Genetic differentiation of chum salmon (*Oncorhynchus keta*) in Alaska influenced by glacial history

Eleni L. Petrou

A thesis
submitted in partial fulfillment of the
requirements for the degree of
Master of Science
University of Washington
2012

Committee members:
Lisa Seeb
Jim Seeb
Lorenz Hauser
Robin Waples

Program Authorized to Offer Degree:
School of Aquatic and Fishery Sciences

Table of Contents

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

Page

LIST OF TABLES

Page

ACKNOWLEDGEMENTS

This research would not have been possible without the help and encouragement of many people. The author thanks her advisor, Dr. Lisa Seeb, and committee members, Drs. Jim Seeb, Lorenz Hauser, and Robin Waples for providing valuable guidance and sharing ideas. Colleagues from the Alaska Department of Fish and Game (Bill Templin, Mark Witteveen, Nick Decovich, Lisa Fox, Penny Crane, and Judy Berger) collected samples from the field, organized data sharing, and assisted with DNA extractions. Carita Pascal and Heather Peterson gave indispensable laboratory support. The author thanks past and present Seeb lab members, especially Dr. Daniel Gomez-Uchida. This research was funded by the Alaska Sustainable Salmon Fund under Study 45919 from the National Oceanic and Atmospheric Administration, US Department of Commerce, administered by the Alaska Department of Fish and Game. Infrastructure was provided by a grant from the Gordon and Betty Moore Foundation. A National Science Foundation GK-12 graduate teaching fellowship provided the author with funding during the last year of research. Finally, many thanks are extended to family and friends, whose company made life sweet.

DEDICATION

This thesis is dedicated to Drs. Ethel and Athos Petrou: mad scientists and beloved parents.

CHAPTER 1: Genetic differentiation of chum salmon influenced by isolation by distance and glacial history of Alaska

INTRODUCTION

One major goal of fisheries management is to ensure the sustainable harvest of wild populations so that productivity is stable over time. Information on population connectivity and individual dispersal is necessary for effective fishery management because these parameters describe the appropriate spatial scale for management units (Palsbøll et al. 2007), and identify populations that are small or isolated and need protection from overharvest (Allendorf et al. 2008). Management units have been defined using various criteria, including information on morphological and/or demographic characters, dispersal patterns of individuals, fishing patterns, and information on the genetic structure of populations (Cadrin et al. 2005). Genetic studies offer valuable insights into population connectivity because they integrate data over many generations and estimate long-term estimates of connectivity (Palumbi 2003).

Multiple factors influence the genetic structure of populations, including mutation, migration, genetic drift, and natural selection. In some aquatic and marine species, individuals occupy a one-dimensional linear structure, such as a river or coastline. In this case, genetic structure can be modeled by a one-dimensional stepping stone model (Kimura and Weiss 1964) where dispersal and reproduction occur more frequently between adjacent populations. If dispersal and reproduction are spatially limited in this way, genetic differentiation is positively correlated with geographic distance separating populations. This pattern is known as isolation by distance (IBD) (Wright 1943).

Theoretical models describing IBD are well established (Wright 1943; Slatkin 1993; Rousset 1997), and predict that in the absence of spatially heterogeneous selection, populations eventually reach equilibrium between gene flow and genetic drift. Gene flow is usually described as the number of reproductively successful migrants exchanged between populations at each generation. This value is the product of the effective population size (N_e) and the proportion of each population that are reproducing migrants (m). IBD has been observed in marine and aquatic populations (Koizumi et al. 2006; Olsen et al. 2008; Cunningham et al. 2009), but a strong IBD relationship does not necessarily imply migration-drift equilibrium between populations (Bradbury and Bentzen 2007).

In linear systems, theory predicts that populations at migration-drift equilibrium will display a monotonic increase between geographic distance and genetic differentiation at all spatial scales (Hutchison and Templeton 1999). However, previous research on aquatic and marine species (Bradbury and Bentzen 2007) has shown that IBD relationships are often non-linear depending on the species and spatial scale examined. Unequal gene flow between populations, natural selection, and historical events (e.g. bottlenecks, range expansions) are examples of processes that may contribute to departures from IBD models and/or migration-drift equilibrium. Understanding how these different processes contribute to the population structure of a species can be useful when interpreting estimates of population connectivity.

Patterns of genetic connectivity have implications for the management or conservation of populations. Taylor and Dizon (1999) have suggested that an important step towards effective management is identifying populations most impacted by anthropogenic mortality and evaluating whether enough connectivity exists to sustain those populations over time. This information is important for the management of fisheries, which often target a mixture of individuals or species originating from different populations or geographical regions. When directed harvest or bycatch occurs in mixtures with unknown population composition, it can lead to inadvertent overexploitation of less abundant populations (Utter and Ryman 1993). Genetic data have been used to estimate mixture composition through mixed stock analysis, a procedure in which individuals originating from different regions are assigned to their population of origin by comparing them to a genetic baseline of known population allele frequencies (Pella and Milner 1987). Mixed stock analysis has also been used to identify stocks most susceptible to bycatch mortality (Laurent et al. 1998; Seeb et al. 2011; Templin et al. 2011), and appropriate management actions can be taken to minimize negative impacts of a fishery on a non-target species.

Chum salmon (*Oncorhynchus keta*) is a species targeted by both commercial and subsistence fisheries. The species is also susceptible to unintended interception as bycatch by the walleye pollock (*Theragra chalcogramma*) fishery (Haflinger et al. 2009). Chum salmon is one of the most abundant Pacific salmonids and is widely distributed across the Northern Pacific Rim. Spawning populations can be found in Korea, Japan and Russia in the western Pacific Ocean, and from Alaska to Oregon in the eastern Pacific (Salo 1991). Chum salmon are anadromous; they hatch in freshwater, travel to open ocean to feed and mature, and return to freshwater to spawn and die. In addition, chum salmon have been observed to home to their natal site for spawning (Tallman and Healey 1994). Fisheries harvest chum salmon nearshore or in rivers as mixed aggregations of individuals. In addition to increasing commercial demand for chum salmon, the species has cultural and subsistence importance to Native Alaskan communities (Wolfe 2004). For example, in 2007 nearly 300,000 chum were harvested in Alaska by the subsistence fishery (Fall et al. 2007). Participating in this fishery is an important economic and social activity for Native Alaskans in rural areas, and people cooperatively harvest fish for use in winter.

Estimating fine-scale population structure of chum salmon in Alaska is of increasing concern because of fluctuating census sizes, and the uncertain effects of bycatch on specific populations (Gisclair 2009). In this study we examined the genetic structure of 66 chum salmon populations spawning in waterways on the Alaska Peninsula and Kodiak Island. The purpose was to provide resource managers with a high-resolution genetic baseline for use in mixed stock analyses. Our study area spanned over 1,300 coastal km, and covered diverse spawning habitats, including braided river systems on the north Alaska Peninsula, and glacially fed streams on the south Alaska Peninsula and Kodiak Island.

We explored regional patterns of population structure and IBD using 93 nuclear and mitochondrial single nucleotide polymorphism (SNP) loci. We hypothesized that IBD will be present because chum salmon spawn in coastal rivers that are distributed along a linear array, and gene flow is spatially restricted due to homing. Evidence of population structure and IBD was found throughout the region. In addition, analyses identified outlier populations to regional IBD patterns, suggesting that populations of chum salmon in southwestern Alaska are influenced by diverse evolutionary and/or demographic histories.

MATERIALS AND METHODS

Samples were collected from 1992 to 2010 by the Alaska Department of Fish and Game. Approximately 95 individuals were sampled at 66 sites on the Alaska Peninsula and Kodiak

Island that likely represent distinct breeding populations (Figure 1.1). Temporal replicates of each sampling location were collected when possible, resulting in 105 total collections (Table 1.1). From 1992 to 2002, salmon were captured by beach seine, and heart or liver tissue was collected and frozen in liquid nitrogen until transferred to long-term storage at -80°C. From 2008 onward, individual salmon were captured by beach seine, held momentarily while the axillary process on the ventral fin was removed, and subsequently released. Axillary processes were stored in 95% ethanol.

DNA was extracted from 8,140 individuals using the DNeasy 96 Tissue Kit (QIAGEN, Valencia, CA). SNPs at 96 loci (Supplementary Table 1.1) were assayed using 5'-nuclease reactions (Seeb et al. 2009) on the BioMark 96.96 Dynamic Array (Fluidigm, San Fransisco, CA). As a quality control measure, approximately eight individuals from each population were re-extracted and genotyped at all loci to ensure that genotypes were accurate and reproducible.

