaluates the strength and persistence of isolation by time while contextualizing the population genetic structure with comparisons to other populations within Northern Southeast Alaska (NSEAK). In doing so, the chapter addresses the following questions: (1) Is there genetic divergence between spatially or temporally segregated spawning assemblages within Auke Lake? (2) How does the genetic composition of the run over the course of the return vary within and among years?

The second chapter explores the signal of demographic change detected in chapter 1 by estimating the effective population size (N_e) over six generations with a long-term dataset from the Auke Creek weir. The effective population size is an important parameter in evolution, ecology, and conservation biology. Populations evolve as a consequence of genetic variability generated by mutation while the relative frequencies of genetic variants that arise and change over time because of the systematic forces of migration, and selection, or the random process of genetic drift. In subdivided populations of finite size, the genetic properties (allele and genotypic frequencies, heterozygosity, and number of alleles at a genetic locus) will fluctuate randomly due to the stochastic process of sampling gametes (which parental alleles are successfully transmitted) and contribute to survival of those offspring. The magnitude of these random fluctuations is inversely related to N_e. The effective population size is the size of an idealized population that would show the same amount of drift (change in some genetic property) as the population being considered (Wright, 1969). The parameter N_e allows us to anticipate how genetic variation (heterozygosity) will decline, inbreeding will increase, and the effect of selection will be minimized within a finite population.

The major objectives of chapter 2 were to explore a potential cause of the genetic uniqueness of Auke Creek relative to other NSEAK populations and evaluate the relative strength of different demographic factors on its N_e. Fluctuations in population size, unequal sex ratios, and variance in reproductive success can all make N_e smaller than the census size (N_c; Frankham 1995). Because it is difficult to estimate N_e by measuring these

causal demographic factors, genetic methods are often employed to measure the signal they impart on some genetic property (Wang 2005). Here data from the National Marine Fisheries Service Auke Creek weir were combined with three years of genetic samples from the return to estimate the effective number of breeders (N_b, the number of breeders over one spawning return) and N_e with both demographic and genetic methods. Using these data I (1) estimated single generation and multigenerational N_e with demographic methods for years 1980 to 2017, (2) estimated N_b with genetic methods (from collections in 2008, 2009 and 2011), and (3) evaluated factors that may contribute to reductions in N_e.

The third chapter of this dissertation developed and evaluated methods to detect the genetic effects of hatchery supplementation on depressed wild stocks. The goal of this chapter was to produce a forward-in-time simulation program that can be applied to large-scale analyses for parentage and sibship studies. The first two chapters provide extensive genetic analyses that describe the natural genetic variation that exists within the Auke Lake Sockeye Salmon population, which has had little hatchery influence. From 2013 to 2015, a portion of the wild return was used as broodstock to supplement the population to investigate the genetic effects of in-basin supplementation. The ability to evaluate differences in relative reproductive success of wild and hatchery fish hinges on accurately resolving parentage. Financial constraints often dictate the balance between genotyping more genetic markers or more individuals in the population; genotyping more markers than needed to accurately resolve relationships represents wasted resources. Simulations allow for the rapid evaluation of different sets of markers. The information gained from these simulations enable us to better tailor our suite of markers so that future genotyping is as efficient as possible while still providing sufficient information to assign individuals back to their parents. With this program two key questions were addressed: Can the full panel of genetic markers be efficiently replaced by a subset of those markers to identify returning fish as either wild- or hatchery-produced? Can individuals be assigned back to their correct parent pair? Specifically, I developed 'PseudoBabies,' a flexible simulation package in the R statistical language, to assess the accuracy of parentage and sibship inference from genetic markers.

Together these chapters contribute to an improved understanding of Sockeye Salmon population structure and evolution. Additionally, this work provided guidance on an experiment to evaluate the genetic effects of hatchery supplementation while describing the baseline genetic structure to which post-supplementation comparisons can be made.

References

ADFG. 2021. Statewide Salmon Gross Earnings by Species. Commercial Salmon Fisheries. Retrieved from https://www.adfg.alaska.gov/index.cfm?adfg=commercialbyfisherysalmon.salmon_grossearnings_byspecies.

Beacham, T., B. McIntosh, C. MacConnachie, K. Miller, R. Withler, and N. Varnavskaya. 2006. Pacific rim population structure of Sockeye Salmon as determined from microsatellite analysis. Transactions of the American Fisheries Society 135(1):174–187.

Beacham T. D., and R. E. Withler. 2017 Population structure of sea-type and lake-type Sockeye Salmon and Kokanee in the Fraser River and Columbia River drainages. PLoS ONE 12(9): e0183713.

Bilby R. E., B. R. Fransen, and P. A. Bisson. 1996 Incorporation of nitrogen and carbon from spawning Coho Salmon into the trophic system of small streams: Evidence from stable isotopes. Canadian Journal of Fisheries and Aquatic Science 53(1):164–173.

Brown, C. L., D. Caylor, J. Dizard, J. A. Fall, S. Georgette, T. Krauthoefer, and M. Turek. 2005. Alaska subsistence salmon fisheries 2003 annual report. Alaska Department of Fish and Game, Division of Subsistence Technical Paper No. 316, Juneau, Alaska.

Burgner, R. L. 1991. The life history of Sockeye Salmon (*Oncorhynchus nerka*). Pages 3–117 *in* C. Groot and L. Margolis, editors. Pacific Salmon Life Histories. University of British Columbia Press, Vancouver.

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.

Cederholm, C. J., M. D. Kunze, T. Murota, and A. Sibatani. 1999. Pacific salmon carcasses: Essential contributions of nutrients and energy for aquatic and terrestrial ecosystems. Fisheries 24:6–15.

Dittman, A. H., and T. P. Quinn. 1996. Homing in Pacific salmon: Mechanisms and ecological basis. Journal of Experimental Biology 199:83–91.

Eiler, J., B. Nelson, and R. Bradshaw. 1992. Riverine spawning by Sockeye Salmon in the Taku River, Alaska and British Columbia. Transactions of the American Fisheries Society 121(6):701–708.

Fall, J. A., D. Caylor, M. Turek, C. Brown, T. Krauthoefer, B. Davis, and D. Koster. 2007. Alaska subsistence salmon fisheries 2004 annual report. Alaska Department of Fish and Game Division of Subsistence, Technical Paper No. 317, Juneau, Alaska.

Fall, J. A., N. M. Braem, C. L. Brown, S. S. Evans, L. Hutchinson-Scarborough, B. Jones, R. La Vine, T. Lemons, M. A. Marchioni, E. Mikow, J. T. Ream, and L. A. Sill. 2014. Alaska subsistence and personal use salmon fisheries 2012 annual report. Alaska Department of Fish and Game Division of Subsistence, Technical Paper No. 406, Anchorage, Alaska.

Fall, J. A., A. Godduhn, G. Halas, L. Hutchinson-Scarbrough, B. Jones, B. McDavid, E. Mikow, L. A. Sill, A. Wiita, and T. Lemons. 2019. Alaska subsistence and personal use salmon fisheries 2016 annual report. Alaska Department of Fish and Game Division of Subsistence, Technical Paper No. 446, Anchorage, Alaska.

Fillatre, E., P. Etherton, and D. Heath. 2003. Bimodal run distribution in a northern population of Sockeye Salmon (*Oncorhynchus nerka*): Life history and genetic analysis on a temporal scale. Molecular Ecology 12(7):1793–1805.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: A review. Genetical Research 66(2):95–107.

Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen, and E. B. Taylor. 2011. Extent and scale of local adaptation in salmonid fishes: Review and meta-analysis. Heredity 106:404–420.

Fremier, A., B. Yanites, and E. Yager. 2018. Sex that moves mountains: The influence of spawning fish on river profiles over geologic timescales. Geomorphology 305:163–172.

Gharrett, A. J., J. E. Joyce, and W. W. Smoker. 2013. Fine-scale temporal adaptation within a salmonid population: mechanism and consequences. Molecular Ecology 22(17):4457–4469.

Gilbert, C. H. 1913. Age at maturity of the Pacific coast salmon of the genus *Oncorhynchus*. Report of the British Columbia Commissioner of Fisheries: 57–70.

Gustafson, R., and G. Winans. 1999. Distribution and population genetic structure of river- and sea-type Sockeye Salmon in western North America. Ecology of Freshwater Fish 8(3):181–193.

Hendry, M. A., J. K. Wenburg, K. W. Myers, and A. P. Hendry. 2002. Genetic and phenotypic variation through the migratory season provides evidence for multiple populations of wild Steelhead in the Dean River, British Columbia. Transactions of the American Fisheries Society 131(3):37–41.

Hendry, A. P., C. Castric, M. T. Kinnison, and T. P. Quinn. 2004. The evolution of philopatry and dispersal versus straying in salmonids. Pages 52–91 *in* A. P. Hendry and S. C. Stearns, editors. Evolution illuminated: salmon and their relatives. Oxford University Press, New York.

Hendry, T., and A. P. Day. 2005. Population structure attributable to reproductive time: Isolation by time and adaptation by time. Molecular Ecology 14(4): 901–16.

Holtgrieve, G. W., and D. E. Schindler. 2011. Marine derived nutrients, bioturbation, and ecosystem metabolism: reconsidering the role of salmon in streams. Ecology 92(2):373–385.

Manhard, C. V., J. E. Joyce, and A. J. Gharrett. 2017. Evolution of phenology in a salmonid population: a potential adaptive response to climate change. Canadian Journal of Fisheries and Aquatic Science 74(10):1519–1527.

McGregor, A., S. Lane, M. Thomason, L. Zhivotovsky, W. Smoker, and A. J. Gharrett. 1998. Migration timing, a life history trait important in the genetic structure of Pink Salmon. North Pacific Anadromous Fish Commission Bulletin 1:262–273.

McPhee, M. V, T. H. Tappenbeck, D. C. Whited, and J. A. Stanford. 2009. Genetic diversity and population structure in the Kuskokwim River drainage support the recurrent evolution hypothesis for Sockeye Salmon life histories. Transactions of the American Fisheries Society 138(6):1481–1489.

Quinn, T. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fisheries Research 18(1-2):29–44.

Quinn, T., M. Unwin, and M. Kinnison. 2000. Evolution of temporal isolation in the wild: Genetic divergence in timing of migration and breeding by introduced Chinook Salmon populations. Evolution 54(4):1372–1385.

Quinn, T. P., G. R. Pess, B. J. Sutherland, S. J. Brenkman, R. E. Withler, K. Flynn, and T. D. Beacham. 2021. Resumption of anadromy or straying? Origins of Sockeye Salmon in the Elwha River. Transactions of the American Fisheries Society. Early version online. https://doi.org/10.1002/tafs.10294.

Smoker, W. W., A. J. Gharrett, and M. S. Stekoll. 1998. Genetic variation of return date in a population of Pink Salmon: A consequence of fluctuating environment and dispersive selection? Alaska Fisheries Research Bulletin 5(1):46–54.

Taylor, E. 1991. A review of local adaptation in salmonidae, with particular reference to Pacific and Atlantic salmon. Aquaculture 98(1–3):185–207.

Varnavskaya, N., C. Wood, R. Everett, R. Wilmot, V. Varnavsky, V. Midanaya, and T. Quinn. 1994. Genetic differentiation of subpopulations of Sockeye Salmon (*Oncorhynchus nerka*) within lakes of Alaska, British Columbia, and Kamchatka, Russia. Canadian Journal of Fisheries and Aquatic Science 51(1):147–157.

Wang, J. L. 2005. Estimation of effective population sizes from data on genetic markers. Philosophical Transactions of the Royal Society B 360:1395–1409.

Wang, S. Z., J. J. Hard, and F. Utter. 2002. Genetic variation and fitness in salmonids. Conservation Genetics 3:321–333.

Waples, R. S., and D. J. Teel. 1990. Conservation genetics of Pacific salmon: I. Temporal changes in allele frequency. Conservation Biology 4(2):144–156.

Willson, M. F., S. M. Gende, and B. H. Marston. 1998. Fishes and the forest: Expanding perspectives on fish–wildlife interactions. BioScience 48:455–462.

Wood, C. C. 1995. Life history variation and population structure in Sockeye Salmon. American Fisheries Society Symposium 17:195–216.

Wood, C. C., B. E. Riddell, and D. T. Rutherford. 1987. Alternative juvenile life histories of Sockeye Salmon (*Oncorhynchus nerka*) and their contribution to production in the Stikine River, northern British Columbia. Proceedings of the International Sockeye Salmon Symposium 96:12–24.

Wood, C., B. Riddell, D. Rutherford, and R. Withler. 1994. Biochemical genetic survey of Sockeye Salmon (*Oncorhynchus nerka*) in Canada. Canadian Journal of Fisheries and Aquatic Science 51(1):114–131.

Wood, C. C., J. W. Bickham, R. John Nelson, C. J. Foote, and J. C. Patton. 2008, Recurrent evolution of life history ecotypes in Sockeye Salmon: implications for conservation and future evolution. Evolutionary Applications 1:207–221.

Wright, S. 1969. Evolution and the genetics of populations. Vol. 2: The Theory of Gene Frequencies. University of Chicago Press, Chicago.

Chapter 1: Spatio-temporal population genetic structure of sockeye salmon *Oncorhynchus nerka* in Auke Lake Alaska[1]

Abstract

Spatial and temporal patterns in population structure of sockeye salmon, *Oncorhynchus nerka* (Walbaum, 1792) from Auke Lake, a coastal lake in northern Southeast Alaska, were investigated with microsatellite and single nucleotide polymorphism (SNP) markers for three return years. Auke Lake sockeye salmon differ genetically from sockeye salmon in 15 other northern Southeast Alaska populations (average $F_{ST} = 0.2249$) for SNP loci. Straying into Auke Lake varied among years; the majority of strays did not originate from nearby sources. Mixture models identified full sibling families in Auke Lake as the primary source of genetic structure. After we accounted for family effects, we identified a genetically distinct cluster in the late portion of the 2008 return. Individuals from this group clustered with late-returning individuals from the 2009 and 2011 return years, which suggests the persistence of this group among years. A significant signal of isolation-by-time was detected in the 2008 and 2011 collections. The late returning group may represent a spatially segregated spawning group, but because fish were sampled as they passed through the weir, their ultimate spawning location was not known, which makes it difficult to evaluate the relative importance of spatial and temporal separation. Our results indicate that the sockeye salmon run within Auke Lake is at least partially reproductively partitioned as a result of temporal, spatial, or both isolating mechanisms.

Introduction

The delineation of genetically distinct populations over their distribution is fundamental to recognizing and conserving genetic diversity. Pacific salmon (*Oncorhynchus* spp.) demonstrate strong philopatry (Dittman and Quinn 1996). Homing to natal streams is generally precise and the resulting reproductive isolation among

---

[1] P.D. Barry, M.V. McPhee, S. Vulstek, S.D. Rogers Olive, W.W. Smoker and A.J. Gharrett. *Formatted for submission to Canadian Journal of Fisheries and Aquatic Sciences*

populations can produce substantial spatial structure. Sockeye salmon (*Oncorhynchus nerka*), in particular the 'lake ecotype,' exhibit strong homing because of the reliance of juveniles on limnetic rearing habitat (Burgner 1991, Quinn 1993, Gustafson and Winans 1999). In contrast, the 'sea/river ecotype' rears in side streams and sloughs for a very brief period (weeks to months) during their seaward migration and is theorized to represent the ancestral colonizing form of *O. nerka* (Wood 1995, Wood et al. 2008, McPhee et al. 2009). Variation in return timing can have a high heritability in Pacific salmon (Smoker et al. 1998, Quinn et al. 2000, Carlson and Seamons 2008), which can lead to temporal structure among subpopulations (McGregor et al. 1998, Fillatre et al. 2003, Hendry and Day 2005) or within populations (Gharrett et al. 2013, Manhard et al. 2017). Over time, selection favors individuals that are well adapted to their natal stream and shapes the genetic variation within the subpopulations (Taylor 1991). The differences that accumulate among populations or segments of populations are presumably important to ensure the long-term viability of populations as well as the species by acting as a buffer against both anthropogenic and environmental changes (Waples and Teel 1990, Wang et al. 2002, Manhard et al. 2017).

Characterizing broad geographic trends in population structure is important for understanding the evolutionary ecology of a species and subsequent management implications. In all species of Pacific salmon, imprecise homing, or straying, is most likely to occur among geographically proximate populations (Olsen et al. 2008), which may result in a pattern of isolation by distance (Beacham et al. 2005, Beacham et al. 2006, Gomez-Uchida et al. 2011, Ackerman et al. 2013). In contrast, weak regional structure that resembles a mosaic pattern of structure has been described, particularly in sockeye salmon populations (Utter et al. 1984, Wood et al.1994, Winans et al. 1996, Nelson et al. 2003), in which genetic variation within regions typically exceeds that among regions (Guthrie et al. 1994, Varnovskaya et al. 1994). This mosaic pattern of genetic structure has been attributed to temporal isolation in spawning times and founder effects (Ramstad et al. 2004), as well as to secondary contact between colonizers from two glacial refugia (Beacham et al. 2006).

Expansive spatial and temporal sampling, combined with more informative molecular markers, has revealed fine-scale divergence within some watersheds with

sockeye salmon populations. Reproductive isolation can develop from spatial, temporal, behavioral, or ecological barriers. Burger and Spearmen (1997) resolved differences in Tustumena Lake, AK where lake and outlet spawners are separated by 10 km and peak spawning occurs 4–6 weeks apart. Fillatre et al. (2003) similarly observed significant differences between early- and late-returning fish in the Klukshu River, Yukon Territory, Canada, which are separated by 30–40 days and may have an additional spatial component (Petkovich 1999). Population divergence has also arisen in the absence of return timing differences. Pavey et al. (2010) observed significant genetic divergence between two spawning ecotypes separated by ~1.5 km despite persistent migration between the two populations. They estimated that the ecological divergence arose over ~100 generations. Hendry et al. (2000) suggest that ecological and genetic divergence could occur an order of magnitude faster than previously described. Following the introduction of sockeye salmon to Lake Washington, WA, a large (100,000–350,000 breeders) tributary spawning population established itself in the Cedar River. In ~13 generations a genetically distinct beach spawning ecotype colonized Pleasure Point, which is ~7 km away. The scale at which reproductive isolation can occur in both space and time is incompletely understood. Most studies on genetic structure focus on large economically important populations, which often occur in large lakes and river systems.

In Southeast Alaska most sockeye salmon production is derived from coastal lakes rather than large river drainages. The Auke Lake system is a model coastal lake system in this region. It is a small lake (71.6 hectares) with six inlet streams, of which Lake Creek is the largest (5.8 km in length with a barrier falls located 2 km from its mouth). Since 1977, escapement of spawning adults into Auke Lake has averaged 2500 sockeye salmon, a decline from an average of 7982 adults between 1963 and 1977 (Taylor and Lum 2003). Previous work identified spatial and temporal segregation in spawning aggregations that appear persistent through time (Bucaria 1968, Nelson 1993, Ray et al. 2015). Spawning occurs predominantly in two tributaries on the northern shore (Lake Creek 86.8% and Lake Two Creek 8.8%), but also occurs in an ephemeral stream on the southern shore, the outlet stream, and within the lake (Ray et al. 2015, Fig 1.1). Separated by 0.6 km of shoreline, no difference in the timing of peak spawning between the two northern tributaries has been detected. Peak spawning occurs three weeks later on

the southern shore spawning sites which are ~1 km away (Ray et al. 2015). There exists potential for genetic divergence to occur among the temporal and spatially isolated spawning aggregations of Auke Lake at a scale much smaller than observed previously.

This study examined the fine-scale genetic structure of sockeye salmon in Auke Lake, Alaska from three return years. We also used single nucleotide polymorphism (SNP) data from northern Southeast Alaska populations that were available from the genetic baseline of the Alaska Department of Fish and Game's (ADF&G; Rodgers Olive et al. 2018) to characterize Auke Lake sockeye salmon in a broader spatial context. Specifically, this research addressed the questions: (1) Is there genetic divergence between spatially or temporally segregated spawning assemblages within Auke Lake? (2) How does the genetic composition of the run over the course of the return vary within and among years?

Methods

Genetic collections

A genetic baseline has been constructed for sockeye salmon in Southeast Alaska (SEAK) that is composed of 171 populations in 9 broad-scale reporting groups (Rogers Olive et al. 2015). We evaluated 16 populations from five reporting groups defined by ADF&G in northern Southeast Alaska [(1) NSEAK – Auke Lake, Berners Bay, Lace River, Cresent Lake, Steep Creek, and Windfall Creek; (2) Taku Mainstem – Taku River, Tulsequah River, Yehring Creek; (3) Speel – Speel Lake / Snettisham Hatchery; (4) Chilkat – Chilkat Lake, Bear Flats, Mule Meadows; and (5) Chilkoot – Chilkoot Lake, Bear Creek, Chilkoot Beaches].

We exhaustively sampled the adult sockeye salmon population of Auke Lake during three return years (2008, 2009, 2011). We included 2011 instead of the 2010 return year so that we could evaluate our genetic markers for parentage inference by assigning age 1.0 (age expressed by the number of freshwater and saltwater annuli) sockeye from 2011 to parents in the 2009 return year. An axillary process was clipped from each returning adult as it passed the weir operated by the National Marine Fisheries Service (2008 n = 1264, 2009 n = 4064, 2011 n = 2427). Axillary processes were stored in 95% ethanol until dried for DNA extraction and permanent storage. Because we used

secondary sexual characteristics (size and shape of vent and kype), sex was not unequivocally determined, particularly for fish that arrived early in the return. Scales were collected opportunistically for aging from a subsample of each return (2008 n = 88, 2009 n = 221, 2011 n = 247) for which sex was determined for each individual.

DNA extraction and genotyping

For Auke Lake samples, DNA was isolated from axillary process samples with Qiagen DNeasy™ kits (Qiagen, Valencia, CA) or a proteinase K and ammonium acetate procedure (Puregene DNA™ isolation protocol - Gentra Systems, Minneapolis, MN). Isolated DNA was hydrated in TE buffer solution (10 mM Tris, 1.0 mM EDTA adjusted to pH 8.0 with HCl) and stored in 1.5 ml tubes at -20°C.

Three overlapping sets of genetic markers were used for all analyses. The 16 NSEAK baseline populations had been genotyped for 96 SNP loci by ADFG. Auke Lake collections were genotyped for both microsatellite (SSR) and SNP genetic markers. Samples from 2008 were genotyped for 110 markers (14 SSR and 96 SNPs). The SSR loci included 12 tetranucleotide loci — Oki10, Oki16, Oki1a, Oki1b, Oki29 (Smith et al. 1998), Oki100 (K. Miller, Canada Department of Fisheries and Oceans, unpublished data), One102, One109, One114 (Olsen et al. 2000), Ots100, Ots103 (Beacham et al. 1998), and Ssa419 (Cairney et al. 2000) — and two dinucleotide loci — Omy77 (Morris et al. 1996) and One8 (Scribner et al. 1996). A reduced panel of markers was used to genotype samples from years 2009 and 2011 (12 SSR and 48 SNPs). Both Oki29 and Ots100 were removed from the microsatellite loci.

Two methods were used to genotype SSR loci. Samples from 2008 were amplified in 10 uL reactions, which included ~1 unit Taq polymerase, 1X PCR buffer (50 mM KCl_2, 10 mM Tris buffer adjusted to pH 9.0 with HCl, 0.1% Triton X-100; Promega Corp., Madison, WI), 0.5 µM deoxyribonucleotide triphosphates (dNTPs), 1.5–1.875 mM $MgCl_2$, 0.025–0.1 µg DNA, 0.2–0.4 µM unlabeled forward and reverse primers, and 0.02–0.04 µM fluorescently labeled forward primer with an IRDye® infrared dye (LI-COR, Lincoln, Nebraska). The amplification profile was: 95°C for 5 min; 25–35 cycles of 0.5 min at 95°C, 0.5 min at locus-specific annealing temperatures (46°C to 60°C), and 0.75 min at 72°C; and a final extension was at 72°C for 10 min. Gel electrophoresis was performed by loading a mixture of 0.5 µL of PCR product and 0.5 µL of stop buffer (95%

formamide, 0.1% bromophenol blue) into a 0.25 mm thick 6% polyacrylamide gel (PAGE-PLUS™, Amresco, Solon, OH) on the LI-COR 4300 System™. All gels were run in 1X TBE buffer (0.09M Tris-Borate, 2 mM EDTA, pH 8.3) at 40 W, 1500 V, 40 mA, and 45°C plate temperature. Run times were determined by the fragment size of the locus and lasted between 1 and 4 hours. Size fragments were scored with SagaGT (Ver. 3.2.1, LI-COR) analysis software by comparing the fragments with either IRD700 or IRD800 standard ladders (LI-COR, Biotechnology Division) or custom-designed size ladders. Microsatellite loci for the 2009 and 2011 collections were amplified at the ADF&G Gene Conservation Laboratory (ADF&G GCL) in Anchorage, Alaska. Loci were amplified in 10 µL reaction volumes that included 0.5 units Taq DNA polymerase (Promega, Madison, WI), 10 mM Tris buffer adjusted to pH 9.0 with HCl, 50 mM KCl, 0.2 mM each dNTP, 0.06–0.20 µM primers. Fragment-size analysis was performed on an ABI 3730 capillary DNA sequencer by loading 0.5 µL PCR product, 0.4 µL of GeneScan™ 600 LIZ™ dye size standard, and 9.0 uL of Hi-Di formamide (Applied Biosystems, Waltham, MA). Size fragments were visualized and separated into bin sets with Applied Biosystems (ABI) GeneMapper software v4.0. All individuals that failed to amplify on the ABI system were rerun with the LI-COR sequencer. Size fragment conversions between the two platforms were accomplished by running all unique alleles for each microsatellite locus from 2009 and 2011 on the LI-COR systems.

Single nucleotide polymorphism loci, for the baseline and Auke Lake collections, were amplified and scored at the ADF&G GCL. The SNP markers were screened with Biomark Dynamic Arrays™ (Fluidigm, South San Francisco, CA) following the methods of Seeb et al. (2009).

Quality control

Allele frequencies and F_{IS} values were estimated with Genepop v4.6 (Roussett 2008). Suspected scoring mistakes for SSR alleles were re-genotyped and run in line-ups against alleles identical in size to confirm their identity. For SNP genotypes, 8% of the fish were re-extracted and re-genotyped. Discrepancy rates, the proportion of conflicting genotypes relative to the total number of genotypes examined, were calculated.

Genotypic data from each collection were tested for conformance to Hardy-Weinberg proportions (HWP) at individual loci and for linkage equilibrium (LE) at pairs of loci with Genepop v4.6 (Roussett 2008). For HWP tests, the Markov chain method was used with a dememorization of 10,000 and 5,000 batches of 100,000 iterations. The log-likelihood ratio statistic was used to evaluate linkage disequilibrium with a 10,000-step dememorization followed by 1,500 batches of 10,000 iterations. The false discovery rate correction method (Benjamini and Yekutieli 2001) was applied to HWP and LE tests to correction for multiple testing.

We took genetic samples from the entire return of sockeye salmon that passed through the Auke Creek weir in return years 2008, 2009, and 2011. Because of this exhaustive sampling, we anticipated that full-sib families would be present and could produce departures from HWP and linkage equilibrium. Full sibling family groups were reconstructed with the program COLONY and families that had high probabilities of not being split ($P_{Exc} > 0.80$, where the P_{Exc} is the probability that all the full siblings deduced for a family are actually full siblings and that no other individuals included in the sample are full siblings with individuals in the deduced family) were sampled for HWP testing.

Spatial genetic structure of northern Southeast Alaska sockeye salmon

We evaluated the genetic relationship of 16 populations from the sockeye salmon baseline from NSEAK. We constructed an unrooted tree with the restricted maximum likelihood method (ContML) in the program Phylip (Felsenstein 1989). We searched for the best tree by randomizing the input order of populations 200,000 times. Support for each node in the maximum likelihood tree was evaluated by bootstrapping the dataset over loci 100 times and producing a majority rule consensus tree in Phylip (Felsenstein 1989). The tree was visualized and edited in the program FigTree (A. Rambaut: https://github.com/rambaut/figtree/). A principal component analysis (PCA) was conducted on the correlation matrix of centered allele counts for each individual. The correlation matrix-based approach standardizes the variables such that each variable has a similar scale. For this and succeeding PCAs, the adgenet package (Jombart 2008, Jombart and Ahmed 2011) was used for scaling and centering and the ade4 package (Chessel et al. 2004, Dray et al. 2007, Dray and Dufour 2007) was used to perform the PCA. Pairwise

F_{ST} values among populations were estimated with the Hierfstat package (Goudet 2005) in R. Significance of the pairwise values was evaluated by performing bootstrapping over loci. Homogeneity tests (pseudo-exact G) and a population-based isolation by distance (IBD) analysis were conducted with Genepop v4.6 (Roussett 2008) with linear geographic distances estimated as the minimum distance between baseline collections connected by water. The significance of the slope (m) of the IBD analysis was evaluated with a Mantel test (Mantel 1967) between the geographic and genetic distance with 100,000 permutations.

Sockeye salmon straying into Auke Lake

In order to determine if alleles were drawn from the same distribution in all Auke Lake collections (the three years sampled in this study and the ADF&G baseline population sampled from Lake Creek spawners in 2013), we conducted homogeneity tests and computed pairwise F_{ST} as described for the baseline population comparisons.

To visually evaluate the correspondence among multilocus genotypes among the Auke Lake collections and the NSEAK baseline collections, a PCA was conducted on the covariance matrix of centered allele counts for each individual. The covariance matrix-based approach incorporates the scale of variation of the variables into the analysis. Each allele at bi-allelic SNP loci inherently has the same variance, which should only affect the scale of the analysis. Asymptotes in the scree plot determined the number of principle components retained in each analysis. The first and second principal components for each individual were plotted. Putative strays were identified as fish collected at the Auke Creek weir whose PC scores clustered them with an alternative NSEAK baseline population.

We identified the probable sources of each putative stray with a genetic stock identification approach (GSI) in which the putative strays comprised a mixture and the SEAK collections was the baseline. Because we used a subset of both the baseline collections and genetic markers, we conducted simulations in ONCOR (Kalinowski et al. 2007) to estimate the potential accuracy of mixture analysis. We conducted leave-one-out (LOO) and 100% simulations to evaluate our reduced NSEAK sockeye baseline composed of 16 populations in eight reporting groups genotyped at 41 SNP loci. Putative

strays were assigned back to their stock of origin with the program BAYES (Pella and Masuda 2001). We ran four chains of 1,500,000 MCMC samples with a thinning interval of 10 for both the stock proportion and baseline allele frequencies and a thinning interval of 100 for individual assignment. The Dirichlet prior on stock proportions was uniform. The Gelman and Rubin (Gelman and Rubin 1992) shrink factor was used to evaluate convergence among the chains. Assignment of individuals can occur even when their population of origin is not included in the baseline. To consider the possibility of omitting a population from the baseline, the z-statistic (log-likelihood accounting for missing data) for each fish to the population with the highest posterior probability was plotted. This was accomplished with the program *rubias* (Moran and Anderson 2019). The z-scores should approximate the normal distribution if the stock of origin for all of the fish is represented in the baseline. Individuals identified as strays into Auke Lake were removed from analyses that considered the inter- and intra-annual genetic structure of Auke Lake sockeye salmon.

Genetic structure within Auke Lake

A centered PCA was conducted on the correlation matrix of allele counts for individuals in each return year with the full dataset (microsatellite and SNP loci). Outliers were identified with Hotelling's T^2 statistic (Hotelling 1931) and the BY false discovery rate correction (Benjamini and Yekutieli 2001) was used to account for multiple tests. Microsatellite genotypes of apparent outliers were reconfirmed by additional genotyping.

Clustering programs can be useful tools for detecting population structure. The program *structure* (Pritchard et al. 2000) was used to cluster individuals into K subpopulations. We conducted 10 replicate runs of the admixture model with the correlated allele frequencies (Falush et al. 2003), and sampling location, or in this case sampling date, as a prior to aid in clustering (LOCPRIOR model; Hubisz et al. 2009). We used a 100,000 step burn-in and 1,000,000 samples from the posterior distribution. The program Clumpp (Jakobsson and Rosenberg 2007) was used to deal with label switching among the 10 replicate runs (i.e., results that only differed in the population label designated to a given group of individuals in the program *structure*) and the program Distruct (Rosenberg 2004) was used to graph the relabeled individual membership

coefficients (Q_i values), from Clumpp. Under the admixture model, the Q_i values represent the proportion of an individual's genome that originated from each of the K subpopulations. In addition to visually inspecting the Q_i plots, the mean log-probability of K (Ln[P(K)]) and DeltaK, the rate of change in Ln[P(K)] between successive K values (Evanno et al. 2005) were evaluated.

The presence of family groups can influence the results from unsupervised Bayesian clustering algorithms (Anderson and Dunham 2008, Rodriguez-Ramilo and Wang 2012). To evaluate the effect of full-sibling (FS) family groups in our temporal *structure* analysis, we conducted FS inference with the program Colony (Jones and Wang 2010) in which we used the full likelihood model, medium precision and run lengths. Subsequently we repeated the structure analysis on data sets that resulted from removing all but a single member from each FS family. Only families for which the probability of exclusion exceeded 0.5 were included to prevent inclusion of related individuals. The sample sizes for each year were reduced; however, individuals from throughout the return period were represented in each year (2008 n = 789, 2009 n = 1539, 2011 n = 914).

To evaluate the effect of including return timing as a prior to assist in clustering with *structure* (LOCPRIOR model), the parameter r, which describes the amount of information carried by return timing, was evaluated for each run. Values of r less than or equal to 1 would indicate that return date is important, while values much larger than 1 would indicate no structure, or clustering independent of return date.

To test for isolation by time (IBT), which could arise if natural variability in return timing is coupled with a propensity to mate with individuals that have similar return timing, we tested for a relationship between temporal distance and genetic distance with the program Genepop v4.6 (Roussett 2008). We first conducted a test of IBT among the return days that had five or more individuals for each of the three years with a linear temporal scale. The significance of the slope of the regression was evaluated with a Mantel test with 200,000 permutations. For the individual-based approach, we tested for a significant relationship between genetic distances [$\hat{a}$ (Rousset 2000) and $\hat{e}$ (Watts et al. 2007)] and return timing with a linear temporal scale. Significance of the slope (m) of the regression was evaluated with a Mantel test with 1,000,000 permutations. It has been suggested that within a breeding season, the temporal dispersal of male Pacific salmon

should exceed that of females (Fleming and Reynolds 2004, Hendry and Day 2005), so we repeated the analyses with females only. Mean dispersal distance was estimated in years for which a significant signal of IBT among return days was detected. The inverse of the slope of the regression of pairwise $F_{ST}/(1-F_{ST})$ provided an estimate of neighborhood size ($D\sigma^2$), the product of population density and mean squared parent-offspring axial distance. The density was estimated as the average number of fish per day of the return for each return year. If we assume a normal dispersal function, an estimate of the mean parent-offspring dispersal distance in a linear habitat is $\mu_x = \sigma * (2/\pi)^{1/2}$ (Puebla et al. 2009) in which σ is the square root of the mean squared parent-offspring axial distance. Confidence intervals for the mean parent-offspring dispersal distance were made from the approximate bootstrap confidence intervals of the slope of the regression.

Autocorrelation of genotypes can develop under restricted gene flow (Turner et al. 1982, Epperson 1995a; 1995b). While IBT tests for a correlation between genetic and temporal distance, autocorrelation analyses test for autocorrelation between genotypes in collections of distance classes. We used the multivariate method developed by Smouse and Peakall (1999) and implemented in the R package PopGenReport (Adamack and Gruber 2014). We used 30 bins to define our distance classes. The significance of the autocorrelation coefficient was evaluated by permuting the genetic distance matrix 1,000 times to create a null distribution of no temporal structure.

Alternative, and potentially more sensitive methods to detect subtle genetic structure, apply an allele-based approach to test whether the distances between the co-occurrence of alleles are distributed at random (Kelly et al. 2010). The spatial analysis of shared alleles (SAShA; Kelly et al. 2010) was used to measure the dispersal of alleles by comparing the observed distribution of alleles by return day with the null distribution expected under random migration. The significance of the test was determined by permuting the observed allele by return-day matrix, maintaining row and column sums, and recalculating the statistic 1000 times to create a null distribution. Each of the return years was analyzed separately. Putative sex of fish was included to see if trends in allele co-occurrence differed between sexes.

Instead of focusing on trends over the entire run duration or distance classes, we focused on individual return days to test the hypothesis that individuals in a given return

day are more related to one another than expected in a random mating population. Mean relatedness estimates (Identity) were computed for each day of the return with the program Identix (Belkhir et al. 2002). If there were fewer than 10 individuals in a return day, they were grouped with contiguous days. The null distribution was constructed by resampling genotypes 1000 times without replacement. If the mean Identity for the day exceeded 95% of the null distribution, individuals are more related to one another than what would be expected under random mating. If the variance of pairwise relatedness coefficient for the day exceeded 95% of the null distribution, multiple independent groups of related individuals may have been sampled.

To detect past demographic changes, we applied two methods to test for genetic bottlenecks. Both methods take advantage of the fact that rare alleles will be lost during a population bottleneck. The method of Cornuet and Luikart (1996), implemented in the program BOTTLENECK (Piry et al. 1999), relies on transient excess in heterozygosity. Because rare alleles contribute little to heterozygosity, allelic diversity is reduced faster than heterozygosity during a bottleneck. Given the observed number of alleles and sample size, the observed heterozygosity from the sample is compared to the expectation under mutation-drift equilibrium generated from coalescent simulations with a given mutational model. We performed analyses with the step-wise mutational model (SMM) and the two-phase mutational model (TPM) with 95% single step mutations and 5% multi-step mutations and a variance among the multiple step mutations of 12 (as recommended by Piry et al. 1999). These two parameterizations of the model are reasonable considering that the majority of our loci are tetranucleotide repeats and only 1% of mutations are multi-step for tetranucleotide microsatellites compared to 32% for dinucleotide loci (Sun et al. 2012). Each year was run individually; also, a random sample of 5000 fish from all three years was run. The significance of deviations from mutation-drift equilibrium was assessed with the Wilcoxon's signed rank test because it may be the most powerful and robust when a moderate number of loci are genotyped (Piry et al. 1999).

A second method, the M-ratio test (Garza and Williamson 2001) relies on the expectation that during a bottleneck the number of allelic states is reduced faster than the size range of those alleles, which results in an M-ratio (number of states divided by size

range) that is smaller than that expected for a population under mutation-drift equilibrium. The M-ratio may recover more slowly than heterozygosity and be more sensitive to detecting bottlenecks that occurred many generations ago (Garza and Williamson 2001). Coalescent simulations are used to create a distribution of M-ratio values expected under mutational drift equilibrium. We performed simulations with the proportion of non-one-step mutations (1-p_s) equal to 0.05, the average size of non-one-step mutations (Δ_g) set to 3.1, and pre-bottleneck θ ($4N_e\mu$) set to 24. The pre-bottleneck θ value was estimated with an average mutational rate from pink salmon of $8.1*10^{-4}$ (Steinberg et al. 2002, adjusted by Panagiotopoulou et al. 2017), which is similar to the tetra-nucleotide mutation rates reported in humans (10.01×10^{-4}; Sun et al. 2012) and an N_e of 7400. The estimate of N_e was based on a 5.3-year generation time estimated from scale ages, an average pre-bottleneck return of 4500 fish, and an N_e/N_c ratio of 0.31 (average estimates for *Oncorhynchus* from Frankham 1995). The type I error rate of bottleneck tests is highly sensitive to assumptions of the mutational model used (Peery et al. 2012) so we evaluated the sensitivity of our results to variation in θ, 1-p_s, and Δ_g by performing analyses with values of θ ranging from 2.4 to 24, 1-p_s ranging from 0 to 0.4, and Δ_g ranging from 0 to 4. Because of prohibitively long run times, each sensitivity analysis was performed with a subsample of 1000 individuals from each return year as well as all three years combined. The effect of subsampling each return year has the potential to induce the very signal used to detect a bottleneck, so we conducted three replicate runs of each analysis to ensure any bottleneck detected is not an artifact of reducing the dataset.

Results

Hardy-Weinberg and linkage equilibrium

The ADF&G baseline for northern SEAK is composed of 96 SNP markers; three that are mitochondrial loci — physically linked and haploid were removed from these analyses. Of the 1416 tests for departures from HWP, two loci (One_lpp1-44 in Auke Lake and One_c3-98 in five populations) remained significant after BY adjustment for multiple comparisons (Benjamini and Yekutieli 2001). Of 4278 tests for linkage equilibrium between loci pairs across all populations, four pairs of loci (One_MHC2_190

and One_MHC2_251, One_GPDH and One_GPDH2, One_TF_ex10-750 and One_TF_ex3-182, and One_RF-112 and One_RF-295) remained significant after BY adjustment (Benjamini and Yekutieli 2001). We removed One_lpp1-44, One_c3-98, One_MHC2_190, One_GPDH2, One_TF_ex3-182, One_RF-295 from the ADF&G baseline collections for all of our analyses.

The 2008 collection showed deviations from HWP at 18 loci; one remained significant after BY correction. In collections from 2009 and 2011, six loci deviated from HWP; a single locus in 2009 and two loci in 2011 remained significant after BY correction. The single locus that deviated from HWP for all collections was One_lpp1-44, which also was significant in the ADF&G baseline collections. We observed significant linkage between 141, 98, and 115 locus pairs, of which four, five, and three locus pairs remained significant after BY correction in collections 2008, 2009 and 2011, respectively. Two pairs of loci were linked in all three collections: One_MHC2_190 with One_MHC2_251 and One8 with One_U1010-81. One locus pair, not genotyped in the 2008 collection or the ADF&G baseline collections deviated from linkage equilibrium in both 2009 and 2011 collections; One_GPH-414 with One_GTHa. Loci One_lpp1-44, One_MHC2_190, One_GTHa, and One_U1010-81 were removed from subsequent analyses.

The successful genotyping rate varied by locus and year. For all loci and years, we failed to amplify 2.02% of the genotypes. One SNP locus had especially high rates of failure (One_U1012-68 = 7.24%). The discrepancy rate for SNP loci was 0.31%. Microsatellite markers had higher successful amplification rates over all three years (99.5–100%) than SNP markers (92.8–98.1%).

Spatial genetic structure of NSEAK sockeye salmon

Pairwise F_{ST} values between collections in the ADF&G baseline ranged from 0.0007 to 0.2585 (Table S1.1). The two most dissimilar collections were Auke Lake and Windfall Lake. Only a single pairwise comparison, between Chilkoot Lake and Chilkoot Beach spawners, was not significant (bootstrap values overlapping 0). Average pairwise F_{ST} values between Auke Lake and all other baseline collections were the highest ($\bar{F}_{ST}$ = 0.2249, sd = 0.0183). The maximum likelihood tree was similar to the results of the

pairwise F_{ST} estimates; Auke Lake differed substantially from all other collections (Fig. 1.2). The most closely related group of populations was that of the Chilkoot drainage. Similarly, the two other collections on the Juneau, AK road system (Steep Creek and Windfall Lake) formed a group distinct from the other collections. Results from the PCA paralleled those from the maximum likelihood tree. The first principal component separated Auke Lake from all other collections (Fig 1.3A). When the Auke Lake collection was removed from the PCA, three groupings emerged. The first principal component separated the three Chilkoot collections from the Chilkat, Taku, and Berners Bay collections while the second principal component separated Steep Creek and Windfall Lake from all other collections (Chilkat, Taku, Berners Bay, Speel and Crescent Lake). Generally, collections clustered by drainage as was observed in the maximum likelihood tree, but there was no general trend of grouping by geographic proximity of the drainages.