Our goal was to obtain a sample size of at least 95 individuals from each spawning location. When study sites included multiple samples taken from different years, we tested for homogeneous allele frequencies across different years using a chi squared test in the program CHIFISH (Ryman 2006) . If no statistically significant difference ($\alpha = 0.01$) over all loci was found between collection years, individuals sampled from the same geographic location were pooled (Table 1.1) to achieve a larger sampling size as recommended by Waples (1990).

Allele frequencies and heterozygosity were calculated in GENALEX 6.4 (Peakall and Smouse 2006). Deviations from Hardy-Weinberg equilibrium (HWE) at each population were detected using a χ^2 test ($\alpha = 0.05$) in GENALEX 6.4. Patterns of linkage equilibria were examined for all locus pairs using exact tests in GENEPOP 4.0. If a pair of loci was in linkage disequilibrium in more than 75% of the populations, only one SNP locus was retained for the remainder of the analyses to avoid redundant information in the data set. We compared patterns of linkage disequilibrium in each population by calculating the statistic r^2 that describes the overall correlation coefficient averaged across multiple loci and alleles. The calculations were done with the program LDNE (Waples and Do 2008)and r^2 was subsequently used to estimate the effective population size (N_e) of each sample collection.

Pairwise genetic differentiation (F_{ST}) between all populations (Weir and Cockerham 1984) was estimated in GENEPOP 4.0. Patterns of population differentiation were further explored with correspondence analysis based on genotypes using the ADEGENET package (Jombart 2008) in R (R Development Core Team 2011). Statistical significance of population differentiation was evaluated using 1,000 permutations in ARLEQUIN 3.5 (Excoffier and Lischer 2010). We tested for IBD to investigate the relationship between genetic differentiation and geographic distance. The shortest waterway distance between sampling locations was estimated using a least cost path analysis in ARCGIS ver. 10 (ESRI, Inc). Linear F_{ST} estimates ($F_{ST}/(1-F_{ST})$) were regressed to waterway distance (Rousset 1997), and statistical significance was evaluated based on least squares regression in R, as in Koizumi et al. (2006). Hellberg (1994) has shown that for pairwise data, p values are almost identical between Mantel tests and ordinary least squares regression with degrees of freedom designated by the number of populations. This procedure was conducted first with all populations considered together. Subsequently, each geographic region (north Alaska Peninsula, south Alaska Peninsula, and Kodiak Island) was analyzed separately.

Putative outlier populations within each geographic region (north Alaska Peninsula, south Alaska Peninsula, and Kodiak Island) were identified by examining mean regression residuals for each population. If a population is divergent from regional IBD patterns, then its regression residuals are biased towards values more different from zero than populations following regional IBD. Following the decomposed pairwise regression analysis (DPRA) method of Koizumi *et al.* (2006), we sequentially removed populations from the analysis when 95% confidence intervals of the mean residual did not include zero. The fit of IBD models considering different putative outlier populations was evaluated using Akaike's information criteria (*AIC*, Burnham and Anderson 2002). *AIC* was calculated as $AIC = 2K + n\ln(RSS/n)$, where K is the number of parameters (in this case the only parameter is waterway distance), n is the number of populations, and *RSS* is the residual sum of squares (Koizumi et al. 2006). Corrected *AIC* (AIC_C) was used because of small sample size ($n/K < 40$). Models were considered equally likely if the difference in AIC_C values (ΔAIC) was less than two.

We examined the impact of scale on linearity of the IBD slope by partitioning the data into increasing bin sizes (0-50, 0-100, 0-150 km, etc) and doing iterative regressions of genetic to waterway distance (following Bradbury and Bentzen 2007) for each geographic region. IBD slope was subsequently examined at incremental spatial scales for deviations from linearity. This was done to test for deviations from migration-drift equilibrium, since populations in a one-dimensional stepping stone are expected to have monotonic and positive IBD slope at all spatial scales when in migration-drift equilibrium.

Finally, the effect of N_e on within-region genetic differentiation was investigated. We expected that populations with small N_e might show relatively high levels of differentiation due to genetic drift. The correlation between genetic differentiation and effective population size was evaluated by regressing $1/N_e$ against mean population F_{ST} for all within-region comparisons (north Alaska Peninsula, south Alaska Peninsula, and Kodiak Island), and statistical significance was evaluated using least squares regression in R.

RESULTS

Quality control analysis showed that 0.05% of the genotypes contained discrepancies between the original genotypes and the re-extracted samples, indicating that our genotypes were reproducible and accurate. A small number of individuals were missing more than 15% of genotypes (suggesting poor quality DNA) and were removed from the data set, leaving 8,103 individuals for the remaining analyses.

No significant differentiation across all loci was found for samples collected from the same geographic location in different years, and temporal replicates were pooled as indicated in Table 1. Two populations, St. Catherine’s Cove ($p = 0.02$) and Sturgeon River ($p = 0.002$) exhibited significantly more deviations from HWE than would be expected by chance alone. Deviations from HWE were not localized in any one SNP, so we did not remove any loci from our analysis.

We tested for linkage disequilibrium by conducting tests between all possible pairs of loci, and found 14,246 (5%) to be statistically significant at $\alpha = 0.05$. Three locus pairs

(*Oke_gdh-162* and *Oke_gdh-191*; *Oke_pgap-111* and *Oke_pgap-92*; *Oke_U1021-102* and *Oke_1022-139*) were found to be in linkage disequilibrium in almost all of the populations. We retained only one locus from each pair for the remainder of the analyses based on highest global F_{ST} (Supplementary Table 1.1). Patterns of linkage disequilibrium within populations varied. The Karluk Lagoon population (47) exhibited many more deviations from linkage disequilibrium than expected by chance alone ($p < 0.0001$). The Karluk Lagoon population had an estimated N_e= 54, the smallest of any population.

Global F_{ST} was 0.04, and pairwise population F_{ST} values ranged from 0 to 0.19 (Supplementary Table 1.2). Correspondence analysis of genotypic data allowed us to visualize population differentiation. When SNPs from both nuclear DNA and mtDNA were considered, most of the variation in the data set was explained by the first axis and was driven by differences in mtDNA haplotypes (Figure 2A). All but three populations (18- Little John Lagoon, 19- Sandy Cove, and 21- Delta Creek) were monomorphic at the composite mtDNA SNP. These three populations originating from the south Alaska Peninsula had an alternative haplotype at frequency greater than 0.25. Interestingly, the three populations were not differentiated when only nuclear SNP loci were considered (Figure 1.2B).

Isolation by distance and decomposed pairwise regression analysis

IBD was examined for all populations, and very weak correlation between genetic and waterway distance was found (Figure 3, $r^2 = 0.06$, $p < 0.0001$). Poor fit of the IBD model to the data was caused by the presence of two populations with very high genetic differentiation: the Sturgeon River (48) and Kitoi Bay Hatchery (66) populations (Figure 1.2B). When these populations were removed, we observed higher correlation between genetic and waterway distance ($r^2 = 0.17$, $p < 0.0001$).

Decomposed pairwise regression analysis identified ten populations that were outliers from regional IBD patterns: Nelson River (10), Sandy Cove (19), Ruby’s Lagoon (24), Zachary Bay (26), Karluk Lagoon (47), Sturgeon River (48), Big Sukhoi (49), Sitkinak Island (51), Gull Cape Lagoon (57), and Kitoi Bay Hatchery (66). In all cases, fit of the IBD model to data improved when outlier populations were removed from each region, as indicated by r^2 and ΔAIC

values (Table 2). Populations partitioned by geographic region (north Alaska Peninsula, south Alaska Peninsula, and Kodiak Island) had very different patterns of regional IBD (Figure 1.4).

The north Alaska Peninsula contained highly differentiated populations that strongly conformed to IBD (slope = 1×10^{-4}, $r^2 = 0.52$, $p < 0.0001$). However, IBD slope was non-linear when examined at increasing spatial scales (Figure 1.5A). IBD slope was smallest at spatial scales smaller than 100 km. At pairwise population distances greater than 100 km, there was an increase in IBD slope and F_{ST} variance (Figure 1.5A). This indicates that populations on the North Alaska Peninsula show a strong but non-linear pattern of increasing genetic differentiation with distance.

Populations from the south Alaska Peninsula were less differentiated than populations from the north Alaska Peninsula (Figure 1.4). Decomposed pairwise regression analysis identified three outlier populations on the south Alaska Peninsula (Table 1.2). When these populations were removed from the data set, there was stronger correlation between genetic and waterway distance (slope = 8×10^{-6}, $r^2 = 0.08$, $p <0.0001$). IBD slope on the south Alaska Peninsula was also non-linear, but in contrast with the north Alaska Peninsula, largest slope was observed at small spatial scales (<100 km) (Figure 1.5B).

Kodiak Island contained genetically diverse populations, despite the small geographic extent of the region. We found evidence of two highly differentiated populations in this area, the Sturgeon River (48), and Kitoi Bay Hatchery (66) (Figure 1.2B). Mean pairwise population F_{ST} for these populations was 0.13 and 0.14, respectively (Supplementary Table 1.2). Decomposed pairwise regression analysis identified six outlier populations on Kodiak Island (Table 1.2). Correlation between genetic and waterway distance increased when outlier populations were removed from the dataset, and relatively strong IBD was observed (slope = 4×10^{-5}, $r^2= 0.48$, $p < 0.0001$). IBD slope was fairly consistent over increasing spatial scales (Figure 1.5B).

Effective population sizes (N_e) ranged from 54 to 17,789 individuals, and the median value of N_e over all populations was 609 (Table 1.1). We tested the hypothesis that populations with small N_e are outliers from regional IBD patterns by regressing the inverse of effective population size ($1/N_e$) to mean population F_{ST} in each region. On the north Alaska Peninsula there was no correlation between these two quantities ($r^2= 6*10^{-5}$, $p = 0.98$), and the same thing was observed on the south Alaska Peninsula (r^2= 0.05, p = 0.24). Populations from Kodiak Island showed no correlation between N_e and average F_{ST} (r^2= 0.001, p = 0.83). When the highly differentiated Sturgeon River (48) and Kitoi Bay Hatchery (66) populations were removed from the analysis, a statistically significant correlation emerged (r^2 = 0.35, p = 0.01). However, this relationship was dominated by the presence of the Karluk Lagoon (47) population that had a very small estimated N_e= 54 (Table 1.1).

DISCUSSION

The purpose of this study was to investigate population structure and patterns of genetic connectivity in chum salmon for applications in fisheries management. We confirmed that populations of chum salmon in southwestern Alaska are influenced by diverse evolutionary and/or demographic histories. In addition, we found evidence of IBD at different spatial and regional scales.

IBD was initially explored using all populations, and very weak correlation between genetic and waterway distance was found due to the presence of highly differentiated Sturgeon River (48) and Kitoi Bay Hatchery (66) populations (mean pairwise F_{ST} = 0.13 and 0.14, respectively).This amount of genetic differentiation at small spatial scales is extremely unusual for the species (Figure 1.2B). Previous research (Seeb and Crane 1999) using allozymes reported similar high genetic divergence between the Sturgeon River and other populations, and this pattern was attributed to the glacial history of southwestern Alaska.

During the Last Glacial Maximum, circa 25,000 years ago, the majority of the southwestern Alaska coast was covered by the Cordilleran Ice Sheet (Clark et al. 2009), making the region inhospitable to anadromous fish. However, there exists geological and palynological evidence that the southwest portion of Kodiak Island remained ice-free and contained proglacial lakes (Karlstrom and Ball 1969; Mann and Peteet 1994; Mann and Hamilton 1995). It is possible that these lakes might have been a glacial refugium, supporting populations of land-locked aquatic organisms during the last ice age. Interestingly, the Sturgeon River drainage is found in the part of southwest Kodiak Island that remained unglaciated (Figure 1.6), and the Kitoi Bay Hatchery was stocked from this population. Both our findings and previous work (Seeb and Crane 1999) suggest that the unusual and extreme genetic divergence of Sturgeon River salmon are caused by many generations of isolation, and this might be an indication that these salmon are descendants of individuals surviving the last ice age in proglacial lakes. If that is the case, then this population reverted to anadromy once the ice sheet receded approximately 14,000 years before present (Mann and Hamilton 1995).

If the Sturgeon River population has been connected by gene flow to other populations since Kodiak Island was colonized by chum salmon, why is it still possible to detect stark genetic differentiation today? One possibility is that not enough generations have passed since gene flow to adjacent populations began, and differences will continue to erode over time. A second possibility is that a barrier to gene flow exists between the Sturgeon River (and Kitoi Bay Hatchery) and other populations. Previous research on salmonids has shown that temporal differences in spawning date are correlated to genetic differentiation (reviewed in Hendry and Day 2005). Sturgeon River spawners were collected in early July, whereas the majority of other Kodiak Island samples were collected from August spawners (Table 1.1). It is possible that successful reproduction between populations is restricted by difference in spawning date. Finally, it is possible that migrants to the Sturgeon River population experience reduced fitness because of selection. Future research could use a landscape genomics approach to explore these interesting patterns of genetic differentiation (Allendorf et al. 2010) and uncover possible environmental factors influencing gene flow.

Identifying outlier populations with decomposed pairwise regression analysis

Decomposed pairwise regression analysis of genetic and waterway distances revealed ten outlier populations that did not conform to IBD patterns in each geographic subregion (north Alaska Peninsula, south Alaska Peninsula, and Kodiak Island). In all cases, fit of regional IBD models to the data improved when outlier populations were removed based on r^2 and ΔAIC scores (Table 1.2).

Theory predicts that small populations experience greater genetic drift. In a one-dimensional system, small populations might be outliers from IBD if genetic drift >> gene flow (case III in Hutchison and Templeton 1999). We tested the hypothesis that populations with small N_e have greater genetic differentiation, and only found support for this hypothesis on Kodiak Island when the highly differentiated Sturgeon River and Kitoi Bay Hatchery populations were not considered in the analysis. However, this correlation was driven by single population (Karluk Lagoon - 47) that has very small estimated N_e of 54 individuals (Table 1.1). These results indicate that differences in N_e may not be the major cause of genetic differentiation between proximate populations on the Alaska Peninsula and Kodiak Island.

Previous studies have shown increased genetic differentiation between proximate populations of sockeye salmon (*O. nerka*) utilizing different ecotypes as spawning grounds (Lin et al. 2008; McGlauflin et al. 2011). Water temperature is an important factor in egg development (Murray and McPhail 1988) that can be variable between spawning habitats. It has been shown that genetically distinct summer and fall spawning chum salmon populations in the Yukon River drainage experience different thermal regimes at spawning sites, suggesting localized adaptation to spawning environment (Zimmerman and Finn 2012) Six out of ten populations identified as outliers by decomposed pairwise regression analysis (Nelson River– 10, Ruby's Lagoon- 24, Karluk Lagoon– 47, Big Sukhoi- 49, Sitkinak Island- 51, and Gull Cape Lagoon-57) were sampled in spawning grounds near or adjacent to lagoons, that might experience warmer water temperatures. It is feasible that spawning habitat selectivity restricts gene flow between adjacent populations. At present we lack environmental data to test this hypothesis because of the remoteness of our sampling locations. Future research efforts could collect ecological data (water temperature, dissolved oxygen, sediment loads, etc) on spawning grounds and test for correlations between differences in environmental variables and genetic differentiation (Manel et al. 2003).

Linearity of IBD slope when outliers are omitted

Patterns of IBD are affected by the spatial arrangement of populations, time since a region has been colonized, and the magnitude of gene flow between populations. In a linear array of populations, if conditions following colonization allow localized gene flow between populations of limited size, then patterns of IBD will begin to form between adjacent populations (case IV in Hutchison and Templeton 1999). Over time, IBD will form at larger spatial scales and populations might reach migration-drift equilibrium. Theory predicts that when linear populations are in migration-drift equilibrium, genetic distance between populations will increase monotonically over all spatial scales (Hutchison and Templeton 1999).

When IBD was explored at increasing spatial scales, different regional patterns emerged (Figures 1.4 and 1.5). On the north Alaska Peninsula populations were highly differentiated and IBD slope was smallest at small spatial scales (Figure 1.5A). This result is the opposite of what is expected in a linear environment recently colonized by a homogeneous source population (Hutchison and Templeton 1999). Instead, it suggests that populations on the North Alaska Peninsula are not in migration-drift equilibrium. Our research has shown that this lack of equilibrium might be caused by recent colonization of the Alaska Peninsula by populations from the northern and southern parts of the species range (Seeb and Crane 1999; this thesis, second chapter). Geological evidence supports that coastal spawning habitats on the north Alaska Peninsula were destroyed by sea level rise circa 2,100 years ago (Jordan 2001), and contemporary spawning sites likely became available only within the last 1,300 years (Maschner 1999). Thus it is possible that the north Alaska Peninsula was colonized relatively recently by heterogeneous founder populations.