We did not detect any signal of isolation by distance among the SEAK baseline populations ($P = 0.940$). The largest genetic distances were between Auke Lake and the watersheds closest to it, which resulted in a negative slope of the regression of genetic distance on geographic distance ($m = -2*10^{-4}$). When Auke Lake was removed from the analysis there was still no significant relationship between genetic and geographic distance ($m = -4.7*10^{-5}$, $P = 0.798$).

Sockeye salmon straying into Auke Lake

Overall tests of homogeneity comparing Auke Lake collections (ADFG baseline, 2008, 2009, and 2011) at 41 SNP loci indicated that alleles were not drawn from the same distribution in all four samples ($X^2 = 336.34$, df = 82, $P < 10^{-5}$). Pairwise tests revealed that differences in allele frequencies existed among all return years. The observed genetic divergence among Auke Lake temporal samples was predictably much less than that among the SEAK baseline collections. The largest pairwise difference was between the 2008 and 2011 collections ($F_{ST} = 0.001$, bootstrap values 0.001–0.003).

We identified putative strays into Auke Lake by conducting PCA that included the ADF&G baseline populations and the three years of complete sampling. The first principal component separated all four temporal samples of Auke Lake (2008, 2009,

2011, and the ADF&G baseline collection from 2013) from all other collections. Thirteen putative strays into Auke Lake (individuals that clustered with other baseline collections; Fig 1.3B) were identified: eight fish in 2008, four fish in 2009, and one in 2011.

A genetic stock identification (GSI) approach was used to evaluate the probability that fish were strays into Auke Lake. We first tested the ability of the 41 SNPs to resolve differences among the NSEAK sockeye populations. We retained the Taku mainstem, Speel Lake / Snettisham Hatchery, Chilkat, and Chilkoot reporting groups. We redefined the NSEAK reporting groups used by Gilk-Baumer et al. (2015) into five separate reporting groups: (1) Auke Lake; (2) Juneau Road system – Steep Creek and Windfall Lake; (3) Berners Bay – Berners River and Lace River; (4) Speel Lake / Snettisham Hatchery; (5) Crescent Lake. The reduced baseline appeared to delineate stocks within the study area. Results of the 100% simulations in ONCOR yielded population estimates ranging from 0.7076 to 0.9999 and reporting group estimates of 0.9622 to 0.9999. Simulations composed of Auke Lake were the most accurate while those from the Taku River were the least accurate. Leave-one-out (LOO) assignment tests that evaluated the ADF&G baseline were similarly optimistic about our power to assign fish that originated in Auke Lake back to their natal system. Nearly all the Auke Lake fish assigned correctly back to Auke Lake in LOO tests (99.4%). One fish from the Auke Lake collection was assigned back to Speel Lake. Assignment accuracy of other collections ranged from 24.2% for the Taku River to 74.4% for the Steep Creek collection. The largest misidentifications were predominantly into populations in the same reporting groups. Average assignment accuracy increased from 61.5% to 78.8% when assignment was to reporting group instead of population. The proportion of baseline individuals correctly assigned to their reporting group ranged from 53.2% to 99.4%. The lowest assignment was to the Taku River, where the largest misidentification was into the Chilkoot reporting group. Two fish from other collections (Chilkat Lake and Yehring River) were assigned back to Auke Lake.

Results of individual assignment from ONCOR and BAYES supported the idea that most of the 13 Auke Lake PCA outliers were strays from other watersheds. Seven and six individuals were assigned back to baseline populations with probability greater than 90% from ONCOR and BAYES, respectively (Table 1.1). Six fish assigned likely

originated from Speel Lake / Snettisham Hatchery ($\bar{P}$ = 0.95 and $\bar{P}$ = 0.94 for ONCOR and BAYES). The remaining fish had high support for assignment to the Berners Bay reporting group from ONCOR (P = 0.96), but only showed moderate support for assignment from BAYES (P = 0.65). Of the six remaining fish, five assigned to Speel / Snettisham Hatchery and one assigned to the Taku River, all with moderate support. No fish showed any support for assignment to Auke Lake. Assignments from the program *rubias* were concordant with those of ONCOR and BAYES and no fish had abnormally low z-scores. Concerned that we might be overly confident in identifying these PCA outliers from the Auke Lake collection as strays, we conducted mixed stock analysis on the full collection of individuals from each of the return years to see if any non-outliers assigned to stocks other than Auke Lake. No fish identified as a non-outlier in the PCA was assigned to a baseline population other than Auke Lake. Individuals identified as strays were removed from the datasets for subsequent analyses.

Fine-scale genetic structure of Auke Lake

Over the three-year study period, 7710 sockeye salmon returned to Auke Lake (Fig. 1.4). The return in 2008 was the smallest (n = 1264) and lasted 69 days (7/7/08–9/14/08). We observed an approximately equal sex ratio of returning fish over the whole run: 627 males and 637 females (Fig 1.4A). The 2009 return was 10 days shorter (6/30/09–8/28/09), but over the 59 days 1878 males and 2186 females (n = 4064) returned (Fig 1.4B). In 2011, 2382 sockeye salmon returned over 81 days (6/30/1–9/19/11). The return had 1106 males and 1273 females (Fig 1.4C). In each of the return years, the run timing was multimodal. In each year we observed an earlier average arrival time of males relative to females (Fig 1.2).

The PCA conducted with the 13 putative strays removed, and on the full marker panel that included both microsatellite and SNP markers for each year, showed two extreme outliers in 2008 (Fig 1.5A). The genotypes of these individuals had multiple low frequency microsatellite alleles, many of which were shared with individuals previously identified as strays with the SNP panel. After removing those individuals, a scatterplot of the principal components showed a large central cluster of individuals and a diffuse cloud of individuals with large positive PC1 scores (Fig 1.5B). These individuals were

primarily from 12 full sibling families (see below). The PCA of collections from 2009 and 2011 revealed a single cluster of individuals.

Consistent with the PCA results, two distinct groups were identified with *structure* in only the 2008 collection. These two groups were present after accounting for the existence of full-sibling groups in each year, therefore only the results for the sibling-culled analyses are presented. The mean LnP(K) peaked at K=2 for the 2008 collection as did the deltaK value. For both 2009 and 2011, the mean LnP(K) was largest for K=1. The inclusion of return date had little effect on the clustering of individuals. The mean value of r among all years was 15.10 (± 3.54). There was, however, limited evidence of an association between Q_i value and run timing. The distinct group of individuals (n = 22), identified from the 2008 return year was present late in the return (Fig. 1.5C). These individuals, mainly from 12 full-sibling families, belonged to the diffuse group previously identified in the PCA (Fig. 1.5b). The first three alleles from two microsatellite loci, which were omitted from the panel used to genotype the 2009 and 2011 collections, accounted for 8% of the variance in the first principal component. A PCA for 2008, which included only the loci genotyped in the other two collections exhibited the same groupings, although the separation was not as pronounced between clusters. When *structure* inference was made on all three years combined, only 15 fish had more than 50% of their genome assigned to the minority cluster: 10 from 2008, 5 from 2009, and 1 from 2011.

We evaluated the relationship between genetic distance and the number of days between return timing of individuals. First, we conducted group-based tests of IBT for each return day within a year. We observed a significant correlation between genetic distance (F_{ST}/1-F_{ST}) and the number of days between the return groups in years 2008 (m = 0.0003, $P < 0.00001$) and 2011 (m = 0.00008, P = 0.00062; Fig. 1.6). A positive slope was observed for the 2009 collection, but it was not significant (m = 0.00006, P = 0.0714). Analyses that included only females did not change the results for collections from 2008 and 2009; however, the correlation in 2011 was not significant (P = 0.2590). Results of individual- based tests of IBT were similar to those of the group-based tests. The slopes of the regressions of genetic distance ($\hat{a}$) on number of days between return

were small but significant in 2008 and 2011 (Fig 1.6; 2008 – $m = 0.0007$, $P < 0.0001$; 2011 – $m = 0.0001$, $P = 0.0051$). When the $\hat{e}$ estimator was used, the slope of each correlation decreased and was significant in collections from 2009 and 2011 (2009 – $m = 2.7*10^{-6}$, $P = 0.0089$; 2011 – $m = 1.9*10^{-5}$, $P < 0.0001$). The return of sockeye salmon into Auke Lake was episodic (Fig 1.4): the majority of pairwise comparisons were between fish that returned within three days of one another and eight to ten days apart in 2008; between fish that returned within four days of one another and 25 to 27 days apart in 2011; and between fish that returned within four days of one another and 23 to 26 days of one another in 2009. Individual-based analyses of females were concordant with the results from the $\hat{a}$-based analyses for pooled sexes (Table 1.2).

We estimated the dispersal rate from the slope of the regression line of group-based analyses of IBT and mean dispersal distance (in days) for years with a significant signal of IBT (2008 and 2011). The density, average number of fish returning per day, was 22.57 for 2008 and 53.84 for 2011. The dispersal rate (σ^2) for the 2008 return year was 837.13 fish*day (411.14 – 4034.92) and was 3130.98 fish*day (1638.70 – 13007.22) for the 2011 return year. Assuming a normal dispersal function, we estimated the mean dispersal distance (in days) of the return (μ_x) was 4.86 (3.41 – 10.67) days for the 2008 return year and 6.08 (4.40 – 12.40) days for the 2011 return year.

We considered temporal genetic structure on a smaller scale with autocorrelation analysis. There was no evidence for positive autocorrelation between genotypes of individuals in any of the years. In 2008, the average autocorrelation coefficient was -0.0042. Three bins had lower correlation coefficients than expected at random; however, all three fell just below the 95% CI. The three significant bins compared the autocorrelation of genotypes of individuals that returned to Auke Lake 37–39, 44–46 and 60–62 days apart. A similar trend was observed in both 2009 and 2011. The average autocorrelation coefficients for 2009 and 2011 were -0.0004 and -0.0085 respectively. Only one autocorrelation coefficient, 54–55 days between return, fell below the 95% lower bound in 2009. Three coefficients were significant for 2011; 44–45, 52–54, and 71–72 days between return. The 95% CIs expand when considering longer gaps between return time of fish as a result of the decreasing number of comparisons to evaluate.

With little evidence of correlation among genotypes, we used an allele-based analysis to consider temporal genetic structure at small scales. We evaluated the temporal arrangement of allelic co-occurrence with the program SAShA. In all three years, the mean distance between shared alleles did not differ from what we would expect at random. Separating each year by putative sex and reanalyzing the data also yielded statistically non-significant results.

We observed no strong evidence of relatedness among days of the return, so we considered relatedness within a day of the return for the entire dataset. Tests for some degree of relatedness among the members of a return day against the alternative of complete un-relatedness as we would expect from a random sample of a panmictic population produced equivocal results. Mean Identity (and variance) among the day groupings was 0.6735 (0.002) in 2008, 0.6274 (0.0046) in 2009, and 0.6398 (0.0029) in 2011. After adjusting for multiple testing, only a single test of mean identity remained significant in the 2008 collection. In each of the three years, only a single test of variance remained significant after adjusting for multiple tests.

Although the contribution of different brood years overlapped among the Auke Lake collections, tests of homogeneity that included the full marker panel suggested that alleles were not drawn from the same distribution in all return years ($P < 10^{-5}$). Scale data showed that seven age classes comprised of four brood years made up each yearly return. The two primary age classes were 2.2 and 2.3 fish (ages five and six). Over the three years sampled we observed a reduction of the 2.2 age class from 63.6% in 2008 to 31.2% in 2011 and an increase in the 2.3 age class from 17.05% in 2008 to 56.7% in 2011.

Tests for demographic changes within the population exhibited contrasting results: expected heterozygosity-based tests suggested the population is expanding, whereas M-ratio tests suggest that the population has undergone a genetic bottleneck. For both the mutational models (SMM and TPM), we observed a significant deficiency of heterozygosity as compared to a population under mutation-drift equilibrium. The TPM is a more conservative model for inferring a population expansion, so those results are presented. For the 2008 collection, 10 of 14 loci showed heterozygosity deficiency ($P = 0.001$). For the 2009 and 2011 collections, 8 of 12 loci showed heterozygosity deficiency ($P = 0.046$ and $P = 0.017$ respectively). The same eight loci were significant in 2009 and

2011, of which seven were significant in the 2008 collection. The M-ratio in Auke Lake was significantly smaller than expected under mutation-drift equilibrium for both the SMM and TPM indicating that the population has undergone a genetic bottleneck. Inferences made from the M-ratio test were sensitive to variation in θ, 1-p_s, and Δ_g (Fig. 1.7). In general, a bottleneck was detected if $\Delta_g < 4$ and 1-$p_s < 0.4$. The average standard deviation in the M-ratio across replicate samples was 0.01, which suggested that the signal that we detected was not entirely caused by subsampling the data.

Discussion

Spatial divergence

The genetic composition of Auke Lake sockeye salmon differed substantially from other NSEAK sockeye populations, despite apparent straying into the system. Based on the ADF&G GCL SNP baseline, Auke Lake was genetically distinct from other populations considered, which is consistent with results obtained from an allozyme study (Guthrie et al. 1994). Similar to Nelson et al. (2003), we observed a mosaic pattern of genetic structure in NSEAK, but with some regional groupings.

The distinctiveness of Auke Lake allowed us to identify sockeye salmon that were strays into the system in three return years. The majority of fish that were identified as strays were from Speel Lake / Snettisham Hatchery (Speel Lake was used as broodstock for the Snettisham Hatchery), which is 66 km south of Auke Lake. No migrants originated from the two closest watersheds: Windfall Lake and Steep Creek. Bett and Hinch (2015) suggest that straying could be a result of sockeye salmon being attracted to conspecifics when imprinted natal cues are weak or absent. Windfall Lake and Steep Creek may be close enough to Auke Lake that natal cues for all three watersheds are present all along a large part of the migratory pathway. Tagging studies on pink (*O. gorbuscha*) and Chinook salmon (*O. tshawytscha*) showed that Icy Strait is the primary migratory pathway used to access the inside waters of northern Southeast Alaska, although a small proportion of the southern populations may use a pathway into Frederick Sound (Hoffman 1982, N. Frost ADF&G per. communication). It is likely that sockeye salmon use these same migration corridors. The 1763 sockeye salmon that escaped to the Speel weir in 2008 was the third lowest recorded since 1983; for that time

escapement averaged 6952 fish. The extent of straying from Speel Lake / Snettisham Hatchery might be inversely correlated with the size of the escapement. Irrespective of the population of origin of the migrants, Auke Lake is distinct from its neighboring populations. If the stray fish failed to produce offspring, gene flow would be zero. Contrarily, these fish could have tried 'backing-out' to salt water, a behavior observed by Eiler et al. (1992), but were prevented by the weir.

Auke Lake is a relatively young system with a small and variable population size. The lake was formed by isostatic rebound around 6000 years ago (Connor and Monteith 2006). Given a 5-year generation time, Auke Lake sockeye salmon would be at most 1200 generations old. The small size of the population and variable access to quality spawning habitat make it susceptible to genetic drift although the overlapping generations should buffer against single year-class failures. The Chilkat and Chilkoot watersheds are characterized by long river systems that include large lakes. Escapement for both watersheds for the years included in the study averaged 44,245 and 97,193 sockeye salmon, respectively. We do not have estimates for each of the spawning aggregations within these two watersheds but assume that they are large and there is appreciable gene flow among populations within each of the watersheds because of their low pairwise F_{ST} values and clustering in PCA and the ML tree. Similarly, the Taku River system is a long transboundary river that often has escapements in excess of 70,000 fish through the lower river; however, finer scale escapement estimates are not available. Escapement into Speel Lake has ranged from 299 to 18,095 fish between 1983 and 2016, but far more have been produced at the Snettisham Hatchery. At the Snettisham Hatchery, the current permit allows for 12,500,000 eggs to be fertilized and 9 million smolts to be released. Steep Creek is a short creek that feeds a large glacial lake; its annual escapement averages 3379 fish. Windfall Lake is a small shallow lake that has an average escapement of 2426 sockeye salmon. While it is the most similar to Auke Lake, it differs substantially in access. Entry to Windfall Lake is through a 1 km long low gradient stream up to 2 m deep that flows to a series of glacially occluded rivers 6 km in length. Entry into Auke Lake is through a high-gradient creek, rising 17 m over 0.48 km where the depth is often less than 20 cm. In Auke Lake, without adequate rain, access to its primary spawning tributary, Lake Creek, can be cutoff entirely (Nelson 1993, Ray et al. 2015). Limited

access could occur for only a few days, or potentially much longer. In the absence of access to the primary tributary, fish would likely elect to spawn in low quality lake beach habitat or other tributaries. Large swings in the number of breeders could drive short-term changes in allele frequencies (Waples and Teel 1990). Long-term loss of genetic variability due to genetic drift is a function of the effective population size. It has been shown that in Pacific salmon populations that are semelparous with overlapping generations, the effective population size is approximately the effective number of breeders (N_b) times the generation length (Hill 1979, Waples 1990, Adkison 1995). The overlap in age structure of Auke Lake sockeye salmon should thus provide some buffer against long-term loss of genetic variation provided the effective number of spawners remains large.

In the past 100 years, the number of spawners that have escaped to Auke Lake may have been substantially reduced by the presence of two canneries next to Auke Lake in the early 1900s (Bower and Aller 1917, Bower 1926). Despite operating for only eight years, the presence of a stationary impoundment net near Auke Lake could have dramatically depressed the run size for an entire generation. Results from the M-ratio suggest that there may have been a historic bottleneck that could explain the genetic distinctiveness of Auke Lake. Inferences made from the M-ratio were sensitive to parameter values; however, the results were consistent under the most likely values. If a historic bottleneck did occur, it appears that heterozygosity has rebounded faster than the M-ratio. The significant deficiencies of heterozygosity that we observed could be indicative of a more recent population expansion (Cornuet and Luikart 1996). Weir records from the last forty years show a recent reduction in both smolt outmigration and adult immigration through time (Taylor and Lum 2003), but it is possible that there has been an increase in the effective population size despite a reduction in census size. The signal of an expanding population may also be a result of violations of the assumptions of no substructure and no recent immigration. Substructure that is not accounted for in the population can lead to a deficiency of heterozygosity through the Wahlund effect. Generally, isolation by distance has been shown to increase the probability of erroneously detecting a population expansion (Leblois et al. 2006), and isolation by time in theory would act similarly. While we have accounted for first generation migrants into Auke

Lake, if there are hybrid offspring, the number of rare alleles could be increased without noticeably affecting heterozygosity mimicking an increase in population size (Cornuet and Luikart 1996, Paz-Vinas et al. 2013). The magnitude of the divergence between Auke Lake and other neighboring watersheds could compound this effect, because many of the introduced alleles would be rare. Based on the two outliers in 2008 that shared low frequency microsatellite alleles with fish identified as strays, we have reason to believe that first-generation hybrids exist in the 2008 collection; however, the signal of population expansion was consistent in all return years.

In addition to the effects of genetic drift or a severe bottleneck, there is also an uncertain history of transplantation into the system. In the 1920s, A. J. Sprague, a hatchery superintendent, introduced sockeye salmon eggs from Afognak Lake on Kodiak Island to Auke Lake (Roppel 1982). The success of these transplants was considered low (Guthrie et al. 1994) but it could have drastically reshaped the genetic composition of the system. Moreover, local traditional knowledge from the Auk'w Kwaan (Tlingit) people posits that Auke Lake sockeye salmon originated from the transplantation of mature male and female sockeye salmon from the Chilkat River (Steve Langdon, UAA personal communication). Ethnographic research suggests that Pacific salmon translocation among the Tlingit may have been a typical practice (Thornton et al. 2015). An anthropological description of the social structure and life of the Tlingit in Alaska suggested the Auk'w Kwaan were held in low esteem by other Tlingit tribes due to their lack of a highly productive sockeye salmon stream near their village, which gives some motivation for the translocation of sockeye into Auke Lake (Olson 1967). Our ability to compare the Chilkat River populations and Auke Lake is restricted to bi-allelic SNP markers in the ADF&G baseline. The presence of shared alleles at highly polymorphic microsatellite markers would have given us more insight into their potential shared ancestry. The population's origin by way of transplantation, however, is contradicted in Chief Phillip Joseph's description of the migration of the Auk'w Kwaan from the Wrangell area near the Stikine River to Auke Bay in the 15th century (Joseph 1967). He wrote,

> The old man told the people to land and he said this is going to be our new home. It didn't take them long to start building later on they built

> the first Dipper House they were busy people. They soon find the Auk Lake and they find out it was a Sockeye creek. (pg. 3–4)

The production of sockeye salmon from Auke Lake is dwarfed by the Stikine, Taku, Chilkat, and Chilkoot rivers. The small size of the sockeye salmon run may have been conveyed by Olson as if there were no sockeye near the village and efforts to boost the size of the run were made by the Auk'w Kwaan through supplementation with Chilkat River fish. Archeological evidence corroborates Chief Joseph's account that the Auk'w Kwaan inhabited the area ~1564 (Monteith 2007) and Auke Lake has experienced a long history of resource use and development that continues to present day. Historically, sockeye salmon spawned at the mouth of Hanna Creek at the northeastern end of Auke Lake; however, construction of a road in the 1960s deposited sediment in the spawning habitat (Bethers et al. 2012). Sedimentation from other construction projects could have destroyed shore spawning habitat along the northwest shore (Bethers et al. 2012).

Interannual genetic variation in Auke Lake

We detected significant heterogeneity in allele frequencies as well as a large shift in the age classes represented among the three return years that we examined. Some population genetic studies on sockeye salmon that sample multiple years from the same system have reported stable allele and genotypic frequencies (Wood et al. 1994, Ramstad et al 2004), while others have detected significant differences (Winans et al. 1996, Withler et al. 2000, Nelson et al. 2003). Typically, the variation observed among years is much less than the variation among populations (Beacham et al. 2005, Beacham et al. 2006), which is consistent with our collections as compared to the ADF&G baseline populations. In the Auke Lake system, sockeye salmon can be three to seven years old when they return to spawn. In any given year returning adults will be composed of five separate brood years. This overlap among brood years will dampen variation in allele frequencies among years (Waples and Teel 1990), although the majority of fish return at age-five or -six. Another consequence of overlapping generations is that the magnitude of change in allele frequencies is not solely determined by the effective population size, but also the age-specific survival and birth rates in addition to the effects of N_e (Jorde and

Ryman 1995). Since 1978, the proportion of age-two smolts has increased (Joyce et al. 2015). Age-five (2.2) sockeye salmon that returned to Auke Lake in 2008 would be from the 21,007 smolt produced in 2003, whereas age-six (2.3) sockeye would be from the 6118 age two-smolt produced in 2002. Similarly, age-five sockeye salmon that returned to Auke Lake in 2011 would be composed of survivors from the 7546 smolt produced in 2006, whereas age-six sockeye salmon originated from the 8827 age-two smolt that survived from brood year 2005 (Joyce et al. 2015). It appears that the poor return in 2006, which produced a small age-two smolt class - only 36% of the 2003 brood year size - was responsible for the substantial shift in age classes observed between 2008 and 2011. Marine survival for the 2000 and 2001 brood years appears to be responsible for the poor 2006 adult return year (Joyce et al. 2015).

Intra-annual genetic variation in Auke Lake

Within each return year, we observed persistent deviations from Hardy-Weinberg expectations. This may be attributed to Wahlund effects among age classes (Waples and Teel 1990), family groups (Castric et al. 2002), or unaccounted-for spatial or temporal structure. Waples and Teel (1990) demonstrated that for salmonid populations composed of individuals that mature at various ages, allele frequencies will fluctuate as a function of the number of breeders per year. Without age information for each individual it is impossible to disentangle these effects. Given the exhaustive sampling scheme, we undoubtedly sampled some FS families. We identified FS families and reanalyzed our collections for conformance to HWP. It appeared that sampling multiple large FS families contributed to the deviations from HWP.

Genetic mixture models demonstrated that that the majority of genetic structure in Auke Lake is a result of family structure; however, once the effect of family structure was removed, we still identified a unique population segment in 2008, 2009, and 2011. In each of these years there was a small group of individuals that entered the Auke Lake weir late in the run. This group of individuals may represent the later spawning southern shore-spawning segment identified by Ray et al. (2015). If there is a spatial component to the structure observed, the ephemeral nature of the southern shore sites likely causes substantial variation in the size and quality of the spawning habitat. The fact that Auke

Lake sockeye salmon have a long generation time and mature at multiple ages creates a buffer against extirpation of this group because it would require the habitat to be inaccessible for more than the maximum generation time (seven years) and a failure in any given return year would be buffered by previous or successive brood years. Mixture models showed that there is a clear temporal component of genetic structure in Auke Lake; future work should collect spawned individuals in the southern shore habitats to confirm the potential spatial component as well. The consistency between the PCA and mixture models suggests that the identification of the genetically distinct group in 2008 was not the result of the algorithms tendency to overestimate genetic structure when the true genetic variation is characterized by isolation by distance or time (Frantz et al. 2009). Despite the consistent presence of this genetically similar late return group, there may be no spatial component and the temporal pattern observed may have arisen through heritability in return timing and temporally limited dispersal of genetic variation.

Return/spawn timing has a high heritability in Pacific salmon (0.29–1; Smoker et al. 1998, Quinn et al. 2000) and consequently temporal structure may accumulate within a population over time (see for example, Gharrett et al. 2013). Isolation by time has been observed in a number of sockeye salmon populations characterized by variation in breeding times (Woody et al. 2000, Fillatre et al. 2003, Ramstad et al. 2004, Hendry et al. 1995, Hendry and Day 2005). While previous studies relied on sizable sampling intervals, our ability to sample the entire population at entry into the system allowed us to analyze daily changes. We observed isolation by time in both 2008 and 2011 for both return day-based and individual-based analyses. There was also a positive but non-significant slope in 2009 for the return day based analysis and $\hat{a}$ individual analysis, but a significant slope for the $\hat{e}$ individual-based analysis. Simulations suggest that despite being asymptotically biased, the $\hat{e}$ estimator performs better when dispersal is large relative to the distances between individuals (Watts et al. 2007). In 2009, it did appear that dispersal could be larger relative to other years and relative to the duration of the return as observed in the tri-modal distribution of fish returning (pairwise comparisons between fish returning ~50 days apart); however, $\hat{e}$ and return day-based analyses on only females resulted in negative slope estimates. Autocorrelation analyses supported IBT analyses and suggested that the longer the time between the entrance of individuals into

Auke Lake, the more negative the autocorrelation coefficient. While only a few of the bins were significant, all years show a downward sloping trend. Prior to entering the lake, sockeye salmon stage in Auke Bay. Increases in stream flow presumably encourage fish to ascend the creek and pass the weir. Dry periods, characterized by high stream temperatures, can make the creek hazardous and difficult to negotiate. Multiple pulses in the return of sockeye salmon to Auke Lake were dictated by precipitation. Assuming a normal dispersal function, estimates of mean temporal dispersal distance were three to twelve days which were considerably shorter than the average duration of the Auke Lake sockeye return (~70 days). Adaptation by time, that is, adaptive divergence in heritable phenotypic traits, may result if isolation-by-time exists and selection changes through the reproductive season (Hendry and Day 2005). Return timing is likely related to fitness because low stream levels often prohibit access to spawning habitat and the associated high creek temperatures can be lethal. Similarly, fry emergence and migration to lake rearing habitat should match production in the lake. Fry that migrate too early will be met with low food availability and an abundance of Dolly Varden (*Salvelinus malma*) and cutthroat trout (*Oncorhynchus clarkii*) predators. Fry that migrate late will be at a competitive disadvantage with fry that migrated earlier and experienced rapid growth. Spatial isolating mechanisms may also impact these trends on a micro-geographic scale because early entrants into Lake Creek tended to spawn higher in the creek (Ray et al. 2015).

Implications

Large declines in Pacific salmon populations across their range coupled with increased hatchery production dictate the careful study of genetic changes in populations. Small populations are at an increased likelihood of random changes from drift, while gene flow from hatchery programs can alter the genetic composition over time. Defining criteria for Pacific salmon stocks of conservation concern has received much attention (Waples 1995, Allendorf et al. 1997, Wood and Holtby 1998) that typically focuses on the genetic, evolutionary, and ecological consequences of extinction of the stock. An important measure of genetic consequences hinges on the level of genetic divergence from other stocks because it may reflect adaptive genetic differences (Waples 1991).

Auke Lake is quite different from other nearby populations, but does this reflect adaptive divergence? Local adaptation is favored when the effective population size is large, that is, the selection coefficient (ϕ) is large relative the reciprocal of the effective population size ($\phi > 1/2N_e$) and the distance among migrants is small (Adkison 1995, Wood and Holtby 1998). If the N_e/N_c ratio for sockeye salmon is ~0.2 (Allendorf et al. 1997) and there has been an average yearly return of 2777 sockeye salmon to Auke Lake since 1980 and an average generation time of 5.3 years then $N_e \sim 2944$ fish. This is well above the value of 500 estimated to maintain long-term adaptive potential proposed by Franklin (1980; see also Jamison and Allendorf 2012).

The use of hatchery propagation to supplement catches to avoid overharvesting wild stocks has added a level of complexity and concern in managing wild stocks of Pacific salmon. Hatchery releases can compete effectively with, withstand higher exploitation rates than, and carry diseases harmful to wild stocks; but their most insidious effect is modifying the genetic composition of wild stocks. Genetic modifications are not as easy to detect as the immediate threats that hatchery fish pose to wild stocks, but they nonetheless can reduce fitness and productivity. Within Southeast Alaska, sockeye salmon are produced as part of the Pacific Salmon Treaty for the Stikine and Taku transboundary rivers. These enhancement activities are conducted according to the principle that they are, 'built upon a good knowledge base of existing wild stocks of salmon' (Pacific Salmon Treaty Appendix to Annex IV 1b). In order to understand what the potential effects of supplementation are, it is of paramount importance to describe how populations are structured in their absence. This study adds to our understanding of the among- and within-population genetic structure of NSEAK sockeye salmon populations.

References

Ackerman, M., Templin, W., Seeb, J., and Seeb, L. 2013. Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. Conserv. Genet. **14**(2): 483–498.

Adamack, A.T., and Gruber, B. 2014. PopGenReport: Simplifying basic population genetic analyses in R. Methods Ecol. Evol. **5**(4): 384–387.

Adkison, M. 1995. Population differentiation in Pacific salmons: Local adaptation genetic drift, or the environment? Can. J. Fish. Aquat. Sci. **52**(12): 2762–2777.

Allendorf, F., Bayles, D., Bottom, D.L., Currens, K.P., Frissell, C.A., Hankin, D., Lichatowich, J.A., Nehlsen, W., Trotter, P.C., and Williams, T.H. 1997. Prioritizing Pacific salmon stocks for conservation. Conserv. Bio. **11**(1): 140–152.

Anderson, E., and Dunham, K. 2008. The influence of family groups on inferences made with the program structure. Mol. Ecol. Resour. **8**(6): 1219–1229.

Beacham, T. D., Margolis, L., and Nelson, R.J. 1998. A comparison of methods of stock identification for sockeye salmon (*Oncorhynchus nerka*) in Barkley Sound, British Columbia. North Pac. Anadromous Fish Comm. Bull. **1**: 227–239.

Beacham, T., McIntosh, B., and MacConnachie, C. 2005. Population structure and stock identification of sockeye salmon (*Oncorhynchus nerka*) in coastal lakes in British Columbia, Canada. Can. J. Zool. **83**(6): 834–844.

Beacham, T., McIntosh, B., MacConnachie, C., Miller, K., Withler, R., and Varnavskaya, N. 2006. Pacific rim population structure of sockeye salmon as determined from microsatellite analysis. Trans. Am. Fish. Soc. **135**(1): 174–187.

Belkhir, K., Castric, V., and Bonhomme, F. 2002. IDENTIX, a software to test for relatedness in a population using permutation methods. Mol. Ecol. Resour. **2**(4): 611–614.

Benjamini, Y., and Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. **29**: 1165–1188.

Bethers, M., Munk, K., and Seifert, C. 2012. Juneau fish habitat assessment. Alaska Department of Fish and Game Regional Information Report 5J12-03.

Bett, N., and Hinch, S. 2015. Attraction of migrating adult sockeye salmon to conspecifics in the absence of natal chemical cues. Behav. Ecol. **26**(4): 1180–1187.

Bower, H.D., Ward T., and Aller, H.D. 1917. Alaska fisheries and fur industries in 1916. Document, U.S. Bureau of Fisheries.

Bower, W.T. 1926. Alaska fisheries and fur industries in 1925. Document, U.S. Bureau of Fisheries.

Bucaria, G. 1968. Growth of juvenile sockeye salmon, *Oncorhynchus nerka* (Walbaum), in Auke Lake, Alaska. Master's thesis, Oregon State University.

Burger, C., and Spearman, W. 1997. Genetic differentiation of sockeye salmon subpopulations from a geologically young Alaskan lake system. Trans. Am. Fish. Soc. **126**: 926–938.

Burgner, R. 1991. Life history of sockeye salmon (*Oncorhynchus nerka*). *In* Pacific Salmon Life Histories. *Edited by* C. Groot and L. Margolis. UBC Press. pp. 3–117.

Cairney, M., Taggart, J. B., and Hoyheim, B. 2000. Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. Mol. Ecol. **9**: 2175–2178.

Carlson, S.M., and Seamons, T.R. 2008. A review of quantitative genetic components of fitness in salmonids: Implications for adaptation to future change. Evol. Appl. **1**: 222–238.

Castric, V., Bernatchez, L., Belkhir, K., and Bonhomme, F. 2002. Heterozygote deficiencies in small lacustrine populations of brook charr *Salvelinus fontinalis* Mitchill (pisces, salmonidae): A test of alternative hypotheses. Heredity **89**: 27–35.

Chessel, D., Dufour, A., and Thioulouse, J. 2004. The ade4 package-I: One-table methods. R News **4**: 5–10.

Connor, C.L., and Monteith, D. 2006 Uncovering the past through geology and archeology. [videorecording]. University of Alaska Southeast.

Cornuet, J., and Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. **144**(4): 2001–2014.

Dittman, A., and Quinn, T. 1996. Homing in Pacific salmon: Mechanisms and ecological basis. J. Exp. Biol. **199**(1): 83–91.

Dray, S., and Dufour, A. 2007. The ade4 package: Implementing the duality diagram for ecologists. J. Stat. Softw. **22**(4): 1–20.

Dray, S., Dufour, A., and Chessel, D. 2007. The ade4 package-II: Two-table and K-table methods. R News **7**(2): 47–52.

Eiler, J., Nelson, B., and Bradshaw, R. 1992. Riverine spawning by sockeye salmon in the Taku River, Alaska and British Columbia. Trans. Am. Fish. Soc. **121**(6): 701–708.

Epperson, B. 1995a. Spatial structure of two-locus genotypes under isolation by distance. Genetics. **140**: 365–375.

Epperson, B. 1995b. Spatial distribution of genotypes under isolation by distance. Genetics **140**: 1431–1440.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. Mol. Ecol. **14**(8): 2611–2620.

Falush, D., Stephens, M., and Pritchard, J. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics **164**(4): 1567–1587.

Felsenstein, J. 1989. PHYLIP - phylogeny inference package (version 3.2). Cladistics **5**: 164–166.

Fillatre, E., Etherton, P., and Heath, D. 2003. Bimodal run distribution in a northern population of sockeye salmon (*Oncorhynchus nerka*): Life history and genetic analysis on a temporal scale. Mol. Ecol. **12**(7): 1793–1805.

Fleming, I.A., and Reynolds, J.D. 2004. Salmonid breeding systems. *In* Evolution illuminated: Salmon and their relatives. *Edited by* A. Hendry. Oxford University Press. pp. 264–294.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: A review. Genetical Res. **66**(2): 95–107.

Franklin, I.R. 1980. Evolutionary change in small populations. *In* Conservation Biology: An evolutionary ecological perspective. *Edited by* M.E. Soule and B.A. Wilcox. Sinauer Associates, Sunderland, Mass. pp. 135–140.

Frantz, A., Cellina, S., Krier, A., Schley, L., and Burke, T. 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? J. Appl. Ecol. **46**(2): 493–505.

Garza, J., and Williamson, E. 2001. Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. **10**: 305–318.

Gelman, A., and Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. **7**: 457–511.

Gharrett, A.J., Joyce, J., and Smoker, W.W. 2013. Fine-scale temporal adaptation within a salmonid population: mechanism and consequences. Mol. Ecol., **22**(17): 4457–4469.

Gilk-Baumer, S., Rodgers Olive, S., Harris, D., Heinl, S., Fox, E., and Templin, W. 2015. Genetic mixed stock analysis of sockeye salmon harvested in selected northern Chatham Strait commercial fisheries, southeast Alaska, 2012-2014. Alaska Department of Fish and Game Fishery Data Series No. 15-03.

Gomez-Uchida, D., Seeb, J., Smith, M., Habicht, C., Quinn, T., and Seeb, L. 2011. Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. BMC Evol. Biol. **11**: 48.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes **5**(1): 184–186.

Gustafson, R., and Winans, G. 1999. Distribution and population genetic structure of river- and sea-type sockeye salmon in western North America. Ecol. Freshw. Fish **8**(3): 181–193.

Guthrie, C.M., Helle, J.H., Aebersold, P., Winans, G.A., and Gharrett, A.J. 1994. Preliminary report on the genetic diversity of sockeye salmon populations from southeast Alaska and northern British Columbia. Alaska Fisheries Science Center report 94–03.

Hendry, A.P., and Day, T. 2005. Population structure attributable to reproductive time: Isolation by time and adaptation by time. Mol. Ecol. **14**(4): 901–916.

Hendry, A., Leonetti, F., and Quinn, T. 1995. Spatial and temporal isolating mechanisms: The formation of discrete breeding aggregations of sockeye salmon (*Oncorhynchus nerka*). Can. J. Zool. **73**(2): 339–352.

Hendry, A., Wenburg, J., Bentzen, P., Volk, E., and Quinn, T. 2000. Rapid evolution of reproductive isolation in the wild: Evidence from introduced salmon. Science **290**(5491): 516–518.

Hill, W. 1979. A note on effective population size with overlapping generations. Genetics **92**: 317–322.

Hoffman, S.H. 1982. Northern southeastern Alaska pink salmon (*Oncorhynchus gorbuscha*) tagging investigations, 1977-1980. Informational Leaflet, Alaska Department of Fish and Game, Ketchikan, Alaska.

Hotelling, H. 1931. The generalization of student's ratio. Ann. Math. Stat. **2**(3): 360–378.

Hubisz, M., Falush, D., Stephens, M., and Pritchard, J. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resources **9**(5): 1322–1332.

Jakobsson, M., and Rosenberg, N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. **23**(14): 1801–1806.

Jamieson, I.G., and Allendorf, F.W. 2012. How does the 50/500 rule apply to MVPs? Trends Ecol. Evol. **27**(10): 578–84.

Jombart, T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics. **24**(11): 1403–1405.

Jombart, T., and Ahmed, I. 2011. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics **27**(1): 3070–3071.

Jones, O., and Wang, J. 2010. COLONY: A program for parentage and sibship inference from multilocus gentype data. Mol. Ecol. Resour. **10**(3): 551–555.

Jorde, P.E., and Ryman, N. 1995. Temporal allele frequency change and estimation of effective size in populations with overlapping generations. Genetics **139**(2): 1077–1090.

Joseph, P. 1967. The history of the Aukquwon. Retrieved from Alaska State Library https://vilda.alaska.edu/digital/collection/cdmg22/id/2588/.

Joyce, J.E., Vulstek, S.C., Bell, D.A., Russell, J.R., Taylor, S.G., Echave, J.D., and Gray, A.K. 2015. Auke Creek research station report: Data summary for 2015 and historical trends from 1980 to 2015. Technical Report, National Marine Fisheries Service Auke Bay Laboratory.

Kalinowski, S., Manlove, K., and Taper, M. 2007. ONCOR: Software for genetic stock identification. https://www.montana.edu/kalinowski/software/oncor.html.

Kelly, R., Oliver, T.A, Sivasundar, A., and Palumbi, S. 2010. A method for detecting population genetic structure in diverse, high gene-flow species. J. Heredity **101**(4): 423–436.

Leblois, R., Estoup, A., and Streiff, R. 2006. Genetics of recent habitat contraction and reduction in population size: Does isolation by distance matter? Mol. Ecol. **15**(12): 3601–3615.

Manhard, C.V., Joyce, J.E., and Gharrett, A.J. 2017. Evolution of phenology in a salmonid population: a potential adaptive response to climate change. Can. J. Fish. Aquat. Sci. **74**(10): 1519–1527.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. **27** (2): 209–220.

McPhee, M.V, Tappenbeck, T.H., Whited, D.C., and Stanford, J.A. 2009. Genetic diversity and population structure in the Kuskokwim River drainage support the recurrent evolution hypothesis for sockeye salmon life histories. Trans. Am. Fish. Soc. **138**(6): 1481–1489.

McGregor, A., Lane, S., Thomason, M., Zhivotovsky, L., Smoker, W.W., and Gharrett, A.J. 1998. Migration timing, a life history trait important in the genetic structure of pink salmon. North Pac. Anadromous Fish Comm. Bull. **1**: 262–273.

Monteith, D. 2007. Auke Lake Trail Archeological Report University of Alaska Southeast.

Moran, B.M., and E.C. Anderson. 2019. Bayesian inference from the conditional genetic stock identification model. Can. J. Fish. Aquat. Sci. **76** (4): 551–560.

Morris, D.B., Richard, K.R., and Wright, J.M. 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. Can. J. Fish. Aquat. Sci. **53**: 120–126.

Nelson, B. 1993. Movements of radio-tagged sockeye salmon in Auke Lake, southeast Alaska. Draft Technical Memorandum, National Marine Fisheries Service Auke Bay Laboratory.

Nelson, R., Wood, C., Cooper, G., Smith, C., and Koop, B. 2003. Population structure of sockeye salmon of the central coast of British Columbia: Implications for recovery planning. N. Am. J. Fish. Manag. **23**(3): 703–720.

Olsen, J.B., Wilson, S.L., Kretschmer, E.J., Jones, K.C., and Seeb, J.E. 2000. Characterization of 14 tetranucleotide microsatellite loci derived from sockeye salmon. Mol. Ecol. **9**: 2185–2187.

Olsen, J., Flannery, B., Beacham, T., Bromaghin, J., Crane, P., Lean, C., Dunmall, K., and Wenburg, J. 2008. The influence of hydrographic structure and seasonal run timing on genetic diversity and isolation-by-distance in chum salmon (*Oncorhynchus keta*). Can. J. Fish. Aquat. Sci. **65**(9): 2026–2042.

Olson, R. 1967. Anthropological records social structure and social life of the Tlingit in Alaska. University of California Press.

Panagiotopoulou, H., Austin, J.D. Zalewska, K., Gonciarz, M., Czarnogórska, K., Gawor, J., Weglenski, P., and Popović, D. 2017. Microsatellite mutation rate in Atlantic sturgeon (*Acipenser oxyrinchus*), J. Heredity **108** (6): 686–692.

Pavey, S., Nielsen, J., and Hamon, T. 2010. Recent ecological divergence despite migration in sockeye salmon (*Oncorhynchus nerka*). Evolution **64**(6): 1773–1783.

Paz-Vinas, I., Quéméré, E., Chikhi, L., Loot, G., and Blanchet, S. 2013. The demographic history of populations experiencing asymmetric gene flow: Combining simulated and empirical data. Mol. Ecol. **22**(12): 3279–3291.

Peery, M., Kirby, R., Reid, B., Stoelting, R., Doucet-Beer, E., Robinson, S., Vasquez-Carillo, C., Pauli, J., and Palsboll, P. 2012. Reliability of genetic bottleneck tests for detecting recent population declines. Mol. Ecol. **21**: 3403–3418.

Pella, J., and Masuda, M. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. Fish. Bull. **99**: 151–167.

Petkovich, D. 1999. Champagne and Aishihik first nations: Klukshu river sockeye salmon radio telemetry project. Technical Report, Department of Fisheries and Oceans Canada.

Puebla, O., Bermingham, E., and Guichard, F. 2009 Estimating dispersal from genetic isolation by distance in a coral reef fish (*Hypoplectrus puella*). Ecology **90**(11): 3087–3098.

Piry, S., Luikart, G., and Cornuet, J.M. 1999. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. J. Heredity **90**(4): 502–503.

Pritchard, J., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics **155**: 945–959.

Quinn, T. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fish. Res. **18**(1–2): 29–44.

Quinn, T., Unwin, M., and Kinnison, M. 2000. Evolution of temporal isolation in the wild: Genetic divergence in timing of migration and breeding by introduced Chinook salmon populations. Evolution **54**(4): 1372–1385.

Ramstad, K., Woody, C., Sage, G., and Allendorf, F. 2004. Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. Mol. Ecol. **13**(2): 277–290.

Ray, J., Sethi, S., Eiler, J., and Joyce, J. 2015. Prespawning movements and spawning distribution of sockeye salmon in an urbanizing Alaskan lake. J. Fish Wildl. Manag. **6**(2): 472–485.

Rodgers Olive, S.D., Fox, E.K.C., and Gilk-Baumer, S.E. 2018. Genetic baseline for mixed stock analyses of sockeye salmon harvested in Southeast Alaska for Pacific Salmon Treaty applications, 2018. Fishery Manuscript No. 18–03, Alaska Department of Fish and Game, Anchorage.

Rodríguez-Ramilo, S., and Wang, J. 2012. The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. Mol. Ecol. Resour. **12**(5): 873–884.

Roppel, P. 1982. Alaska's salmon hatcheries 1891-1959. National Marine Fisheries Service, Portland, OR.

Rosenberg, N. 2004. DISTRUCT: A program for the graphical display of population structure. Mol. Ecol. Notes. **4**(1): 137–138.

Rousset, F. 2000. Genetic differentiation between individuals. J. Evol. Biol. **13**(1): 58–62.

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the genepop software for windows and linux. Mol. Ecol. Resour. **8**(1): 103–106.

Seeb, J., Pascal, C.E., Ramakrishnan, R., and Seeb, L.W. 2009. SNP genotyping by the 5'-nuclease reaction: Advances in high throughput genotyping with non-model organisms. *In* Methods in molecular biology, single nucleotide polymorphisms. *Edited by* A.A. Komar. Humana Press, New York. pp. 277–292.

Scribner, K.T., Gust, J.R., and Fields, R.L. 1996. Isolation and characterization of novel salmon microsatellite loci: Cross-species amplification and population genetic applications. Can. J. Fish. Aquat. Sci. **53**: 833–841.

Smith, C.T., Koop, B.F., and Nelson, R.J. 1998. Isolation and characterization of coho salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids. Mol. Ecol. **7**: 1614–1616.

Smoker, W., Gharrett, A., and Stekoll, M. 1998. Genetic variation of return date in a population of pink salmon: A consequence of fluctuating environment and dispersive selection? Alsk. Fish. Res. Bull. **5**(1): 46–54.

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

Steinberg, E., Lindner, K., Gallea, J., Maxwell, A., Meng, J., and Allendorf, F. 2002. Rates and patterns of microsatellite mutations in pink salmon. Mol. Biol. Evol. **19**: 1198–1202.

Sun, J., Helgason, A., Masson, G., Ebenesersdóttir, S., Li, H., Mallick, S., Gnerre, S., Patterson, N., Kong, A., Reich, D., and Stefansson, K. 2012. A direct characterization of human mutation based on microsatellites. Nat. Genet. **44**(10): 1161–1165.

Taylor, E. 1991. A review of local adaptation in salmonidae, with particular reference to Pacific and Atlantic salmon. Aquaculture **98**(1–3): 185–207.

Taylor, S.G., and Lum, J.L. 2003. Auke creek weir 2001 annual report operations, fish counts, and historical summaries. Unpublished Report, National Marine Fisheries Service Auke Bay Laboratory, 11305 Glacier Hwy, Juneau, Alaska 99801.

Thornton, T., Deur, D., and Kitka Sr., H. 2015. Cultivation of salmon and other marine resources on the northwest coast of North America. Hum. Ecol. **43**: 189–199.

Turner, M., Stephens, J., and Anderson, W. 1982. Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. Proc. Nat. Acad. Sci. **79**: 561–573.

Utter, F., Aebrsold, P., Helle, J., and Winans, G. 1984. Genetic characterization of populations in the southeastern range of sockeye salmon. *In* Proceedings of the Olympic wild fish conference. *Edited by* J.M. Walton and D.B. Houston. Peninsula College; Fisheries Technology Program, Port Angeles, Washington. pp. 17–32.

Varnavskaya, N., Wood, C., Everett, R., Wilmot, R., Varnavsky, V., Midanaya, V., and Quinn, T. 1994. Genetic differentiation of subpopulations of sockeye salmon (*Oncorhynchus nerka*) within lakes of Alaska, British Columbia, and Kamchatka, Russia. Can. J. Fish. Aquat. Sci. **51**(1): 147–157.

Wang, S., Hard, J., and Utter, F. 2002. Genetic variation and fitness in salmonids. Conserv. Genet. **3**(3): 321–333.

Waples, R. 1990. Conservation genetics of Pacific salmon. III. Estimating effective population size. J. Heredity **81**(4): 277–289.

Waples, R. 1991. Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the endangered species act. Mar. Fish. Rev. **53**: 11–22.

Waples, R. 1995. Evolutionarily significant units and the conservation of biological diversity under the endangered species act. *In* Evolution and the aquatic ecosystem: Defining unique units in population conservation, Symposium 17. *Edited by* J.L. Nielsen and D.A. Powers. American Fisheries Society, Bethesda Maryland. pp. 8-27.

Waples, R., and Teel, D. 1990. Conservation genetics of Pacific salmon I. Temporal changes in allele frequency. Conserv. Bio. **4**(2): 144–156.

Watts, P., Rousset, F., Saccheri, I.J., Leblois, R., Kemp, S.J., and Thompson, D. 2007. Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion mercuriale*: Zygoptera) populations: analysis of 'neighborhood size' using a more precise estimator. Mol. Ecol. **16**: 737–751.

Withler, R.E., Le, K.D., John Nelson, R., Miller, K.M., and Beacham, T.D. 2000. Intact genetic structure and high levels of genetic diversity in bottlenecked sockeye salmon (*Oncorhynchus nerka*) populations of the Fraser River, British Columbia, Canada. Can. J. Fish. Aquat. Sci. **57**: 1985–1998.

Winans, G.A., Aebersold, P.B. and Waples, R.S. 1996. Allozyme variability of *Oncorhynchus nerka* in the Pacific Northwest, with special consideration to populations of Redfish Lake, Idaho. Trans. Am. Fish. Soc. **125**: 645–663.

Wood, C.C. 1995. Life history variation and population structure in sockeye salmon. *In* Evolution and the aquatic eco system: defining unique units in population conservation, Symposium 17. *Edited by* J.L. Nielsen and D.A. Powers. American Fisheries Society, Bethesda, Maryland. pp. 195–216.

Wood, C., and Holtby, L. 1998. Defining conservation units for Pacific salmon using genetic survey data. *In* Action before extinction: An international conference on conservation of fish genetic diversity. *Edited by* B. Harvey, C. Ross, D. Greer, and J. Carolsfeld. World Fisheries Trust, Victoria, Canada. pp. 233–250.

Wood, C., Riddell, B., Rutherford, D., and Withler, R. 1994. Biochemical genetic survey of sockeye salmon (*Oncorhynchus nerka*) in Canada. Can. J. Fish. Aquat. Sci. **51**(1): 114–131.

Wood, C.C., Bickham, J.W., John Nelson, R., Foote, C.J., and Patton, J.C. 2008. Recurrent evolution of life history ecotypes in sockeye salmon: implications for conservation and future evolution. Evol. Appl. **1**(2): 207–221.

Woody, C.A., Olsen, J., Reynolds, J., and Bentzen, P. 2000. Temporal variation in phenotypic and genotypic traits in two sockeye salmon populations, Tustumena Lake, Alaska. Trans. Am. Fish. Soc. **129**: 1031–1043.

Figure Captions

Figure 1.1: Map of Southeast Alaska with insets of Alaska and the Auke Lake system (adapted from Lum and Taylor 2006). The two main tributaries, Lake Creek and Lake Two Creek, as well as three smaller inlet streams enter the lake on the northern shore. The dark black line represents 15 m; the deepest part of the lake is 31 m. Asterisks on creek names, denote unofficial names.

Figure 1.2: (Upper) Temporal distribution of sockeye salmon returning to Auke Creek, AK for years 2008, 2009, and 2011. (middle) Precipitation (mm) records for the Juneau, AK area for each return year. (lower) Cumulative return of sockeye salmon for each year by sex.

Figure 1.3: Unrooted tree of northern Southeast Alaska collections of sockeye salmon from the Alaska Department of Fish and Game baseline collections for mixed stock analysis was constructed with the restricted maximum likelihood method (ContML) in the program Phylip (Felsenstein 1989).

Figure 1.4: Results of principal components analysis (PCA) conducted on (A) the correlation matrix of centered allele counts for each individual sockeye salmon in the northern Southeast Alaska (NSEAK) collections from the Alaska Department of Fish and Game baseline collections (87 SNP loci) and (B) on the covariance matrix for individuals in the sockeye salmon baseline and the three return years from this study (2008, 2009, and 2011; 41 SNP loci). Sockeye salmon identified as Auke Lake outliers in PCA are represented by triangles. The first principal component separates all Auke Lake collections from all other stocks and the majority of outliers identified in individual PCA overlap with NSEAK collections.

Figure 1.5: After putative strays into Auke Lake, AK were removed, principal component analysis was conducted on the covariance matrix of allele counts for the 2008 return year of sockeye salmon with both microsatellite and single nucleotide polymorphism markers. (A) Two outliers were identified with Hotelling's T^2 test statistic. (B) After removal of outlier individuals, two groups of individuals were identified separated by the first principal component. These individuals primarily came from 12 full sibling families. (C) Plot of Q_i values for 2008 return year inferred with the program *structure* for K = 2 with the admixture model with correlated allele frequencies and sampling location prior. Black vertical lines denote separation between return days.

Figure 1.6: Isolation by time (IBT) plots for sockeye salmon in Auke Lake, AK. Analyses were conducted for each year by considering genetic distances between individuals (A = 2008, B = 2009, C = 2011) and between days of the return (D = 2008, E = 2009, F = 2011). Individual-based plots are represented by a density plot where overlapping dots are represented by heat mapping (red indicates a high density of points). There is a significant IBT signal in years 2008 and 2011 for both individual and day of return distance comparisons.

Figure 1.7: Sensitivity analysis of bottleneck tests with M-ratio for sockeye salmon in Auke Lake, AK. Conclusions were sensitive to variation in the parameters for the proportion and size of non-one step mutations (1-p_s, and Δ_g) and the pre-bottlenecked θ ($4N_e\mu$).

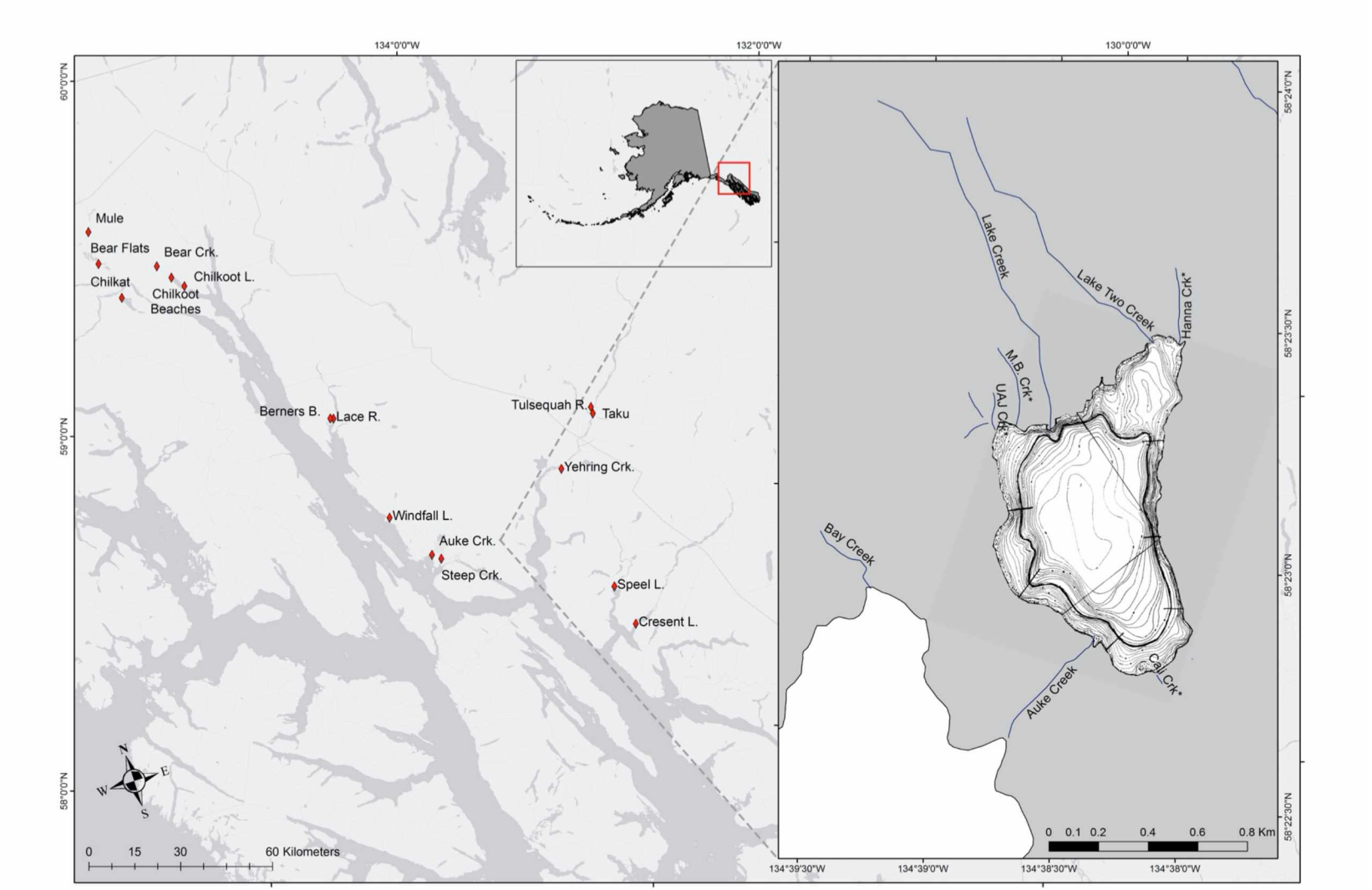

134°0'0"W
132°0'0"W
130°0'0"W
60°0'0"N
59°0'0"N
58°0'0"N
Mule
Bear Flats
Bear Crk.
Chilkoot L.
Chilkat
Chilkoot
Beaches
Berners B.
Lace R.
Tulsequah R.
Taku
Yehring Crk.
Windfall L.
Auke Crk.
Steep Crk.
Speel L.
Cresent L.
N
E
W
S
0 15 30 60 Kilometers
58°24'0"N
58°23'30"N
58°23'0"N
58°22'30"N
Lake Creek
Lake Two Creek
Hanna Crk*
M.B. Crk*
UAJ Crk*
Bay Creek
Auke Creek
Cecil Crk*
0 0.1 0.2 0.4 0.6 0.8 Km
134°39'30"W
134°39'0"W
134°38'30"W
134°38'0"W

Figure 1.1

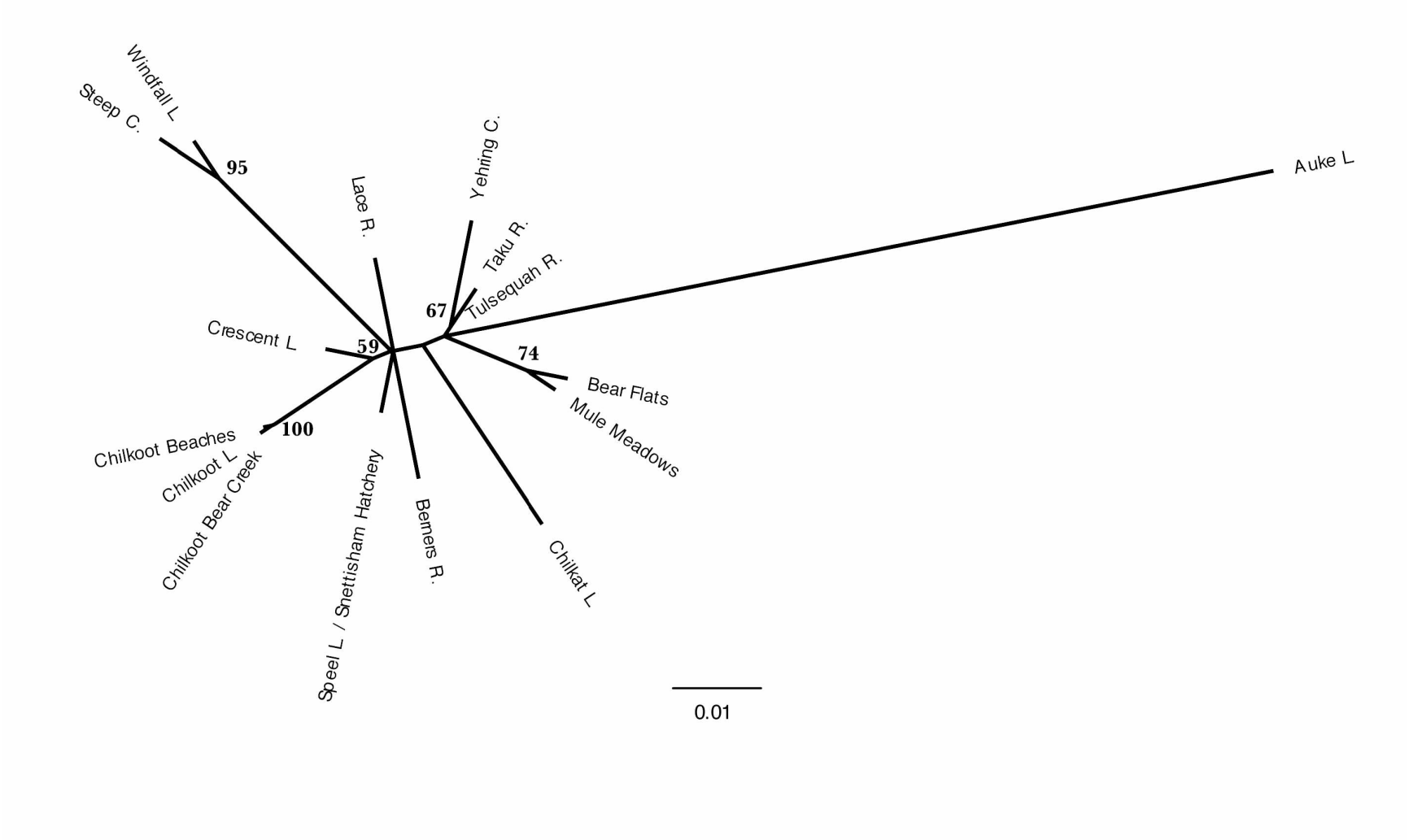

Windfall L
Steep C.
95
Lace R.
Yehning C.
Taku R.
Tulsequah R.
67
Auke L
Crescent L
59
74
Bear Flats
Mule Meadows
Chilkoot Beaches
Chilkoot L
Chilkoot Bear Creek
100
Speel L / Snettisham Hatchery
Berners R.
Chilkat L
0.01

Figure 1.2

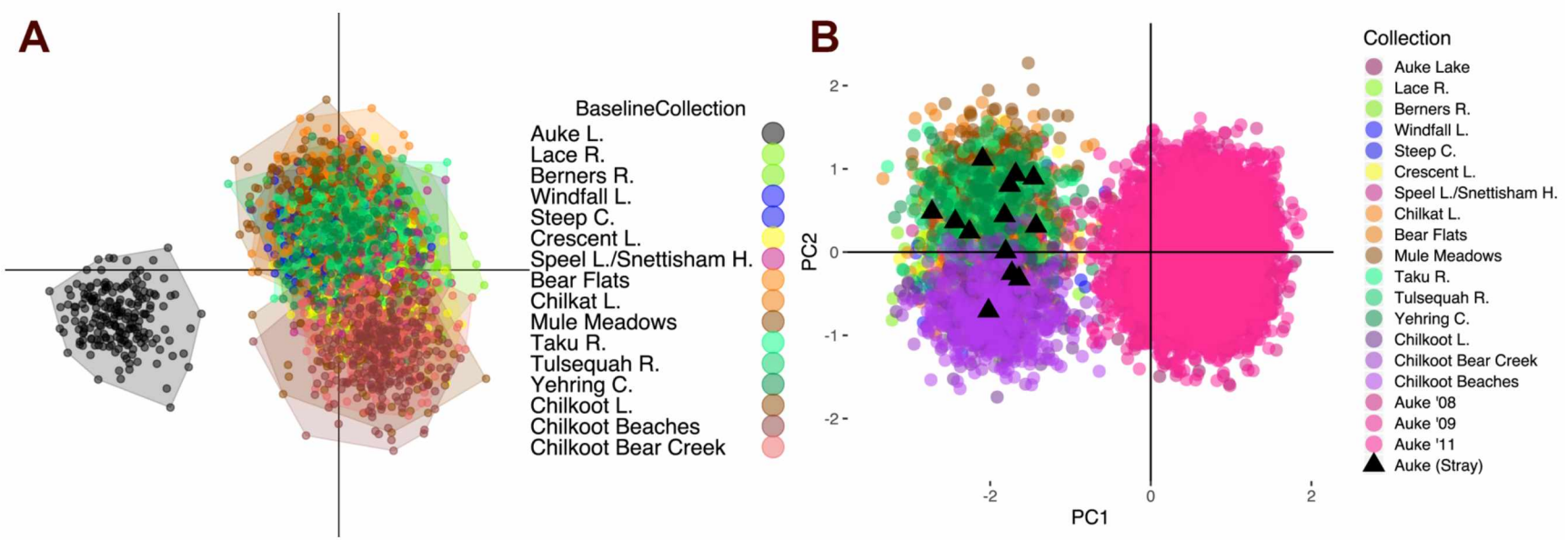

A
BaselineCollection
Auke L.
Lace R.
Berners R.
Windfall L.
Steep C.
Crescent L.
Speel L./Snettisham H.
Bear Flats
Chilkat L.
Mule Meadows
Taku R.
Tulsequah R.
Yehring C.
Chilkoot L.
Chilkoot Beaches
Chilkoot Bear Creek
B
PC2
PC1
2
1
0
-1
-2
-2
0
2
Collection
Auke Lake
Lace R.
Berners R.
Windfall L.
Steep C.
Crescent L.
Speel L./Snettisham H.
Chilkat L.
Bear Flats
Mule Meadows
Taku R.
Tulsequah R.
Yehring C.
Chilkoot L.
Chilkoot Bear Creek
Chilkoot Beaches
Auke '08
Auke '09
Auke '11
Auke (Stray)

Figure 1.3

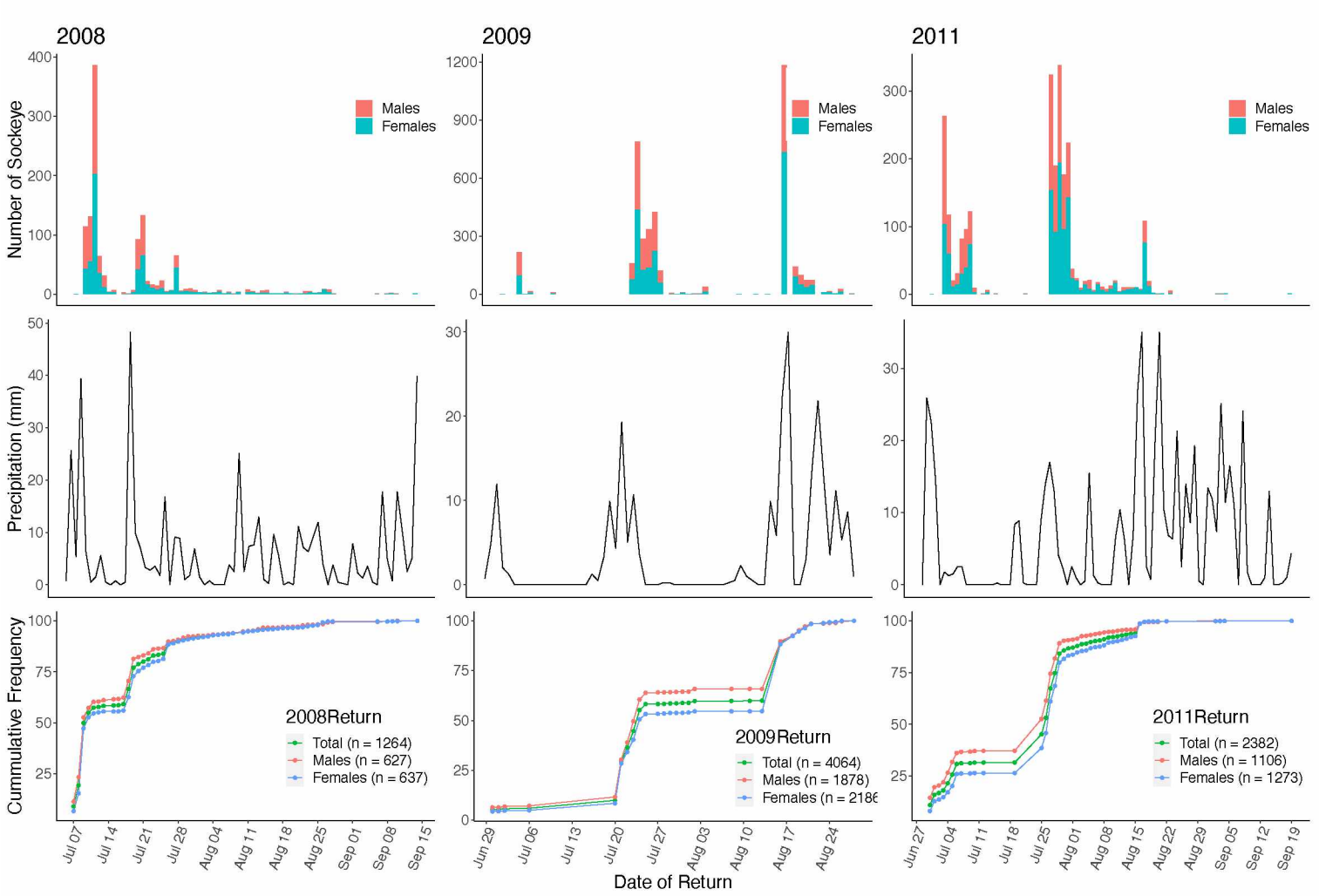

2008
2009
2011
Number of Sockeye
Males
Females
Precipitation (mm)
Cummulative Frequency
2008Return
Total (n = 1264)
Males (n = 627)
Females (n = 637)
2009Return
Total (n = 4064)
Males (n = 1878)
Females (n = 218
2011Return
Total (n = 2382)
Males (n = 1106)
Females (n = 1273)
Date of Return

Figure 1.4

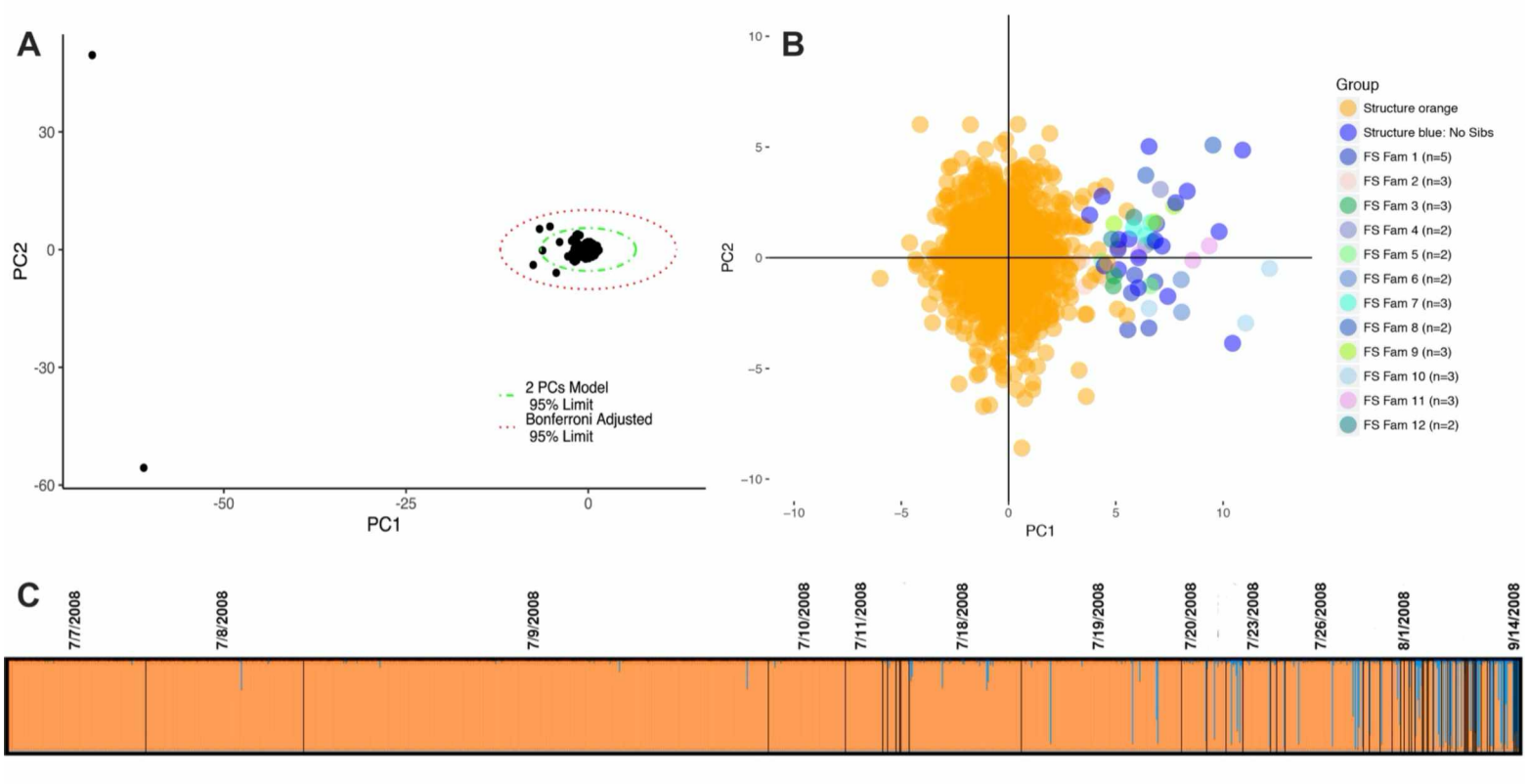

A
30
0
-30
-60
PC2
-50
-25
0
PC1
2 PCs Model
95% Limit
Bonferroni Adjusted
95% Limit
B
10
5
0
−5
−10
PC2
−10
−5
0
5
10
PC1
Group
Structure orange
Structure blue: No Sibs
FS Fam 1 (n=5)
FS Fam 2 (n=3)
FS Fam 3 (n=3)
FS Fam 4 (n=2)
FS Fam 5 (n=2)
FS Fam 6 (n=2)
FS Fam 7 (n=3)
FS Fam 8 (n=2)
FS Fam 9 (n=3)
FS Fam 10 (n=3)
FS Fam 11 (n=3)
FS Fam 12 (n=2)
C
7/7/2008
7/8/2008
7/9/2008
7/10/2008
7/11/2008
7/18/2008
7/19/2008
7/20/2008
7/23/2008
7/26/2008
8/1/2008
9/14/2008

Figure 1.5

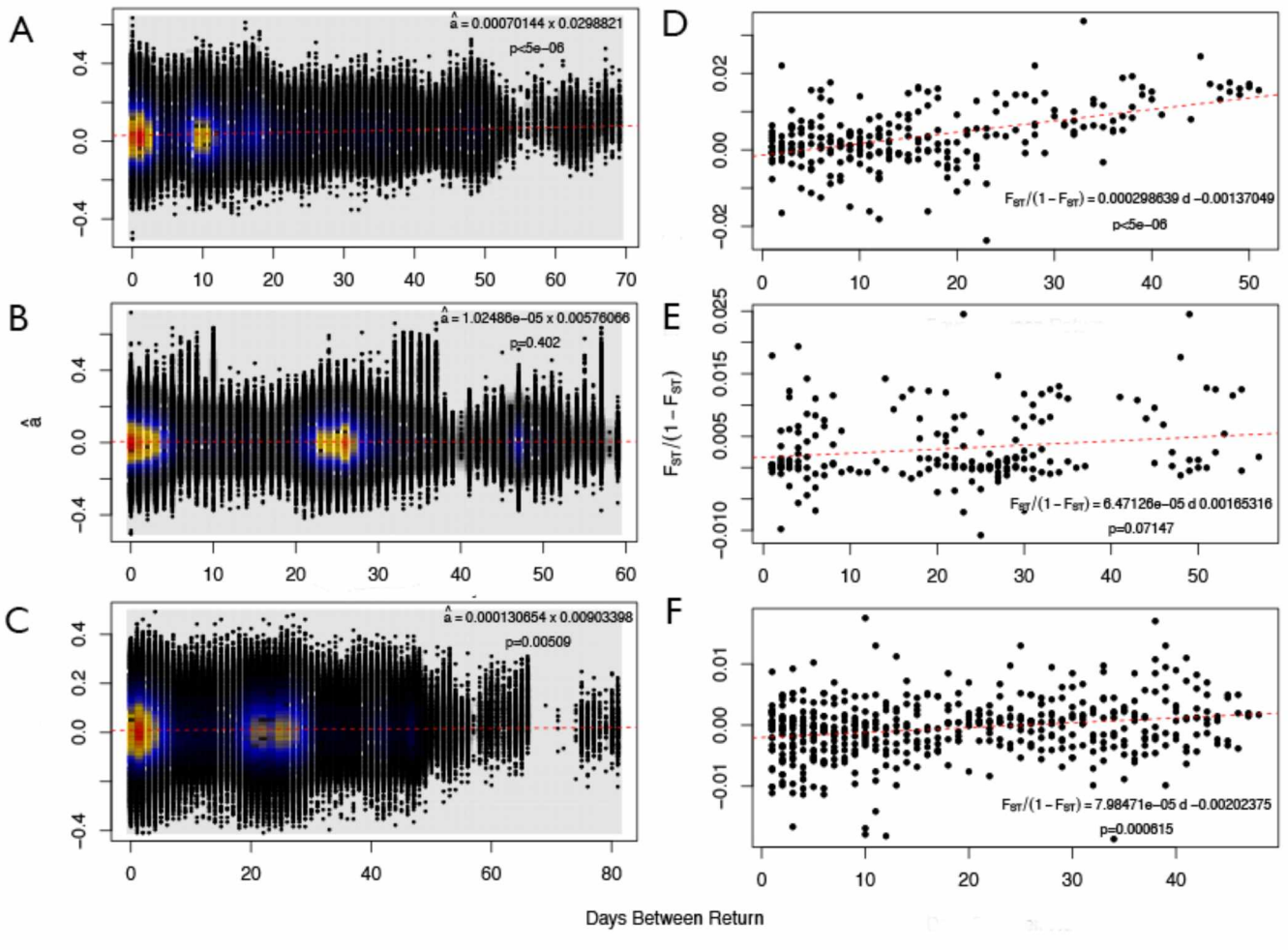

A
â = 0.00070144 x 0.0298821
p<5e−06
B
â = 1.02486e−05 x 0.00576066
p=0.402
C
â = 0.000130654 x 0.00903398
p=0.00509
D
F_ST/(1 − F_ST) = 0.000298639 d −0.00137049
p<5e−06
E
F_ST/(1 − F_ST) = 6.47126e−05 d 0.00165316
p=0.07147
F
F_ST/(1 − F_ST) = 7.98471e−05 d −0.00202375
p=0.000615
â
F_ST/(1 − F_ST)
Days Between Return

Figure 1.6

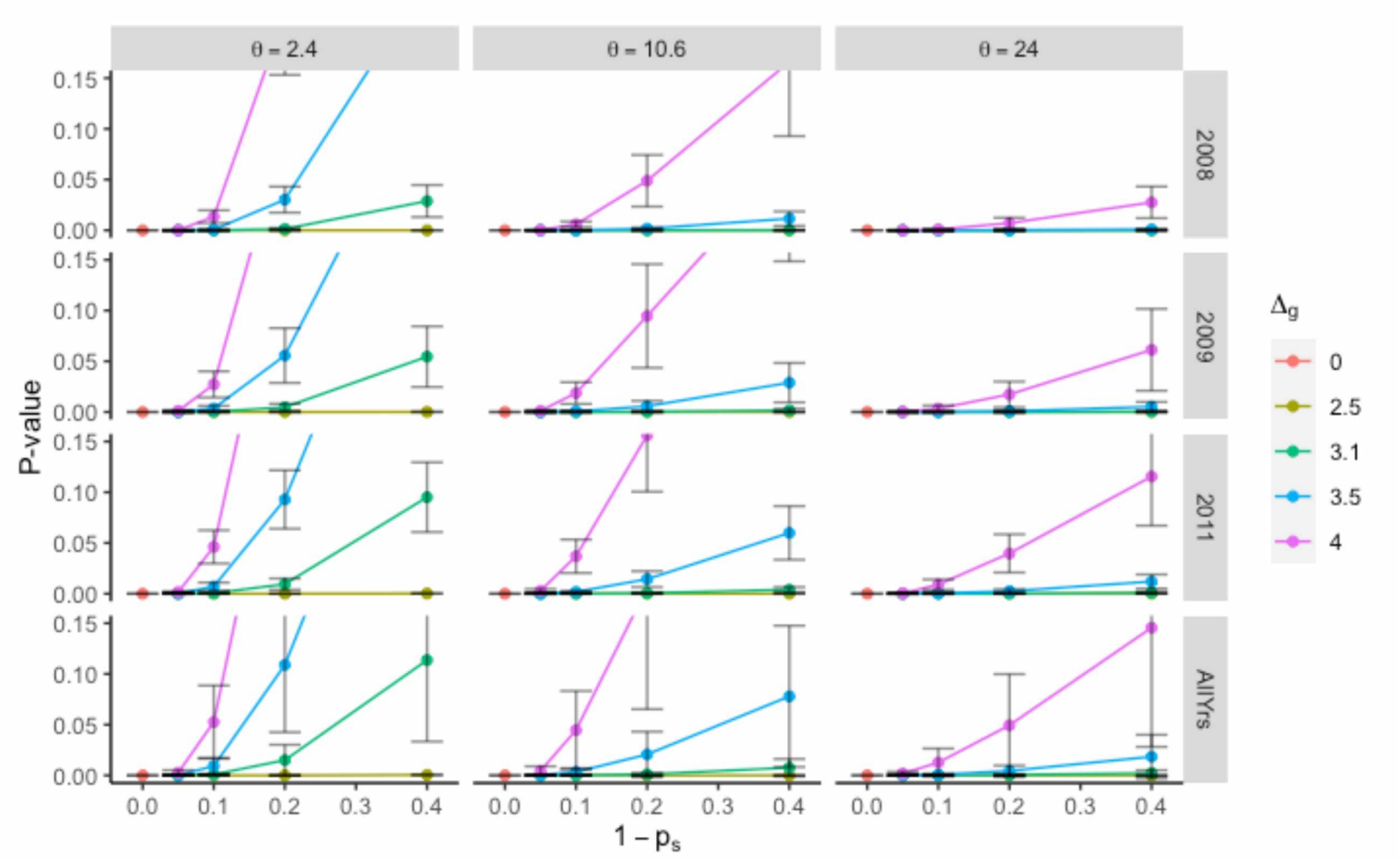

θ = 2.4
θ = 10.6
θ = 24
2008
2009
2011
AllYrs
P-value
0.15
0.10
0.05
0.00
0.0
0.1
0.2
0.3
0.4
1 – ps
Δg
0
2.5
3.1
3.5
4

Figure 1.7

Table 1.1: Assignment of sockeye salmon identified as putative strays with the program BAYES. Populations, other than Auke Lake, AK, that were unlikely to contribute strays (less than 10% probability) have been removed. No fish were assigned back to Auke Lake. Six fish had high probabilities of assignment back to the Speel Lake / Snettisham Hatchery collection which was consistent with the program ONCOR. The largest assignment probability for each stray is denoted in bold.

	Baseline Population								
Individual	Auke L.	Lace R.	Berners R.	Crescent R.	Speel L.	Mule M.	Taku R.	Tulsequah R.	Yehring C.
SAUKE08_204	0.00	0.19	**0.46**	0.02	0.27	0.00	0.03	0.03	0.00
SAUKE08_208	0.00	0.06	0.01	0.18	**0.55**	0.00	0.04	0.07	0.00
SAUKE08_229	0.00	0.00	0.00	0.00	**0.94**	0.02	0.02	0.00	0.01
SAUKE08_254	0.00	0.00	0.00	0.02	**0.52**	0.12	0.23	0.05	0.01
SAUKE08_282	0.00	0.01	0.04	0.00	**0.95**	0.00	0.00	0.00	0.00
SAUKE08_283	0.00	0.00	0.00	0.00	**0.99**	0.00	0.00	0.00	0.00
SAUKE08_287	0.00	0.04	0.01	0.02	**0.90**	0.00	0.02	0.01	0.00
SAUKE08_1246	0.00	0.00	0.02	0.00	**0.97**	0.00	0.00	0.00	0.01
SAUKE09_3381	0.00	0.00	0.00	0.06	**0.67**	0.11	0.04	0.04	0.00
SAUKE09_4364	0.00	0.06	0.01	0.00	**0.91**	0.00	0.00	0.02	0.00
SAUKE09_5645	0.00	0.01	0.05	0.09	0.28	0.01	**0.30**	0.09	0.11
SAUKE09_6039	0.00	0.11	0.23	0.06	**0.50**	0.00	0.04	0.02	0.00
SAUKE11_9121	0.00	0.00	0.02	0.01	**0.81**	0.01	0.10	0.04	0.02

Table 1.2: Tests of isolation-by-time (IBT) for Auke Lake, AK sockeye salmon. Tests were either individual-based or group-based (each return day had five or more individuals). Each test included all fish collected or putative females only, excluding individuals identified as putative strays.