In contrast, populations from the south Alaska Peninsula were characterized by much lower genetic differentiation and IBD slopes that were two orders of magnitude smaller than those found on the north Alaska Peninsula (Figure 1.4). On the south Alaska Peninsula, IBD slope was smallest at large spatial scales. Other studies have shown similar fading of IBD relationship at larger geographic distances (Hellberg 1995; Castric and Bernatchez 2003). In a one-dimensional array of populations, the area over which IBD is present depends on the equation $\sqrt{2Nm\tau}$, where τ = time (Slatkin 1993). In recently colonized environments, IBD relationship is expected to be weak over larger spatial scales. Thus, the observed IBD pattern on the south Alaska Peninsula is consistent with a recently colonized linear system that has been settled by a homogeneous source population (case II; Hutchison and Templeton 1999).

IBD slope on Kodiak Island was fairly consistent across spatial scales, suggesting that non-outlier populations were in migration-drift equilibrium. Pairwise waterway distances between populations on Kodiak were smaller than those on the Alaska Peninsula because of the island's geography. These patterns suggest that IBD slope is influenced by the genetic diversity of colonizing populations, time since colonization, and spatial extent of the region examined.

Application of genetic data to fisheries management

Our study was funded by the Western Alaska Salmon Stock Identification Program, a multiparty collaborative effort to sample salmon spawning sites in Alaska and better understand genetic diversity and composition of harvest and/or bycatch mixtures. SNP loci used in this study appear to have stable allele frequencies over several generations, and should make a reliable genetic baseline for future use. A detailed genetic baseline will enable stock-specific research on the effects of chum salmon bycatch in different geographic regions.

By incorporating loci from both nuclear and mitochondrial genomes we were able to gain unexpected insights into the genetic diversity of chum salmon. The majority of populations surveyed in the region were fixed for a single haplotype at two mtDNA SNPs (*Oke_Cr30* and *Oke_Cr386*). However, in three westernmost populations (18-Little John Lagoon, 19- Sandy Cove, and 21- Delta Creek) we found a high proportion of individuals with a haplotype that is characteristic of Asian population (Sato et al. 2004). Interestingly, these populations showed little differentiation at nuclear DNA loci (Figure 1.2B). Discordance between nuclear and mitochondrial genomes has been documented in a diversity of animal taxa (Hoffman et al. 2009; Humphries and Winker 2011; Toews and Brelsford 2012). There are multiple factors that can uncouple the evolutionary history of nuclear and mtDNA, including selection, genetic drift, sex-biased gene flow, and mtDNA capture. At present, it is unclear which of these factors are responsible for the observed genetic patterns but these mtDNA loci will provide increased resolution for mixed stock analyses.

Conclusion

We have shown that chum salmon populations on the Alaska Peninsula and Kodiak Island contain considerable genetic diversity for the species. Patterns of differentiation were influenced by the geographic distance separating populations and the glacial history of the region. The fine-scale genetic baseline that we have assembled will be useful to fisheries managers as they continue to evaluate the impacts of bycatch and monitor the long-term productivity of chum salmon populations.

TABLES

Table 1. 1. List of sampling locations with associated collection information. Population groupings (north Alaska Peninsula, south Alaska Peninsula, and Kodiak Island) correspond to map colors on Figure 1. *Ne* was estimated using the program Ldne (Waples and Do 2008).

Pop#	Location	Collection dates	Latitude	Longitude	N	N_e
North Alaska Peninsula						
1	Wiggly Creek	08/04/1993, 08/13/2009	56.98	-157.66	178	951
2	Plenty Bear Creek	08/02/1993, 08/11/2009	56.71	-158.30	138	2939
3	Meshik River	08/01/1992, 08/11/2009	56.61	-158.50	172	1208
4	Ilnik River	07/30/2002	56.55	-159.64	49	486
5	Cape Seniavin	08/28/2001, 08/16/2009, 08/2010	56.43	-160.03	96	717
6	Moller Bay	08/16/2009	55.78	-160.35	94	523
7	Lawrence Valley Creek	08/01/1992, 08/16/2009	55.83	-160.66	190	547
8	Coal Valley	08/27/2008	55.85	-160.76	94	605
9	Deer Valley	08/27/2008	55.72	-160.79	91	2669
10	Nelson River	08/01/1992, 08/27/2008	55.73	-161.11	143	675
11	Moffet Creek	09/05/1996	55.46	-162.51	95	1388
12	Joshua Green	08/18/2009	55.37	-162.47	92	2659
13	Frosty Creek	09/01/1992, 08/20/2009	55.19	-162.86	189	1226
14	Alligator Hole	09/04/1996, 08/20/2009	55.16	-162.94	183	1385
15	Trader's Cove	08/01/1992	54.93	-163.28	75	197
16	St. Catherine Cove	08/19/2009	55.01	-163.53	93	262
17	Peterson Lagoon	08/01/1992,	54.93	-164.17	180	449

Pop#	Location	Collection dates	Latitude	Longitude	N	N_e
		08/19/2009				
South Alaska Peninsula						
18	Little John Lagoon	08/01/1992, 08/19/2009	55.01	-162.91	172	1213
19	Sandy Cove	08/26/1996, 08/20/2009	55.14	-163.04	188	707
20	Russell Creek	08/30/1993, 08/20/2009	55.19	-162.68	187	2144
21	Delta Creek	08/29/1996	55.11	-162.35	94	831
22	Belkovski River	08/01/1992	55.15	-162.14	87	203
23	Volcano Bay	08/15/2009	55.24	-162.00	95	653
24	Ruby's Lagoon	08/31/1996	55.57	-161.68	92	331
25	Canoe Bay	08/01/1992, 08/15/2009	55.53	-161.14	181	4107
26	Zachary Bay	8/13/1992	55.38	-160.69	77	396
27	Foster Creek	08/20/1992, 08/14/2009	55.57	-160.72	181	464
28	Coleman Creek	09/06/1996	55.61	-160.57	95	7443
29	Chichagof Bay	08/17/2009	55.69	-160.24	92	211
30	Big River	08/17/2009	55.86	-159.75	94	406
31	Stepovak River	08/18/1992, 08/17/2009	55.93	-159.65	144	192
32	Ivanoff River	08/23/1993, 08/14/2009	55.91	-159.43	181	907
33	Portage Creek	8/21/1993	56.20	-158.62	95	1053
34	Kujulik Bay	08/22/1993, 08/12/2009	56.68	-157.76	163	521
35	Aniakchak River	8/3/1993	56.80	-157.66	94	4566
36	Amber Bay	08/12/2009	56.86	-157.46	85	433
37	Northeast Creek	08/23/2008	56.88	-157.39	94	8867

Pop#	Location	Collection dates	Latitude	Longitude	N	N_e
38	Ocean Bay	08/13/2009	56.91	-157.02	79	442
39	Nakililock River	08/23/2008	56.99	-156.94	95	756
40	Chiginigak Bay River	08/20/1993, 08/12/2009	57.04	-156.76	160	465
41	Kialagvik River	08/11/1993, 08/25/2009	57.28	-156.80	178	1088
42	Pass Creek	08/24/2009	57.40	-156.36	94	592
43	Dry Bay River	08/24/2009	57.64	-155.78	71	300
44	Bear Bay Creek	08/11/1993, 08/19/2009	57.87	-155.25	188	1244
45	Alagogshak River	08/12/1993	58.02	-155.04	94	427
46	Big River, Hallo Bay	09/02/2009	58.63	-153.96	95	385
Kodiak Island						
47	Karluk Lagoon	08/31/2009	57.57	-154.42	84	54
48	Sturgeon River	07/01/1992, 07/10/2009	57.50	-154.51	158	684
49	Big Sukhoi	08/01/1992, 08/07/2009	56.97	-154.31	190	17789
50	Deadman River	08/18/2009	57.18	-153.77	95	401
51	Sitkinak Island	08/07/2009	56.58	-154.09	93	148
52	Northeast Portage	08/18/2009	57.00	-153.76	94	1707
53	Barling Bay Creek	08/06/2009	57.20	-153.40	93	490
54	West Kiliuda Creek	09/12/2009	57.31	-153.20	83	777
55	Dog Bay	08/01/1992	57.31	-153.16	95	612
56	Coxcomb Creek	08/17/2009	57.34	-153.00	90	334
57	Gull Cape Lagoon	09/14/1993, 09/23/2009	57.33	-152.64	186	324
58	Eagle Harbor	08/17/2009	57.42	-152.73	95	165
59	Rough Creek	09/12/2009	57.51	-152.77	76	628

Pop#	Location	Collection dates	Latitude	Longitude	N	N_e
60	American River	09/01/1992	57.65	-152.51	94	2006
61	Russian River	08/17/2007, 08/17/2009	57.70	-152.59	185	547
62	Kizhuyak River	08/01/1992, 08/19/2009	57.72	-152.88	175	542
63	Uganik River	08/01/1992, 08/20/2009	57.68	-153.42	175	943
64	Spiridon River - Upper	08/18/2009	57.51	-153.42	89	1307
65	Zachar River	07/21/2009	57.49	-153.64	66	521
66	Kitoi Bay Hatchery	07/23/1993, 07/29/2009	58.19	-152.37	192	373

Table 1. 2. Fit of alternative IBD models in each geographic region with different outlier populations excluded from the analysis. The number of populations in the analysis are indicated by *n*, *K* is the number of parameters, R^2 is from the linear regression of linear F_{ST} to waterway distance, p values are from a least squares regression, and *AIC* is calculated as in Koizumi et al. (2006).