Data Included	Year	Group	Estimator	Probability	y-intercept	Slope
All Fish	2008	Ind.	$\hat{a}$	<0.001	0.030	$7.01*10^{-4}$
All Fish	2008	Ind.	$\hat{e}$	0.632	$-1.99*10^{-5}$	$4.53*10^{-7}$
All Fish	2008	Return Day	$F_{ST}/(1-F_{ST})$	<0.001	-0.001	$2.99*10^{-4}$
All Fish	2009	Ind.	$\hat{a}$	0.402	0.006	$1.02*10^{-5}$
All Fish	2009	Ind.	$\hat{e}$	0.009	$-2.96*10^{-5}$	$2.72*10^{-6}$
All Fish	2009	Return Day	$F_{ST}/(1-F_{ST})$	0.072	0.002	$6.47*10^{-5}$
All Fish	2011	Ind.	$\hat{a}$	0.005	0.009	$1.31*10^{-4}$
All Fish	2011	Ind.	$\hat{e}$	<0.001	$-2.87*10^{-4}$	$1.92*10^{-5}$
All Fish	2011	Return Day	$F_{ST}/(1-F_{ST})$	0.001	-0.002	$7.98*10^{-5}$
Females	2008	Ind.	$\hat{a}$	<0.001	0.038	$8.77*10^{-4}$
Females	2008	Ind.	$\hat{e}$	<0.001	-0.001	$9.76*10^{-5}$
Females	2008	Return Day	$F_{ST}/(1-F_{ST})$	0.004	$-3.21*10^{-4}$	$4.70*10^{-4}$
Females	2009	Ind.	$\hat{a}$	0.284	0.004	$2.18*10^{-5}$
Females	2009	Ind.	$\hat{e}$	0.645	$8.54*10^{-5}$	$-3.21*10^{-6}$
Females	2009	Return Day	$F_{ST}/(1-F_{ST})$	0.864	0.003	$-4.95*10^{-5}$
Females	2011	Ind.	$\hat{a}$	<0.001	0.015	$3.60*10^{-7}$
Females	2011	Ind.	$\hat{e}$	<0.001	$2.91*10^{-5}$	$4.55*10^{-9}$
Females	2011	Return Day	$F_{ST}/(1-F_{ST})$	0.259	0.002	$8.87*10^{-8}$

Table S1.2: Pairwise F_{ST} values between northern Southeast Alaska collections from the Alaska Department of Fish and Game baseline collections for mixed stock analysis (below diagonal) with homogeneity P values (above diagonal). The smallest differences occurred between populations within a drainage system (indicated above the population labels). * represents a value of $P < 10^{-5}$.

				Berners Bay			
	Auke C.	Windfall L.	Steep C.	Lace R.	Berners R.	Crescent L.	Speel L.
Auke C.	-	*	*	*	*	*	*
Windfall L.	0.258	-	*	*	*	*	*
Steep C.	0.256	0.023	-	*	*	*	*
Lace R.	0.222	0.086	0.099	-	*	*	*
Berners R.	0.246	0.103	0.101	0.046	-	*	*
Crescent L	0.208	0.083	0.08	0.032	0.042	-	*
Speel L	0.207	0.077	0.084	0.033	0.036	0.033	-
Chilkat R.	0.236	0.132	0.13	0.087	0.098	0.086	0.064
Bear Flats	0.234	0.125	0.118	0.078	0.083	0.063	0.055
Mule R.	0.204	0.11	0.106	0.076	0.078	0.06	0.053
Taku R.	0.22	0.085	0.087	0.039	0.05	0.024	0.033
Tulsequah R.	0.202	0.075	0.076	0.036	0.049	0.023	0.03
Yehring C.	0.206	0.088	0.09	0.056	0.062	0.04	0.041
Chilkoot L.	0.225	0.106	0.107	0.057	0.063	0.039	0.045
Chilkoot B.	0.228	0.104	0.107	0.055	0.064	0.041	0.054
Chilkoot Bear Ck.	0.221	0.097	0.098	0.056	0.062	0.037	0.042

Table S1.1 (continued)

	Chilkat River				Taku River		
	Chilkat R.	Bear Flats	Mule R.		Taku R.	Tulsequah R.	Yehring C.
Auke C.	*	*	*	*	*	*	*
Windfall L.	*	*	*	*	*	*	*
Steep C.	*	*	*	*	*	*	*
Lace R.	*	*	*	*	*	*	*
Berners R.	*	*	*	*	*	*	*
Crescent L	*	*	*	*	*	*	*
Speel L	*	*	*	*	*	*	*
Chilkat R.	-	*	*	*	*	*	*
Bear Flats	0.073	-	*	*	*	*	*
Mule M.	0.069	0.014	-		*	*	*
Taku R.	0.075	0.029	0.032		-	*	*
Tulsequah R.	0.066	0.03	0.03		0.007	-	*
Yehring C.	0.085	0.044	0.043		0.018	0.022	-
Chilkoot L.	0.084	0.089	0.082		0.053	0.052	0.058
Chilkoot B.	0.095	0.102	0.099		0.061	0.061	0.069
Chilkoot Bear Ck.	0.083	0.09	0.083		0.053	0.051	0.059

Table S1.1 (continued)

	Chikoot River		
	Chilkoot L.	Chilkoot B.	Chilkoot Bear Ck.
Auke C.	*	*	*
Windfall L.	*	*	*
Steep C.	*	*	*
Lace R.	*	*	*
Berners R.	*	*	*
Crescent L	*	*	*
Speel L	*	*	*
Chilkat R.	*	*	*
Bear Flats	*	*	*
Mule M.	*	*	*
Taku R.	*	*	*
Tulsequah R.	*	*	*
Yehring C.	*	*	*
Chilkoot L.	-	*	0.682
Chilkoot B.	0.009	-	*
Chilkoot Bear Ck.	0.001	0.009	-

Chapter 2: Demographic and genetic estimators of effective population size for Sockeye Salmon in Auke Lake, Alaska [2]

[A] Abstract

The conservation and management of Pacific salmon populations is often based on estimates of the census size (N_c). However, in small populations the rate of loss of genetic variation (heterozygosity), increase in inbreeding, and decrease in population fitness is a function of the effective population size (N_e), which can be much smaller than N_c. A long-term dataset on Sockeye Salmon *Oncorhynchus nerka* in Auke Creek, Alaska allowed us to estimate N_e with demographic methods over a 38-year time frame and the effective number of breeders (N_b) with genetic methods for three return years. Demographic estimates of N_e demonstrate substantial variability among the seven generations; initially low values peaked by the fourth generation, followed by a gradual decline. The major demographic factors that determine N_e were variance in family size, variable contribution to the next generation by brood years within a generation, and fluctuations in population size. Freshwater productivity (adult to smolt) appeared to influence population size fluctuations more than marine survival (smolt to adult) over the 38-year timeframe. Genetic estimates of N_e were smaller than demographic estimates. The lower-bound estimate for N_e/N_c was between 0.21 and 0.37, which is consistent with values reported for other salmonid species.

[A] Introduction

The amount of genetic variation existing within and among populations promotes long-term stability by providing a buffer to natural environmental and anthropogenic perturbations (Booy et al. 2000; De Meester et al. 2018). In small populations genetic variation declines, inbreeding increases, and population fitness can be reduced by the accumulation of deleterious recessive alleles as a result of genetic drift (Hartl and Clark 1997). The rate of drift in a population is a function of its effective population size (N_e), the size of an idealized population that would show the same genetic response to drift or inbreeding as the population being considered (Wright 1931). This idealized population

---

[2] P.D. Barry, M.V. McPhee, S. Vulstek, J. Joyce, D.A. Tallmon and A.J. Gharrett. *Formatted for submission to Transactions of the American Fisheries Society*

has discrete generations, constant size through time, random mating, a 1:1 sex ratio, and family sizes that follow a Poisson distribution (Wright 1931). Natural populations differ from this ideal, so that N_e is usually smaller than the census size (N_c; Frankham 1995; Vucetich et al. 1997; Palstra and Ruzzante 2008; but see Waples et al. 2013).

For conservation and management of wild populations, the ratio of N_e/N_c is used in criteria for endangerment (Mace and Lande 1991) and for setting minimum viable population sizes (Nunney and Campbell 1993; Jamieson and Allendorf 2012). The 50/500 rule (Franklin 1980) suggests that an N_e of less than 50 represents an immediate risk of extinction. Populations smaller than this size may enter an extinction vortex because of demographic stochasticity and inbreeding depression (Gilpin and Soulé 1986). An N_e exceeding 500 is needed to maintain the evolutionary potential of a population (Franklin and Frankham 1998). The 50/500 rule has been criticized as an overly optimistic estimate of the minimum number of individuals required (Lande 1995; Reed and Bryant 2000), and it has been suggested that the rule be updated to 100/1000 (Frankham et al. 2014; but see Franklin et al. 2014).

While N_e is a critical parameter, its estimation and, therefore, its use in conservation and management is often challenging. Demographic estimators (reviewed in Caballero 1994) are based on the causal parameters that reduce N_e, such as fluctuations in census size, unbalanced sex ratios, and the variance in family size, which are difficult to measure accurately. In addition, N_e estimates based on demographic parameters are often biased upward because rarely are all the factors that reduce N_e incorporated into the estimate (Frankham 1995). Genetic methods (reviewed in Wang 2005), which use the signal that demographic factors impart on genetic variation in the population, are based on easily obtainable data, but interpretation of the estimate is not always straightforward (Waples 2005). It is often easier to estimate the effective number of breeders (N_b) because only a single season of data is necessary. For Pacific salmon populations with variable size, the two effective sizes are related in that the harmonic mean of the N_b values in individual years times the generation length approximates N_e (Hill 1979; Waples 1990).

Pacific salmon (*Oncorhynchus* spp.) are ecologically and economically important around the northern Pacific Rim. Widespread declines in the sizes and numbers of

populations (Nehlsen et al. 1991; Slaney et al. 1996) have prompted the development of the theory and methods for estimating N_e for Pacific salmon populations (Waples 1990, 2002, 2006). Despite the fact that Alaska is lauded for its successful management of Pacific salmon populations and abundances are at historically high levels for Sockeye Salmon *O. nerka*, Chum Salmon *O. keta*, and Pink Salmon *O. gorbuscha* (Van Alen 2000; Ruggerone and Irvine 2018), global climate change is projected to disproportionately affect northern latitudes, including Alaska, and creates uncertainty in the future health of Pacific salmon populations (Schoen et al. 2017). In Southeast Alaska, wild Sockeye Salmon originate from more than 200 drainages. While a few large watersheds account for the majority of the commercial harvest in drift-gillnet and seine fisheries (Gilk-Baumer et al. 2015), the collective production from the numerous small stocks is substantial. The small size of these populations puts them at higher risk of temporal instability in abundance (Einum et al. 2003) while isostatic rebound throughout the region has been shown to drastically alter the hydrology and vegetation of rivers thereby reducing productivity (Faber 2008).

Auke Creek in Alaska represents a unique Sockeye Salmon population in which to evaluate N_e. The genetic differences observed between Auke Creek and other Southeast Alaska Sockeye Salmon populations along with reduced heterozygosity, suggest that N_e may be small relative to N_c (Chapter 1). The National Oceanic and Atmospheric Administration (NOAA) Auke Creek Research Station has a permanent weir that has been in full-time operation since 1980. The weir is located at the head of the tidewater connecting Auke Creek to Auke Bay. Complete enumeration of smolt and adults over 38 years (> 6 generations) provides an opportunity to evaluate how the population has changed over time. The objectives of this study were to evaluate how different demographic factors influence the effective population size of Sockeye Salmon in Auke Creek, Alaska. Specifically, we (1) estimated single generation and multigenerational N_e with demographic methods for years 1980 to 2017, (2) estimated N_b with genetic methods (from collections in 2008, 2009 and 2011), and (3) evaluated factors that may contribute to reductions in N_e.

[A] Methods

[C] *Auke Lake system.* — Located approximately 17.7 km northwest of Juneau, Alaska, Auke Lake is 1.6 km long by 1.2 km wide and has a maximum depth of 31.4 m. Two major tributaries, Lake Creek and Lake Two Creek, feed the lake and are the main Sockeye Salmon spawning habitat (Ray et al. 2015). Auke Creek is the only outlet into Auke Bay (Figure 2.1).

Sockeye Salmon typically return to the Auke Creek weir in June and July and hold in Auke Lake until they mature. Spawning occurs primarily in Lake Creek in August and September. Fry migrate down to the lake in March and April and spend one or two years in Auke Lake. Seaward emigration of smolt occurs in May and June. A small proportion of males may mature within the lake foregoing emigration to the ocean (Kovach et al. 2014), but it is assumed that this life history tactic is uncommon and should not substantially alter estimates of N_b and N_e (Scott Vulstek personal observation). Migrating individuals spend one to three years in the ocean (Kovach et al. 2014).

A permanent weir structure allows the complete enumeration of out-migrating smolts in the spring and returning adults in the summer. The weir is installed in late winter before ice-out of the lake and the downstream configuration is used to count all emigrating fry and smolts from early March until June. The weir is changed to its upstream configuration in late June and kept in place until October to count all immigrating adults. Subsets of the return have been sampled for age, length, and weight each year.

[C] *Age composition.* — Two age classes of smolts outmigrate from Auke Creek each year; the midpoint of migration of age-2 smolts was, on average over the 38 years, five days prior to that of age-1 smolts. The proportion of age-1 and age-2 smolts emigrating on each day has been estimated from a binomial logistic regression of smolt age on outmigration date ($\bar{r}^2_{pseudo}$= 0.75, S. Vulstek unpublished data). Scale samples have been collected opportunistically from smolt during the outmigration, ranging from 2 to 15 days of the outmigration per year and averaging 38 smolts sampled per day.

The adult age composition for each year was estimated from scale samples (Clutter and Whitesel 1956). Scale samples have been opportunistically taken since 1980

with an average of 10.2% (± 6.4 sd) of the returning adults sampled. Yearly sample sizes of scales ranged from 52 to 468 and had an average of 235 samples.

[C] *Demographic N_e.* — Pacific salmon populations can be characterized by large fluctuations in population size (Eninum et al. 2003). When temporal variation in population abundance occurs, the effective population size is determined, to a large extent, by the smallest N_c in a population's demographic history. The effective size can be estimated as the harmonic mean of the population sizes (Motro and Thomson 1982; Wright 1938). For Pacific salmon with variable age at maturity, N_e is a function of the harmonic mean of the number of breeders in individual return years ($\widetilde{N}_b$) that comprise a generation ($N_e \approx T\widetilde{N}_b$) where T is the generation length; Hill 1979; Waples 1990, 2002). While this method accounts for variability among years in return size, it assumes that the number of recruits (maturing fish that returned to the weir) per spawner (λ) is equal among years within the generation. Inter-annual variation in λ can cause N_e to approach N_b for a single year (Waples 2002). If recruitment data are available, differential reproductive success among return years within a generation can be incorporated into estimates of N_e (Ryman and Laikre 1991; Waples 2002),

$$N_e = 1/\sum(\mathrm{X}_t^2/N_{bt})$$

where X_t is the proportional contribution of spawners in year t to the next generation, N_{bt} is the effective number of breeders in year t, and the summation is over all years within the generation. To evaluate the effect of fluctuations in population size and variable contribution among years within a generation on N_e, we assumed that all returning fish contributed to the next generation (N_{bt} was equal N_t the total return size for year t). We can estimate X_t from data available on ages of fish and the number of spawners and express the above equation in terms of per generation recruitment

$$N_e = 1/\sum\left(\frac{\mathrm{R}_t^2}{N_t * \mathrm{R}_T^2}\right)$$

where R_t is the number of recruits for brood year t and R_T is the total recruitment for the generation. This equation can be re-arranged to emphasize the relation to variation in recruits per spawner ($\lambda_t = R_t/N_t$),

$$N_e = \frac{1}{\left(1/{R_T}^2\right)\sum \lambda_t R_t}$$

In this study, the total number of spawners (N_t), successful and unsuccessful breeders, was determined by direct count at the Auke Creek weir. Scale samples provided estimates of age composition by year, which combined with N_t were used to estimate R_t and λ_t. The generation length T was estimated from scale-aged fish from the entire time series as $T = \sum_{i=1}^{M} iA_i$ where A_i is the probability of maturing at age i and M is the maximum age at reproduction (Waples 2006). Here we make the simplifying assumption that all fish that return as spawners contribute equally to N_b and that the average age of the return is equal to the average age at maturity weighted by age specific reproductive success. The generation length was rounded to the nearest integer value (T*) to produce a nominal generation length and estimates of N_e were scaled by a factor of T/T* for an unbiased estimate (Waples 2006). The choice of a reference year that started a generation was arbitrary, so we estimated the results of three different reference years for grouping the 38 years into generations to evaluate the effect of different arrangements of years.

Fluctuations in N_e for anadromous Pacific salmon may result from variable mortality in the freshwater environment, the marine environment, or both. Freshwater survival could not be estimated directly because fish were not sampled on the spawning grounds. From estimates of smolt age and the observed return year abundance, we calculated the number of smolts to potential spawners (herein referred to as smolts per spawner) as a proxy for freshwater survival. Trends in freshwater productivity were investigated with a breakpoint analysis in the R package changepoint (Killick and Eckley 2014). We tested for changes in the mean and variance with the At Most One Change (AMOC) method with distribution-free test statistics. Female fecundity and potential eggs deposited (PED) to smolt survival were estimated from broodstock collected for a hatchery experiment between 2011 and 2013 (2011 n = 30, 2012 n = 23, 2013 n = 24). For each female sampled, 100 eggs were weighed to estimate the average weight of an individual egg. The total weight of all eggs was then divided by the average weight of a single egg to estimate fecundity for the individual. The relationship between length and fecundity was evaluated with a linear model. Lastly, marine survival was estimated for each brood year. Stock-specific estimates of catch composition are not available to

estimate fishery exploitation rate in Auke Creek Sockeye Salmon, so our estimate of marine survival includes fisheries mortality.

An unbalanced sex ratio is a crucial demographic factor that reduces N_e below the number of spawners (Wright 1931; Hill 1979; Frankham 1995). If there are unequal numbers of males and females, stochastic variation in the genotypes of gametes passed into the next generation will cause drift from the allele frequencies in the previous generation. Sex identification of Sockeye Salmon returning to Auke Creek from 2008 to 2017 was based on the shape and size of the vent and kype, but misidentifications were possible, particularly in the early portion of the run when fish were not sexually mature. We evaluated the effect that the estimated sex ratio would have on N_e for the two most recent generations of the time series.

Variation in the number of progeny produced by individuals can also reduce N_e below N_c. Often the variance in family size exceeds the mean and N_e is smaller than the census size; however, N_e can exceed the census size when the variance is smaller than the mean. With genetic data we estimated the mean and variance of family sizes for return years 2008, 2009, and 2011. We genotyped each fish for a panel of genetic markers and deduced full sibling family groups with the program Colony V2 (Jones and Wang 2010; see genetic methods). We estimated the mean ($\bar{k}$), variance (V_k), and the index of variability ($V_k/\bar{k}$), which under the assumption of random survival of offspring (the Poisson expectation) should be 1 (Crow and Morton 1955). The observed family sizes were compared to those expected under a Poisson distribution with a log-likelihood ratio (G-test) goodness of fit test. We estimated the effective number of breeders as

$$N_b = \frac{\bar{k}N - 1}{\bar{k} - 1 + {V_k}/{\bar{k}}}$$

(Crow and Denniston 1988). If the mean and variance in offspring number is consistent among years, the ratio of N_e to N_c can be approximated by

$$\frac{N_e}{N_c} \approx \frac{4}{V_k + 2}$$

(Wright 1938).

[C] *DNA extraction and amplification.* — We exhaustively sampled the population during three return years by clipping the axillary process from each returning adult as it passed through the weir (2008, n = 1264; 2009, n = 4064; 2011, n = 2427). Axillary processes were stored in 95% ethanol until dried for DNA extraction and archived. Samples were genotyped for both microsatellite and single nucleotide polymorphism (SNP) genetic markers (Supplementary Table S2.1). Samples from 2008 were genotyped for 82 markers (14 microsatellites and 68 SNPs). A reduced panel of markers was used to genotype samples from years 2009 and 2011 (12 microsatellites and 45 SNPs). DNA was isolated from axillary process samples with Qiagen DNeasy™ kits (Qiagen, Valencia, California) or a proteinase K and ammonium acetate procedure (Puregene DNA™ isolation protocol - Gentra Systems, Minneapolis, MN). Isolated DNA was hydrated in 0.1X TE buffer (TE is 1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) and stored in 1.5 ml tubes at -20°C.

Two methods were used to genotype microsatellite loci. Samples from 2008 were amplified in 10 uL reactions, which contained ~1 unit Taq polymerase, 1X PCR buffer (50 mM KCl_2, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100; Promega Corp., Madison, WI), 0.5 µM of each deoxyribonucleotide triphosphates (dNTPs), 1.5–1.875 mM $MgCl_2$, 0.025–0.1 µg DNA, 0.2–0.4 µM unlabeled forward and reverse primers, and 0.02–0.04 µM fluorescently labeled forward primer with an IRDye® infrared dye (LI-COR, Lincoln, Nebraska). The amplification profile was: 95°C for 5 min; 25–35 cycles of 0.5 min at 95°C, 0.5 min at locus specific annealing temperature (46°C to 60°C), and 0.75 min at 72°C; and a final extension of 72°C for 10 min. Gel electrophoresis was performed by loading a mixture of 0.5 µL of PCR product and 0.5 µL of stop buffer (95% formamide, 0.1% bromophenol blue) into a 0.25 mm thick 6% polyacrylamide gel (PAGE-PLUS™, Amresco, Solon, OH) on the LI-COR 4300 System™. All gels were run in 1X TBE buffer (0.09M Tris-Borate, 2mM EDTA, pH 8.3) at 40 W, 1500 V, 40 mA, and 45°C plate temperature. Run length, determined by the fragment size of the locus, lasted between 1 and 4 hours. Size fragments were scored with SagaGT (Ver. 3.2.1, LI-COR) analysis software by comparing the fragments with either IRD700 or IRD800 standard ladders (LI-COR, Biotechnology Division) or custom-designed size ladders.

Microsatellite loci for the 2009 and 2011 collections were amplified at the Alaska Department of Fish and Game Gene Conservation Laboratory (ADF&G GCL) in Anchorage, Alaska. Loci were amplified in 10 µL reaction volumes that contained 0.5 units Taq DNA polymerase (Promega, Madison, WI), 10 mM Tris-HCl, 50 mM KCl, 0.2 mM each dNTP, 0.06–0.20 µM primers. Fragment size analysis was performed on an ABI 3730 capillary DNA sequencer by loading 0.5 µL PCR product, 0.4 µL of GeneScan™ 600 LIZ™ dye size standard, and 9.0 uL of Hi-Di formamide (Applied Biosystems, Foster City, CA). Fragments were visualized and separated by size into bin sets with Applied Biosystems (ABI) GeneMapper software v4.0. All individuals that failed to amplify on the ABI system were rerun with the LI-COR. Size fragment conversions between the two platforms were accomplished by running all unique alleles at each microsatellite locus from 2009 and 2011 on the LI-COR systems.

Single nucleotide polymorphism loci were amplified and scored at the ADF&G GCL. The SNP markers were screened with Biomark Dynamic Arrays™ (Fluidigm, South San Francisco, CA) following the methods of Seeb et al. (2009). All genetic loci used in this study were previously shown to be in Hardy-Weinberg and linkage equilibrium (Chapter 1).

[C] *Genetic N_e.* — The linkage (gametic) disequilibrium (LD) method estimates N_e from the non-random association of alleles between loci, because the amount of LD between two neutral loci is a function of N_e and the recombination rate (Hill 1981; Waples 1991). The LD method assumes that the population being considered is panmictic and stable and that the evolutionary forces of selection, migration, and mutation are negligible. We estimated N_b with the LD approach of LDNe (Waples and Do 2008) implemented in NeEstimator V2.1 (Do et al. 2014) with a $P_{crit} = 0.02$ to control for the upward bias introduced by low frequency alleles (Waples and Do 2010). We estimated N_b from the full return year (Plan 1) and from a subset of individuals which we were able to assign to their brood year based on scale ages (Plan 2; Waples 2005). In Plan 1 sampling, the mixture of multiple cohorts within a single return year generates additional disequilibrium at pairs of loci (Sinnock 1975; Waples 2005), which makes estimates of N_b from LD estimators intermediate between N_b and N_e (Waples 2005). In Plan 2

sampling, this issue was alleviated by arranging fish into their respective brood years; however, the number of scale-aged fish was a small fraction of the overall return.

We also estimated N_b with the sibship assignment approach in Colony 2 (Jones and Wang 2010). This approach estimates N_b based on the probabilities that a pair of offspring taken at random from the population are full or half-siblings (Wang 2009):

$$\frac{1}{N_b} = \frac{1-3\alpha}{4}(Q_1 + Q_2 + 2Q_3) - \frac{\alpha}{2}\left(\frac{1}{N_M} + \frac{1}{N_F}\right)$$

where Q_1 and Q_2 are the probability of drawing a pair of maternal or paternal half-siblings, respectively, at random from the population, Q_3 is the probability of drawing two full-siblings from the population, N_F and N_M are the numbers of male and female parents, and α is equivalent to Wright's (1969) F_{IS} statistic. If a population is small, the probability that two individuals sampled at random share a father, mother, or both increases. This method assumes that the sample of individuals is taken at random from a single cohort of the population. Samples from an individual return year from Auke Creek are not completely random with respect to kinship. The variable age at maturity means that many individuals will come from three or more brood years. We analyzed each return year individually, and combined each return year (2008, 2009, and 2011) which represent three of five years from a single generation. Early-maturing Sockeye Salmon sampled in 2011 (age-3 'jacks,' or males that spend one winter less in the ocean than the youngest females) could have parents from just the 2008 return year. The inclusion of parents in the analysis on the combined return years introduces negative bias on N_b when parent-offspring relationships are inferred as full-sibling relationships (Wang 2009). We identified age-3 jacks that had parents included in the 2008 return year by conducting parentage analysis with the program FRANz (Riester et al. 2009) accounting for unsampled parents from 2005 to 2007. Only individuals that were recorded as jacks during genetic sampling (visually, based on small size) and had a high (> 0.95) posterior probability of assignment to 2008 parents were omitted from the combined year dataset before N_b was estimated. Age-3 jacks comprise a very small proportion of the return (0–6.6%), based on scale-aged data, and removing them should have a minor influence on the estimates of probability that two individuals sampled randomly share a single (Q_1 and Q_2) or pair (Q_3) of parents. For each analysis with Colony, we conducted a single

medium length run (5000N reconfigurations or 100N successful reconfigurations of the simulated annealing step, where N is the number of offspring) with medium precision (the number of elements in the transition matrix used for likelihood calculations) of the full-likelihood method with no sibship prior, a genotyping error rate of 1%, and either a polygamous or polygynous mating system. Estimates were made with both mating systems to evaluate the sensitivity of estimates of N_b as well as family size mean and variance to the choice of mating system. It is assumed that polygyny is the most common mating system in salmonids because females typically construct and defend a single redd fertilized predominantly by a single male, whereas males compete amongst each other and may leave a redd after mating to seek additional reproductive opportunities (Bentzen et al. 2001; Ackerman et al. 2017). The degree to which a dominant male can exclude subordinate males determines the frequency of polyandry. Jack Sockeye Salmon were observed to fertilize on average 47% of eggs deposited when spawning in the presence of a single dominant male (Foote et al. 1997). The departure from polygyny may be density dependent because increased densities promote alternative reproductive tactics (Fleming and Gross 1994) and reproductive dominance by a single male may be relaxed favoring a polygamous mating structure.

[C] *Simulations.* — Simulations were conducted within R (R Core Team, 2014) with the package PseudoBabies (Chapter 3) to evaluate the accuracy in estimating the mean and variance of the family sizes. Similarly, simulations were used to determine the accuracy of inferring full- and half-sibling dyads for estimating N_b with the sibship approach. Ten simulations were conducted based on data from each of the return years. Each simulation was initialized with all of the individuals sampled from a return year. These individuals were then randomly mated in a both a polygynous and polygamous mating design. The number of offspring that each pair produced was drawn from a logarithmic distribution (estimated from full-sibling family reconstruction with Colony for the focal years). Each simulation was run for thirteen years with the number of breeders equal to the return year size. Data were given 1% genotyping error and 1% missing data to reflect the uncertainty estimated in our datasets.

The uncertainty in full-sibling families deduced with Colony is expressed as the probability of inclusion and the probability of exclusion (P_{Inc} and P_{Exc} respectively). The P_{Inc} describes the probability that all members of the full-sibling family deduced are truly full-siblings. Small values of P_{Inc} indicate that an inferred family should be split into multiple full-sibling families. The P_{Exc} describes the probability that all the full siblings deduced for a family are actually full siblings and that no other individuals included in the sample are full siblings with individuals in the deduced family. Small values of P_{Exc} indicate that the deduced full-sibling family may be split. Families that are split cannot be combined; however, deduced families that show evidence that they should be split could lead to an overestimate of the mean and variance of family size and accepting only highly likely full-sibling families may increase our accuracy in estimating these parameters. We evaluated the effect that different P_{Inc} (0, 0.75, 0.95) have on resolving families and estimating the distribution parameters of family size with simulations.

The accuracy of the sibship approach for estimating N_b is determined by the ability of the marker panel to identify simulated full- and half-sibling dyads accurately. Half-sibling relationships are more difficult to resolve than full-siblings. In the absence of highly informative markers, loosely related individuals such as cousins may be inferred as half-siblings. Similarly, avuncular relationships have the same kinship coefficient as half-siblings. Jacks that spent a single year in freshwater (age 3) may have aunts and uncles returning in the same year (age 6); however, the frequency of these jacks is relatively low. We evaluated the simulations detailed above for their accuracy in full- and half-sibling assignment. We compared the full and half-sibling inference for the maximum likelihood configuration at the end of the analysis. Unfortunately, there is no straightforward way to use the probability associated with inferred full- and half-sibling dyads to determine if cutoff values increase accuracy of the estimator. During the search for a maximum likelihood arrangement of individuals into family groups, Colony stores configurations with high likelihoods, defined as having a log likelihood value greater than the maximum likelihood configuration minus 10. The uncertainty in full- and half-sibling dyads is estimated by evaluating how often the maximum likelihood dyads were sampled in the stored high likelihood configurations. While we could estimate Q_1, Q_2,

and Q_3 after applying a cutoff for P_{Inc} to produce a modified estimate of N_b, it is not possible to apply that cutoff to the configuration archive to develop confidence intervals.

We also estimated N_b for all of the simulated datasets with Colony and LDNe to evaluate the utility of each method for estimating N_b from a return year. Up to five separate brood years can contribute spawners to a single return year. The true N_b values to which estimates were compared were calculated by incorporating the variance in family size by estimating the effective number of males and females (with mean and variance used to simulate the data), for each of the five years contributing offspring to the return year sampled. The effective numbers of male and female spawners were then combined with Wright's sex ratio adjustment and the harmonic mean was used as the true N_b. True N_e was calculated as N_b times the generation length.

[A] Results

[B] Demographic estimates

The number of Sockeye Salmon returning to Auke Creek each year varied by an order of magnitude. The mean number of adults returning each year to Auke Creek since the installation of the modern weir was 2778 (± 1237 SD) with a maximum of 6123 and a minimum of 325 (Figure 2.2A). There was also substantial variation in the age distribution among return years (Figure 2.3). Five- and six-year-olds were the predominant age classes and were 5.4% to 86.4% of the return. The relative proportion of age-5 and age-6 fish was more variable prior to 1998, after which the proportions of age-6 fish never exceeded 60% and the proportion of age-5 fish was never lower than 32.5%. Three-year-old jacks were the least frequent age class and accounted for between 0 and 6.6% of the return. The generation length (average age at spawning) was 5.3 years.

Fluctuations in population size and variation in proportional contribution of a return year to the next generation made N_e smaller than N_c. With a nominal generation length of 5 years, return years could be grouped into seven generations in three ways that had reference years beginning in 1980, 1981, or 1982. Trends in N_e were consistent among the reference years; N_e began low, peaked in the middle of the time series, and declined gradually over the last two generations (Figure 2.2B). Low initial N_e values for the harmonic mean method stemmed from the multiple low return years from 1983 to

1986 (Figure 2.2A). Early low-return years combined with a large variance in λ further decreased estimates based on the Ryman and Laikre method (Table 2.1). The multigenerational N_e estimate over all reference years for the Ryman and Laikre method was 8603 fish with a mean N_e/N_c over generations of 0.73 (min = 0.41, max = 0.88).

Freshwater productivity and marine survival were estimated to evaluate their effect on fluctuations in the return size. The mean number of smolts per spawner was 7.76 (± 6.07 sd) with a maximum of 23.13 and a minimum of 0.43 (Figure 2.4A). A breakpoint analyses supported a 2-segment optimal partition with a breakpoint at brood year 1993 for the mean and a breakpoint at brood year 1991 for the variance. Between 1980 and 1993 spawners produced on average 10.89 (± 8.00 sd) smolts, which declined to 5.51 (± 2.66 sd) smolts per spawner after 1993. Female fecundity had a positive linear relationship with body length ($r_{adj}^2 = 0.208$, $P < 0.001$). The mean fecundity was 2805 (± 528 sd) eggs. If we assume a 19:20 ratio of males to females (see below) from the 1993 brood year on, the average number of smolts per female spawner was 10.76 (± 5.15 sd). The estimate of survival from PED to smolt was 0.4%. Marine survival was high for all years. The average marine survival was 17.09% (± 5.75% sd), and only twice fell below 10% (Figure 2.4B).

The effective population size was not appreciably affected by the sex ratio. The average proportion of male spawners per generation was 47.5%. The reduction in N_e caused by such a marginally skewed sex ratio would be less than 1%, whereas the average reductions over all generations caused by fluctuations in the population size was 17% and variation in productivity within a generation was 30%. Only a single year, prior to the installation of the modern weir, had a largely skewed sex ratio. In 1974, there were purportedly 3.4 males to every female. Even then, N_e would be reduced less than 1% because the two years preceding and following had approximately equal sex ratios; from 1972 to 1976 the return was 47% female (S. Vulstek, unpublished data).

In the absence of pedigree data, understanding the degree to which N_e is reduced because of differences in the mean and variance of the family size depends on our ability to reconstruct families from genetic data. Results from simulations showed that we had sufficient power to reconstruct full-sibling families and to accurately estimate the mean and variance of family size of the individuals in the sample. For simulations under both

mating structures, a 0.75 cutoff value for P_{Inc} for accepting a full sibling family group minimized the difference between the true mean and variance of family size, while still resolving more than 90% of the true families. On average there was a 1.5% difference between the true and inferred mean family size and a 6.7% difference between the true and inferred variance. Increasing the cutoff value of P_{Inc} resulted in larger discrepancies in the mean and variance estimate, but a larger proportion of the true families were reconstructed correctly (Supplemental Table S2.2). This tradeoff occurred because when a single individual was incorrectly included in a large family, the probability of inclusion was low and that family was split which decreased the mean family size, but only accounted for a single incorrectly identified full true family relationship. For these reasons, we reconstructed full-sibling families with a P_{Inc} cutoff value of 0.75 to estimate the mean and variance of family size for the empirical data. Despite the fact that we could estimate the mean and variance of the families from a single return year, simulations indicated that because individuals within a full-sibling family can mature at different ages, the true mean and variance were underestimated.

The distribution of full-sibling family sizes reconstructed with the program Colony did not fit a Poisson distribution (Table 2.2). Before estimating family sizes for the combined year analysis, potential offspring of spawners in 2008 were removed from the 2011 dataset. Eight of the 118 jacks sampled in the 2011 return year were identified as age-3 jacks (age 1.1) and removed before sibship analysis. The remaining 110 jacks were likely age-4 (two year freshwater and 1 year ocean). For each return year and the combined years, full-sibling family size more closely followed a logarithmic distribution (negative binomial with a size parameter of zero). Differences in the mean and variance in family sizes between analyses with polygamous or polygynous mating systems differed by an average of 1.6% and 1.7% respectively. Estimates of N_b based on the mean and variance of family size differed by 1.2% between the two mating systems. Because of the similarities, we report results from the polygamous mating system below. The mean family size across all data sets was 1.77 with a variance of 3.05 (Table 2.2). Results from all collections except 2008 were similar. The inequality between the mean and variance in family size for these years reduced N_e relative to the population size by 32% on average. Unexpectedly, the variance for the 2008 data was smaller than the mean. Family

sizes for the 2008 return year ranged from 1 to 7 fish, while families from 2009 and 2011 ranged from 1 to 31 fish. The vast majority of fish (72%) that returned in 2008 were five years of age and came from the 2003 brood year. Six-year-old fish from the same brood year accounted for a smaller overall proportion of the 2009 return (40%); but because the return was so large, almost twice as many fish from that brood year were predicted to have matured as age six fish. When analyzed together with fish from the 2009 and 2011 return years, 14 full-sibling families included fish from 2008 and 2011, 343 families included fish from 2008 and 2009, and only two families had fish that returned in all three years. Estimates of the mean and variance in family size for the 2008 dataset were robust to the markers used in family reconstruction. The results were consistent when the dataset for 2008 was reduced to the same marker panel used in 2009 and 2011. Similarly, results were robust to the P_{Inc} cutoff value used to identify families. On average, estimates of N_b made with a P_{Inc} cutoff of 0 (retaining all inferred FS families) was 10% higher than those made with a cutoff value of 0.95 (retaining FS families with a high probability of including only true FS).

[B] Genetic estimates

Simulations indicated that estimates of N_b from both the sibship and LD approach from Plan 1 sampling were much closer to N_b than N_e rather than intermediate between the two. Estimates made with the LD method typically exceeded those from the sibship method and were closer to the true N_b. There was a 28% difference between the true and estimated N_b with the sibship approach. In 43% of the simulations, the true value was within the sibship 95% confidence interval. Estimates made with the LD approach were closer to the true value, only differing by 12% on average and the true value of N_b was within the 95% confidence interval in 62% of the simulations. The largest differences were for simulations based on the 2008 data, where the variance in family size was smaller than the mean. If we omit those simulations, the differences between the true and estimated N_b were 5.5% and 8.7% for the sibship and LD methods respectively.

The accuracy of the sibship approach is determined mainly by the ability to accurately estimate the probability of drawing a pair of full- and half-siblings from the population. Results from simulations suggest that the ability to correctly infer half-sibling

dyads was low. The mean proportion of true half-sibling dyads that were inferred correctly was 48.4% and 50.0% for the polygynous and polygamous simulations, respectively. Of the half-sibling dyads inferred by the program Colony, on average only 14.5% and 26.1% were correct for the polygynous and polygamous simulations, respectively. Colony inferred, on average, 2.7 times the number of half-sibling dyads that were actually present in the dataset and 1.7 times the number of true full-sibling dyads. The tendency of Colony to infer more full- and half-sibling dyads resulted in a negative bias in the estimation of N_b. If the Q values were known without error, estimates of N_b fell between N_b and N_e as is expected from a collection of multiple cohorts in a return year. The negative bias effectively drives the estimate toward the N_b for each year (Supplementary Table 2.3S).

Based on simulation results, estimates of N_b from Plan 1 sampling may best represent lower bound estimates. While the choice of mating structure within the program Colony had minimal effect on estimates of N_b, substantial differences existed between the sibship and LD approaches (Table 2.3). Estimates of N_b from the sibship approach of Colony were on average twice as large as those from the linkage disequilibrium method. The lowest estimate of N_b was for the 2011 collection, despite the fact that the 2008 return was nearly half the size. A lower bound for N_e for both these methods was made with the harmonic mean of the individual collection estimates of N_b multiplied by the generation length. For the sibship approach of Colony N_e was 5322, with a ratio of N_e/N_c of 0.37. For the LD method N_e was 2994, with a ratio of N_e/N_c of 0.21.

Of the 7683 fish genotyped, 525 could be assigned to a brood year based on scale age. On average, 8% of each brood year was aged and genotyped. Sample sizes ranged from 90 individuals from 2006 to 152 individuals from 2003 and 2005 (Table 2.4). There was substantial variation among brood years in N_b with a harmonic mean of 590. The effective population size, calculated as the harmonic mean of the brood year N_b estimates multiplied by the generation length, from Plan 2 sampling was 2945 with a ratio of N_e/N_c of 0.21.

[A] Discussion

[B] Demographic estimates

Within the Auke Creek population, reduced run sizes in the early 1980s and variable contribution of each return year within the earliest generation led to the smallest estimates of N_e. While N_e for each generation increased over the study period, the multigenerational estimate showed a persistent effect of the reduced run size from early in the time series. In only two instances were the estimates from the harmonic mean smaller than those based on the Ryman and Laikre method. These results were in agreement with Waples's (2002) conclusion that high variance in recruits per spawner results in a larger inbreeding coefficient and a smaller effective population size.

The high variance in recruits per spawner observed in Auke Lake in the early 1980s likely resulted from variation in freshwater survival. Mean freshwater productivity and the variance around it decreased in later generations. Previous work indicated that increased precipitation during the incubation period had the largest positive effect on Sockeye Salmon smolt production in Auke Lake, presumably by increasing oxygen circulation and preventing the streambed from freezing (Fukushima 1996). Fukushima (1996) only evaluated the spawner-smolt relationship with environmental variables through the 1991 brood year. The shift in production we observed after that year warrants a reevaluation of the environmental variables that may govern freshwater productivity. Our estimate of potential egg deposition (PED) to smolt survival was five times lower than the 2% egg to smolt survival for Sockeye Salmon reported by Bradford (1995). In the estimates of smolt and recruits per spawner, we made the implicit assumption of negligible pre-spawning mortality. In 2012, pre-spawn mortality (PSM) in Auke Lake Sockeye Salmon was estimated to be 11.3% based on radio-tagged fish (Ray et al. 2015). If we applied this estimate of pre-spawn mortality, the average smolts per spawner since 1993 would increase to 6.37 and PED to smolt survival would increase to 0.5%. Pre-spawn mortality is unlikely to be constant through time. If PSM in Auke Lake has increased since the early 1990s, our estimates of smolts per spawner would have a negative bias. Auke Creek has significantly warmed over the study period (Kovach et al. 2013). High water temperatures have been linked to increased PSM in Sockeye Salmon (Crossin et al. 2008) and the effect can be intensified by large numbers of returning fish

(Tillotson and Quinn 2017). Although Pink and Sockeye salmon within Lake Creek typically favor dissimilar spawning habitat, in years of high spawner abundance, the amount of spatial overlap between the two species increased (Fukushima and Smoker 1997). Years of high Pink Salmon abundance could increase the density-dependent effect that high temperatures have on lowering the dissolved oxygen of the creek and increase PSM (Sergeant et al. 2017).

Estimates of smolt to adult (marine) survival were high throughout the time series. They were, on average, twice the survival reported for Sockeye Salmon from Washington state to coastal Western Alaska (Bradford 1995). While marine survival estimates appear to be quite high for each brood year, without such high marine survival the population would likely crash. A marine survival of 20% is required to maintain a stable population if 10 smolts are produced per female spawner (5 smolts per spawner assuming a 50:50 sex ratio). Marine survival for Coho Salmon in Auke Creek is similarly high, averaging 28% over the same time period (Russell 2019). Fishery mortality for Coho Salmon was, on average, 11% and survival to escapement averaged 17%. The high rates of marine survival may be due to the fact that smolts pass through only 0.65 km of Auke Creek to reach the ocean without a complex estuary. The seemingly similar marine survival between Coho and Sockeye Salmon in the system is somewhat surprising given that Coho Salmon have up to twice the survival of Sockeye Salmon from smolt to adult (Bradford 1995). The discrepancy in Auke Creek Sockeye Salmon may arise from the fact that the nearshore marine waters, creek, and lake have been closed to sport fishing since 1980. Similarly, adult Sockeye Salmon migrate earlier than Coho Salmon and likely do not suffer the same mortality from commercial and recreational fisheries, although the extent to which Auke Creek Sockeye Salmon contribute to these fisheries is unknown.

The sex ratio of spawners in Auke Creek has been approximately equal over the study period. The multiple ages at maturity and long generation length of Auke Creek Sockeye Salmon buffer it from the effects of a single year of highly skewed sex ratios. It does not appear that an unbalanced sex ratio is a major demographic factor affecting N_e/N_c. However, sex was not assigned without error, particularly early in the run when returning fish were not fully mature and secondary sexual characteristics were less well developed.

The observed deviation from a Poisson distribution of offspring appeared to have as large an effect as the variation in λ among years within a generation, and the effect was much larger than the effect of simple fluctuations in population size. The long generation time of Sockeye Salmon and variation in age at maturity ensure that a single poor brood year does not result in large fluctuations in the population size; however, variability in the freshwater environment may drive large fluctuations in annual recruits per spawner. Since 1993, the variability and overall productivity have decreased. The distribution of offspring can be non-random because of genetic factors, life history characteristics, or environmental perturbations. In Pink Salmon, correlated marine survival among relatives increased variance in family size (Geiger et al. 1997) and family-correlated survival in hatchery reared Coho Salmon reduced N_e by 20% relative to that expected given random family survival (Moyer et al. 2007). Variable habitat for redd construction, or spawn and/or emergence timing, could lead to differential familial freshwater survival and increased variance in family size. While these factors may cause inter-annual variation in V_k, because each generation is composed of multiple breeding events, the variance within a generation will be dampened relative to that of a single year. The average index of variability observed (2.17) is consistent with those of other salmonid species: 2.96 for Pink Salmon (Geiger et al. 1997), 2.56 for Steelhead (average of unscaled values; Araki et al. 2007), and 1.64 for Coho Salmon (Moyer et al. 2007). Of note was the small index of variability observed for the 2008 return year. It is unlikely that more uniform survival across families resulted in this low value, but rather that the full sibling families from brood year 2003 were divided between two return years. If our estimates of the mean and variance are biased high because of the individuals within full-sibling families maturing at different ages, then N_b will be overestimated. Current and future genotyping efforts will improve the assignment of fish to brood year and the estimation of accurate family size parameters to assess the amount of bias that arises from making estimates from individual return years.

[B] Genetic estimates

Estimates of N_b made with genetic methods likely represent lower bound values based on simulations. It has been proposed that estimates made with the LD method from

a sample of multiple cohorts should be intermediate between N_b and N_e (Waples 2005; Luikart et al. 2010), but our simulations indicated that estimates with both LDNe and the sibship method were closer to the true N_b. For LDNe, N_b is likely underestimated because of additional LD introduced by the Wahlund effect of multiple cohorts and structure within returns (Chapter 1; Waples et al. 2014) exacerbated by the complete sampling of the population ensuring that each cohort was represented in the genetic samples. Similarly, demographic estimates indicated that the effective population size increased from 1980 to 2005. Residual disequilibrium from the first two generations, when the population size was smaller, may still contribute to downward bias in the contemporary estimates despite the fact that residual disequilibrium will be filtered through multiple years due to variation in age at maturity (Waples 2005). The underestimation of N_b with the sibship approach of Colony resulted from an overestimation of the number of full and half-sibling dyads. Without highly informative markers, distantly related individuals (first and second cousins) may have been incorrectly identified as half-siblings. While estimates of N_b from the sibship method were consistently lower than those of the LD method in simulations, we observed the opposite in the empirical data. This likely is a result of additional linkage disequilibrium that was not accounted for in the simulations such as non-random mating and temporal population structure (Chapter 1).

It is undoubtedly preferable to estimate N_b using genetic data from a single cohort, but estimates made from Plan 1 sampling (full return year) may be convenient and useful. Although estimates of N_b from Plan 1 sampling are not intuitively linked to a single brood year, they instead represent a composite N_b over the last generation. Despite this somewhat convoluted interpretation of N_b estimates from Plan 1 sampling, estimates of N_e and N_e/N_c between the two sampling plans were quite similar. It is often easy to sample adult fish as they pass through a weir, but scale ages may be impractical to collect. Similarly, available genetic datasets may have not been collected with N_e estimation in mind, but could prove useful in understanding region-wide patterns or identifying populations of specific management concern.

[B] Outlook for Auke Creek

The lower bound estimate of N_e from genetic methods is well above the revised minimum effective size of 100/1000 proposed by Frankham et al. (2014) suggesting that for now Auke Creek Sockeye Salmon should retain their evolutionary potential and be able to adapt to a constantly changing environment. Current estimates of N_e/N_c based on three years of returns were much higher than the median values reported by Frankham (1995) and Palstra and Ruzzante (2008); however, they are similar to those of other highly fecund fishes (mean = 0.31, Frankham 1995). Continued genetic sampling of Auke Creek Sockeye Salmon will facilitate a more accurate estimate of N_b and N_e through the construction of full pedigrees.

Freshwater production of smolts appears to have decreased over the study period, which may have been caused by a reduction in spawning habitat, increased juvenile mortality, increased pre-spawning mortality, climate change, or interactions among these factors. At least 50% of the Auke Lake shoreline has been urbanized, with increasing sedimentation and pollution resulting from greater human presence in the watershed (Moles and Marty 2005; Rice et al. 2008; Blanc et al. 2010; Bethers et al. 2012). Climate warming could also be changing lake productivity and food-web structure, juvenile growth, and age at smoltification (Kovach et al. 2013, 2014; Carter et al. 2017).

In conclusion, variance in family size, fluctuations in population size, and variable contribution to the next generation by different brood years within a generation were the major demographic factors that determine N_e in Auke Creek Sockeye Salmon. Even lower bound estimates of N_e for the Auke Creek population suggest that this population should be resilient to stochastic events; however, chronic stressors in the freshwater environment may limit freshwater productivity and will only intensify with projected changes from climate change and increased urbanization of the area.

[A] References

Ackerman, M., B. Hand, R. S. Waples, G. Luikart, R. Waples, C. Steele, B. Garner, J. McCane, and M. Campbell. 2017. Effective number of breeders from sibship reconstruction: Empirical evaluations using hatchery Steelhead. Evolutionary Applications 10(2):146–160.

Araki, H., R. S. Waples, W. R. Ardren, B. Cooper, and M. S. Blouin. 2007. Effective population size of Steelhead Trout: Influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. Molecular Ecology 16(5):953–966.

Beacham, T., N. Varnavskaya, K. Le, and M. Wetklo. 2008. Determination of population structure and stock composition of Chum Salmon (*Oncorhynchus keta*) in Russia determined with microsatellites. Fishery Bulletin 106(3):245–256.

Bethers, M., K. Munk, and C. Seifert. 2012. Juneau fish habitat assessment (Revised). Alaska Department of Fish and Game Regional Information Report 5J12-03, Douglas, Alaska.

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(2):127–136.

Blanc, A., L. Holland, S. Rice, and C. Kennedy. 2010. Anthropogenically sourced low concentration PAHs: In situ bioavailability to juvenile Pacific salmon. Ecotoxicology and Environmental Safety 73(5):849–857.

Booy, G., R. J. J. Hendriks, M. J. M. Smulders, J. M. Van Groenendael and B. Vosman. 2000. Genetic diversity and the survival of populations. Plant Biology 2: 379–395.

Bradford, M. 1995. Comparative review of Pacific salmon survival rates. Canadian Journal of Fisheries and Aquatic Sciences 52(6):1327–1338.

Caballero, A. 1994. Developments in the prediction of effective population size. Heredity 73(6):657–679.

Cairney, M., J. Taggart, and B. Hyheim. 2000. Characterization of microsatellite and minisatellite loci in Atlantic Salmon (*Salmo salar* l.) and cross-species amplification in other salmonids. Molecular Ecology 9(12):2175–2178.

Campbell, N., and S. Narum. 2011. Development of 54 novel single-nucleotide polymorphism (SNP) assays for Sockeye and Coho salmon and assessment of available SNPs to differentiate stocks within the Columbia River. Molecular Ecology Resources 11(1):20–30.

Carter, J. L., D. E. Schindler, and T. B. Francis. 2017. Effects of climate change on zooplankton community interactions in an Alaskan lake. Climate Change Responses 4:3.

Clutter, R. and L. Whitsel. 1956. Collection and interpretation of Sockeye Salmon scales. International Pacific Salmon Fisheries Commission, New Westminster, Canada.

Crossin, G., S. Hinch, S. Cooke, D. Welch, D. Patterson, S. Jones, A. Lotto, R. Leggatt, M. Mathes, J. Shrimpton, G. Van Der Kraak, and A. Farrell. 2008. Exposure to high temperature influences the behaviour, physiology, and survival of Sockeye Salmon during spawning migration. Canadian Journal of Zoology 86(2):127–140.

Crow, J. F., and C. Denniston. 1988. Inbreeding and variance effective population numbers. Evolution 42(3):482–495.

Crow, J., and N. Morton. 1955. Measurement of gene frequency drift in small populations. Evolution 9:202–214.

De Meester, L., R. Stoks, and K. Brans. 2018. Genetic adaptation as a biological buffer against climate change: Potential and limitations. Integrative Zoology 13:372–91.

Do, C., R. Waples, D. Peel, G. Macbeth, B. Tillett, and J. Ovenden. 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. Molecular Ecology Resources 14(1):209–214.

Einum, S., I. A. Fleming, I. M. Cote, and J. D. Reynolds. 2003. Population stability in salmon species: effects of population size and female reproductive allocation. Journal of Animal Ecology 72:811–821.

Elfstrom, C., C. Smith, and J. Seeb. 2006. Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of Sockeye Salmon. Molecular Ecology Notes 6(4):1255–1259.

Faber, D. 2008. Rising land falling fishery: the effects of isostatic rebound and rapid succession on east Alsek River Sockeye Salmon. Master's thesis. University of Alaska Fairbanks, Alaska.

Fleming, I., and M. Gross. 1994. Breeding competition in a Pacific salmon (Coho: *Oncorhynchus kisutch*): Measures of natural and sexual selection. Evolution 48(3):637–657.

Foote, C. J., G. S. Brown, and C. C. Wood. 1997. Spawning success of males using alternative mating tactics in Sockeye Salmon, *Oncorhynchus nerka*. Canadian Journal of Fisheries and Aquatic Sciences 54(8):1785–1795.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: A review. Genetical Research 66(2):95–107.

Frankham, R., C. Bradshaw, and B. Brook. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, red list criteria and population viability analyses. Biological Conservation 170:56–63.

Franklin, I. 1980. Evolutionary change in small populations. Pages 135–149 *in* M. Soule and B. Wilcox, editors. Conservation biology - an evolutionary-ecological perspective. Sinauer Associates, Sunderland, Massachusetts

Franklin, I., F. Allendorf, and I. Jamieson. 2014. The 50/500 rule is still valid - reply to Frankham et al. Biological Conservation 176:284–285.

Franklin, I., and R. Frankham. 1998. How large must populations be to retain evolutionary potential? Animal Conservation 1(1):69–70.

Fukushima, M. 1996. Effects of density-dependence, environment and species interaction during spawning and incubation on population dynamics of Pink and Sockeye salmon in the Auke Lake system, southeast Alaska. Doctoral dissertation. University of Alaska Fairbanks, Alaska.

Fukushima, M., and W. Smoker. 1997. Determinants of stream life, spawning efficiency, and spawning habitat in Pink Salmon in the Auke Lake system, Alaska. Canadian Journal of Fisheries and Aquatic Science 54(1):96–104.

Geiger, H. J., W. W. Smoker, L. A. Zhivotovsky, and A. J. Gharrett. 1997. Variability of family size and marine survival in Pink Salmon (*Oncorhynchus gorbuscha*) has implications for conservation biology and human use. Canadian Journal of Fisheries and Aquatic Sciences 54(11):2684–2690.

Gilk-Baumer, S., S. Rodgers Olive, D. Harris, S. Heinl, E. Fox, and W. Templin. 2015. Genetic mixed stock analysis of Sockeye Salmon harvested in selected northern Chatham Strait commercial fisheries, southeast Alaska, 2012-2014. Alaska Department of Fish and Game Fishery Data Series No. 15-03, Anchorage, AK.

Gilpin, M. E., and M. Soulé. 1986. Minimum viable populations: Process of species extinctions. Pages 19–34 in M. Soulé, editor. Conservation biology: The science of scarcity and diversity. Sinauer Associates, Sunderland, Massachusetts.

Habicht, C., L. Seeb, K. Myers, E. Farley, and J. Seeb. 2010. Summer-fall distribution of stocks of immature Sockeye Salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. Transactions of the American Fisheries 139(4):1171–1191.

Hartl, D., and A. Clark. 1997. Principles of population genetics 3rd edn. Sinauer Associates, Sunderland, Massachusetts.

Hill, W. 1979. A note on effective population size with overlapping generations. Genetics 92:317–322.

Hill, W. 1981. Estimation of effective population size from data on linkage disequilibirum. Genetics Research 38:209–216.

Jamieson, I., and F. Allendorf. 2012. How does the 50/500 rule apply to MVPs? Trends in Ecology and Evolution 27(10):578–584.

Jones, O., and J. Wang. 2010. COLONY: A program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10(3):551–555.

Killick, R. and I. Eckley. 2014. Changepoint: An R package for changepoint analysis. Journal of Statistical Software 58(3):1–19.

Kovach, R., J. Joyce, J. Echave, M. Lindberg, and D. Tallmon. 2013. Earlier migration timing, decreasing phenotypic variation, and biocomplexity in multiple salmonid species. PLoS ONE 8(1).

Kovach, R., J. Joyce, S. Vulstek, E. Barrientos, and D. A. Tallmon. 2014. Variable effects of climate and density on the juvenile ecology of two salmonids in an Alaskan lake. Canadian Journal of Fisheries and Aquatic Sciences 71(6):799–807.

Lande, R. 1995. Mutation and conservation. Conservation Biology 9(4):782–791.

Luikart, G., N. Ryman, D. A. Tallmon, M. K. Schwartz, and F. W. Allendorf. 2010. Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. Conservation Genetics 11(2):355–373.

Mace, G., and R. Lande. 1991. Assessing extinction threats: Toward a reevaluation of IUCN threatened species categories. Conservation Biology 5(2):148–157.

Miller, K., and T. Beacham. 2007. Single nucleotide polymorphisms identified in Sockeye Salmon. Report to Northern Endowment Fund, Project NF 2006 I 4.

Moles, A., and G. Marty. 2005. Physiological changes in Prickly Sculpin (*Cottus asper*) inhabiting a lake used by jet-propelled watercraft. Bulletin of Environmental Contamination and Toxicology 74(6):1151–1158.

Morris, D., K. Richard, and J. Wright. 1996. Microsatellites from Rainbow Trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. Canadian Journal of Fisheries and Aquatic Sciences 53(1):120–126.

Motro, U. and G. Thomson. 1982. On heterozygosity and effective size of populations subject to size change. Evolution 36:1056–1060.

Moyer, G. R., M. S. Blouin, and M. A. Banks. 2007. The influence of family-correlated survival on N_b/N for progeny from integrated multi- and single-generation hatchery stocks of Coho Salmon (*Oncorhynchus kisutch*). Canadian Journal of Fisheries and Aquatic Sciences 64(9):1258–1265.

Nehlsen, W., J. E. Williams, and J. A. Lichatowich. 1991. Pacific salmon at the crossroads: Stocks at risk from California, Oregon, Idaho, and Washington. Fisheries 16(2):4–21.

Nelson, R., and T. Beacham. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. Animal Genetics 30(3):228–229.

Nunney, L., and K. Campbell. 1993. Assessing minimum viable population size: Demography meets population genetics. Trends in Ecology and Evolution 8(7):234–239.

Olsen, J., S. Wilson, E. Kretschmer, K. Jones, and J. Seeb. 2000. Characterization of 14 tetranucleotide microsatellite loci derived from Sockeye Salmon. Molecular Ecology 9(12):2185–2187.

Palstra, F., and D. Ruzzante. 2008. Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity for wild population persistence? Molecular Ecology 17(15):3428–3447.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ray, J., S. Sethi, J. Eiler, and J. Joyce. 2015. Prespawning movements and spawning distribution of Sockeye Salmon in an urbanizing Alaskan lake. Journal of Fish and Wildlife Management 6(2):472–485.

Reed, D., and E. Bryant. 2000. Experimental tests of minimum viable population size. Animal Conservation 3(1):7–14.

Rice, S., L. Holland, and A. Moles. 2008. Seasonal increases in polycyclic aromatic hydrocarbons related to two-stroke engine use in a small Alaskan lake. Lake and Reservoir Management 24(1):10–17.

Riester, M., P. F. Stadler, and K. Klemm. 2009. FRANz: Reconstruction of wild multi-generation pedigrees. Bioinformatics 25(16):2134–2139.

Ruggerone, G., and J. Irvine. 2018. Numbers and biomass of natural- and hatchery-origin Pink Salmon, Chum Salmon, and Sockeye Salmon in the North Pacific Ocean, 1925–2015. Marine and Coastal Fisheries 10(2):152–168.

Russell, J. 2019. Evaluation of marine and freshwater growth and survival of Auke Creek Coho Salmon. Master's thesis. University of Alaska Fairbanks, Alaska.

Ryman, N., and L. Laikre. 1991. Effects of supportive breeding on the genetically effective population size. Conservation Biology 5(3):325–329.

Schoen, E., M. Wipfli, E. Trammell, D. Rinella, A. Floyd, J. Grunblatt, M. McCarthy, B. Meyer, J. Morton, J. Powell, A. Prakash, M. Reimer, S. Stuefer, H. Toniolo, B. Wells, and F. Witmer. 2017. Future of Pacific salmon in the face of environmental change: Lessons from one of the world's remaining productive salmon regions. Fisheries 42(10):538–553.

Scribner, K., J. Gust, and R. Fields. 1996. Isolation and characterization of novel salmon microsatellite loci: Cross-species amplification and population genetic applications. Canadian Journal of Fisheries and Aquatic Sciences 53(4):833–841.

Seeb, J. E., C. E. Pascal, R. Ramakrishnan, and L. W. Seeb. 2009. SNP genotyping by the 5'-nuclease reaction: Advances in high-throughput genotyping with non-model organisms. Methods in Molecular Biology 578:277–292.

Sergeant, C. J., J. R. Bellmore, C. McConnell, and J. W. Moore. 2017. High salmon density and low discharge create periodic hypoxia in coastal rivers. Ecosphere 8(6):e01846.

Sinnock, P. 1975. The Wahlund effect for the two-locus model. American Naturalist 109:565–570.

Slaney, T. L., K. D. Hyatt, T. G. Northcote, and R. J. Fielden. 1996. Status of anadromous salmon and trout in British Columbia and Yukon. Fisheries 21(10):20–35.

Smith, C., C. Elfstrom, L. Seeb, and J. Seeb. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14(13):4193–4203.

Smith, C., B. Koop, and R. Nelson. 1998. Isolation and characterization of Coho Salmon (*Oncorhynchus kisutch*) microsatellites and their use in other salmonids. Molecular Ecology 7(11):1614–1616.

Storer, C., C. Pascal, S. Roberts, W. Templin, L. Seeb, and J. Seeb. 2012. Rank and order: Evaluating the performance of SNPs for individual assignment in a non-model organism. PLoS ONE 7(11).

Tillotson, M., and T. Quinn. 2017. Climate and conspecific density trigger pre-spawning mortality in Sockeye Salmon (*Oncorhynchus nerka*). Fisheries Research 188:138–148.

Van Alen, B. 2000. Status and stewardship of salmon stocks in southeast Alaska. Pages 161–193 *in* E. Knudsen, C. Steward, D. D. MacDonald, J. Williams, and D. Reiser, editors. Sustainable fisheries management: Pacific salmon. Lewis Publishers, Boca Raton, Florida.

Vucetich, J., T. Waite, and L. Nunney. 1997. Fluctuating population size and the ratio of effective to census population size. Evolution 51(6):2017–2021.

Wang, J. 2005. Estimation of effective population sizes from data on genetic markers. Philosophical Transactions of the Royal Society B: Biological Sciences 360(1459):1395–1409.

Wang, J. 2009. A new method for estimating effective population sizes from a single sample of multilocus genotypes. Molecular Ecology 18(10):2148–2164.

Waples, R. 1990. Conservation genetics of Pacific salmon. iii. Estimating effective population size. Journal of Heredity 81(4):277–289.

Waples, R. 1991. Genetic methods for estimating the effective size of cetacean populations. Report - International Whaling Commission, Special Issue 13:279–300.

Waples, R. 2002. Effective size of fluctuating salmon populations. Genetics 161(2):783–791.

Waples, R. 2005. Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? Molecular Ecology 4:3335 –3352.

Waples, R. 2006. Seed banks, salmon, and sleeping genes: Effective population size in semelparous, age-structured species with fluctuating abundance. American Naturalist 167(1):118–135.

Waples, R., T. Antao, and G. Luikart. 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics 197(2):769–780.

Waples, R., and C. Do. 2008. LDNE: A program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8(4):753–756.

Waples, R., and C. Do. 2010. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. Evolutionary Applications 3(3):244–262.

Waples, R., G. Luikart, J. Faulkner, and D. Tallmon. 2013. Simple life-history traits explain key effective population size ratios across diverse taxa. Proceedings of the Royal Society B: Biological Sciences 280(1768).

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Wright, S. 1938. Size of population and breeding structure in relation to evolution. Science 87:430–431.

Wright, S.1969. Evolution and the genetics of populations. V. II. The theory of gene frequencies. University of Chicago Press, Chicago.

Table 2.1: Spawner-recruit data for Auke Creek, AK Sockeye Salmon from 1980 to 2009. The data are arranged into six five-year generations beginning in 1980.

Return Year	Spawners N_t	Recruits R_t	R_t/N_t (λ_t)	X_t (R_t/R_T)
1980	4570	195	0.043	0.019
1981	4089	3436	0.840	0.330
1982	1334	1245	0.934	0.119
1983	1805	1913	1.060	0.184
1984	975	3636	3.729	0.349
N_T, R_T	*12773*	*10425*		
1985	325	729	2.243	0.056
1986	1033	3114	3.014	0.237
1987	2896	2468	0.852	0.188
1988	1392	3399	2.442	0.259
1989	2807	3418	1.218	0.260
N_T, R_T	*8453*	*13128*		
1990	3452	3300	0.956	0.178
1991	2764	8912	3.224	0.481
1992	1668	3377	2.025	0.182
1993	3058	971	0.317	0.052
1994	3869	1986	0.513	0.107
N_T, R_T	*14811*	*18546*		
1995	3371	2378	0.705	0.146
1996	6123	3528	0.576	0.216
1997	4705	4598	0.977	0.281
1998	2139	2282	1.067	0.140
1999	1681	3551	2.112	0.217
N_T, R_T	*18019*	*16336*		
2000	2513	1963	0.781	0.154
2001	4009	3005	0.750	0.236
2002	3012	1405	0.466	0.110
2003	3397	2800	0.824	0.220
2004	2978	3562	1.196	0.280
N_T, R_T	*15909*	*12735*		
2005	3019	2415	0.800	0.193
2006	1868	1607	0.860	0.129
2007	2942	1387	0.471	0.111
2008	1263	1589	1.258	0.127
2009	4048	5507	1.360	0.440
N_T, R_T	*13140*	*12504*		

Table 2.2: Full-sibling families of Auke Creek, AK Sockeye Salmon were reconstructed with Colony v2 to estimate the mean (k) and variance (V_k) in family size. A probability cutoff of 0.75 was used to accept full-sibling families. The distribution of family sizes was tested for conformance to the Poisson distribution with a G-test with degrees of freedom (df) and probability (P). The effective number of breeders (N_b) and the ratio of N_e to the population size (N) were estimated. The AllYears collection does not equal the sum of years 2008, 2009, and 2011 because jacks from 2011 have been removed.

				Family Size							
	Collection	N	Families	$\bar{k}$	V_k	$V_k/\bar{k}$	G	df	P	N_b	N_e / N
Polygamy	2008	1254	994	1.26	0.50	0.40	1154.50	6	< 0.001	2405	1.60
	2009	4057	2130	1.90	3.15	1.65	2045.71	14	< 0.001	3021	0.78
	2011	2380	1230	1.93	4.23	2.18	1404.87	15	< 0.001	1476	0.64
	AllYears	7683	3848	2.00	4.33	2.17	4337.30	22	< 0.001	4849	0.63
Polygyny	2008	1254	971	1.29	0.52	0.40	1049.27	6	< 0.001	2342	1.59
	2009	4057	2097	1.93	3.15	1.63	1901.75	16	< 0.001	3059	0.78
	2011	2380	1226	1.94	4.26	2.19	1399.46	15	< 0.001	1473	0.64
	AllYears	7683	3763	2.04	4.42	2.17	4052.77	22	< 0.001	4891	0.62

Table 2.3: Estimates of the effective number of Auke Creek Sockeye Salmon breeders made with the sibship approach of Colony and the linkage disequilibrium approach of LDNe for Plan 1 sampling.

		Colony - Polygamy			Colony - Polygyny			LDNe		
Collection	N	N_b	95%CI(L)	95%CI(U)	N_b	95%CI(L)	95%CI(U)	N_b	95%CI(L)	95%CI(U)
2008	1254	991	900	1083	1132	1034	1244	654	603	711
2009	4057	1619	1510	1734	1767	1657	1879	783	740	828
2011	2380	839	761	923	900	820	981	454	429	481
AllYrs	7683	2530	2397	2669	2671	2524	2813	1023	972	1076

Table 2.4: Estimates of the effective number of Auke Creek Sockeye Salmon breeders (N_b) made with the linkage disequilibrium approach of LDNe for Plan 2 sampling of scale-aged fish. The sample size of genotyped fish of known age (n), parametric 95% confidence intervals, the census size of the brood year (N_c) and the ratio of the effective number of breeders to the census size (N_b / N_c) are given.

Brood Year	n	N_b	95% CI	N_c	N_b / N_c
2003	152	988	574 - 3136	3397	0.29
2004	131	624	405 - 1287	2978	0.21
2005	152	307	243 - 409	3019	0.10
2006	90	1089	448 - Inf.	1868	0.58

Figure Captions

Figure 2.1: Map of Auke Lake located in Southeast Alaska. The primary spawning tributaries of Lake Creek and Lake Two Creek are located along the Northern shoreline. Asterisks denote unofficial stream names.

Figure 2.2: Auke Creek, AK Sockeye Salmon population size through time. (A) Annual total return to the Auke Creek weir from 1980 to 2017. (B) Generational total return (*N*) and the effective population sizes based on the harmonic mean and Ryman and Laikre (RL_91) methods. The nominal generation length was 5 years. Over the 38-year data set, three years were omitted from either the start or end of the time series. The reference year, or year that the first generation began, is listed above the panel.

Figure 2.3: The age distribution of Sockeye Salmon returning to Auke Lake, AK determined from scale aging (*n* = sample size of scale aged fish). Fish returned as 3, 4, 5, and 6-year-old fish.

Figure 2.4: (A) Freshwater productivity of Auke Creek, AK Sockeye Salmon was characterized by the number of smolts per spawner produced from brood years 1980 to 2014. The horizontal dashed lines represent the mean smolts per spawner produced in each time period, and the dark grey shading represents the variance around the mean. (B) Marine survival was estimated for brood years 1980 to 2011 with a LOESS fit (blue line) with standard error (gray shading). The estimate from 1981 has been removed because of a sampling issue that led to marine survivals in excess of 100%.

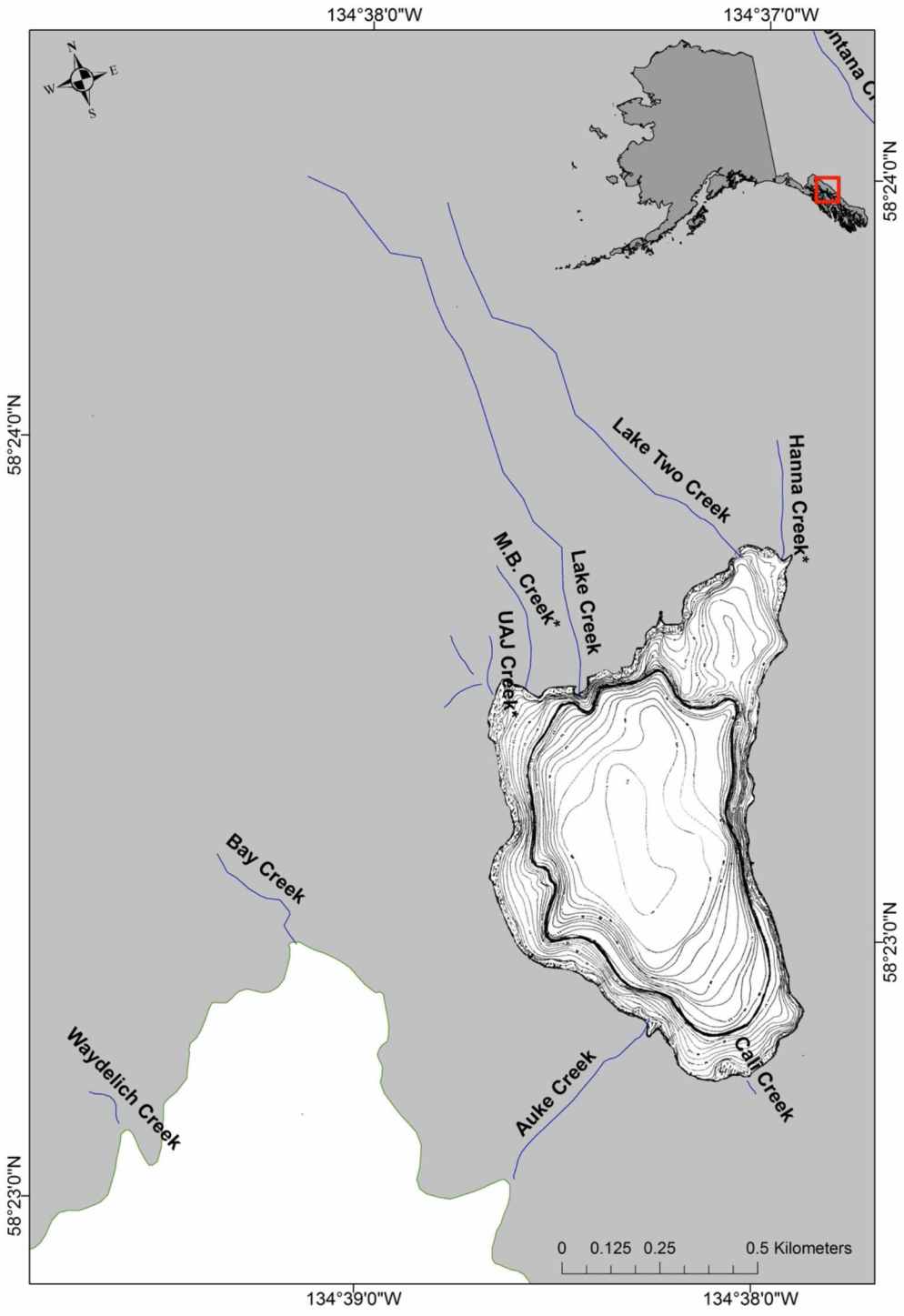

134°38'0"W
134°37'0"W
58°24'0"N
58°23'0"N
Lake Two Creek
Hanna Creek*
M.B. Creek*
Lake Creek
UAJ Creek*
Bay Creek
Waydelich Creek
Auke Creek
Cain Creek
0 0.125 0.25 0.5 Kilometers
134°39'0"W
134°38'0"W

Figure 2.1

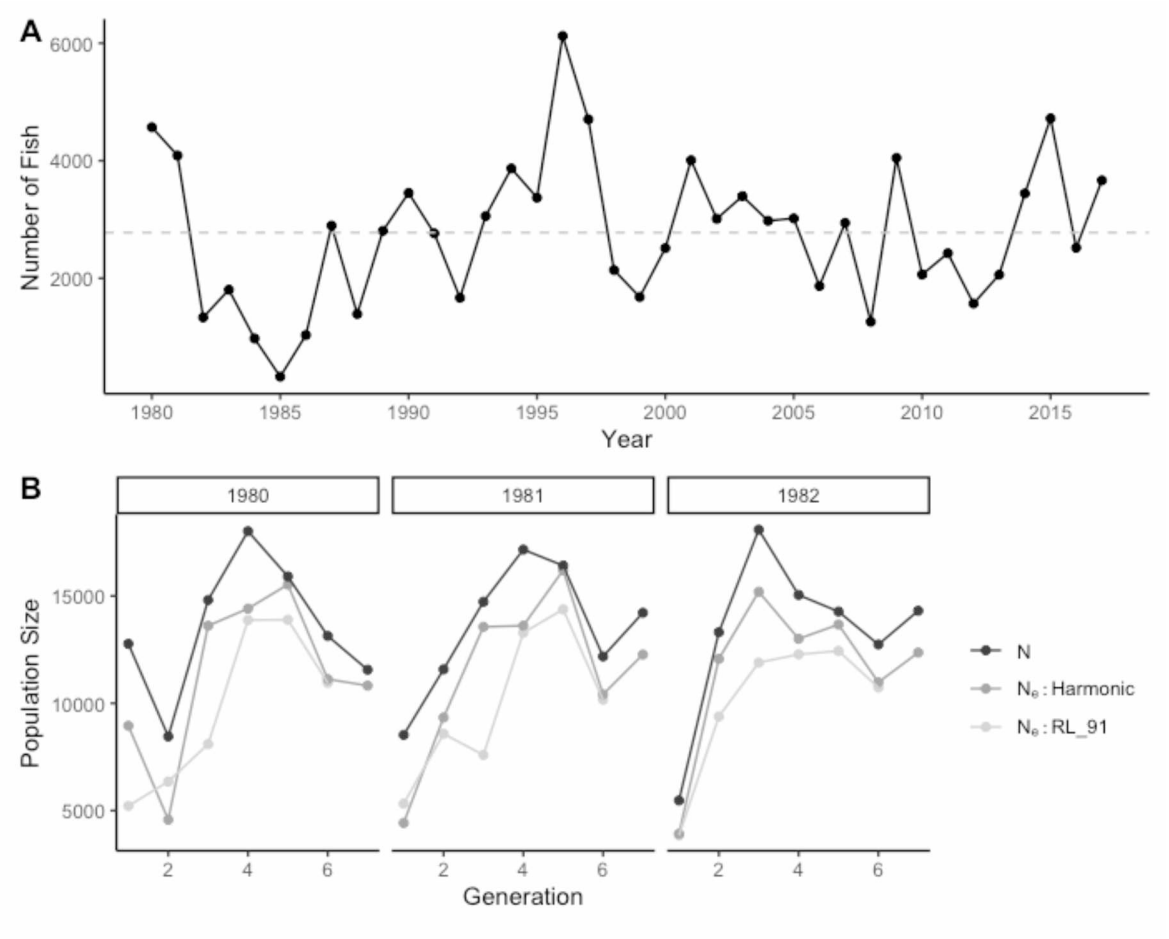

A
Number of Fish
6000
4000
2000
1980
1985
1990
1995
2000
2005
2010
2015
Year
B
1980
1981
1982
Population Size
15000
10000
5000
2
4
6
Generation
N
N_e : Harmonic
N_e : RL_91

Figure 2.2

Figure 2.3

Proportion
0.00
0.25
0.50
0.75
1.00
Year
1980 (n = 130)
1981 (n = 114)
1982 (n = 120)
1983 (n = 73)
1984 (n = 220)
1985 (n = 52)
1986 (n = 337)
1987 (n = 147)
1988 (n = 304)
1989 (n = 213)
1990 (n = 336)
1991 (n = 388)
1992 (n = 392)
1993 (n = 292)
1994 (n = 145)
1995 (n = 250)
1996 (n = 402)
1997 (n = 242)
1998 (n = 201)
1999 (n = 167)
2000 (n = 241)
2001 (n = 468)
2002 (n = 243)
2003 (n = 157)
2004 (n = 110)
2005 (n = 353)
2006 (n = 284)
2007 (n = 235)
2008 (n = 88)
2009 (n = 258)
2010 (n = 244)
2011 (n = 217)
2012 (n = 256)
2013 (n = 149)
2014 (n = 316)
2015 (n = 285)
2016 (n = 236)
2017 (n = 277)
6
5
4
3

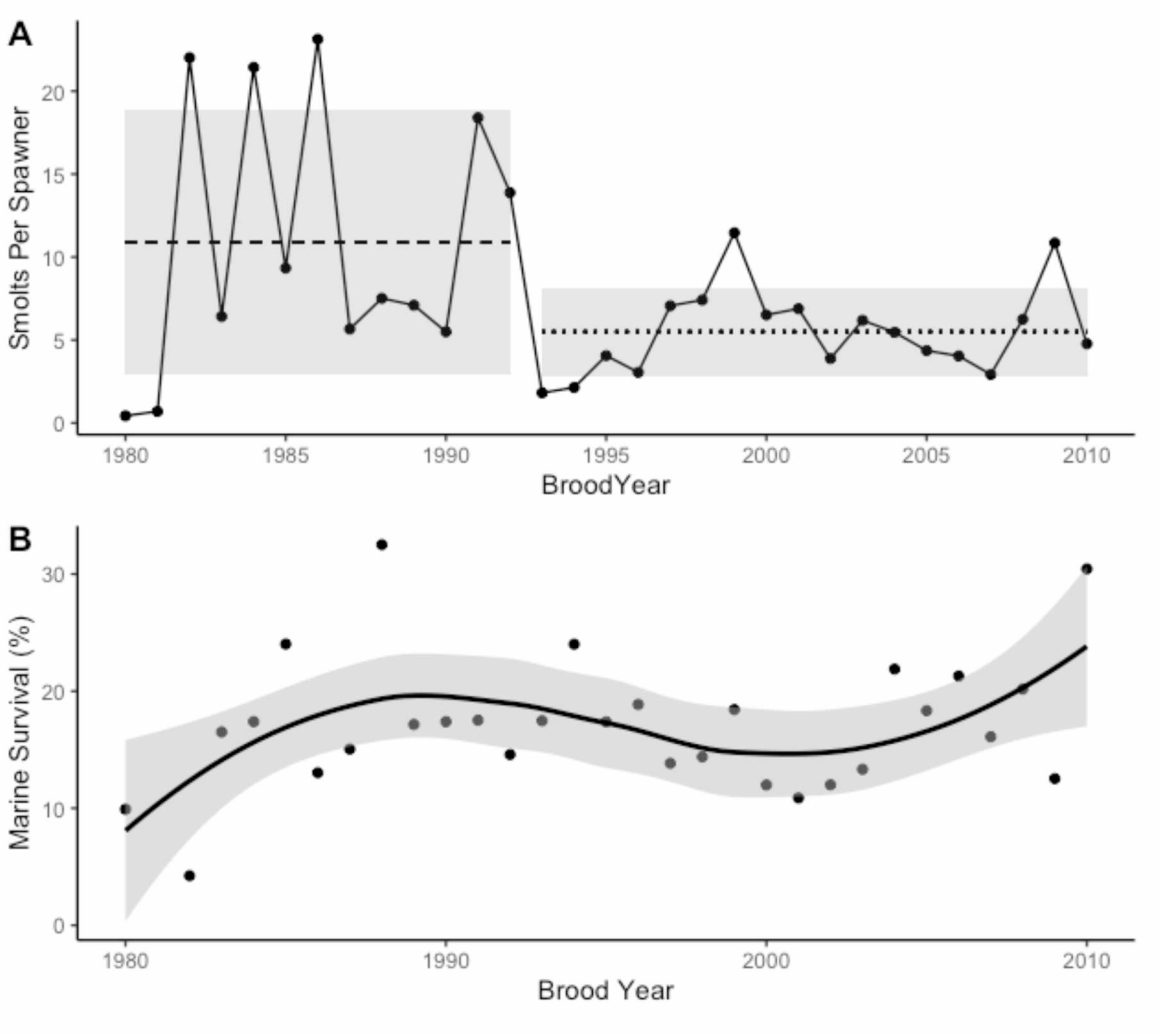

A
Smolts Per Spawner
0
5
10
15
20
1980
1985
1990
1995
2000
2005
2010
BroodYear
B
Marine Survival (%)
0
10
20
30
1980
1990
2000
2010
Brood Year

Figure 2.4

Supplementary Tables:

Table 2.1S: Table of genetic markers for Auke Creek, AK Sockeye Salmon genotyped for each return year and for the combined all year collection. The first 14 markers are microsatellites and markers 15 through 72 are single nucleotide polymorphisms.