Population excluded	**n**	**K**	**R^2**	**intercept * 10^2**	**slope* 10^2**	**p value**	**AICc**	**ΔAICc**
North Alaska Peninsula								
10	16	1	0.517	0.926	0.012	< 0.0001	-101	1
none	17	1	0.394	1.350	0.011	< 0.0001	-100	
South Alaska Peninsula								
24, 26, 19	26	1	0.084	0.892	0.001	< 0.0001	-211	10
24, 26	27	1	0.086	0.926	0.001	< 0.0001	-200	7
24	28	1	0.034	1.141	0.001	0.0003	-194	9
none	29	1	0.015	1.361	0.001	0.0132	-185	
Kodiak Island								
66, 48, 47, 51, 49, 57	14	1	0.485	0.274	0.004	< 0.0001	-131	14
66, 48, 47, 51, 49	15	1	0.095	0.709	0.003	0.0013	-117	8
66, 48, 47, 51	16	1	0.046	0.987	0.003	0.0187	-109	2
66, 48, 47	17	1	0.024	1.347	0.003	0.0746	-108	0
66, 48	18	1	0.024	1.555	0.003	0.0542	-108	32
66	19	1	0.007	2.672	0.005	0.2605	-76	7
none	20	1	0.009	3.490	0.007	0.1847	-69	

FIGURES

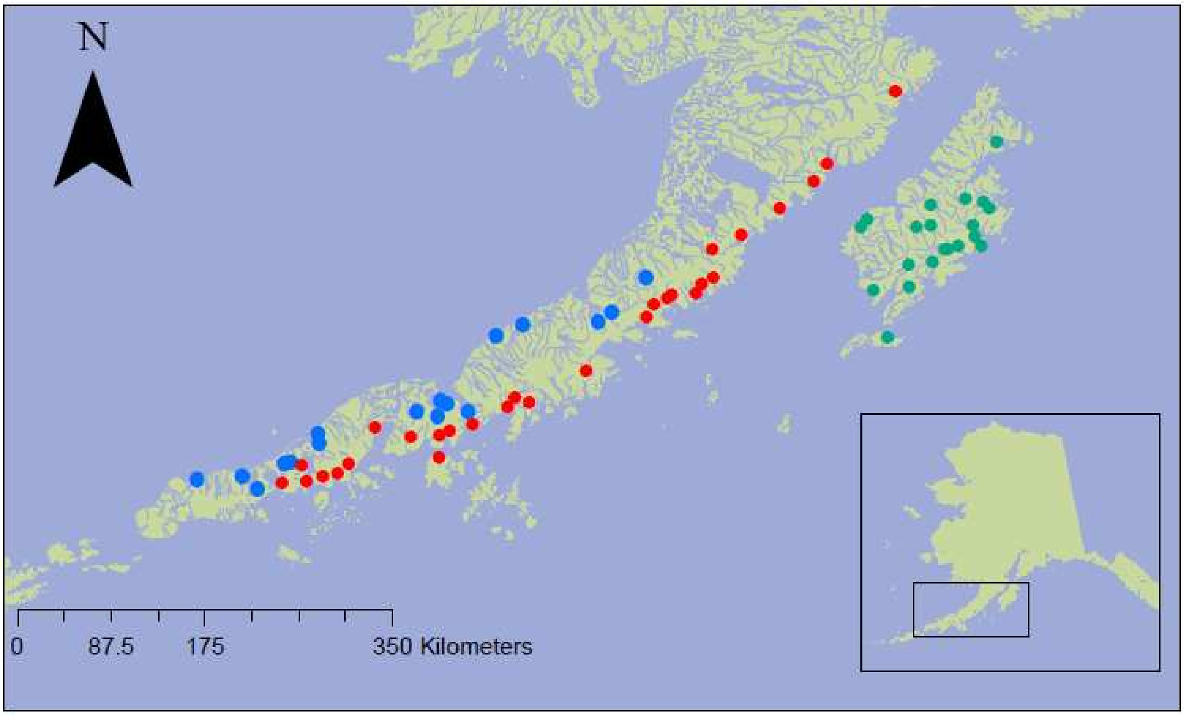


Figure 1. 1. Map of sampling locations. Populations collected from the north Alaska Peninsula (n = 17) are in blue, south Alaska Peninsula populations (n = 29) are in red, and Kodiak Island populations (n = 20) are in green.

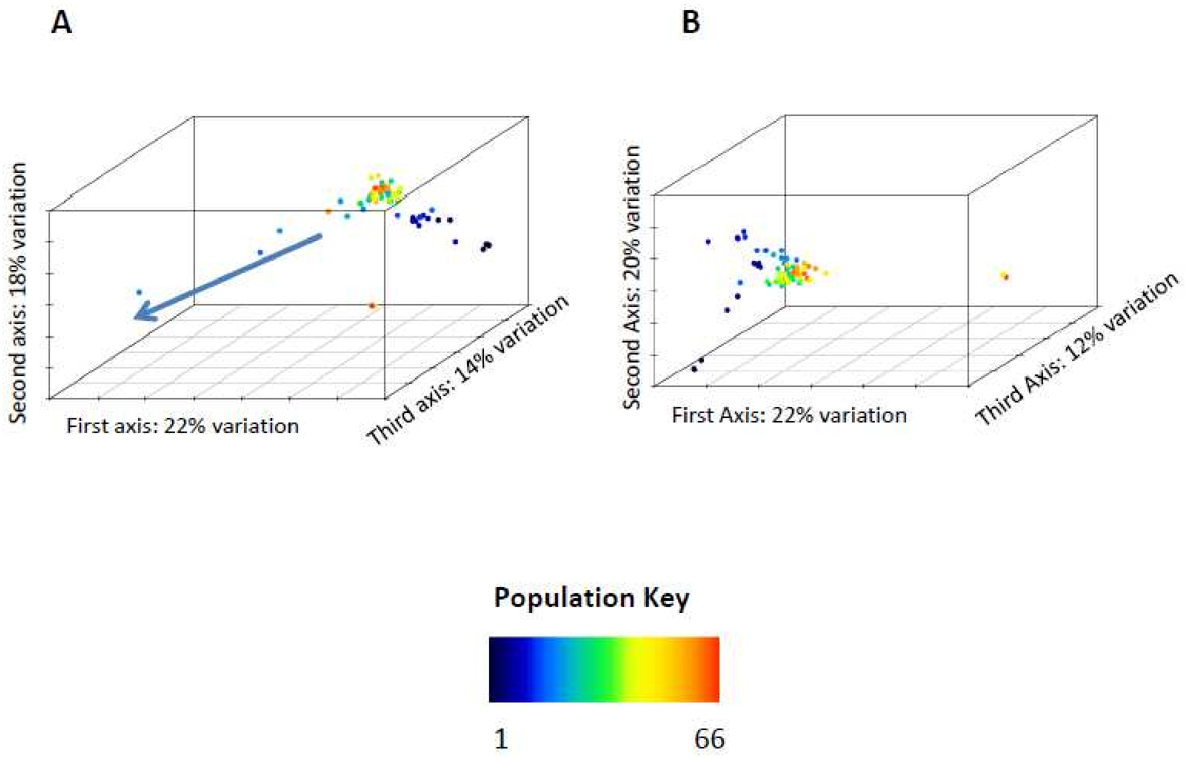


Figure 1. 2. Correspondence analysis using mean population data. Each dot represents a population and color indicates the population number as in Table 1. A) Analysis using 93 nuclear SNP loci and composite mtDNA haplotype. The arrow is a schematic representation of differentiation driven by mtDNA composite haplotype. Three populations (18- Little John Lagoon, 19- Sandy Cove, and 21- Delta Creek) contain alternative mtDNA haplotype. B) Little John Lagoon, Sandy Cove, and Delta Creek populations appear similar to adjacent populations when only nuclear SNP loci are considered. Highly differentiated north Alaska Peninsula populations are depicted in dark blue, while the Sturgeon River and Kitoi Bay Hatchery populations are found at the far right of the plot.

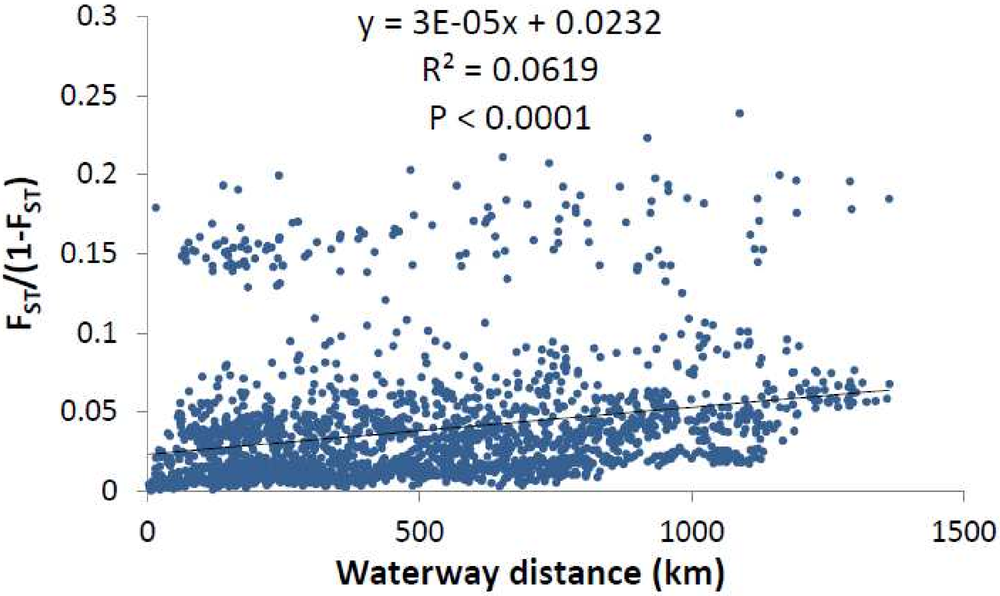


Figure 1. 3. Regression of waterway to genetic distance considering all populations. The cloud of highly differentiated points is composed of pairwise comparisons between the Sturgeon River and Kitoi Bay Hatchery to all other populations.

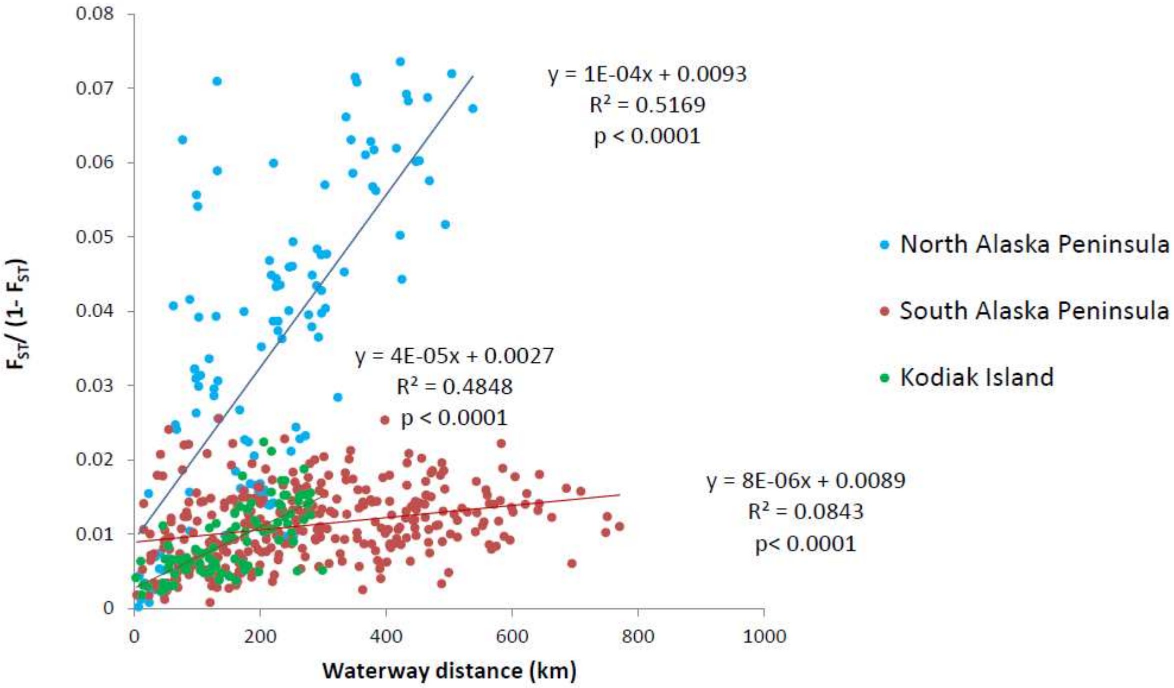


Figure 1. 4. Regression of waterway to genetic distance with populations partitioned by geographic region. Populations identified as outliers by decomposed pairwise regression analysis were removed from this analysis.

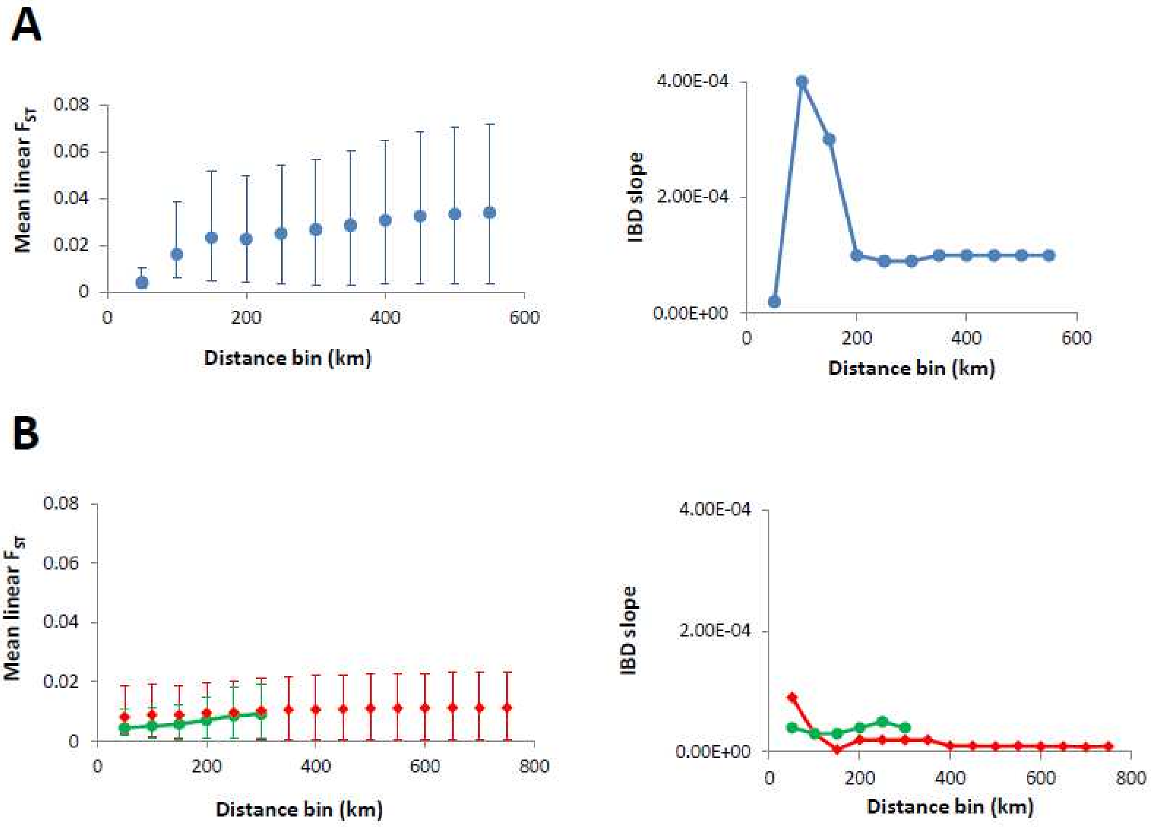


Figure 1. 5. Mean F_{ST} with 95% confidence intervals and IBD slope at increasing spatial scales for populations from the A) North Alaska Peninsula (blue) and B) South Alaska Peninsula (red) and Kodiak Island (green).