	2008	2009/2011	All Years	Citation
	Microsatellite Loci			
1	Oki10	Oki10	Oki10	Smith et al. 1998
2	Oki16	Oki16	Oki16	Smith et al. 1998
3	Oki29			Smith et al. 1998
4	Ots100			Nelson and Beacham 1999
5	Ots103	Ots103	Ots103	Nelson and Beacham 1999
6	One109	One109	One109	Olsen et al. 2000
7	Oki1a	Oki1a	Oki1a	Smith et al. 1998
8	Oki1b	Oki1b	Oki1b	Smith et al. 1998
9	One8	One8	One8	Scribner et al. 1996
10	Omy77	Omy77	Omy77	Morris et al. 1996
11	Oki100	Oki100	Oki100	Beacham et al. 2008
12	One102	One102	One102	Olsen et al. 2000
13	One114	One114	One114	Olsen et al. 2000
14	Ssa419	Ssa419	Ssa419	Cairney et al. 2000
	Single Nucleotide Polymorphism Loci			
15	One_ACBP-79	One_ACBP-79	One_ACBP-79	Elfstrom et al. 2006
16	One_agt-132	One_agt-132	One_agt-132	Storer et al. 2012
17	One_aldB-152			Storer et al. 2012
18	One_apoe-83	One_apoe-83	One_apoe-83	Storer et al. 2012
19	One_CD9-269			Storer et al. 2012
20	One_cetn1-167	One_cetn1-167	One_cetn1-167	Storer et al. 2012
21	One_cin-177	One_cin-177	One_cin-177	Campbell & Narum 2011
22		One_dds-529		Campbell & Narum 2011
23		One_DDX5-86		Storer et al. 2012
24	One_E2	One_E2	One_E2	Smith et al. 2005
25	One_gdh-212	One_gdh-212	One_gdh-212	Campbell & Narum 2011
26	One_GHII-2461	One_GHII-2461	One_GHII-2461	Elfstrom et al. 2006
27	One_GPDH			Smith et al. 2005
28	One_GPH-414	One_GPH-414	One_GPH-414	Elfstrom et al. 2006
29	One_hcs71-220	One_hcs71-220	One_hcs71-220	Elfstrom et al. 2006
30	One_HGFA	One_HGFA	One_HGFA	Smith et al. 2005
31	One_HpaI-436			Elfstrom et al. 2006
32	One_HpaI-99	One_HpaI-99	One_HpaI-99	Elfstrom et al. 2006
33	One_Hsp47	One_Hsp47	One_Hsp47	Miller & Beacham (2007)
34	One_IL8r-362	One_IL8r-362	One_IL8r-362	Habicht et al. 2010

Table 2.1S Continued

35	One_KCT1-453	One_KCT1-453	One_KCT1-453	Storer et al. 2012
36	One_KPNA-422			Elfstrom et al. 2006
37	One_LEI-87	One_LEI-87	One_LEI-87	Elfstrom et al. 2006
38	One_MHC2_251	One_MHC2_251	One_MHC2_251	Elfstrom et al. 2006
39	One_Mkpro-129	One_Mkpro-129	One_Mkpro-129	Campbell & Narum 2011
40	One_ODC1-196	One_ODC1-196	One_ODC1-196	Storer et al. 2012
41	One_Ots208-234	One_Ots208-234	One_Ots208-234	Campbell & Narum 2011
42	One_Ots213-181	One_Ots213-181	One_Ots213-181	Elfstrom et al. 2006
43	One_pax7-248			Campbell & Narum 2011
44	One_PIP	One_PIP	One_PIP	Miller & Beacham (2007)
45	One_Prl2	One_Prl2	One_Prl2	Elfstrom et al. 2006
46		One_psme2-354		Storer et al. 2012
47	One_redd1-414			Campbell & Narum 2011
48	One_spf30-207			Campbell & Narum 2011
49	One_ssrd-135			Campbell & Narum 2011
50	One_STC-410	One_STC-410	One_STC-410	Elfstrom et al. 2006
51	One_STR07	One_STR07	One_STR07	Elfstrom et al. 2006
52	One_SUMO1-6	One_SUMO1-6	One_SUMO1-6	Campbell & Narum 2011
53	One_sys1-230			Campbell & Narum 2011
54	One_Tf_ex10-750	One_Tf_ex10-750	One_Tf_ex10-750	Elfstrom et al. 2006
55	One_U1004-183	One_U1004-183	One_U1004-183	Storer et al. 2012
56	One_U1012-68	One_U1012-68	One_U1012-68	Storer et al. 2012
57	One_U1016-115	One_U1016-115	One_U1016-115	Storer et al. 2012
58	One_U1024-197	One_U1024-197	One_U1024-197	Storer et al. 2012
59	One_U1201-492	One_U1201-492	One_U1201-492	Storer et al. 2012
60	One_U1202-1052	One_U1202-1052	One_U1202-1052	Storer et al. 2012
61	One_U1204-53			Storer et al. 2012
62	One_U1206-108	One_U1206-108	One_U1206-108	Storer et al. 2012
63	One_U1208-67	One_U1208-67	One_U1208-67	Storer et al. 2012
64	One_U1209-111			Storer et al. 2012
65	One_U1210-173	One_U1210-173	One_U1210-173	Storer et al. 2012
66	One_U1212-106	One_U1212-106	One_U1212-106	Storer et al. 2012
67	One_U1214-107			Storer et al. 2012
68		One_U1215-82		Storer et al. 2012
69	One_U1216-230	One_U1216-230	One_U1216-230	Storer et al. 2012
70	One_U301-92	One_U301-92	One_U301-92	Elfstrom et al. 2006
71	One_U401-224	One_U401-224	One_U401-224	Habicht et al. 2010
72	One_U504-141	One_U504-141	One_U504-141	Habicht et al. 2010

Table 2.2S: Results from full-sibling family reconstruction of Auke Creek, AK Sockeye Salmon with the program Colony of simulations conducted in the R package PseudoBabies (Chapter 3). Simulations were conducted with polygamous and polygynous mating designs. The true mean full sibling family size ($\bar{k}_T$), inferred family size ($\bar{k}_I$), probability that the distributions are the same for the mean ($P_{\bar{k}}$), true variance in family size (V_T), inferred variance in family size (V_I), probability that the distributions are the same for the variance (P_V) and proportion of true families inferred (Correct) are reported for 10 simulations based on return years 2008, 2009, and 2011 for three probability of inclusion (P_{Inc}) values.

Mating	P_{Inc}	Year	Sim	$\bar{k}_T$	$\bar{k}_I$	$P_{\bar{k}}$	V_T	V_I	P_V	Correct
Polygamy	0	2008	1	1.15	1.19	0.11	0.21	0.23	0.05	0.90
Polygamy	0	2008	2	1.17	1.20	0.18	0.24	0.26	0.12	0.90
Polygamy	0	2008	3	1.16	1.20	0.04	0.22	0.26	0.02	0.90
Polygamy	0	2008	4	1.14	1.18	0.02	0.18	0.21	0.01	0.88
Polygamy	0	2008	5	1.17	1.21	0.11	0.24	0.26	0.07	0.90
Polygamy	0	2008	6	1.15	1.19	0.04	0.20	0.23	0.02	0.90
Polygamy	0	2008	7	1.15	1.19	0.08	0.20	0.22	0.04	0.90
Polygamy	0	2008	8	1.17	1.19	0.44	0.24	0.26	0.39	0.90
Polygamy	0	2008	9	1.16	1.19	0.18	0.22	0.25	0.13	0.91
Polygamy	0	2008	10	1.19	1.22	0.18	0.29	0.31	0.10	0.91
Polygamy	0	2009	1	1.55	1.56	0.69	1.41	1.32	0.25	0.85
Polygamy	0	2009	2	1.54	1.57	0.40	1.43	1.37	0.09	0.84
Polygamy	0	2009	3	1.56	1.57	0.57	1.44	1.34	0.21	0.84
Polygamy	0	2009	4	1.53	1.55	0.55	1.22	1.17	0.22	0.83
Polygamy	0	2009	5	1.55	1.57	0.48	1.59	1.49	0.08	0.84
Polygamy	0	2009	6	1.55	1.56	0.61	1.43	1.32	0.15	0.85
Polygamy	0	2009	7	1.54	1.56	0.44	1.41	1.33	0.07	0.83
Polygamy	0	2009	8	1.55	1.57	0.39	1.41	1.32	0.07	0.84
Polygamy	0	2009	9	1.53	1.55	0.36	1.29	1.22	0.08	0.83
Polygamy	0	2009	10	1.52	1.54	0.49	1.24	1.17	0.16	0.84
Polygamy	0	2011	1	1.57	1.57	0.98	1.37	1.33	0.93	0.88
Polygamy	0	2011	2	1.57	1.58	0.90	1.65	1.57	0.64	0.87
Polygamy	0	2011	3	1.49	1.49	0.98	1.09	1.06	0.86	0.90
Polygamy	0	2011	4	1.55	1.55	0.98	1.27	1.19	0.75	0.87
Polygamy	0	2011	5	1.62	1.61	0.88	1.71	1.60	0.93	0.88
Polygamy	0	2011	6	1.59	1.56	0.65	1.54	1.39	0.78	0.88
Polygamy	0	2011	7	1.53	1.55	0.71	1.30	1.28	0.53	0.88
Polygamy	0	2011	8	1.60	1.60	0.96	1.66	1.56	0.86	0.88
Polygamy	0	2011	9	1.57	1.58	0.80	1.26	1.21	0.58	0.87
Polygamy	0	2011	10	1.58	1.58	0.98	1.48	1.40	0.76	0.89
Polygamy	0.75	2008	1	1.15	1.14	0.66	0.21	0.20	0.70	0.95
Polygamy	0.75	2008	2	1.17	1.18	0.92	0.24	0.24	0.90	0.95

Table 2.2S Continued

Polygamy	0.75	2008	3	1.16	1.16	0.75	0.22	0.22	0.67	0.95
Polygamy	0.75	2008	4	1.14	1.14	0.77	0.18	0.18	0.68	0.93
Polygamy	0.75	2008	5	1.17	1.18	0.79	0.24	0.23	0.71	0.94
Polygamy	0.75	2008	6	1.15	1.15	0.82	0.20	0.20	0.75	0.96
Polygamy	0.75	2008	7	1.15	1.15	0.95	0.20	0.20	0.97	0.96
Polygamy	0.75	2008	8	1.17	1.16	0.61	0.24	0.23	0.53	0.93
Polygamy	0.75	2008	9	1.16	1.16	0.91	0.22	0.23	0.85	0.95
Polygamy	0.75	2008	10	1.19	1.19	1.00	0.29	0.29	0.98	0.95
Polygamy	0.75	2009	1	1.55	1.51	0.13	1.41	1.28	0.09	0.90
Polygamy	0.75	2009	2	1.54	1.50	0.11	1.43	1.31	0.06	0.90
Polygamy	0.75	2009	3	1.56	1.51	0.09	1.44	1.29	0.05	0.90
Polygamy	0.75	2009	4	1.53	1.49	0.09	1.22	1.12	0.05	0.89
Polygamy	0.75	2009	5	1.55	1.51	0.13	1.59	1.43	0.09	0.90
Polygamy	0.75	2009	6	1.55	1.51	0.15	1.43	1.28	0.15	0.90
Polygamy	0.75	2009	7	1.54	1.49	0.08	1.41	1.28	0.04	0.90
Polygamy	0.75	2009	8	1.55	1.51	0.11	1.41	1.27	0.09	0.90
Polygamy	0.75	2009	9	1.53	1.49	0.12	1.29	1.18	0.07	0.89
Polygamy	0.75	2009	10	1.52	1.48	0.10	1.24	1.12	0.05	0.89
Polygamy	0.75	2011	1	1.57	1.53	0.30	1.37	1.28	0.19	0.92
Polygamy	0.75	2011	2	1.57	1.53	0.36	1.65	1.52	0.24	0.91
Polygamy	0.75	2011	3	1.49	1.46	0.52	1.09	1.04	0.45	0.93
Polygamy	0.75	2011	4	1.55	1.51	0.30	1.27	1.16	0.23	0.91
Polygamy	0.75	2011	5	1.62	1.56	0.21	1.71	1.55	0.12	0.93
Polygamy	0.75	2011	6	1.59	1.53	0.22	1.54	1.37	0.13	0.92
Polygamy	0.75	2011	7	1.53	1.50	0.55	1.30	1.25	0.46	0.92
Polygamy	0.75	2011	8	1.60	1.56	0.36	1.66	1.52	0.26	0.92
Polygamy	0.75	2011	9	1.57	1.53	0.43	1.26	1.18	0.36	0.92
Polygamy	0.75	2011	10	1.58	1.54	0.36	1.48	1.37	0.27	0.93
Polygamy	0.95	2008	1	1.15	1.13	0.16	0.21	0.18	0.13	0.96
Polygamy	0.95	2008	2	1.17	1.15	0.35	0.24	0.23	0.26	0.97
Polygamy	0.95	2008	3	1.16	1.14	0.41	0.22	0.20	0.37	0.97
Polygamy	0.95	2008	4	1.14	1.11	0.08	0.18	0.14	0.06	0.97
Polygamy	0.95	2008	5	1.17	1.15	0.36	0.24	0.21	0.31	0.96
Polygamy	0.95	2008	6	1.15	1.13	0.43	0.20	0.18	0.39	0.97
Polygamy	0.95	2008	7	1.15	1.13	0.33	0.20	0.18	0.26	0.97
Polygamy	0.95	2008	8	1.17	1.14	0.07	0.24	0.21	0.03	0.96
Polygamy	0.95	2008	9	1.16	1.14	0.33	0.22	0.21	0.22	0.97
Polygamy	0.95	2008	10	1.19	1.17	0.41	0.29	0.27	0.31	0.98
Polygamy	0.95	2009	1	1.55	1.46	<0.01	1.41	1.24	<0.01	0.93
Polygamy	0.95	2009	2	1.54	1.46	<0.01	1.43	1.27	<0.01	0.92
Polygamy	0.95	2009	3	1.56	1.47	<0.01	1.44	1.25	<0.01	0.92

Table 2.2S Continued

Polygamy	0.95	2009	4	1.53	1.44	<0.01	1.22	1.07	<0.01	0.92
Polygamy	0.95	2009	5	1.55	1.47	<0.01	1.59	1.39	<0.01	0.92
Polygamy	0.95	2009	6	1.55	1.46	<0.01	1.43	1.24	<0.01	0.93
Polygamy	0.95	2009	7	1.54	1.45	<0.01	1.41	1.24	<0.01	0.93
Polygamy	0.95	2009	8	1.55	1.47	<0.01	1.41	1.23	<0.01	0.93
Polygamy	0.95	2009	9	1.53	1.44	<0.01	1.29	1.13	<0.01	0.92
Polygamy	0.95	2009	10	1.52	1.43	<0.01	1.24	1.08	<0.01	0.92
Polygamy	0.95	2011	1	1.57	1.48	0.04	1.37	1.24	0.01	0.94
Polygamy	0.95	2011	2	1.57	1.49	0.08	1.65	1.48	0.01	0.93
Polygamy	0.95	2011	3	1.49	1.43	0.13	1.09	1.01	0.05	0.95
Polygamy	0.95	2011	4	1.55	1.46	0.03	1.27	1.12	<0.01	0.93
Polygamy	0.95	2011	5	1.62	1.51	0.03	1.71	1.50	<0.01	0.93
Polygamy	0.95	2011	6	1.59	1.49	0.04	1.54	1.33	0.01	0.93
Polygamy	0.95	2011	7	1.53	1.46	0.09	1.30	1.21	0.02	0.94
Polygamy	0.95	2011	8	1.60	1.51	0.05	1.66	1.47	<0.01	0.93
Polygamy	0.95	2011	9	1.57	1.49	0.06	1.26	1.15	0.01	0.94
Polygamy	0.95	2011	10	1.58	1.51	0.12	1.48	1.35	0.03	0.95
Polygyny	0	2008	1	1.15	1.21	0.02	0.22	0.27	0.01	0.88
Polygyny	0	2008	2	1.16	1.23	<0.01	0.19	0.24	<0.01	0.86
Polygyny	0	2008	3	1.15	1.23	0.01	0.22	0.27	<0.01	0.84
Polygyny	0	2008	4	1.19	1.24	0.05	0.24	0.28	0.03	0.87
Polygyny	0	2008	5	1.16	1.23	0.01	0.19	0.24	<0.01	0.88
Polygyny	0	2008	6	1.16	1.25	<0.01	0.23	0.28	<0.01	0.85
Polygyny	0	2008	7	1.15	1.20	0.05	0.20	0.23	0.02	0.88
Polygyny	0	2008	8	1.22	1.27	0.14	0.37	0.40	0.06	0.88
Polygyny	0	2008	9	1.18	1.24	0.05	0.27	0.27	0.01	0.86
Polygyny	0	2008	10	1.17	1.24	0.01	0.21	0.26	<0.01	0.86
Polygyny	0	2009	1	1.58	1.68	<0.01	1.43	1.42	<0.01	0.78
Polygyny	0	2009	2	1.54	1.65	<0.01	1.26	1.25	<0.01	0.76
Polygyny	0	2009	3	1.57	1.66	<0.01	1.41	1.37	<0.01	0.77
Polygyny	0	2009	4	1.56	1.66	<0.01	1.45	1.42	<0.01	0.76
Polygyny	0	2009	5	1.51	1.63	<0.01	1.22	1.22	<0.01	0.75
Polygyny	0	2009	6	1.56	1.66	<0.01	1.44	1.41	<0.01	0.77
Polygyny	0	2009	7	1.55	1.65	<0.01	1.33	1.31	<0.01	0.76
Polygyny	0	2009	8	1.53	1.63	<0.01	1.25	1.25	<0.01	0.76
Polygyny	0	2009	9	1.52	1.63	<0.01	1.22	1.19	<0.01	0.76
Polygyny	0	2009	10	1.53	1.63	<0.01	1.29	1.25	<0.01	0.76
Polygyny	0	2011	1	1.57	1.64	0.14	1.77	1.74	0.01	0.83
Polygyny	0	2011	2	1.59	1.66	0.14	1.45	1.41	0.02	0.83
Polygyny	0	2011	3	1.60	1.64	0.35	1.41	1.34	0.11	0.85
Polygyny	0	2011	4	1.54	1.61	0.10	1.35	1.31	0.01	0.82

Table 2.2S Continued

Polygyny	0	2011	5	1.54	1.62	0.09	1.38	1.32	0.01	0.80
Polygyny	0	2011	6	1.53	1.58	0.25	1.28	1.22	0.05	0.83
Polygyny	0	2011	7	1.57	1.63	0.19	1.41	1.35	0.04	0.84
Polygyny	0	2011	8	1.57	1.65	0.13	1.42	1.40	0.02	0.80
Polygyny	0	2011	9	1.59	1.66	0.16	1.61	1.54	0.01	0.83
Polygyny	0	2011	10	1.54	1.60	0.13	1.28	1.20	0.01	0.83
Polygyny	0.75	2008	1	1.15	1.16	0.69	0.22	0.23	0.63	0.96
Polygyny	0.75	2008	2	1.16	1.16	0.72	0.19	0.19	0.65	0.96
Polygyny	0.75	2008	3	1.15	1.15	1.00	0.22	0.21	0.92	0.95
Polygyny	0.75	2008	4	1.19	1.18	0.74	0.24	0.22	0.73	0.95
Polygyny	0.75	2008	5	1.16	1.17	0.56	0.19	0.20	0.50	0.96
Polygyny	0.75	2008	6	1.16	1.17	0.74	0.23	0.23	0.65	0.97
Polygyny	0.75	2008	7	1.15	1.16	0.89	0.20	0.19	0.80	0.96
Polygyny	0.75	2008	8	1.22	1.21	0.91	0.37	0.36	0.91	0.96
Polygyny	0.75	2008	9	1.18	1.18	1.00	0.27	0.23	0.84	0.95
Polygyny	0.75	2008	10	1.17	1.17	0.94	0.21	0.20	0.96	0.96
Polygyny	0.75	2009	1	1.58	1.57	0.63	1.43	1.36	0.75	0.88
Polygyny	0.75	2009	2	1.54	1.53	0.75	1.26	1.18	1.00	0.87
Polygyny	0.75	2009	3	1.57	1.55	0.52	1.41	1.30	0.74	0.87
Polygyny	0.75	2009	4	1.56	1.54	0.40	1.45	1.34	0.52	0.88
Polygyny	0.75	2009	5	1.51	1.51	0.81	1.22	1.15	0.90	0.87
Polygyny	0.75	2009	6	1.56	1.55	0.92	1.44	1.34	0.65	0.88
Polygyny	0.75	2009	7	1.55	1.52	0.23	1.33	1.22	0.22	0.88
Polygyny	0.75	2009	8	1.53	1.51	0.38	1.25	1.17	0.42	0.88
Polygyny	0.75	2009	9	1.52	1.51	0.59	1.22	1.12	0.84	0.87
Polygyny	0.75	2009	10	1.53	1.50	0.33	1.29	1.17	0.43	0.88
Polygyny	0.75	2011	1	1.57	1.55	0.67	1.77	1.66	0.70	0.92
Polygyny	0.75	2011	2	1.59	1.56	0.51	1.45	1.35	0.50	0.93
Polygyny	0.75	2011	3	1.60	1.56	0.38	1.41	1.29	0.34	0.92
Polygyny	0.75	2011	4	1.54	1.51	0.49	1.35	1.24	0.47	0.92
Polygyny	0.75	2011	5	1.54	1.50	0.44	1.38	1.24	0.44	0.90
Polygyny	0.75	2011	6	1.53	1.49	0.41	1.28	1.16	0.37	0.91
Polygyny	0.75	2011	7	1.57	1.54	0.50	1.41	1.29	0.46	0.93
Polygyny	0.75	2011	8	1.57	1.54	0.45	1.42	1.33	0.35	0.89
Polygyny	0.75	2011	9	1.59	1.57	0.57	1.61	1.48	0.63	0.92
Polygyny	0.75	2011	10	1.54	1.52	0.65	1.28	1.14	0.88	0.91
Polygyny	0.95	2008	1	1.15	1.13	0.39	0.22	0.20	0.27	0.98
Polygyny	0.95	2008	2	1.16	1.13	0.35	0.19	0.17	0.31	0.97
Polygyny	0.95	2008	3	1.15	1.13	0.34	0.22	0.19	0.29	0.97
Polygyny	0.95	2008	4	1.19	1.16	0.18	0.24	0.20	0.14	0.97
Polygyny	0.95	2008	5	1.16	1.14	0.46	0.19	0.17	0.44	0.98

Table 2.2S Continued

Polygyny	0.95	2008	6	1.16	1.15	0.69	0.23	0.21	0.67	0.98
Polygyny	0.95	2008	7	1.15	1.13	0.32	0.20	0.17	0.33	0.98
Polygyny	0.95	2008	8	1.22	1.18	0.31	0.37	0.33	0.20	0.97
Polygyny	0.95	2008	9	1.18	1.15	0.31	0.27	0.21	0.32	0.97
Polygyny	0.95	2008	10	1.17	1.14	0.23	0.21	0.18	0.18	0.97
Polygyny	0.95	2009	1	1.58	1.50	<0.01	1.43	1.30	<0.01	0.93
Polygyny	0.95	2009	2	1.54	1.46	<0.01	1.26	1.13	<0.01	0.93
Polygyny	0.95	2009	3	1.57	1.47	<0.01	1.41	1.23	<0.01	0.91
Polygyny	0.95	2009	4	1.56	1.48	<0.01	1.45	1.28	<0.01	0.92
Polygyny	0.95	2009	5	1.51	1.43	<0.01	1.22	1.08	<0.01	0.93
Polygyny	0.95	2009	6	1.56	1.49	0.01	1.44	1.29	<0.01	0.94
Polygyny	0.95	2009	7	1.55	1.46	<0.01	1.33	1.17	<0.01	0.92
Polygyny	0.95	2009	8	1.53	1.45	<0.01	1.25	1.12	<0.01	0.92
Polygyny	0.95	2009	9	1.52	1.44	<0.01	1.22	1.06	<0.01	0.93
Polygyny	0.95	2009	10	1.53	1.44	<0.01	1.29	1.11	<0.01	0.93
Polygyny	0.95	2011	1	1.57	1.50	0.16	1.77	1.61	0.04	0.95
Polygyny	0.95	2011	2	1.59	1.51	0.07	1.45	1.31	0.02	0.94
Polygyny	0.95	2011	3	1.60	1.51	0.05	1.41	1.25	0.01	0.94
Polygyny	0.95	2011	4	1.54	1.45	0.05	1.35	1.19	0.01	0.94
Polygyny	0.95	2011	5	1.54	1.46	0.07	1.38	1.20	0.02	0.93
Polygyny	0.95	2011	6	1.53	1.44	0.03	1.28	1.12	<0.01	0.93
Polygyny	0.95	2011	7	1.57	1.49	0.08	1.41	1.25	0.02	0.95
Polygyny	0.95	2011	8	1.57	1.48	0.04	1.42	1.28	<0.01	0.92
Polygyny	0.95	2011	9	1.59	1.52	0.13	1.61	1.45	0.05	0.95
Polygyny	0.95	2011	10	1.54	1.46	0.07	1.28	1.10	0.03	0.94

Table 2.3S: Genetic estimates of the effective number of breeders (N_b) of Auke Creek, AK Sockeye Salmon from datasets simulated with the R package PseudoBabies (*Chapter 3*). Data were simulated with a polygynous or polygamous mating structure, with allele frequencies from the 2008, 2009, and 2011 return year. Each simulation represents a single collection of fish sampled that are derived from spawners in multiple years. The census size of the years contributing to the sampled return year (N_c), the harmonic mean of the effective number of breeders for years contributing to the sampled return year ($\widetilde{N}_b$), the effective population size (N_e), the number of spawners that contributed to the genetic sample (N_{sp}), and the effective number of breeders ($\widehat{N}_b$) with 95% confidence intervals from Colony and LDNe.

							Colony			LDNe		
Mating	Year	Sim	N_c	$\widetilde{N}_b$	N_e	N_{sp}	$\widehat{N}_b$	95L	95U	$\widehat{N}_b$	95L	95U
Polygyny	2008	1	5515	1873	9366	1201	847	761	945	1852	1462	2484
Polygyny	2008	2	5523	1886	9428	1175	875	789	970	2126	1634	2983
Polygyny	2008	3	5614	1923	9617	1176	827	736	924	1672	1339	2193
Polygyny	2008	4	5541	1897	9485	1198	882	793	987	1163	991	1394
Polygyny	2008	5	5394	1809	9045	1143	846	766	941	2300	1731	3349
Polygyny	2008	6	5590	1882	9409	1185	796	715	895	1892	1481	2572
Polygyny	2008	7	5557	1868	9341	1254	866	782	966	2060	1605	2823
Polygyny	2008	8	5523	1841	9203	1155	861	774	962	2102	1618	2940
Polygyny	2008	9	5534	1865	9325	1106	782	698	886	1699	1351	2255
Polygyny	2008	10	5496	1854	9269	1190	818	729	917	1772	1409	2351
Polygyny	2009	1	28530	3833	19163	6628	3565	3403	3745	3910	3529	4355
Polygyny	2009	2	28675	3865	19325	6782	3645	3485	3825	4433	3964	4995
Polygyny	2009	3	28589	3829	19145	6761	3614	3457	3799	4360	3909	4897
Polygyny	2009	4	29114	3854	19271	6763	3585	3416	3751	3875	3504	4309
Polygyny	2009	5	28652	3802	19009	6844	3664	3503	3854	4210	3782	4718
Polygyny	2009	6	28124	3769	18847	6796	3600	3438	3777	3740	3388	4149
Polygyny	2009	7	28631	3815	19076	6715	3579	3410	3763	4374	3915	4921
Polygyny	2009	8	28359	3797	18983	6750	3626	3457	3815	3894	3515	4337
Polygyny	2009	9	28681	3836	19180	6841	3718	3554	3904	4019	3622	4486
Polygyny	2009	10	28121	3789	18943	6760	3584	3430	3771	4416	3952	4970

Table 2.3S Continued

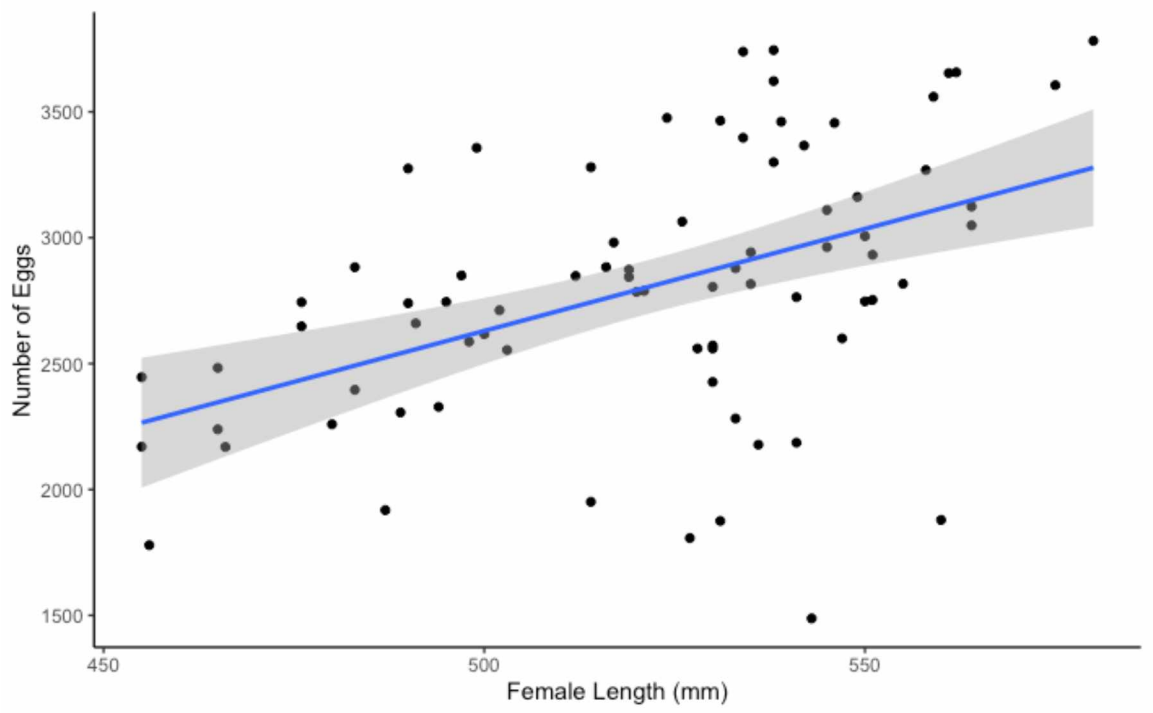
Polygyny	2011	1	10238	1260	6299	2446	1302	1199	1412	1298	1176	1439
Polygyny	2011	2	10260	1268	6340	2260	1311	1208	1416	1440	1286	1625
Polygyny	2011	3	10298	1271	6354	2364	1397	1288	1509	1432	1285	1605
Polygyny	2011	4	10140	1277	6383	2352	1323	1221	1434	1392	1250	1561
Polygyny	2011	5	10195	1257	6284	2304	1309	1210	1422	1471	1312	1661
Polygyny	2011	6	10074	1257	6283	2284	1330	1229	1436	1461	1305	1648
Polygyny	2011	7	10126	1252	6262	2329	1321	1224	1433	1484	1326	1672
Polygyny	2011	8	10258	1272	6360	2244	1289	1185	1399	1306	1172	1463
Polygyny	2011	9	10504	1291	6454	2383	1307	1209	1417	1187	1077	1313
Polygyny	2011	10	10132	1251	6255	2373	1346	1246	1460	1400	1258	1569
Polygamy	2008	1	6269	2168	10838	1434	1002	914	1100	1470	1271	1727
Polygamy	2008	2	6291	2182	10911	1482	1081	986	1186	1650	1422	1949
Polygamy	2008	3	6283	2159	10797	1425	1033	943	1127	1645	1408	1958
Polygamy	2008	4	6190	2147	10736	1403	1037	945	1141	1604	1374	1908
Polygamy	2008	5	6205	2143	10716	1404	1038	945	1143	1826	1544	2210
Polygamy	2008	6	6240	2156	10781	1445	1079	985	1180	1461	1270	1706
Polygamy	2008	7	6180	2158	10792	1385	1004	909	1099	1918	1604	2358
Polygamy	2008	8	6096	2120	10599	1435	1070	971	1174	1816	1540	2189
Polygamy	2008	9	6187	2153	10765	1389	1028	936	1128	1814	1538	2189
Polygamy	2008	10	6280	2161	10803	1435	1059	969	1152	1455	1268	1694
Polygamy	2009	1	30236	3299	16496	5962	3364	3208	3536	3650	3309	4046
Polygamy	2009	2	30046	3290	16451	6049	3377	3229	3549	3547	3232	3907
Polygamy	2009	3	30093	3298	16490	5960	3357	3199	3538	3627	3295	4010
Polygamy	2009	4	30416	3327	16633	6018	3469	3309	3646	3644	3309	4032
Polygamy	2009	5	30466	3298	16490	5952	3270	3123	3438	3170	2901	3475
Polygamy	2009	6	30033	3310	16550	6021	3390	3244	3562	3542	3221	3914

Table 2.3S Continued

Polygamy	2009	7	30070	3308	16540	6039	3390	3229	3573	3569	3251	3935
Polygamy	2009	8	30272	3306	16531	6008	3383	3222	3556	3674	3335	4067
Polygamy	2009	9	30257	3298	16492	6036	3450	3286	3633	3347	3052	3684
Polygamy	2009	10	30500	3316	16581	5948	3391	3233	3562	3517	3196	3887
Polygamy	2011	1	10013	1044	5219	1941	1157	1058	1249	1249	1131	1386
Polygamy	2011	2	10339	1066	5332	2076	1110	1019	1215	1079	987	1184
Polygamy	2011	3	10142	1058	5292	2010	1162	1065	1270	1131	1026	1254
Polygamy	2011	4	10124	1048	5240	1970	1207	1116	1311	1180	1074	1303
Polygamy	2011	5	10514	1067	5337	2068	1149	1060	1247	1158	1054	1278
Polygamy	2011	6	10342	1059	5297	1967	1163	1067	1273	1088	991	1198
Polygamy	2011	7	10376	1059	5294	2059	1205	1110	1311	1162	1056	1283
Polygamy	2011	8	10202	1065	5325	1959	1114	1027	1210	1032	942	1134
Polygamy	2011	9	10300	1059	5293	2040	1242	1147	1350	1298	1171	1446
Polygamy	2011	10	10539	1078	5389	2057	1194	1096	1291	1200	1090	1327

Supplementary Figures:

Figure S2.1: Linear model of female fecundity (blue line) of Auke Creek, AK Sockeye Salmon with 95% confidence interval (gray shaded region) was estimated from fish in years 2011 through 2013. There was a significant positive relationship between fecundity (total number of eggs) with length (b_1 = 8.10, SE = 1.78, r^2_{adj} =0.21).



Chapter 3: PseudoBabies: A flexible simulation package to test the assignment accuracy of genetic markers for parentage inference[3]

[A] Abstract

Genetic data can be used to deduce genealogical relationships among individuals for which pedigree information is unknown. The success of any estimator used to determine relationships depends primarily on the proportion of the population sampled and the informativeness of the genetic loci surveyed. We developed a flexible, forward-in-time simulation package, which was written in the R statistical environment, to aid in assessing the accuracy of genetic marker panels. The program ('PseudoBabies') can be applied to large-scale analyses for parentage and sibship. We used genetic data from a population of Sockeye Salmon (*Oncorhynchus nerka*) in Southeast Alaska to demonstrate its application to genetic marker selection. The program is available from: https://github.com/patbarry6/PseudoBabies.

[A] Introduction

Parentage inference has become increasingly important in revealing ecological and evolutionary processes. Understanding patterns of inheritance enables the estimation of critical population parameters such as heritability (Vandeputte et al. 2004), gene flow (Wang 2014), mean breeding success of alternative reproductive life histories (Neff 2001), and effective population size per generation (N_e; Araki et al. 2007). The identification of highly polymorphic loci (amplified fragment length polymorphisms [AFLPs] and microsatellites) was a boon for parentage analyses and made possible large-scale parentage studies without *a priori* pedigree information. The isolation and characterization of markers is becoming easier and more affordable, a trend that should increase as next-generation sequencing technology becomes more efficient, accurate, and economical. However, *in silico* studies can save time and money from applying existing marker panels, provided there is adequate genetic variation to address the research objectives

---

[3] P.D. Barry, M.V. McPhee, S. Vulstek, J. Joyce, and A.J. Gharrett *Formatted for submission to Transactions of the American Fisheries Society*

An array of types of genetic markers exists, but microsatellites or single tandem-repeats (STRs) and single nucleotide polymorphisms (SNPs) are currently the two most commonly used. Long considered the most powerful tool for parentage analysis, STRs are slowly being supplemented or replaced by SNPs when a large panel of SNPs can be genotyped. The discovery and genotyping of large panels of SNPs has become more realistic with advances in next-generation sequencing even in non-model organisms (Thrasher et al. 2018). Many more SNP loci are needed than STR loci to have the same ability to assign individuals to their parents (Glaubitz et al. 2003), but other attributes make them attractive. Within the genome, SNPs are the most abundant marker (Brumfield et al. 2003). Similarly, detection of SNP variation is relatively robust to genotyping error, amenable to automation (Ranade et al. 2001; Olivier et al. 2002), and can be standardized and compared more easily across labs (Smith et al. 2005). Microsatellites, on the other hand, often amplify in closely related species, which may reduce the investment in marker development and screening for some non-model organisms. The resolution power of any suite of markers is determined by the variability of each individual marker (Chakraborty et al. 1988; Marshall et al. 1998; Glaubitz et al. 2003) necessitating a large number of markers to be surveyed to identify particularly informative markers. Often there is a tradeoff between the number of individuals and the number of loci that can be genotyped economically. Finding an optimal number of markers ensures not only that a high proportion of parents can be included but also that there is sufficient power within the marker panel to identify those parents.

Of the many decisions facing researchers, the number of genetic markers and proportion of the population sampled have the largest effect on parentage assignment error rates. Correct assignment occurs when an offspring is assigned to its correct parent, if they have been sampled, or the offspring remains unassigned if their parents have not been sampled. Errors in assignment can manifest in two distinct ways. Type I errors, or false positives, can occur from false assignment of the offspring to a parent when the true parent was included in the parent sample, or when the true parent was omitted from the sample. Type II errors, or false negatives, occur when there is no assignment but the offspring's correct parent was included in the sample.

The marker panel characterized by the highest power to assign offspring may not be the most efficient in terms of minimizing both costs and error. Because a single analysis can amplify many microsatellites or SNP loci (multiplex reaction; Chamberlain et al. 1988, Hayden et al. 2008), multiplexing minimizes both consumable costs (e.g., reagents, plates, tips) as well as operational costs (e.g., personnel). However, successful multiplexing of microsatellite loci depends on compatible annealing temperatures and concentrations of PCR reagents such as $MgCl_2$ and dNTPs, as well as platform-specific limitations such as the number of channels that can be run simultaneously. Multiplexing SNP loci is less demanding, but the number of markers and individuals that can be run is dictated by the chemistry (GT-seq, TaqMan limitations, etc.) as well as the genotyping platform. Under these constraints, grouping loci by information content described by some metric (allelic richness, minor allele frequencies, Shannon Diversity, etc.) can save both time and money.

Attempts have been made to develop generalized methods for cost-effective marker panel selection by maximizing genetic diversity parameters such as heterozygosity and allelic richness (Bromaghin et al. 2008; Matson et al. 2008; Sanderlin et al. 2012). Exclusion probabilities (Wiener et al.1930; Chakraborty et al. 1988) are often reported as a metric of the utility of a set of markers for a parentage study. The parent-pair exclusion probability is the probability that a specific locus can exclude two randomly chosen putative parents that are unrelated to the offspring being assigned. Other exclusion probabilities describe the probability of excluding one parent when one parental genotype is known (paternity/maternity), or when one parental genotype is unavailable. Despite efforts to refine these estimates (Double et al. 1997; Wang 2007; Baruch and Weller 2008; Vandeputte 2012), they may not be the most appropriate measure of power unless exclusion is used as the parentage inference method (Anderson and Garza 2006). Simulations are a more robust way to measure the power of a panel of markers (Jones et al. 2010) and simulation is incorporated into a variety of programs such as Cervus (Marshall et al. 1998), GERUD (Jones 2001), PAPA (Duchesne et al. 2002), and Colony (Jones and Wang 2010). All of these programs except Colony require manual user input with a graphical user interface that prohibits their use in large-scale analyses. However, the computational time required to analyze many datasets may render Colony

unsuitable (Whiteley et al. 2014). The development of a stand-alone simulation package that is independent of the parentage inference software would facilitate the use of simulation to determine the most appropriate marker panel and most useful program to analyze the data.

The goal of this study was to produce a forward-in-time simulation program that can be applied to large-scale analyses for parentage and sibship studies. Specifically, we developed a flexible simulation package (hereafter referred to as 'PseudoBabies') in the R statistical language (R Core Team, 2014) that allows for variation in age at maturity, number of breeders per year, number of offspring per breeding pair, genotyping error, missing data, and the type and number of loci included in the marker panel. The program was designed for semelparous species such as Pacific salmon *Oncorhynchus* spp., because of the widespread use of parentage analysis in the study and management of hatchery-wild interactions in this commercially important taxon (Reisenbichler and Rubin 1999; Matala et al. 2012; Christie et al. 2014; Venditti et al. 2018). We used empirical data for both STR and SNP loci collected from a small population of Sockeye Salmon *Oncorhynchus nerka* in Southeast Alaska to illustrate the usefulness of our simulation package for marker panel selection.

[A] Methods

[C] *PseudoBabies simulation package.* — We developed the R package PseudoBabies to estimate the assignment accuracy of genetic markers for parentage studies. The functions of the R package are described in detail in its associated vignette, which can be accessed with the command vignette ("PseudoBabies") after the package has been installed. The user can parameterize the simulations with the number of years (or more generally time-steps if reproduction occurs on the scale of months) of reproduction, the number of individuals that reproduce each year, the distribution of the number of offspring for each pair of parents, the number of simulations, and the marker panels to be evaluated. For each year of the simulation, the individuals that reproduce are randomly chosen from the population and crossed according to a mating design: monogamy, polygyny, polyandry, or polygamy. The numbers of offspring for each pair can be drawn from one of three distributions: uniform, Poisson, or negative binomial. Genotypes of offspring are then

constructed by randomly sampling (with replacement) alleles at the multi-locus genotypes of their parents. This ensures that all loci are in linkage and Hardy-Weinberg equilibrium. The ages at which simulated offspring mature are drawn from a user-defined distribution. During the following year (or time-step), individuals that mature are available to be selected as parents. Each set of simulations can accommodate genotyping error at each locus of the marker panel and/or missing data (dataset-wide proportion of missing genotypes). Homogeneity tests can be performed with Genepop (Rousset 2008) within the R statistical environment to determine if the allelic frequencies of the simulated datasets differ substantially from the data from which the simulation was initiated.

Output of the simulation, parental and offspring genotypes, are formatted as input files for two commonly used parentage inference programs: Colony (Jones and Wang 2010) and FRANz (Riester et al. 2009). Functions within the PseudoBabies package facilitate the batch processing of all of the simulations with these two programs. These functions were designed to execute a single run of Colony (V2.0.5.7) in which the program is set to the full-likelihood-based method with medium length runs (5000N reconfigurations or 100N successful reconfigurations of the simulated annealing step, where N is the number of offspring), medium precision (number of elements in the transition matrix used for likelihood calculations), the true mating structure of the simulation, no sibship prior, and the true genotyping error rate for each of the loci specified. The run length determines how many reconfigurations of the pedigree are evaluated for each annealing temperature when searching for the maximum likelihood configuration. Increasing the run length corresponds to a 10-fold increase in the number of configurations considered. The precision of the likelihood calculation allows for the exclusion of minor terms in the likelihood calculation when those terms are small. A similar function performs a single run of the program FRANz. FRANz reduces the number of likelihood calculations of putative parent pairs by estimating the maximum number of mismatches expected between the parent-offspring dyads by simulation. FRANz was run with default parameters, but we specified complete sampling of parents contributing offspring to the sample as well as the maximum genotyping error rate that was used to simulate the data. All candidate parents were assigned identical birth years to

restrict the reconstructed pedigrees to a single generation. Each of the datasets was evaluated based on the percentage of correct parent-offspring dyad assignments.

[C] *Auke Lake Sockeye Salmon sampling and DNA extraction.* — Auke Lake is located approximately 17.7 km northwest of Juneau, Alaska. The lake is 1.6-km long by 1.2-km wide with a maximum depth of 31.4 m. Sockeye salmon spawn in two major tributaries, Lake Creek and Lake Two Creek (Ray et al. 2015). Auke Creek is the only outlet into Auke Bay. The National Marine Fisheries Service (NMFS) operates a permanent weir that allows the complete enumeration of out-migrating smolt in the spring and returning adults in the summer.

In 2008, tissue (axillary process) was sampled from each returning adult as it passed through the weir (n = 1264) and analyzed to determine allele frequencies in the brood year and to evaluate the power of the genetic loci for parentage analysis. Genetic samples were collected from all returning adult Sockeye Salmon in years 2009 (n = 4064) and 2013 (n = 2056). In addition to collecting genetic samples, scales were sampled from a subset of fish on multiple days of the run in each year. After clipping the axillary process, forceps were used to remove a scale from the preferred area (Clutter and Whitsel 1956). Scales were transferred to an acetate card and aged by NMFS staff at the Ted Stevens Marine Research Institute (Clutter and Whitesel 1956). From these data, ages at maturity were estimated for the population for use in the simulations.