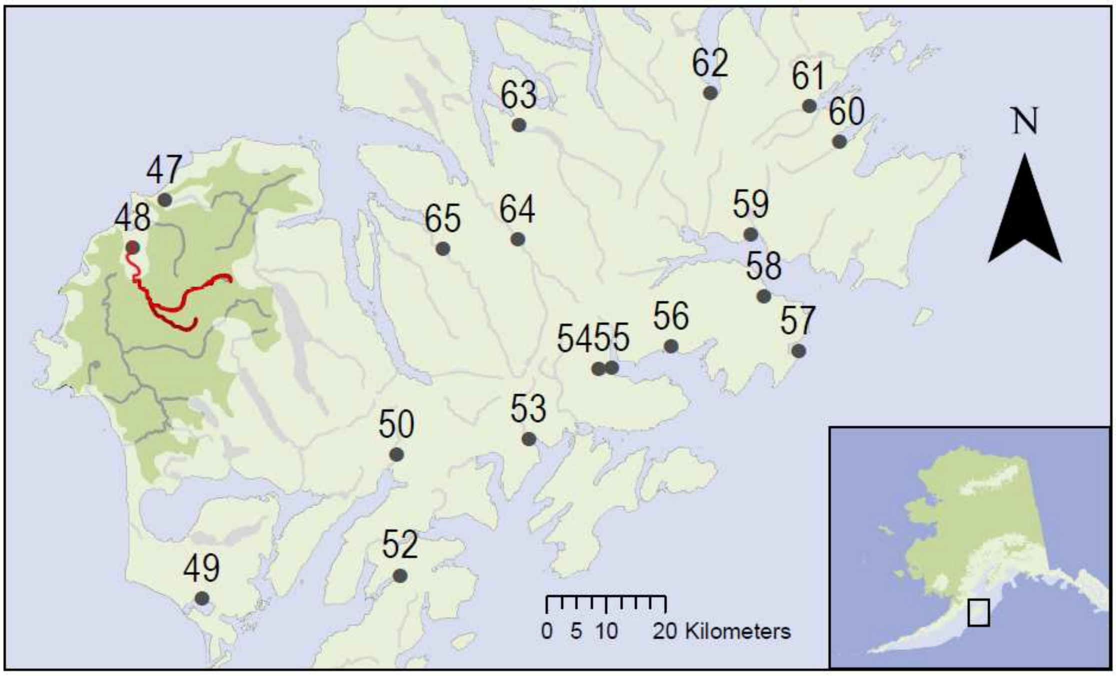


Figure 1. 6. Sampling locations on Kodiak Island. Transparent white overlay represents the Cordilleran Ice Sheet and the ice free pocket in southwest Kodiak Island is discernible in green. The Sturgeon River is indicated in red.

CHAPTER 2: Population structure of chum salmon in Alaska mediated by sea level fluctuations

INTRODUCTION

A main goal of population genetics is to understand environmental and historical processes influencing the genetic diversity of populations. Migration, selection, and genetic drift are factors contributing to either genetic differentiation or homogeneity in a species. Locally restricted dispersal and reproduction are expected to give rise to patterns of isolation by distance (IBD), in which genetic differentiation increases with geographic distance for populations in a one-dimensional array (Wright 1943). Theoretical models predict that populations will reach migration-drift equilibrium over time when uniform demographic and selective conditions exist (Wright 1943; Kimura and Weiss 1964). Restricted gene flow across physical (e.g. waterfall, Koizumi et al. 2006) or temporal (e.g. reproduction time, Ramstad et al. 2004) barriers can cause departures from IBD and migration-drift equilibrium. Similarly, selection across an environmental gradient (Freedman et al. 2010) and/or secondary contact can lead to abrupt genetic differentiation between neighboring populations (Endler 1977; Barton and Hewitt 1985).

It can be difficult to disentangle historical factors (such as secondary contact) and contemporary processes (such as migration patterns, environmental gradients, etc.) in influencing population structure. Genetic differentiation between neighboring populations is often attributed to secondary contact between lineages that were separated during the Pleistocene glaciations, and there exist multiple examples of this in terrestrial species (Swenson and Howard 2005; Hewitt 2011). Pleistocene climatic oscillations likely affected habitats of aquatic and marine species as well, as cycles of glacial advances reached their largest extent during the Last Glacial Maximum (LGM) approximately 25,000 years before present (ybp). This period was characterized by colder temperatures, reduced precipitation, and lower sea levels (Clark et al. 2009). The reduction in sea level and encroachment of ice sheets onto the continental shelf restricted available habitat for coastal species occupying subpolar latitudes. This likely led to fragmentation of populations, and as time passed different alleles became prevalent in different populations because of mutation, genetic drift, and/or selection (Hewitt 2004). Numerous studies report genetic differentiation in aquatic organisms that is attributed to climatic changes of the Pleistocene. Fish species that conform to this pattern include lake whitefish (*Coregonus clupeaformis*, Bernatchez and Dodson 1991), pygmy whitefish (*Prosopium coulteri*, Witt et al. 2011), rainbow trout (*Oncorhynchus mykiss*, Tamkee et al. 2010) , char (*Salvelinus sp.*, Redenbach and Taylor 2002), and coho salmon (*Oncorhynchus kisutch*, Smith et al. 2001).

Here, we examined the fine scale population structure of chum salmon (*Oncorhynchus keta*) in southwestern Alaska to understand factors influencing genetic differentiation in the region. Chum salmon are widely distributed across the Northern Pacific Rim, and spawning populations can be found in Korea, Japan, China, and Russia in the western Pacific, and from Alaska to Oregon in the eastern Pacific (Salo 1991). During Pleistocene glaciations, the range was fragmented by changes in sea level and emergence of the Bering Land Bridge, as well as presence of ice sheets along the coast (Pewé 1975; Dyke et al. 2002). In Alaska, much of the current habitat used by chum salmon was covered by the Cordilleran Ice Sheet, except for central Alaska which remained ice-free (Pewé 1975; Brigham-Grette 2001). These climatic changes could have affected genetic population structure in two ways (Bernatchez and Dodson 1991). First, populations persisting in separate glacial refugia might have diverged into genetically distinct lineages due to geographic isolation. Alternatively, changes in hydrology might have facilitated gene flow between previously isolated populations.

Numerous studies contribute to our understanding of the evolutionary history of chum salmon (Wilmot et al. 1994; Seeb and Crane 1999; Sato et al. 2004; Beacham et al. 2009; Seeb et al. 2011). Genetic diversity exists across the species range, and populations cluster in large-scale geographical groups. Research also indicates that the Alaska Peninsula contains a major zoogeographic break for the species (Wilmot et al. 1994; Seeb & Crane 1999). Using allozymes, these studies demonstrate that populations sampled from the northwestern Alaska Peninsula are similar to populations from the southern range of chum salmon in North America, while populations from the northeastern Alaska Peninsula are similar to populations from Western Alaska. This abrupt genetic differentiation is unusual for chum salmon, and it has been suggested that the region is a secondary contact zone (Seeb and Crane 1999). Theory predicts that genetic differences between populations in secondary contact will decay as a function of gene flow and time since initial contact (Endler 1977). However, secondary contact zones can persist over many generations when maintained by some kind of barrier to gene flow, such selection against individuals with mixed ancestry (Barton and Hewitt 1985).

In this study we investigated the putative contact zone for chum salmon on the Alaska Peninsula and explored whether observed patterns of genetic differentiation are caused by relatively recent secondary contact or contemporary factors restricting gene flow. We screened 161 single nucleotide polymorphisms (SNP) to explore patterns of genetic differentiation in the region. Subsequently, empirical data were compared to data simulated by two different colonization routes. The first modeled colonization of the Alaska Peninsula by two lineages meeting in a secondary contact zone, and the second simulated colonization by a single genetic lineage. The simulation procedure allowed us to test whether the observed genetic structure on the northern Alaska Peninsula is caused by secondary contact between two genetic lineages of chum salmon.

We confirmed evidence of hierarchical population structure and found lack of migration-drift equilibrium on the northern Alaska Peninsula. Simulated data from the secondary contact and single lineage models show different patterns of IBD as the simulated populations reach equilibrium. Empirical data did not strongly resemble either colonization model, and we were not able to discern between the two scenarios based on simulation results alone. However, previous research in geology and archaeology suggests that the most recent colonization of the lower Alaska Peninsula by chum salmon likely occurred only 1,300 ybp due to the destruction of coastal spawning habitats by fluctuations in sea level (Maschner 1999; Jordan 2001). Thus, lack of migration-drift equilibrium in the empirical data is likely caused by recent colonization of the region by chum salmon.