In 2013, a subset of the returning wild fish was retained as broodstock for 26 experimental crosses to provide a test of the accuracy of simulation-based parentage assignment. Sires (n = 11) were crossed with dams (n = 22) in a 1:2 breeding design to produce both siblings and half siblings. Also, three sires and three dams were crossed in a 1:1 breeding design. Genetic samples were collected from 20 progeny of each cross. For each full-sibling family, an average of 8.7 full siblings were genotyped. One sire of the 2013 broodstock was not genotyped; however, the multilocus genotype was reconstructed from the known offspring with the program GERUD (Jones 2001, 2005). All fish that returned in brood years 2008 and 2013 (wild spawners and hatchery broodstock) as well as the subset of offspring from hatchery crosses (n = 201) were genotyped at 12 STR and 93 SNP loci from DNA isolated from the tissue. Fish from 2009 were genotyped at nine

STR and 48 SNPs. DNA was isolated from axillary process samples with Qiagen DNeasy™ kits (Qiagen, Valencia, California) or a proteinase K and ammonium acetate procedure (Puregene DNA™ isolation protocol; Gentra Systems, Minneapolis, MN). Isolated DNA was hydrated in 0.1X TE buffer (TE is 1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) and stored in 1.5 ml tubes at -20°C.

Microsatellite loci were amplified with primers that were fluorescently labeled with an IRDye® infrared dye (LI-COR, Lincoln, Nebraska, Table1). Gel electrophoresis was performed by loading 1 μL of PCR product and stop buffer (95% formamide, 0.1% bromophenol blue) into a 0.25 mm 6% polyacrylamide gel (19:1 acrylamide/bisacrylamide, 1X TBE gel buffer) on the LI-COR 4300 System™. All gels were run in 1X TBE buffer (0.09M Tris-Borate, 2mM EDTA, pH 8.3) at 40 W, 1500 V, 40 mA, and 45°C. Fragment sizes were scored with Saga^{GT} (Ver. 3.2.1, LI-COR) analysis software by comparing the fragments to IRD700 or IRD800 standard ladders (LI-COR, Biotechnology Division). Single nucleotide polymorphisms were amplified and genotyped at the Gene Conservation Laboratory of the Alaska Department of Fish and Game's (ADF&G GCL). The SNP markers were screened with Biomark Dynamic Arrays™ (Fluidigm) following the methods of Seeb et al. (2009).

Marker panels included both STR and SNP loci. Three multiplexes were made for the STR markers. We attempted to group STR loci by information content while keeping PCR reagent concentrations similar. The Fluidigm system used for SNP genotyping accommodated arrays that run 93, 48, or 24 loci. SNPs were rank ordered by their minor allele frequency (MAF) estimated from the 2008 return year and grouped by their rank order for each of the three array sizes. Fifteen marker panels included different combinations of the three STR multiplexes and the three different SNP arrays were evaluated for their utility in assigning parentage to Auke Creek Sockeye Salmon (Table S2). To estimate locus-specific error rates, 8% of the project fish were re-extracted, re-amplified, and re-genotyped for both SNP and STR markers.

[C] *Auke Creek Sockeye evaluation.* — We first determined if there was sufficient power in the marker panel to use exclusion-based parentage methods. Parent-pair exclusion probabilities for each locus and the combined exclusion probabilities for each marker

panel were calculated in Cervus (Marshall et al. 1998) with the 2008 return year data. Parentage inference for the known hatchery offspring from the 2013 brood year (n = 201 offspring) was then conducted with the exclusion-based program WHICHPARENTS (Will Eichert, Bodega Bay Marine Lab, California available at https://marinescience.ucdavis.edu/research-programs/conservation/salmon-research/software). The assignment accuracy of each panel was evaluated when the pool of potential parents included was limited to only the broodstock used to produce the offspring (n = 42) versus the entire 2013 return year (broodstock spawners and wild return; n = 2056). The program WHICHPARENTS uses Mendelian incompatibilities to exclude putative parental genotypes for a given offspring genotype. Errors that resulted in incompatibilities between parents and offspring genotypes, which arise from genotyping mistakes or mutations, were accommodated by allowing up to two mismatches between parent and offspring genotypes in each analysis.

Three series of simulations were then used to estimate the ability of each marker panel to accurately infer parentage with likelihood-based parentage methods (Figure 3.1). These three sets of simulations allowed us to eliminate marker panels that contained insufficient information to deduce parentage with simple simulations and to proceed to more complex simulations that more closely resembled the experimental design of the supplementation project. Marker panels that failed to accurately determine at least 95% of the relationships for any set of simulations were removed from subsequent simulations.

The first series of simulations used the genotypes from the 2008 return year and simulated a number of offspring from random pairings (Figure 3.1A). Ten simulated datasets were evaluated for each of the 15 marker panels. For each simulation, 625 random pairs of parents were selected. A polyandrous mating structure (allowing for multiple sires per dam) was used which produced both full and half siblings in the offspring. The number of offspring that each pair produced was drawn from the Poisson distribution with a mean of two. All simulations assumed perfect genotypic data (i.e., no genotyping errors and no missing data). Parentage inference was conducted with the programs FRANz and Colony. The accuracy of parentage assignment was evaluated with a strict (P = 0.95) cutoff for assignment probability.

Peculiar results from Colony for these simulations (described below in Results – Simulations) led us to perform two additional sets of simulations as above but with minor modifications: one with a monogamous mating structure and one that, in addition to a monogamous mating structure, removed the family structure in the founders by simulating unrelated individuals as the potential parents. Parentage inference for these simulations was conducted only in Colony. We repeated the simulations described above with both 1% genotyping error and 1% missing data to compare to results from empirical hatchery data.

The second series of simulations included age structure and missing data. We simulated 10 datasets and for each dataset we applied three levels of missing data (0%, 2% and 5%) randomly across all individuals and loci. For each simulation, we projected the 2008 return year 13 years into the future under the assumption of no gene flow into the population (Figure 3.1B). This period was chosen because the age structure of the population required randomly generating a number of genotypes from the allele frequency distribution for the first six years. By extending the simulation over 2 generations, we better approximated natural conditions by allowing the possibility of related individuals in the parent set, which reduced parentage assignment accuracy (i.e., the aunt/uncle effect; Olsen et al. 2001). For each year of the simulation, 625 random pairs of parents were selected and crossed with polygyny as the mating structure. Polygyny was chosen because it may be more common than polyandry in Pacific salmon; females typically deposit all of their eggs into a single redd and males compete amongst one another for reproductive access, but provide no parental care (redd construction and maintenance; Bentzen et al. 2001, Ackerman 2017), allowing them to court multiple females. Because the sex of parents was not incorporated into the analyses for either program, the mating structure is the same for polygyny and polyandry. The number of offspring that each pair produced was drawn from the Poisson distribution with a mean of two. The age of each individual was drawn from the empirical distribution determined from scale samples. The final year was sampled as offspring and all the parents from brood years that could have contributed to the offspring collection, determined by the age at maturity, were sampled as putative parents. Parentage inference was conducted with

the programs FRANz and Colony. The accuracy of parentage assignment was evaluated with a strict ($P = 0.95$) cutoff for the posterior probability of assignment to each parent.

The third series of simulations was designed to mimic a hatchery breeding experiment (Figure 3.1C). We simulated ten datasets with two levels of genotyping error (0% and 1%) and two levels of missing data (0% and 1%). Simulations were conducted in a similar manner as previously, but the simulation was run for 16 years and on the 10^{th} year, 13 sires were crossed with 26 dams in a 1:2 breeding design. These simulations were extended to year 16 because offspring from the hatchery crosses in year 10 would be sampled in years 13 through 16 following the age at maturity distribution. The distribution of offspring from the hatchery crosses was the same as that of the wild population (Poisson distributed with mean of 2). While hatchery programs can result in full sibling family groups with a mean and variance that exceed two, large full-sib families typically enhance the performance of pedigree reconstruction algorithms. Results from the first and second set of simulations suggested that Colony would not lead to improved resolution of families so parentage inference was conducted only with the program FRANz. For these simulations, each return year was coded as a separate population and a geographic distance matrix was used to limit the putative parent pairs to fish that returned in the same year. A separate set of simulations was conducted to evaluate the effect that specifying the sex of each parent would have on assignment accuracy. Including sex can reduce the number of putative parent-offspring triads evaluated and may eliminate close relatives of the true parents of the opposite sex. These simulations were run as described above, but the true sex of the parents was made available to the program FRANz during parentage inference.

In addition, we tested the behavior of our simulations and the likelihood-based approaches by comparing with the empirical data from hatchery crosses. Two sets of analyses were conducted on the fry produced from hatchery crosses. The first analysis used only the broodstock that produced the fry as potential parents while the second used the entire 2013 return year as potential parents. Similarly, we genotyped individuals from the 2009 return year for a subset of markers (9 STRs and 48 SNPs) so that we could assign the fish used as broodstock to potential parents from the two return years that should comprise over 50% of the fish contributing to the 2013 return.

[A] Results

[B] Exclusion

We surveyed 105 genetic markers (12 STR and 93 SNP loci) from Sockeye Salmon that returned to the Auke Creek system in 2008. The total number of alleles for each locus ranged from 2 to 36. No locus was fixed and the expected heterozygosities (H_{Exp}) ranged from 0.001–0.894. Single-locus exclusion probabilities were substantially larger for STR markers than SNPs because of their larger number of alleles and high resultant H_{Exp} (Table S2). An exception to this was locus *Oki1a*, which had only three alleles and had a H_{Exp} of 0.467. Parent-pair exclusion probabilities ranged from 0–0.93. Combined exclusion probabilities calculated for the marker panels composed of STR and SNP markers were high, all exceeding 0.98. Decreasing the number of markers of either type led to a decrease in the estimated exclusion probability. A decrease in the number of STR multiplexes included in the panel affected accuracy more than a decrease in the number of SNPs.

The high combined exclusion probabilities for each panel were compared with results from exclusion-based parentage of fry from the 2013 Auke Creek hatchery brood year. When only the broodstock was considered as potential parents (n = 41), five marker panels were able to resolve more than 95% of the relationships (Figure 3.2A), three panels included 12 STRs and two panels included 9 STRs. However, when the entire return year was included as potential parents (n = 2056) the most informative panels only resolved 50% of the parent-offspring dyads. The average number of non-excluded parents increased from 2.06 to 4.11 for the marker panel with 12 STRs and 93 SNPs.

[B] Simulations

The six marker panels that included only a single marker type (STR or SNP) were unable to correctly assign parentage for more than 95% of the relationships with the two programs tested in the simplest of simulations. All nine marker panels composed of a combination of the two marker types performed better when FRANz was used to determine parentage (Figure 3.3A). Increasing the number of included markers increased assignment accuracy for the pairwise approach of FRANz. Unexpectedly, when Colony was used for parentage analyses, an increase in the number of SNP markers beyond the

first 24 included with any STR markers reduced accuracy (Figure 3.3A). The majority of these errors were not the result of an incorrect parent being assigned, but rather the failure of assignment to a parent in the set (Type II error). If simulations were conducted with a monogamous mating structure so that half-siblings were not present, all nine marker panels composed of both marker types yield more than 95% correct assignment, but increasing the number of SNP markers above 24 still resulted in reduced accuracy (Supplemental Figure S3.1). Simulations that removed any family structure among the parents by simulating random genotypes from the allele frequency distribution resulted in median accuracies that exceeded 99%; however, the trend of decreased accuracy with more SNP markers persisted albeit to a much smaller degree (Supplemental Figure S3.2). The minor allele frequency (MAF) of the 24 highest-ranked SNPs ranged from 0.50 to 0.23, while the MAFs of the remaining SNPs ranged from 0.23 to 0.09.

The second series of simulations introduced age structure and the potential for missing data (up to 5%). The effect of adding age structure substantially increased the number of potential parents. For the same number of offspring, four return years instead of a single return year were included as potential parents. Accuracy of parentage assignment decreased with missing data and had the largest effect for panels with fewer markers (Figure 3.3B). The mean accuracy of parentage with Colony was less than 95% for all marker panels. Again, we observed decreased accuracy with the addition of SNP markers after the first 24 were included (Figure S3.3). FRANz performed much better. For a 100% genotyping rate, only a single panel failed to accurately resolve 95% of the relationships (9 STRs and 24 SNPs) with FRANz. If 5% of the genotypes are missing, the accuracy of an additional two panels (12 STRs & 24 SNPs and 5 STRs & 48 SNPs) fell below 95% (Figure 3.3B).

In the hatchery simulation scenario, to identify individuals that were produced from a hatchery experiment in year y-6 (Figure 3.1C), four independent analyses were performed for years y to y-3. Because each analysis is run independently, these simulations do not increase the number of potential parents and should behave similarly to the preceding series. For this series of simulations, we used empirical estimates from the 2013 return year for the amount of missing data (1%) and genotyping error (1%). Results from these simulations did not differ substantially from the second series of

simulations (Figure S3.4). Assignment accuracy exceeded 97% for all five panels. The panel that included 9 STRs and 48 SNPs produced an accuracy of 97.5% while minimizing the total number of markers that need to be analyzed. A small proportion of parent-offspring dyads (1.3%) were inferred correctly as parents, but with the posterior probability for assignment < 0.95, and consequently they remained unassigned (type II errors). FRANz reports the posterior probability for each parent in the likeliest pair. If these probabilities were used instead of the parent-pair posterior probability for accepting parentage assignment, overall accuracy increased to 99.2%. Of the incorrect single-parent assignments, 34% were assigned a parent that was a full sibling to the true parent (Figure 3.4). The posterior probabilities for these avuncular assignments were lower on average than those for true parents, but some exceeded our 95% cutoff. The type I error was low (0.03%) for these simulations as was the type II error (0.75%). The inclusion of sex of parents led to a marginal increase, 0.15%, in the proportion of parents inferred correctly across all marker panels. Unexpectedly, inclusion of sex of the putative parents with the least informative marker panel (5 STRs and 93 SNPs) resulted in a decrease in proportion of true parent pairs being inferred when a posterior probability cutoff of 0.95 was used. When a threshold probability of accepting parentage was removed, all marker panels resolved more than 98.7% of the true parent pairs and the inclusion of sex only increased accuracy by 0.03% across all marker panels.

[C] *Likelihood-based methods on hatchery data*. — We conducted parentage inference on fry with known parents from the 2013 brood year with the program FRANz to determine if inferences from the simulations, such as the exclusion probabilities, were overly optimistic. The broodstock used to produce all fry were included in our candidate parents, so complete sampling of parents was specified for these analyses. The first set of simulations was the most reasonable to compare with these assignments because they used a single return year as the potential parents. Six panels yielded accuracies that exceeded 95% when the pool of potential parents was limited to the broodstock, or included all adults that returned in 2013 (Figure 3.2B). As expected, accuracy decreased as the pool of parents increased, but reducing the number of STRs from 12 to 9, or the number of SNPs from 93 to 48, had little effect. The results were largely concordant with

the simulations (Figure 3.2B). For thirteen of the fifteen marker panels, the median expected accuracy for simulations fell between the results obtained with the empirical data. In two instances, estimates based on the simulations slightly exceeded those observed from the empirical data.

Sockeye Salmon from the 2009 return year were genotyped at 9 STRs and 48 SNPs, which allowed us to test this marker panel for its ability to assign parentage to our 2013 broodstock. Based on age estimates from scales from the 2013 return year we would expect 79% (27% age 4 and 52% age 5) of the fish to be ages 4 and 5, while 20% would be age 6 and unsampled in return year 2007. For these analyses, there was the possibility of un-sampled parents, so the number of sampled candidate parents (n) was set to the number of parents in 2008 and 2009 and the number of unsampled candidate parents (N-n) was estimated to be 2520. The number of unsampled parents was set such that the estimate slightly exceeded half the number of fish that returned in years 2007 and 2010, which are the two other brood years that could contribute offspring to the 2013 return year. Of the 41 sockeye used as broodstock, 48% were assigned back to parents from the 2008 and 2009 return years. Of the fish used as broodstock, 31.7% and 17.1% were inferred to be five- and four-year-old fish respectively. Two fish failed to assign to two parents, but had high posterior probabilities (> 0.95), eleven fish failed to assign to any parents and seven fish assigned to one parent but with low posterior probability (< 0.95).

[A] Discussion

Identifying first-degree relationships (i.e., relationships where approximately half of the genes are shared) among individuals is a simple task if an adequate portion of the genome has been sampled from each individual. While exclusion-based parentage remains 'the paragon of parentage analysis' (Jones and Ardren 2003), often studies must rely on likelihood-based methods because of the tradeoff between sampling many individuals in a population to capture the true relationships and sampling many genetic loci in order to have adequate power to resolve those relationships. There has been substantial effort to develop summary statistics describing the relative utility of a panel of markers, but we found with empirical data from Sockeye Salmon in Auke Creek that

these statistics can sometimes be misleading. The results described here showed that both false exclusion (individuals that are excluded because of scoring errors or mutations) as well as a lack of power in the marker panels to exclude all but the two correct parents was a problem when analyzing a single return year of Sockeye Salmon from Auke Creek. Exclusion probabilities, while still reported despite the use of likelihood-based inference methods, have been criticized for a number of reasons. Double et al. (1997) advised that exclusion probabilities could be misleading because close relatives are included among the putative parents. High levels of philopatry characterize Pacific salmon populations, leading to relatively small breeding populations, and it is not unrealistic to assume that there may be groups of siblings and half siblings included among the putative parents, which results in over confidence in a set of markers. If exclusion probabilities are unable to adequately estimate the accuracy one might expect to obtain with a panel of genetic markers, simulations should be undertaken. We presented a simulation package ('PseudoBabies') written in the R statistical environment and used it to evaluate a suite of markers for a parentage study on Sockeye Salmon in Southeast Alaska. Our simulation package has no restrictions on the number or type of loci analyzed and can be used in conjunction with parentage programs to evaluate their relative performance. Similarly, because PseudoBabies is written as a package in R it is amenable to batch processing and can be run on any computer operating system.

We applied simulations conducted in PseudoBabies to a marker panel being considered for a large-scale hatchery/wild parentage study. Substantial differences in the expected power of our marker panels were observed with the three simulation series, allowing us to reduce the number of markers that would be needed for accurate pedigree reconstruction. Our results suggest that five of the fifteen marker panels evaluated would be sufficient to resolve relationships for the Auke Creek population. We were also able to evaluate the efficacy of two parentage programs. The program Colony appeared to have decreased accuracy when markers with low variability were included. Results from additional simulations suggest that these markers increase the likelihood of some non-parents, which resulted in the failure to assign any parent. This issue is compounded when closely related individuals are included in the pool of parents and when half-sibling relationships are possible. With large numbers of SNP markers becoming available from

next generation sequencing, it is critical to evaluate the bias that may be associated by including all the markers identified in downstream analyses. With complete sampling of the population, the presence of full siblings resulted in a number of type I errors. When full siblings cannot be excluded as putative parents *a priori*, they can have higher log-odds ratio (LOD) score than the true parents (Thompson and Meagher 1987). The LOD score represents how much more likely that the putative mother and father are than random parents drawn from the population (Meagher and Thompson 1986). In our simulations some full siblings of the true parents had higher likelihoods and were mistakenly inferred as parents. While the difference in the LOD scores between the two most likely parentages (ΔLOD score; Marshall 1998), is not reported in FRANz, the posterior probability of the parent-pair and single parent parentage did give some indication of the relative strength. The posterior probability for offspring assigned to full siblings of their true parents was generally smaller than those offspring that were assigned to their true parent with a posterior probability of less than 0.95 (Figure 3.5). Decreasing our cutoff value for the posterior probability increased the number of true full parents identified while only marginally increasing the type I error. The use of statistical cutoffs for parentage assignment is a contentious issue that can bias downstream estimates (Araki and Blouin 2005, Ford and Williamson 2009), but simulations can help evaluate the effect of their use on estimating population parameters (see Chapter 2).

Data from the 2013 hatchery fry and broodstock were used to validate our simulations. Estimates of accuracy based on simulations matched well with the empirical results; however, they were slightly optimistic when compared with the assignments of the 2013 hatchery fry. Occasionally the results of simulations more closely matched analyses with a smaller pool of potential parents. The 2013 return year contained more candidate parents (n = 2056) than the simplest simulations, which were constrained to the size of the 2008 return year (n = 1264), and missing data were not uniformly spread throughout the empirical dataset which may explain these discrepancies. The assignment of the Sockeye Salmon used as broodstock in 2013 to parents in both 2008 and 2009 suggests that 9 STR and 48 SNPs should provide adequate information to assign parentage with a high degree of accuracy. Unfortunately, data were not obtained for the 2007 return year from which half of the 2013 broodstock could have originated.

The inclusion of the true sex of putative parents resulted in marginal gains in accuracy, but identifying the true sex of fish sampled at a weir is problematic. Sex identification based on secondary sexual characteristics (i.e., kype and vent size and shape) can decrease the pool of putative parents, but early in the run when fish have not reached sexual maturity (P. Barry, pers. obs.), misspecification of sex would lead to exclusion of a true parent from the possible parent-offspring dyads. Our results indicated that additional effort to include a sex marker (sdY; Larson et al. 2016) in the genotyping panel might have limited benefit. The reported discordance between phenotypic sex and sdY genotype in some Sockeye Salmon populations (Larson et al. 2016) suggests that potential error may far outweigh the increase in assignment accuracy (< 1%) we observed by including known sex.

These results demonstrate the applicability of the R library package PseudoBabies to marker panel selection for parentage studies as well as the evaluation of software packages used for pedigree reconstruction. While PseudoBabies currently simulates data for semelparous species with variable age at maturity it could be extended to iteroparous species. Similarly, our primary focus was to identify parents from previous generations of returning fish; however, the program has since been adapted to evaluate marker panels for sibship reconstruction. Written as open source code in the R statistical environment the package is flexible and can be adapted to include a wide array for biological and experimental complexity.

[A] References

Ackerman, M., B. Hand, R. S. Waples, G. Luikart, R. Waples, C. Steele, B. Garner, J. McCane, and M. Campbell. 2017. Effective number of breeders from sibship reconstruction: Empirical evaluations using hatchery Steelhead. Evolutionary Applications 10(2):146–160.

Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172(4):2567–2582.

Araki, H., and M. S. Blouin. 2005. Unbiased estimation of relative reproductive success of different groups: Evaluation and correction of bias caused by parentage assignment errors. Molecular Ecology 14:4097–4109.

Araki, H., R. Waples, W. Ardren, B. Cooper, and M. Blouin. 2007. Effective population size of Steelhead Trout: Influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. Molecular Ecology 16(5):953–966.

Baruch, E., and J. Weller. 2008. Estimation of the number of SNP genetic markers required for parentage verification. Animal Genetics 39(5):474–479.

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(2):127–136.

Bromaghin, J. 2008. Bels: Backward elimination locus selection for studies of mixture composition or individual assignment. Molecular Ecology Resources 8(3):568–571.

Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18(5):249–256.

Chamberlain, J., R. Gibbs, J. Rainer, P. Nguyen, and C. Thomas. 1988. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Research 16(23):11141–11156.

Chakraborty, R., T. R. Meagher, and P. E. Smouse. 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118(3):527–536.

Christie, M., M. Ford, and M. Blouin. 2014. On the reproductive success of early-generation hatchery fish in the wild. Evolutionary Applications 7(8):883–896.

Clutter, R. and L. Whitsel. 1956. Collection and interpretation of Sockeye Salmon scales. International Pacific Salmon Fisheries Commission, New Westminster, Canada.

Double, M., A. Cockburn, S. Barry, and P. Smouse. 1997. Exclusion probabilities for single-locus paternity analysis when related males compete for matings. Molecular Ecology 6(12):1155–1166.

Duchesne, P., M. H. Godbout, and L. Bernatchez. 2002. PAPA (package for the analysis of parental allocation): A computer program for simulated and real parental allocation. Molecular Ecology Notes 2(2):191–193.

Ford, M. J., and K. S. Williamson. 2009. The aunt and uncle effect revisited–the effect of biased parentage assignment on fitness estimation in a supplemented salmon population. Journal of Heredity 101(1):33–41.

Glaubitz, J. C., O. E. Rhodes Jr., and J. A. Dewoody. 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Molecular Ecology 12(4):1039–1047.

Hayden, M., T. Nguyen, A. Waterman, and K. Chalmers. 2008. Multiplex-ready PCR: A new method for multiplexed SSR and SNP genotyping. BMC Genomics 9.

Jones, A. 2001. GERUD 1.0: A computer program for the reconstruction of parental genotypes from progeny arrays using multilocus DNA data. Molecular Ecology Notes 1(3):215–218.

Jones, A. 2005. GERUD 2.0: A computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parents. Molecular Ecology Notes 5(3):708–711.

Jones, A. G., and W. R. Ardren. 2003. Methods of parentage analysis in natural populations. Molecular Ecology 12(10):2511–2523.

Jones, A., C. Small, K. Paczolt, and N. Ratterman. 2010. A practical guide to methods of parentage analysis. Molecular Ecology Resources 10(1):6–30.

Jones, O. R., and J. Wang. 2010. COLONY: A program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10(3):551–555.

Larson, W. A., G. J. McKinney, J. E. Seeb, and L. W. Seeb, 2016. Identification and characterization of sex-associated loci in Sockeye Salmon using genotyping by-sequencing and comparison with a sex-determining assay based on the sdY gene. Journal of Heredity 107:559–566.

Marshall, T., J. Slate, L. Kruuk, and J. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7(5):639–655.

Matala, A., S. Narum, W. Young, and J. Vogel. 2012. Influences of hatchery supplementation, spawner distribution, and habitat on genetic structure of Chinook Salmon in the South Fork Salmon River, Idaho. North American Journal of Fisheries Management 32(2):346–359.

Matson, S., M. Camara, W. Eichert, and M. Banks. 2008. P-loci: A computer program for choosing the most efficient set of loci for parentage assignment. Molecular Ecology Resources 8(4):765–768.

Meagher, T., and E. Thompson. 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theoretical Population Biology 29(1):87–106.

Neff, B. 2001. Genetic paternity analysis and breeding success in Bluegill Sunfish (*Lepomis macrochirus*). Journal of Heredity 92(2):111–119.

Olivier, M., L.-M. Chuang, M.-S. Chang, Y.-T. Chen, D. Pei, K. Ranade, A. de Witte, J. Allen, N. Tran, D. Curb, R. Pratt, H. Neefs, M. de Arruda Indig, S. Law, B. Neri, L. Wang, and D. R. Cox. 2002. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. Nucleic Acids Research 30(12):e53–e53.

Olsen, J. B., C. Busack, J. Britt, and P. Bentzen. 2001. The aunt and uncle effect: an empirical evaluation of the confounding influence of full sibs of parents on pedigree reconstruction. Journal of Heredity 92:243–247.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ranade, K., M. S. Chang, C. T. Ting, D. Pei, C. F. Hsiao, M. Olivier, R. Pesich, J. Hebert, Y. D. Chen, V. J. Dzau, D. Curb, R. Olshen, N. Risch, D. R. Cox, and D. Botstein. 2001. High-throughput genotyping with single nucleotide polymorphisms. Genome Research 11(7):1262–1268.

Ray, J., S. Sethi, J. Eiler, and J. Joyce. 2015. Prespawning movements and spawning distribution of Sockeye Salmon in an urbanizing Alaskan lake. Journal of Fish and Wildlife Management 6(2):472–485.

Reisenbichler, R., and S. Rubin. 1999. Genetic changes from artificial propagation of Pacific salmon affect the productivity and viability of supplemented populations. ICES Journal of Marine Science 56(4):459–466.

Riester, M., P. Stadler, and K. Klemm. 2009. FRANz: Reconstruction of wild multi-generation pedigrees. Bioinformatics 25(16):2134–2139.

Roesti, M., W. Salzburger, and D. Berner. 2012. Uninformative polymorphisms bias genome scans for signatures of selection. BMC Evolutionary Biology 12:94.

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the genepop software for windows and linux. Molecular Ecology Resources 8(1):103–106.

Sanderlin, J., N. Lazar, M. Conroy, and J. Reeves. 2012. Cost-efficient selection of a marker panel in genetic studies. Journal of Wildlife Management 76(1):88–94.

Seeb, J. E., C. E. Pascal, R. Ramakrishnan, and L. W. Seeb. 2009. SNP genotyping by the 5'-nuclease reaction: Advances in high-throughput genotyping with non-model organisms. Methods in Molecular Biology 578:277–292.

Smith, C. T., W. D. Templin, J. E. Seeb, and L.W. Seeb. 2005. Single-nucleotide polymorphisms (SNPs) provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook Salmon caught in Yukon River fisheries. North American Journal of Fisheries Management 25: 944–953.

Thompson, E., and T. Meagher. 1987. Parental and sib likelihoods in genealogy reconstruction. Biometrics 43:585–600.

Thrasher, D., B. Butcher, L. Campagna, M. Webster, and I. Lovette. 2018. Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. Molecular Ecology Resources 18(5):953–965.

Vandeputte, M. 2012. An accurate formula to calculate exclusion power of marker sets in parentage assignment. Genetics Selection Evolution 44:36.

Vandeputte, M., M. Kocour, S. Mauger, M. Dupont-Nivet, D. De Guerry, M. Rodina, D. Gela, D. Vallod, B. Chevassus, and O. Linhart. 2004. Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile Common Carp (*Cyprinus carpio* l.). Aquaculture 235(1–4):223–236.

Venditti, D., R. Kinzer, K. Apperson, B. Barnett, M. Belnap, T. Copeland, M. Corsi, and K. Tardy. 2018. Effects of hatchery supplementation on abundance and productivity of natural-origin Chinook Salmon: Two decades of evaluation and implications for conservation programs. Canadian Journal of Fisheries and Aquatic Sciences 75(9):1495–1510.

Wang, J. 2007. Parentage and sibship exclusions: Higher statistical power with more family members. Heredity 99(2):205–217.

Wang, J. 2014. Estimation of migration rates from marker-based parentage analysis. Molecular Ecology 23(13):3191–3213.

Whiteley, A., J. Coombs, B. Letcher, and K. Nislow. 2014. Simulation and empirical analysis of novel sibship-based genetic determination of fish passage. Canadian Journal of Fisheries and Aquatic Sciences 71(11):1667–1679.

Wiener, A. S., M. Lederer, and S. H. Polayes. 1930. Studies in isohemagglutination. IV. On the chances of proving non-paternity with special reference to blood groups. The Journal of Immunology 19(3):259–282.

Figure Captions

Figure 3.1: Schematic of three series of simulations used to evaluate genetic marker panels for parentage inference of Sockeye Salmon in Auke Creek, AK. Three series of simulations were conducted to evaluate 15 genetic marker panels. Genotypes from the 2008 return year were used to simulate offspring. For each year (circle) in the simulation, 625 random pairs of parents were selected. A polyandrous mating structure was used to simulate offspring. The number of offspring that each pair produced was drawn from the Poisson distribution with a mean of 2. A) A single reproduction event. B) Simulations with age structure in the return year. Age at maturity proportions were based on scale samples from the 2008 return year. For a single year sampled as offspring (grey circle), the pool of potential parents included all fish returning 3 to 6 years prior. C) Simulations to mimic assignment of hatchery produced offspring. Hatchery fish can return at four ages so each simulation was analyzed for each of the four return years (grey circles).

Figure 3.2: Proportion of hatchery Sockeye Salmon fry that were correctly assigned to known parents with (A) an exclusion-based approach and allowing for a maximum of two Mendelian incompatibilities between offspring and parent genotypes or (B) a likelihood-based approach (FRANz). The potential parents of the offspring included in the analyses were either the broodstock (true parents) or the entire 2013 return year including the broodstock. The dotted line represents 95% accuracy.

Figure 3.3: Boxplots of the percent of correct assignment of simulated Sockeye Salmon offspring with 15 different marker panels composed of varying numbers of microsatellite (STR) and single nucleotide polymorphism (SNP) loci. The box extends from the first quartile to the third quartile with a horizontal line through the box at the median. The whiskers extend from each quartile to the minimum and maximum. The dotted line represents 95% accuracy. A) Simulations consisting of a single round of reproduction with no missing data or genotyping error. Parentage inference was conducted with the programs FRANz and Colony. B) Simulations consisting of multiple reproductive cycles based on variable age at maturity with missing data ranging from 0 to 5% and no genotyping error. Parentage inference was conducted with FRANz.

Figure 3.4: Boxplot of the posterior probability for single parent assignment of simulated offspring (n is the number of parent-offspring dyads) of Sockeye Salmon from the hatchery experimental design with 1% missing data and 1% genotyping error for a marker panel with 9 STRs and 48 SNPs. Each box extends from the first quartile to the third quartile. The median marks the mid-point of the data and is shown by a horizontal line. The lower and upper whiskers extend from each quartile to the minimum and maximum respectively. Parents were inferred correctly with a posterior probability greater than 0.95, inferred correctly but with a posterior probability less than 0.95, parentage was assigned to a full sibling of a true parent, or parentage was assigned incorrectly.

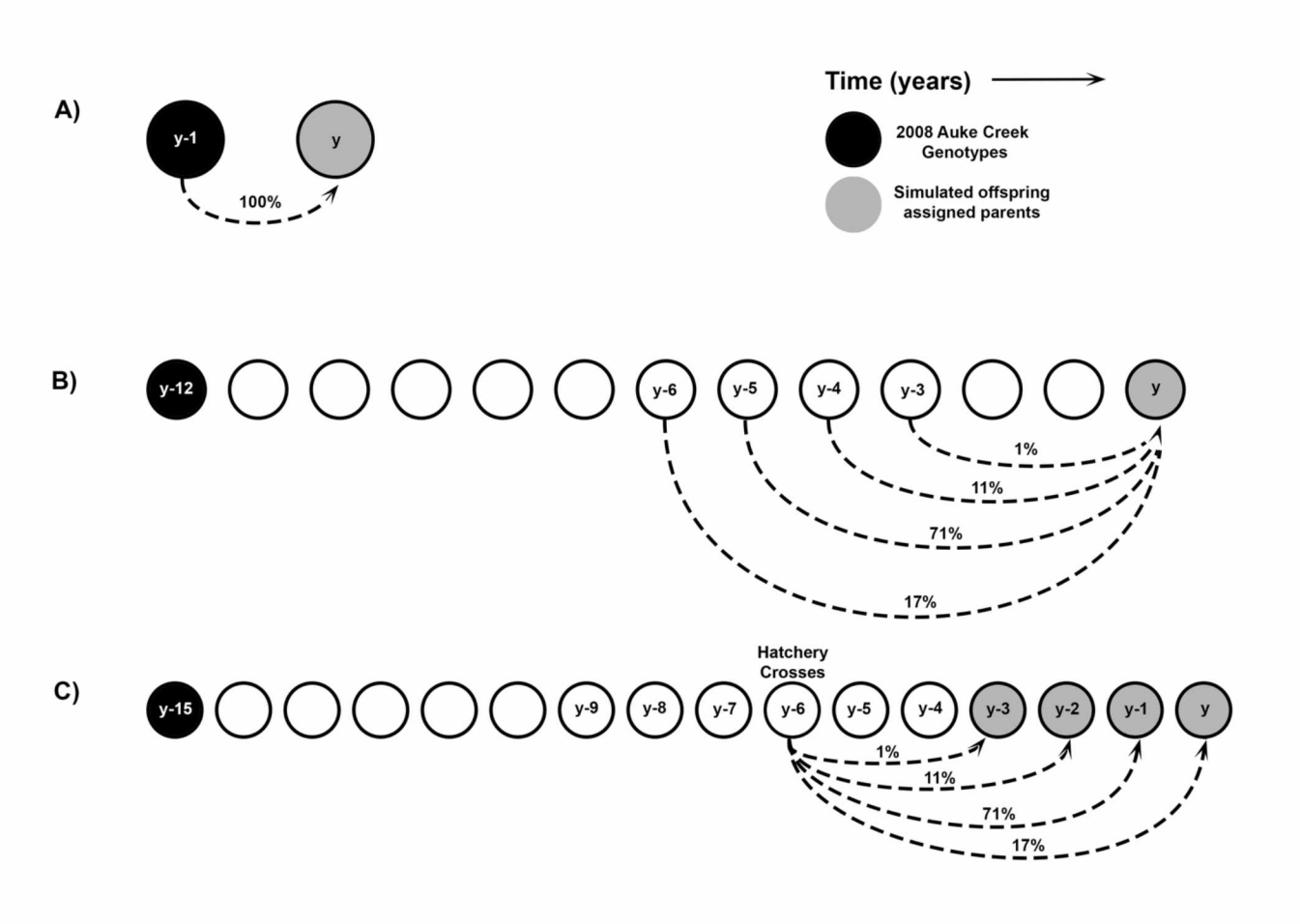

A)
y-1
y
100%
Time (years)
2008 Auke Creek Genotypes
Simulated offspring assigned parents
B)
y-12
y-6
y-5
y-4
y-3
y
1%
11%
71%
17%
C)
Hatchery Crosses
y-15
y-9
y-8
y-7
y-6
y-5
y-4
y-3
y-2
y-1
y
1%
11%
71%
17%

Figure 3.1

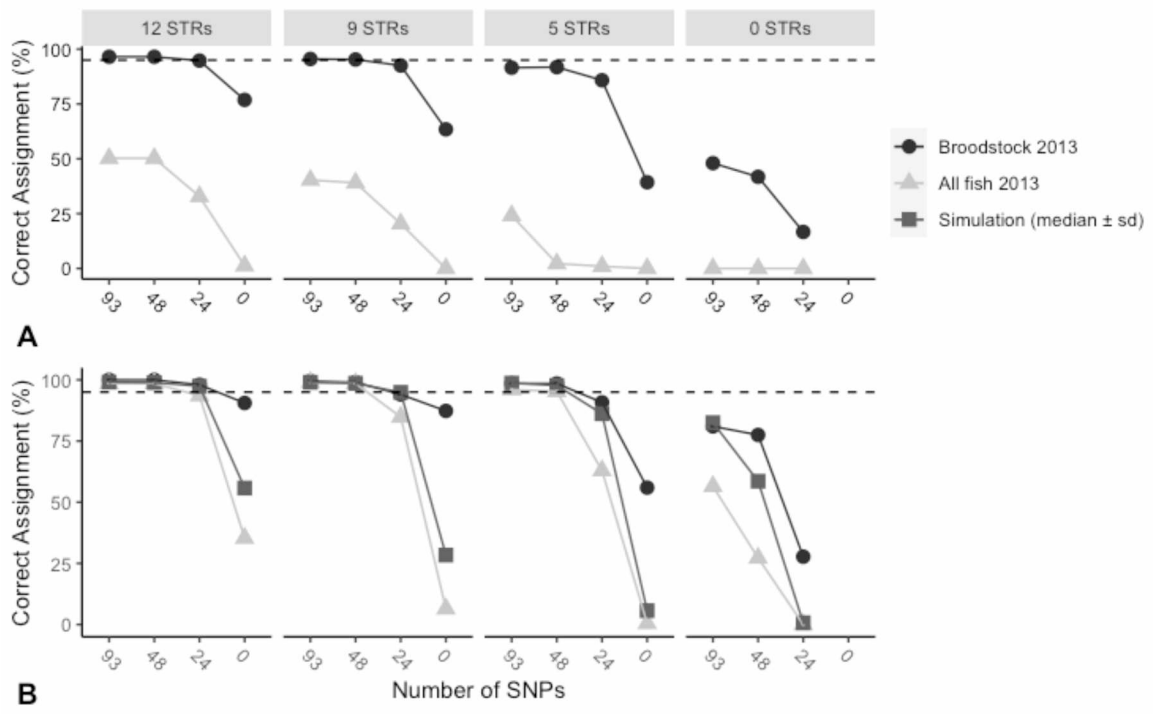

12 STRs
9 STRs
5 STRs
0 STRs
Correct Assignment (%)
100
75
50
25
0
93
48
24
0
A
B
Number of SNPs
Broodstock 2013
All fish 2013
Simulation (median ± sd)

Figure 3.2

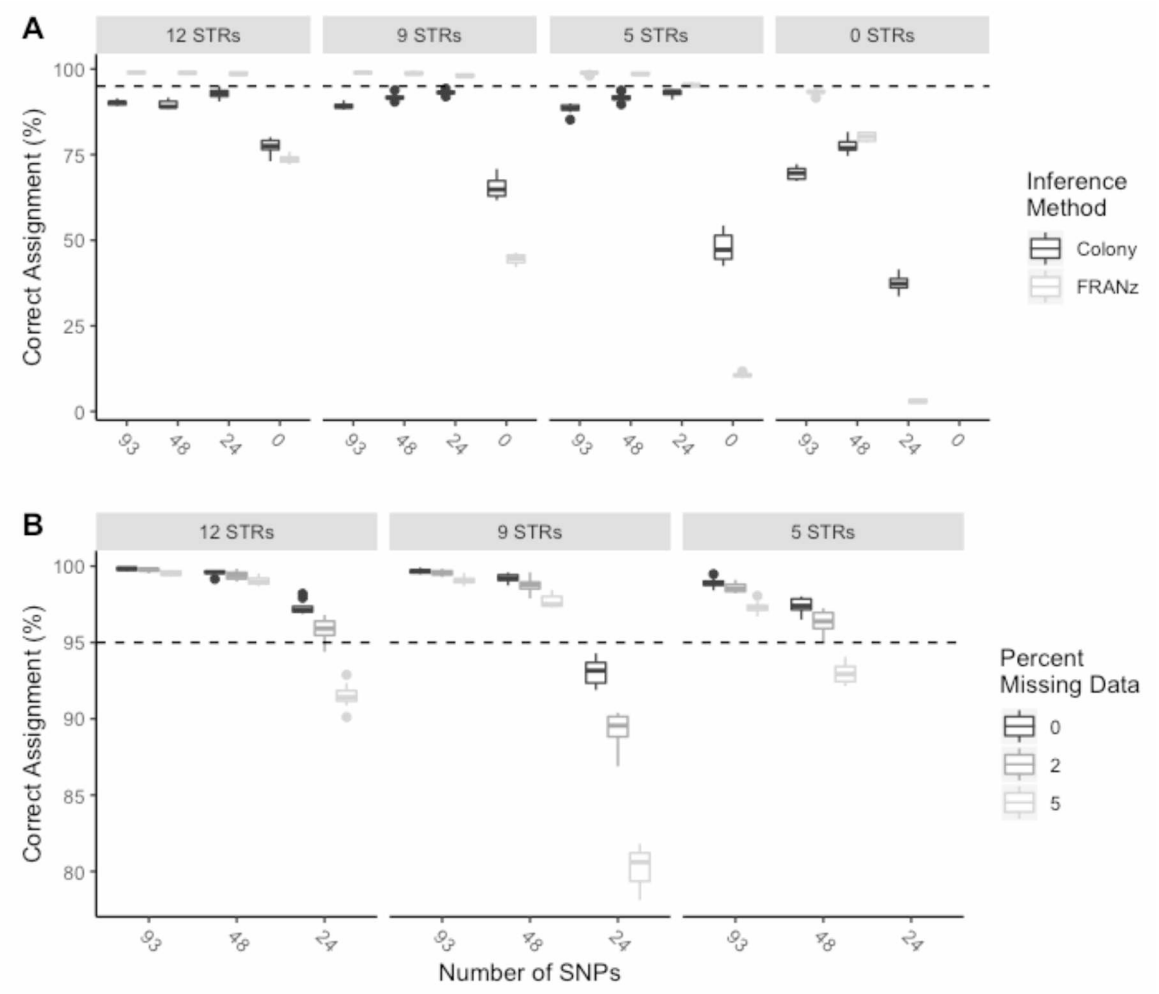

A
12 STRs
9 STRs
5 STRs
0 STRs
Correct Assignment (%)
100
75
50
25
0
93
48
24
0
Inference Method
Colony
FRANz
B
12 STRs
9 STRs
5 STRs
Correct Assignment (%)
100
95
90
85
80
93
48
24
Number of SNPs
Percent Missing Data
0
2
5

Figure 3.3

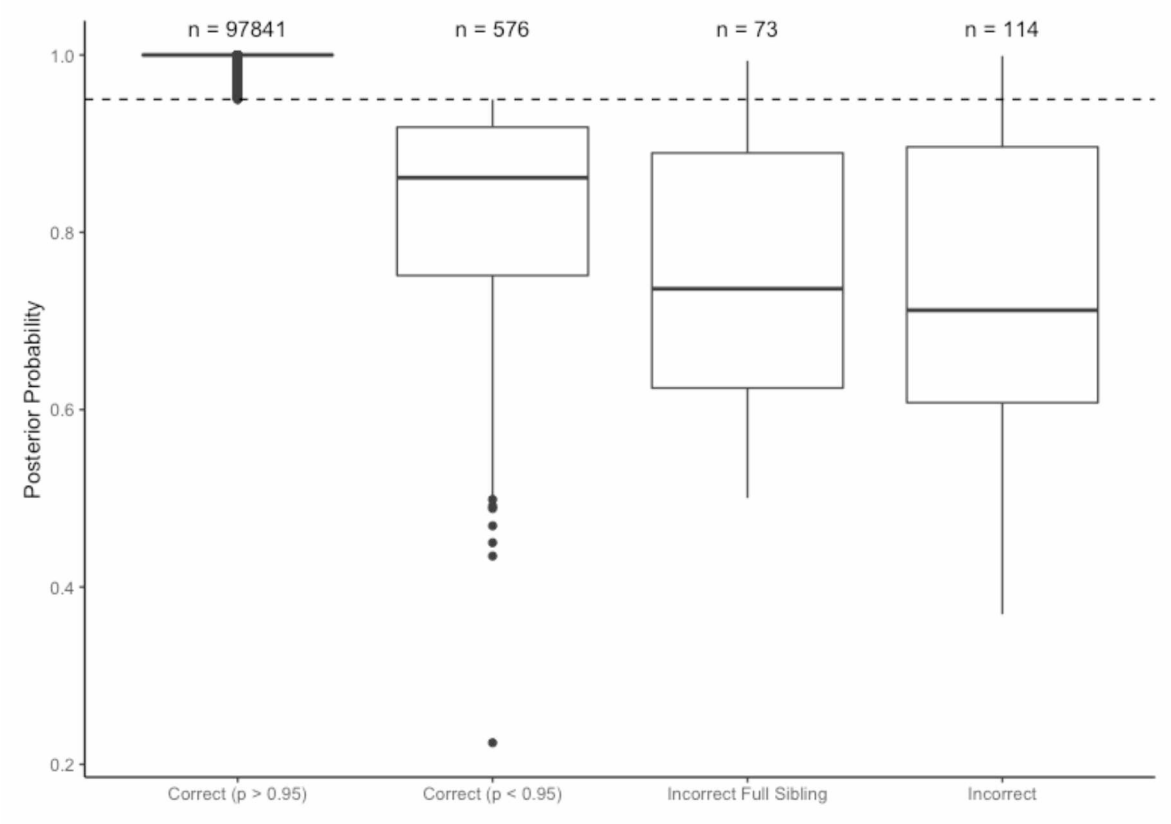

n = 97841
n = 576
n = 73
n = 114
1.0
0.8
0.6
0.4
0.2
Posterior Probability
Correct (p > 0.95)
Correct (p < 0.95)
Incorrect Full Sibling
Incorrect

Figure 3.4

Table S3.1: Microsatellite loci of Sockeye Salmon in Auke Creek, AK with thermocycling amplification conditions for non-multiplexed reactions; annealing temperature (T_a), number of cycles, concentration for $MgCl_2$ and dNTPs.

Locus	T_a	Cycle #	$MgCl_2$ (µM)	dNTPs (µM)	Additive (µM)	Citation
Oki 10	55	35	1.38	0.50		Smith et al. 1998
Oki 100	55	35	1.25	0.70		Miller et al. unpub
Oki 16	46	37	0.50	0.85	BSA – 0.1	Smith et al. 1998
Oki 1a	58	32	1.80	0.72		Smith et al. 1998
Oki 1b	58	32	1.80	0.72		Smith et al. 1998
Oki 29	50	30	1.25	0.50	BSA – 0.1	Smith et al. GenBank #AF055453.1
Omy 77	52	32	1.30	0.70	DMSO – 0.0025	Morris et al. 1996
One 102	55	35	1.25	0.70		Olsen et al. 2000
One 109	56	27	1.15	0.70		Olsen et al. 2000
One 114	57	30	1.00	0.70		Olsen et al. 2000
One 8	60	30	1.38	0.70		Scribner et al. 1996
Ots 100	57	35	1.38	0.70		Nelson and Beacham 1999
Ots 103	46	40	1.80	0.72		Beacham et al. 1998
Ssa 419	57	35	1.00	0.70		Cairney et al. 2000

Table S3.2: Summary statistics for the 105 loci used in genetic marker panels for Sockeye Salmon from Auke Creek, AK. The first twelve (*Ssa419 – Omy77*) are microsatellite loci: k is the number of alleles observed, H_{Obs} is the heterozygosity observed, H_{Exp} is the heterozygosity expected, PIC is the Polymorphic information content, E-1P is the average exclusion probability for one candidate parent, E-2P is the average exclusion probability for one candidate parent given the genotype of a known parent of the opposite sex, E-PP is the average exclusion probability for a candidate parent pair, E-I is the average exclusion probability for identity of two unrelated individuals, and E-S is the average exclusion probability for identity of two siblings. Loci are listed in order of inclusion for each of the 15 marker panels evaluated (i.e., if a panel had 5 STRs and 0 SNPs it would include loci Ssa419 – Oki100).

Table S3.2 Continued

	Locus	k	H_{Obs}	H_{Exp}	PIC	E-1P	E-2P	E-PP	E-I	E-S
Microsatellite Loci										
1	Ssa419	36	0.818	0.894	0.885	0.65	0.79	0.93	0.98	0.69
2	One8	14	0.664	0.659	0.614	0.26	0.43	0.62	0.84	0.54
3	Oki10	23	0.834	0.881	0.870	0.62	0.76	0.92	0.97	0.68
4	One114	16	0.842	0.859	0.844	0.56	0.72	0.88	0.96	0.67
5	Oki100	18	0.688	0.722	0.685	0.33	0.51	0.70	0.89	0.58
6	One102	12	0.649	0.681	0.632	0.28	0.44	0.63	0.85	0.55
7	One109	13	0.600	0.598	0.537	0.19	0.34	0.52	0.78	0.49
8	Oki1a	3	0.439	0.467	0.360	0.11	0.18	0.27	0.61	0.39
9	Oki1b	6	0.655	0.657	0.584	0.22	0.37	0.52	0.81	0.53
10	Oki16	7	0.636	0.676	0.605	0.23	0.39	0.54	0.82	0.54
11	Ots103	7	0.532	0.539	0.496	0.16	0.31	0.48	0.74	0.45
12	Omy77	8	0.603	0.615	0.557	0.20	0.36	0.53	0.79	0.51
Single Nucleotide Polymorphism Loci										
13	One_MHC2_251	2	0.473	0.500	0.375	0.12	0.19	0.28	0.62	0.41
14	One_U1012.68	2	0.458	0.500	0.375	0.12	0.19	0.28	0.62	0.41
15	One_ODC1.196	2	0.498	0.499	0.375	0.12	0.19	0.28	0.62	0.41
16	One_lpp1.44	2	0.268	0.498	0.374	0.12	0.19	0.28	0.62	0.41
17	One_gdh.212	2	0.495	0.497	0.374	0.12	0.19	0.28	0.62	0.40
18	One_MHC2_190	2	0.459	0.497	0.373	0.12	0.19	0.28	0.62	0.40
19	One_cetn1.167	2	0.476	0.493	0.371	0.12	0.19	0.28	0.62	0.40
20	One_U1216.230	2	0.475	0.488	0.369	0.12	0.18	0.28	0.62	0.40
21	One_GHII.2461	2	0.447	0.486	0.368	0.12	0.18	0.28	0.62	0.40
22	One_Ots208.234	2	0.464	0.483	0.367	0.12	0.18	0.28	0.62	0.40
23	One_agt.132	2	0.479	0.482	0.366	0.12	0.18	0.28	0.62	0.40
24	One_U1024.197	2	0.458	0.476	0.363	0.11	0.18	0.27	0.61	0.39
25	One_HpaI.99	2	0.449	0.468	0.358	0.11	0.18	0.27	0.61	0.39
26	One_hcs71.220	2	0.467	0.447	0.347	0.10	0.17	0.26	0.59	0.37
27	One_U1004.183	2	0.417	0.442	0.344	0.10	0.17	0.26	0.59	0.37
28	One_Prl2	2	0.405	0.419	0.331	0.09	0.17	0.26	0.57	0.35
29	One_GPH.414	2	0.411	0.416	0.329	0.09	0.17	0.25	0.57	0.35
30	One_U301.92	2	0.377	0.400	0.320	0.08	0.16	0.25	0.56	0.34
31	One_ACBP.79	2	0.371	0.397	0.318	0.08	0.16	0.25	0.56	0.34
32	One_Hsp47	2	0.356	0.390	0.314	0.08	0.16	0.24	0.55	0.33
33	One_STR07	2	0.354	0.374	0.304	0.07	0.15	0.24	0.54	0.32
34	One_Mkpro.129	2	0.355	0.363	0.297	0.07	0.15	0.23	0.53	0.31
35	One_U401.224	2	0.346	0.359	0.295	0.06	0.15	0.23	0.53	0.31
36	One_HGFA	2	0.336	0.354	0.291	0.06	0.15	0.23	0.52	0.31
37	One_U1212.106	2	0.335	0.352	0.290	0.06	0.15	0.23	0.52	0.31
38	One_SUMO1.6	2	0.303	0.327	0.274	0.05	0.14	0.22	0.49	0.29
39	One_Tf_ex10.750	2	0.316	0.326	0.273	0.05	0.14	0.22	0.49	0.29
40	One_STC.410	2	0.280	0.294	0.251	0.04	0.12	0.20	0.46	0.26
41	One_U1016.115	2	0.271	0.291	0.249	0.04	0.12	0.20	0.45	0.26
42	One_apoe.83	2	0.274	0.283	0.243	0.04	0.12	0.20	0.45	0.25
43	One_cin.177	2	0.275	0.281	0.241	0.04	0.12	0.20	0.44	0.25
44	One_E2	2	0.262	0.273	0.236	0.04	0.12	0.19	0.44	0.24
45	One_U1201.492	2	0.257	0.271	0.234	0.04	0.12	0.19	0.43	0.24
46	One_KCT1.453	2	0.238	0.261	0.226	0.03	0.11	0.19	0.42	0.23
47	One_U1210.173	2	0.244	0.261	0.226	0.03	0.11	0.19	0.42	0.23
48	One_U504.141	2	0.239	0.247	0.216	0.03	0.11	0.18	0.40	0.22
49	One_Ots213.181	2	0.242	0.234	0.207	0.03	0.10	0.18	0.39	0.21
50	One_PIP	2	0.218	0.232	0.205	0.03	0.10	0.17	0.38	0.21
51	One_U1202.1052	2	0.221	0.228	0.202	0.03	0.10	0.17	0.38	0.21
52	One_U1208.67	2	0.201	0.221	0.197	0.02	0.10	0.17	0.37	0.20
53	One_IL8r.362	2	0.216	0.216	0.192	0.02	0.10	0.17	0.36	0.20
54	One_U1206.108	2	0.186	0.198	0.178	0.02	0.09	0.16	0.34	0.18
55	One_LEI.87	2	0.184	0.196	0.177	0.02	0.09	0.15	0.33	0.18

Table S3.2 Continued

	Locus	k	H_{Obs}	H_{Exp}	PIC	E-1P	E-2P	E-PP	E-I	E-S
56	One_KPNA.422	2	0.196	0.192	0.174	0.02	0.09	0.15	0.33	0.18
57	One_aldB.152	2	0.180	0.185	0.168	0.02	0.08	0.15	0.32	0.17
58	One_ssrd.135	2	0.174	0.181	0.164	0.02	0.08	0.14	0.31	0.17
59	One_GPDH	2	0.149	0.159	0.146	0.01	0.07	0.13	0.28	0.15
60	One_spf30.207	2	0.144	0.156	0.144	0.01	0.07	0.13	0.28	0.15
61	One_redd1.414	2	0.144	0.152	0.141	0.01	0.07	0.13	0.27	0.14
62	One_U1214.107	2	0.141	0.143	0.133	0.01	0.07	0.12	0.26	0.14
63	One_U1209.111	2	0.130	0.142	0.132	0.01	0.07	0.12	0.25	0.13
64	One_U1204.53	2	0.118	0.128	0.119	0.01	0.06	0.11	0.23	0.12
65	One_sys1.230	2	0.121	0.125	0.117	0.01	0.06	0.11	0.23	0.12
66	One_pax7.248	2	0.095	0.106	0.100	0.01	0.05	0.09	0.19	0.10
67	One_HpaI.436	2	0.093	0.106	0.100	0.01	0.05	0.09	0.19	0.10
68	One_CD9.269	2	0.100	0.101	0.096	0.01	0.05	0.09	0.19	0.10
69	One_srp09.127	2	0.091	0.090	0.086	0.00	0.04	0.08	0.17	0.09
70	One_VIM.569	2	0.082	0.088	0.084	0.00	0.04	0.08	0.16	0.08
71	One_c3.98	2	0.087	0.083	0.080	0.00	0.04	0.07	0.16	0.08
72	One_RAG1.103	2	0.080	0.083	0.079	0.00	0.04	0.07	0.16	0.08
73	One_ZNF.61	2	0.078	0.081	0.078	0.00	0.04	0.07	0.15	0.08
74	One_GPDH2	2	0.043	0.061	0.059	0.00	0.03	0.06	0.12	0.06
75	One_p53.576	2	0.066	0.065	0.063	0.00	0.03	0.06	0.12	0.06
76	One_RAG3.93	2	0.064	0.064	0.061	0.00	0.03	0.06	0.12	0.06
77	One_U1009.91	2	0.040	0.050	0.049	0.00	0.02	0.05	0.10	0.05
78	One_U1203.175	2	0.043	0.048	0.047	0.00	0.02	0.05	0.09	0.05
79	One_U404.229	2	0.040	0.045	0.044	0.00	0.02	0.04	0.09	0.04
80	One_U1013.108	2	0.043	0.043	0.042	0.00	0.02	0.04	0.08	0.04
81	One_vatf.214	2	0.041	0.040	0.039	0.00	0.02	0.04	0.08	0.04
82	One_Tf_ex3.182	2	0.029	0.035	0.034	0.00	0.02	0.03	0.07	0.04
83	One_metA.253	2	0.036	0.039	0.038	0.00	0.02	0.04	0.07	0.04
84	One_vamp5.255	2	0.034	0.035	0.034	0.00	0.02	0.03	0.07	0.04
85	One_RF.112	2	0.021	0.028	0.028	0.00	0.01	0.03	0.06	0.03
86	One_rpo2j.261	2	0.030	0.031	0.031	0.00	0.02	0.03	0.06	0.03
87	One_U1101	2	0.030	0.031	0.031	0.00	0.02	0.03	0.06	0.03
88	One_U1003.75	2	0.026	0.030	0.030	0.00	0.02	0.03	0.06	0.03
89	One_U1105	2	0.031	0.030	0.030	0.00	0.02	0.03	0.06	0.03
90	One_CFP1	2	0.024	0.027	0.026	0.00	0.01	0.03	0.05	0.03
91	One_U1010.81	2	0.019	0.023	0.023	0.00	0.01	0.02	0.05	0.02
92	One_U503.170	2	0.015	0.015	0.015	0.00	0.01	0.02	0.03	0.02
93	One_ghsR.66	2	0.010	0.013	0.013	0.00	0.01	0.01	0.03	0.01
94	One_sast.211	2	0.013	0.013	0.012	0.00	0.01	0.01	0.03	0.01
95	One_tshB.92	2	0.007	0.007	0.007	0.00	0.00	0.01	0.01	0.01
96	One_txnip.401	2	0.009	0.009	0.009	0.00	0.00	0.01	0.02	0.01
97	One_zP3b	2	0.006	0.008	0.008	0.00	0.00	0.01	0.02	0.01
98	One_ctgf.301	2	0.002	0.002	0.002	0.00	0.00	0.00	0.01	0.00
99	One_U502.167	2	0.006	0.006	0.006	0.00	0.00	0.01	0.01	0.01
100	One_taf12.248	2	0.001	0.001	0.001	0.00	0.00	0.00	0.00	0.00
101	One_U1205.57	2	0.005	0.005	0.005	0.00	0.00	0.01	0.01	0.01
102	One_rab1a.76	2	0.002	0.002	0.002	0.00	0.00	0.00	0.01	0.00
103	One_RF.295	2	0.002	0.002	0.002	0.00	0.00	0.00	0.00	0.00
104	One_U1014.74	2	0.002	0.002	0.002	0.00	0.00	0.00	0.00	0.00
105	One_U1103	2	0.002	0.002	0.002	0.00	0.00	0.00	0.01	0.00
	Mean STR	13.6	0.663	0.687	0.639	0.32	0.47	0.63	0.84	0.55
	Mean SNPs	2	0.191	0.203	0.167	0.03	0.08	0.14	0.30	0.18
	Mean All	3.3	0.245	0.258	0.221	0.07	0.13	0.19	0.36	0.22

Supplemental Figures

Figure S3.1: Boxplot of assignment accuracy of Colony analysis of Sockeye Salmon from Auke Creek, AK from simulations with a monogamous mating structure. Each box extends from the first quartile to the third quartile with a horizontal line through the box at the median. The whiskers extend from each quartile to the minimum and maximum. Parentage inference with the program Colony was substantially better when simulations had a monogamous mating structure. Only a single simulation resulted in fewer than 95% of the parent-offspring dyads inferred correctly. The addition of single nucleotide polymorphism (SNP) loci with a minor allele frequency less than 0.1 (including 93 instead of 48 SNPs) resulted in decreased accuracy.

Figure S3.2: Boxplot of assignment accuracy of Colony analysis of Sockeye Salmon from Auke Creek, AK from simulations with monogamous mating structure and unrelated founding parents. Each box extends from the first quartile to the third quartile. The median marks the mid-point of the data and is shown by a horizontal line. The lower and upper whiskers extend from each quartile to the minimum and maximum respectively. Removal of related individuals among the putative parents, in addition to the monogamous mating structure, resulted in a large increase in correctly inferred parentage.

Figure S3.3: Boxplot of assignment accuracy of Colony analysis of Sockeye Salmon from Auke Creek, AK from simulations with age structure (Figure 1B). Each box encompasses the first to the third quartile with a horizontal line through the median. The lower and upper whiskers extend from the quartiles to the minimum and maximum respectively. Fewer than 95% of all parent-offspring relationships are inferred correctly for all marker panels. The addition of single nucleotide polymorphism (SNP) loci with a minor allele frequency less than 0.2 (including 93 or 48 SNPs) resulted in decreased accuracy.

Figure S3.4: Boxplot of assignment accuracy of FRANz for simulations of Sockeye Salmon from Auke Creek, AK based on a hatchery experiment with 1% missing data and 1% genotyping error. The boxes extend from the first quartile to the third quartile with the median marked by a horizontal line. The lower and upper whiskers extend from each quartile to the minimum and maximum respectively. Inference was conducted with and without the inclusion of the sex of the candidate parents. The inclusion of the true sex of the parents increased accuracy by 0.15%.

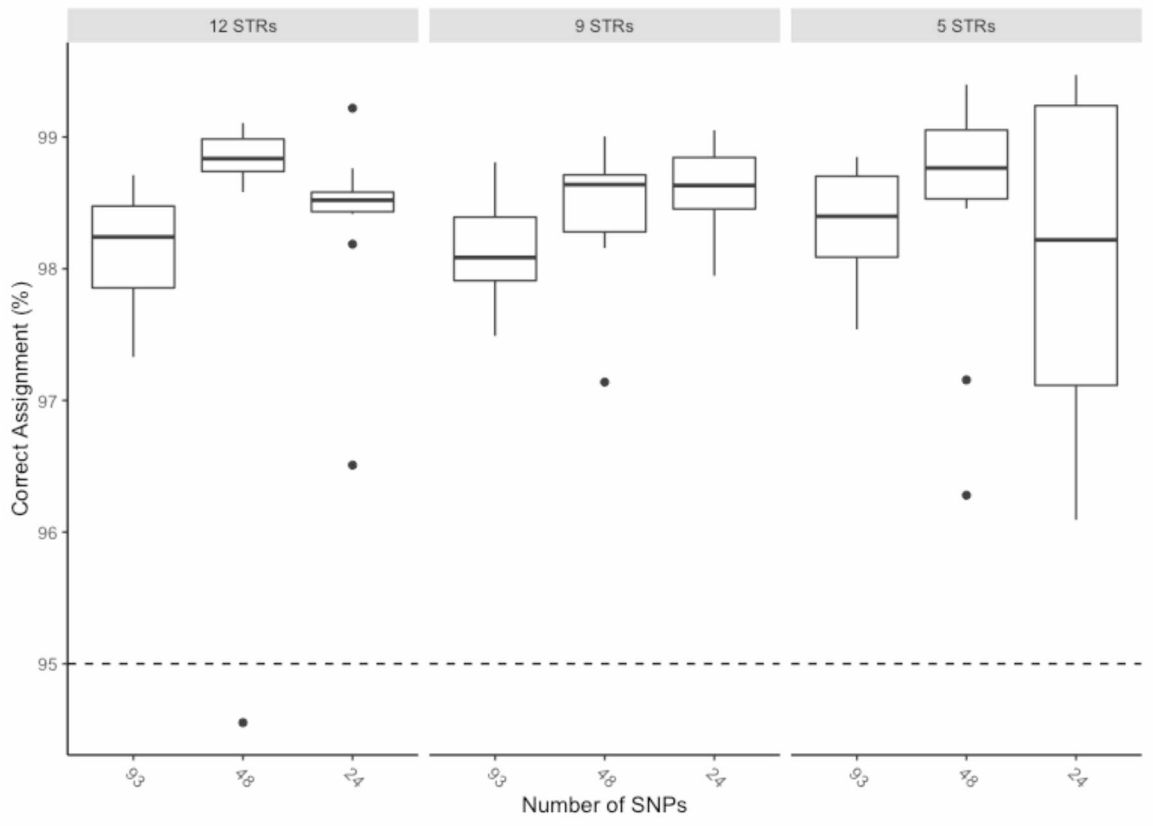

12 STRs
9 STRs
5 STRs
Correct Assignment (%)
99
98
97
96
95
93
48
24
Number of SNPs

Figure S3.1

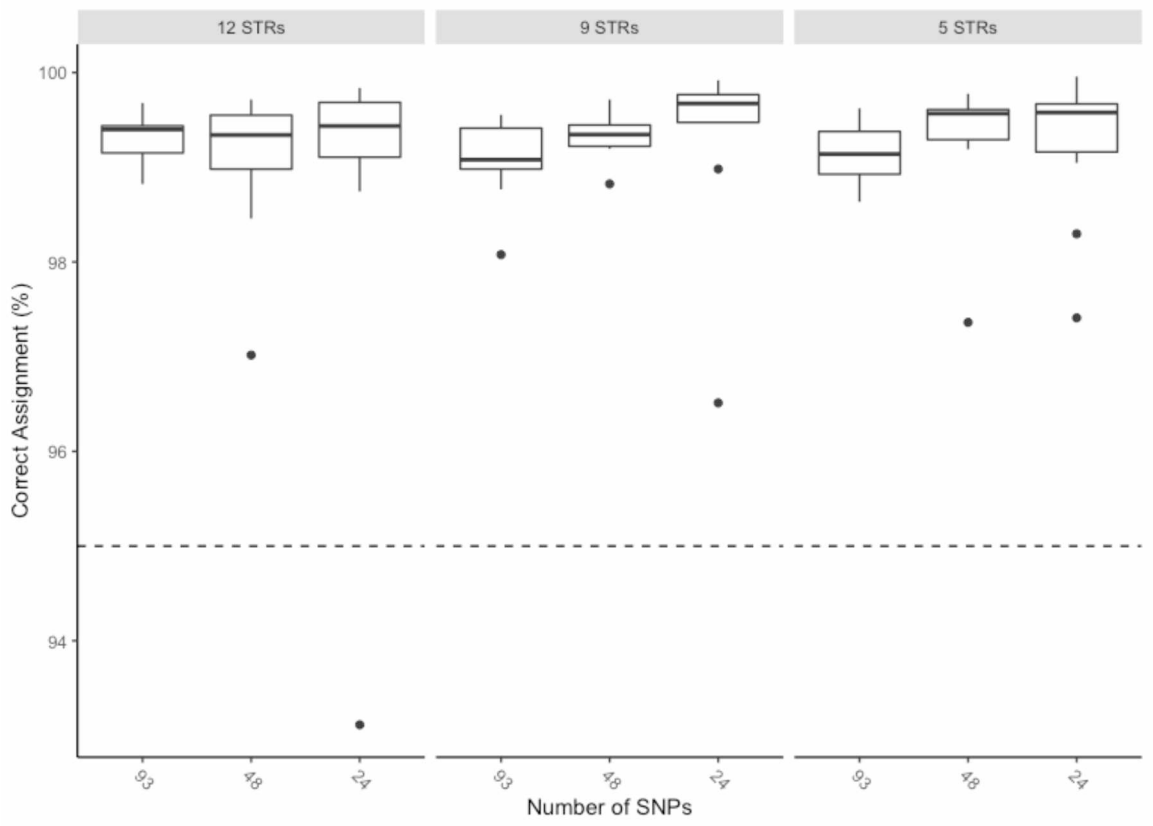

12 STRs
9 STRs
5 STRs
100
98
96
94
Correct Assignment (%)
93
48
24
Number of SNPs

Figure S3.2

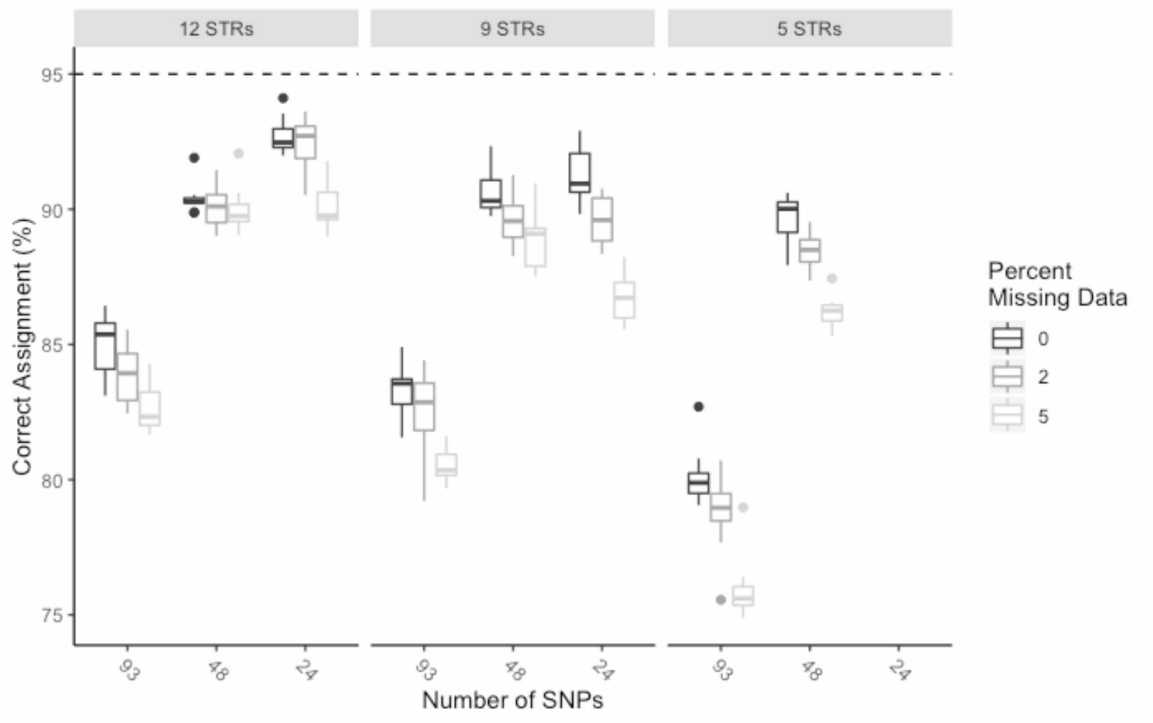

12 STRs
9 STRs
5 STRs
95
90
85
80
75
Correct Assignment (%)
93
48
24
Number of SNPs
Percent
Missing Data
0
2
5

Figure S3.3

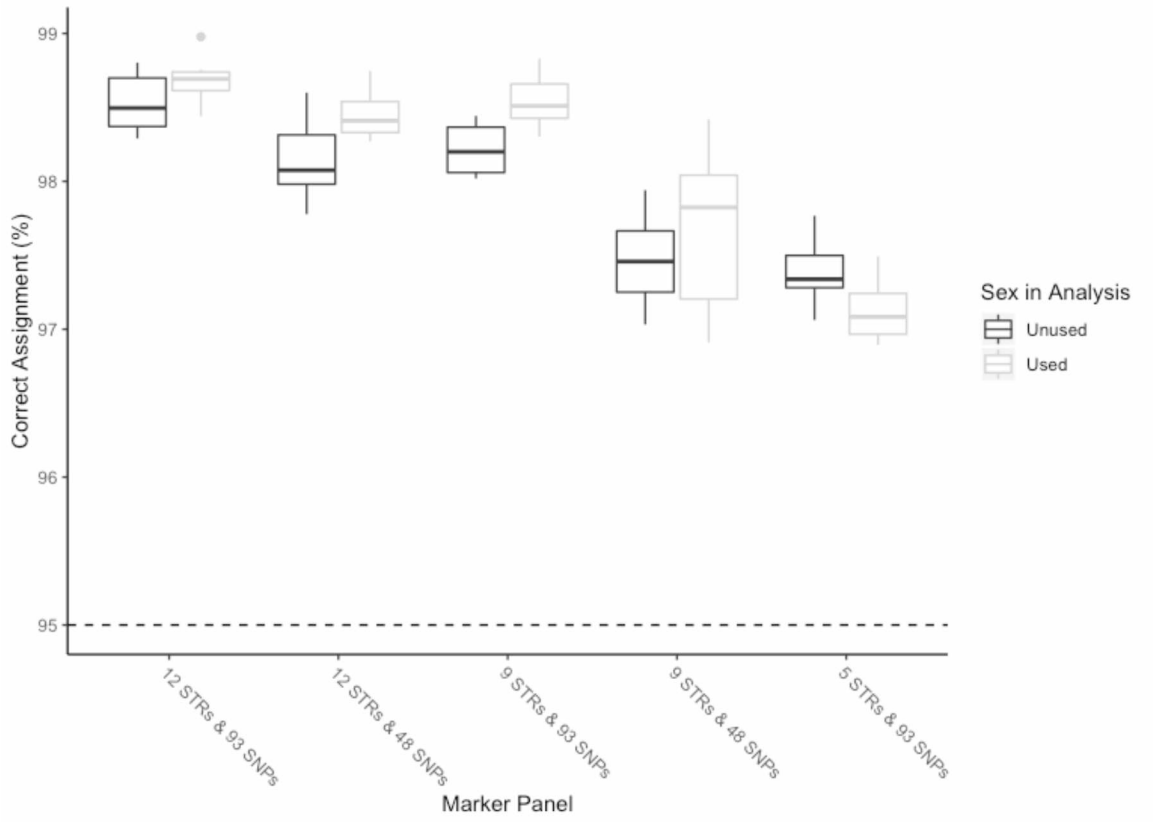

99
98
97
96
95
Correct Assignment (%)
12 STRs & 93 SNPs
12 STRs & 48 SNPs
9 STRs & 93 SNPs
9 STRs & 48 SNPs
5 STRs & 93 SNPs
Marker Panel
Sex in Analysis
Unused
Used

Figure S3.4

Conclusions

Auke Lake provided an ideal system in which to study the fine-scale genetic structure of Sockeye Salmon and compare genetic signals with demographic data. While a few large watersheds (Stikine, Taku, Chilkat and Chilkoot rivers) account for the majority of the commercial harvest in northern Southeast Alaska drift-gillnet and seine fisheries (Gilk-Baumer et al. 2015, Thynes et al. 2021a; 2021b), the collective production from the numerous small stocks is substantial. Auke Lake is one such stock and the genetic studies presented in this dissertation may be illustrative of the genetic structure and demographic factors that affect similar populations within the region. This dissertation contributes to the understanding of Sockeye Salmon population structure and evolution and will hopefully serve as a foundation for future research.

Among this study's most significant observations was the signal of isolation by time (IBT) detected within the Auke Lake Sockeye Salmon population. The strength of the relationship was variable among years, but in both individual and return-day based analyses I observed a positive slope. I detected a unique genetic cluster from the late portion of the 2008 return. Individuals from this group clustered with individuals in 2009 and 2011, suggesting persistence of the group among years. Because genetic samples were collected as fish passed the Auke Creek weir, I cannot disentangle the temporal genetic structure from potential spatial structure. On average, Sockeye Salmon spend 29 days holding in Auke Lake before ascending Lake Creek or Lake Two Creek to spawn (Ray et al. 2015). Connectivity between the lake and major spawning tributaries can be restricted, which may erode any correlation between weir passage and spawn timing (Nelson 1993, Ray et al. 2015). Spatial segregation with a temporal component within Lake Creek in which early entrants into Lake Creek ascend further up the creek to spawn has been observed (Ray et al. 2015). Additionally, as the spawning season proceeds, spawning habitat use shifts from riffle to pool (Fukushima 1996). Decreased precipitation and water flow may promote this shift, or changes in spawning habitat preference may be adaptive. The hydrology of pool habitat is regularly characterized by upwelling (Stuart 1953). Warmer incubation conditions from upwelling, relative to riffle areas, could increase metabolic rates and accelerate hatching and emergence of eggs (Fukushima 1996). If there is selection that varies over the reproductive season, adaptation by time

(ABT) can develop. The effectiveness of selection on any genetic polymorphism that conveys increased fitness will depend on both the selection coefficient (s) and the effective population size (N_e). When the product of both parameters ($N_e s$) is much smaller than 1, the force of selection will have little effect in determining the relative frequency of the polymorphism in the population (Hedrick 2005).

A small effective population size also may contribute to the large genetic differences observed between Auke Lake Sockeye Salmon and other northern Southeast Alaska populations. By using the Alaska Department of Fish and Game Sockeye Salmon genetic baseline, I identified stray fish sampled at the weir. The number and origin of these strays varied by year. One fish showed moderate support for originating from the sea/river type populations in the Taku River watershed, while the majority likely originated from Speel Lake / Snettisham Hatchery. Of the eight strays identified in 2008, five were assigned with high probability to Speel Lake / Snettisham Hatchery. The Snettisham Hatchery broodstock is made up of Speel Lake fish and so I can only speculate as to why this population disproportionately contributed to the strays observed. Hatchery fish may have a higher straying rate than wild fish, particularly when they are outplanted (i.e., transported from their rearing site for released; Quinn 1993); alternatively, straying could be a result of Sockeye Salmon being attracted to conspecifics when imprinted natal cues are weak or absent (Bett and Hinch 2015). A small number of strays into a large population are unlikely to have substantial negative effects; however, they may add to the standing genetic variation and increase a population's stability and resilience. The two outliers from 2008 that were not identified as strays but shared many low frequency alleles with the strays might indicate a low level of introgression into Auke Lake; however, without microsatellite allele frequency data from surrounding populations, I could not quantitatively test if these individuals were first generation hybrids.

Genetic tests for demographic change suggested that the population underwent a bottleneck and has since increased in size. A long-term dataset from the Auke Creek weir provided an opportunity to use demographic and genetic methods to further evaluate the signal of demographic change by estimating N_e over six generations. Genetic estimates of N_e were smaller than demographic estimates. The ratio of N_e to census size was between

0.21 and 0.37, which is consistent with other estimates from salmonid species (Frankham 1995, Allendorf et al. 1997). Demographic estimates of N_e demonstrate substantial variability among the seven generations; initially low values peaked by the fourth generation and were followed by a gradual decline. The major demographic factors that determine N_e were variance in family size, variable contribution to the next generation by brood years within a generation, and fluctuations in population size.

Freshwater productivity (adult to smolt) appeared to influence population size fluctuations more than marine survival (smolt to adult) over the 38-year timeframe. I detected a breakpoint in the freshwater productivity (smolts per spawner) in the early to mid-1990s. The substantial decrease in variance of freshwater productivity within a generation contributed to an increase in N_e. Other northern Southeast Alaska (NSEAK) systems experienced shifts in productivity around the same time. McDonald Lake experienced a severe reduction in recruitment in the mid-1990s (Walker et al. 2018). Similarly, Chilkat Lake experienced a similar decline with recruits per spawner falling below replacement levels between 1995 and 2006, although this decline has been linked to declining marine survival (Golder Associates 2013). Modeling of residuals from a Ricker stock recruit model with environmental variables (flow, precipitation, summer temperature, winter temperature, etc.), predator abundance (Dolly Varden [*Salvelinus malma*] and Cutthroat Trout [*Oncorhynchus clarkii*]), and interspecific competition (Coho Salmon [*Oncorhynchus kisutch*] smolt abundance) will be further explored with NOAA collaborators (Scott Vulstek and Joshua Russell) to identify potential drivers of freshwater productivity.

The results of this dissertation suggest a number of additional important potential avenues for future research. An additional seven years (totaling two full generations) of Auke Lake Sockeye Salmon have been sampled. Analyses conducted in chapter 3 helped refine a panel of genetic markers for parentage analyses. With accurate reconstruction of families, the genetic patterns observed in chapters 1 and 2 can be investigated further. Initial tests for adaptation by time could be done by comparing the reproductive success of matings within and between early and late returning fish. If ABT were present I would expect the reproductive success of the early and late crosses to be lower relative to matings between individuals with similar return timing. I was able to identify putative

strays with principal component analysis and outlier tests, but parentage analysis will permit the quantification of their reproductive success and measure actual introgression. Scale age data facilitated the comparison of estimates of N_e from individual brood years with estimates from return years composed of multiple cohorts. Despite the different interpretation of estimates of the effective number of breeders (N_b) from each analysis, the estimates of N_e were concordant. Chapter 3 of this dissertation provides guidance on marker selection for inferring family groups that would not only provide a direct estimation of N_b, but would also further refine our understanding of N_e in small sockeye populations.

References

Allendorf, F. W., D. W. Bayles, D. L. Bottom, K. P. Currens, C. A. Frissell, D. Hankin, J. A. Lichatowich, W. Nehlsen, and T. H. Williams. 1997. Prioritizing Pacific salmon stocks for conservation. Conservation Biology 11:140–152.

Bett, N., and S. Hinch. 2015. Attraction of migrating adult Sockeye Salmon to conspecifics in the absence of natal chemical cues. Behavioral Ecology 26(4):1180–1187.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: A review. Genetical Research 66(2):95–107.

Fukushima, M. 1996. Effects of density-dependence, environment and species interaction during spawning and incubation on population dynamics of Pink and Sockeye salmon in the Auke Lake system, southeast Alaska. Doctoral dissertation. University of Alaska Fairbanks, Alaska.

Gilk-Baumer, S., S. Rodgers Olive, D. Harris, S. Heinl, E. Fox, and W. Templin. 2015. Genetic mixed stock analysis of Sockeye Salmon harvested in selected northern Chatham Strait commercial fisheries, southeast Alaska, 2012-2014. Alaska Department of Fish and Game Fishery Data Series No. 15-03, Anchorage, Alaska.

Golder Associates. 2013. Data review of the Sockeye Salmon declines in Chilkat and Chilkoot Lakes, Southeast Alaska. Report Number: 1214920097-R-RevC. Available:https://www.hainesalaska.gov/sites/default/files/fileattachments/community_amp_economic_development/page/1630/golder_report.pdf

Hedrick, P. 2005. Genetics of populations. Jones and Bartlett Publishers, Sudbury, Massachusetts.

Nelson, B. 1993. Movements of radio-tagged Sockeye Salmon in Auke Lake, Southeast Alaska. Draft Technical Memorandum, National Marine Fisheries Service Auke Bay Laboratory.

Quinn, T. P. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fisheries Research 18(1–2):29–44.

Ray, J., S. Sethi, J. Eiler, and J. Joyce. 2015. Prespawning movements and spawning distribution of Sockeye Salmon in an urbanizing Alaskan lake. Journal of Fisheries and Wildlife Management 6(2):472–485.

Stuart, T. A. 1953. Water currents through permeable gravels and their significance to spawning salmonids, etc. Nature 172:407–408.

Thynes T., A. Dupuis, D. Harris, B. Meredith, A. Piston, and P. Salomone. 2021a. 2021 Southeast Alaska purse seine fishery management plan. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 1J21-09, Douglas, Alaska.

Thynes, T., N. Zeiser, S. Forbes, T. Kowalske, B. Meredith, and A. Dupuis. 2021b. 2021 Southeast Alaska drift gillnet fishery management plan. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 1J21-07, Douglas, Alaska.

Walker, S., T. Thynes, D. Gray, K. S. Reppert, A. W. Piston, and S. C. Heinl. 2018. McDonald Lake Sockeye Salmon stock status and action plan 2018. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 1J18-03, Douglas, Alaska.

Appendix

(907) 474-7800
(907) 474-5993 fax
uaf-iacuc@alaska.edu
www.uaf.edu/iacuc

Institutional Animal Care and Use Committee
909 N Koyukuk Dr. Suite 212, P.O. Box 757270, Fairbanks, Alaska 99775-7270

May 28, 2019

To:	Megan McPhee, PhD Principal Investigator
From:	University of Alaska Fairbanks IACUC
Re:	[452507-11] Fitness consequences of hatchery supplementation on Auke Creek sockeye salmon

The IACUC reviewed and approved the Closure/Final Report referenced above by Designated Member Review.

Received:	May 22, 2019
Approval Date:	May 28, 2019
Initial Approval Date:	May 29, 2013
Expiration Date:	

This action is included on the June 6, 2019 IACUC Agenda.

PI responsibilities:

- *Acquire and maintain all necessary permits and permissions prior to beginning work on this protocol. Failure to obtain or maintain valid permits is considered a violation of an IACUC protocol and could result in revocation of IACUC approval.*
- *Ensure the protocol is up-to-date and submit modifications to the IACUC when necessary (see form 006 "Significant changes requiring IACUC review" in the IRBNet Forms and Templates)*
- *Inform research personnel that only activities described in the approved IACUC protocol can be performed. Ensure personnel have been appropriately trained to perform their duties.*
- *Be aware of status of other packages in IRBNet; this approval only applies to this package and the documents it contains; it does not imply approval for other revisions or renewals you may have submitted to the IACUC previously.*
- *Ensure animal research personnel are aware of the reporting procedures on the following page.*