MATERIALS AND METHODS

1. Empirical data

Tissue samples were collected from 2001 to 2010 by the Alaska Department of Fish and Game (ADFG) and the University of Washington. Approximately 95 spawning salmon were sampled from each of 11 different locations on the Alaska Peninsula that likely represent distinct breeding populations (Figure 2.1 and Table 2.1). In 2001 and 2002 (Table 2.1), salmon were captured by beach seine, and heart tissue was collected and frozen in liquid nitrogen until transferred to long-term storage at -80°C. From 2008 onward, individual salmon were captured by beach seine, held momentarily while the axillary process on the ventral fin was removed, and subsequently released. Axillary processes were stored in 95% ethanol.

DNA was extracted from 992 individuals using the DNeasy 96 Tissue Kit (QIAGEN, Valencia, CA). SNPs at 192 loci (Supplementary Table 2.1) were assayed using 5'-nuclease reactions (Seeb et al. 2009) on the BioMark 96.96 Dynamic Array (Fluidigm, San Fransisco, CA). Eighty-four loci were previously described (Smith et al. 2005a; Smith et al. 2005b; Elfstrom et al. 2007; Seeb et al. 2011), and novel assays for 108 loci were developed specifically for this project (Supplementary Tables 2.1 and 2.2). As a quality control measure, 8% percent of the samples were genotyped again to ensure reproducibility and identify laboratory errors.

Our goal was to obtain a sample size of 95 individuals from each spawning population. It was not possible to achieve this goal at the Cape Seniavin sampling site due to logistical constraints and the remote location. MCMC exact tests for genic differentiation (GENEPOP 4.0, Raymond and Rousset 1995; Rousset 2008) between collection years at Cape Seniavin were conducted to test for significant temporal variation.

Allele frequencies and expected heterozygosity of each locus were calculated in GENALEX 6.4 (Peakall and Smouse 2006). Deviations from Hardy-Weinberg equilibrium (HWE) at each population were detected using exact P-values ($\alpha = 0.05$) and the MC algorithm in GENEPOP 4.0 (Raymond and Rousset 1995; Rousset 2008). Tests were performed across all loci using the default parameters in the program. We examined patterns of linkage disequilibrium between all locus pairs using exact tests in GENEPOP 4.0. If a pair of loci was in linkage disequilibrium in the majority of populations, we chose one of the SNP loci (based on highest F_{ST}) and retained it for the remainder of the analyses. This was done to avoid redundancy in the data set. We compared patterns of linkage disequilibrium in each population by calculating the statistic r^2 that describes the overall correlation coefficient averaged across multiple loci and alleles. The calculations were performed with the program LDNE (Waples and Do 2008), and r^2 was subsequently used to estimate the effective population size (N_e) of each sample collection.

Genetic differentiation between populations was estimated using the Weir and Cockerham (1984) F_{ST} statistic in GENEPOP 4.0. Statistical significance of pairwise population differentiation was conducted in ARLEQUIN 3.5 using 1,000 permutations (Excoffier and Lischer 2010). Patterns of differentiation were explored with a principal components analysis (PCA) based on individual genotypes using the ADEGENET package (Jombart 2008) in R (R Development Core Team 2011).

The shortest waterway distance between sampling locations was estimated using a least cost path analysis in ARCGIS ver. 10 (ESRI, Inc). A regression of $F_{ST}/(1-F_{ST})$ estimates to waterway distance (Rousset 1997) was computed in R, and Mantel tests with 1,000 permutations were used to evaluate statistical significance ($\alpha = 0.05$). We examined the effect of spatial scale on the linearity of IBD by partitioning the data into increasing bin sizes (0-50, 0-100, 0-150 km, etc.) and iteratively regressing genetic to waterway distance (following Bradbury and Bentzen 2007). The slope of IBD was subsequently examined at these incremental spatial scales for deviations from linearity. Non-linear IBD slope indicates departure from migration-drift equilibrium, as equilibrium populations arranged along a linear transect are expected to have constant and monotonic IBD slopes at all spatial scales (Hutchison and Templeton 1999).

2. Simulated data

To evaluate whether the processes of genetic drift, mutation, and migration could lead to patterns of genetic differentiation similar to those observed on the Alaska Peninsula, we modeled genotypic data for 161 loci in 12 populations using the program Easypop (Balloux 2001). Each population contained an effective population size (N_e) of 1,000 randomly mating, diploid individuals, with even sex ratio. Every locus had two possible allelic states, mimicking the biallelic SNP loci in the empirical data set. Alleles were randomly assigned to the initial populations, and free recombination was allowed between all loci. Each locus mutated at a rate of 10^{-8}, as this approximates the mutation rate of neutral SNPs (Brumfield et al. 2003). Simulated populations were subsequently allowed to colonize a linear habitat by following a: 1) secondary contact or 2) single lineage model.

Secondary contact simulations

For the first part of the secondary contact model (before colonization), a hierarchical island migration pattern was used with two hierarchical groups (Carmelli and Cavalli-Sforza 1976; Sawyer and Felsenstein 1983). Within each group there were six populations, and individuals were allowed to migrate to any other population with a uniform migration rate m = 0.05 but there was no migration between groups. We ran this model for 6,250 generations (approximately 25,000 years, considering that chum salmon most commonly reach sexual maturity at 3 to 4 years of age (Beacham and Murray 1987). This time was chosen to reflect the maximum extent of the Cordilleran Ice Sheet on the Alaska Peninsula about 25,000 ybp (Mann and Hamilton 1995).

For the second part of the secondary contact model, we used a one-dimensional stepping-stone model (Kimura and Weiss 1964) to simulate colonization by genetically distinct lineages. Individuals were allowed to migrate to adjacent populations with a uniform migration rate m = 0.05, which is within the range of estimated adult chum salmon straying between natal streams (reviewed in Johnson et al. 1997). Populations at either end of the one-dimensional array only exchanged migrants with one adjacent population. The one-dimensional stepping stone model is an appropriate approximation for sampled streams along the Alaska Peninsula, as they are distributed along a transect from east to west (Figure 2.1). We ran the one-dimensional stepping stone model for different lengths for time: 100, 250, 500, and 1000 generations. Once simulations were finished, data were tested for IBD using step distance between populations as the independent variable. Linearity of the IBD slope was examined by partitioning the data into increasing bin sizes (1-2 steps, 1-3 steps, 1-4 steps, etc.) and iteratively regressing step distance to genetic distance.

Single lineage simulations

For the first 6,250 generations of the single lineage simulation, individuals were allowed to migrate to any other population with a uniform m = 0.05. Subsequently, all populations were ordered along a one-dimensional stepping stone (as described above), and the simulation ran for 100, 250, 500, and 1000 generations. Once simulations were finished, data were tested for IBD and the linearity of the IBD slope was examined as described for the empirical data.

RESULTS

1. Empirical data

Repeated genotyping of individuals at all loci indicated that genotyping errors occurred at an overall frequency of 0.05%. A small number of individuals from each population were missing more than 15% of genotypes, indicating poor-quality DNA, and were removed from the dataset, leaving 912 individuals out of 993 for the remainder of the analyses. Monomorphic loci and loci with an average minor allele frequency across all populations less than 0.01 were discarded (Supplementary Table 2.1). No significant differentiation across all loci ($\alpha = 0.01$) was found for the Cape Seniavin collections taken from different years (2001, 2009, 2010), and we pooled individuals to achieve a larger sample size for that sampling location.

We conducted 1,772 tests to screen for deviations from HWE and found 56 tests (3.2%) to be statistically significant ($\alpha = 0.05$). There were no consistent patterns in HW disequilibrium across loci or populations. As the number of significant tests was less than could be expected by chance alone, we concluded that the HWE tests support our interpretation that the genetic data reflect simple Mendelian inheritance of codominant markers in populations undergoing approximately random mating.

We tested for linkage disequilibrium by conducting 207,717 tests between pairs of loci and found 9927 (4.8%) to be statistically significant at $\alpha = 0.05$. Although this overall level of LD is slightly less than expected by chance, sixteen locus pairs were found to be in LD in the majority of the populations, and all of these LD pairs were ascertained from the same sequence. We retained only one locus from each pair for the remainder of the analyses, to avoid redundancy in the data. After accounting for LD and low minor allele frequency, 161 nuclear SNP loci were available for use in the downstream analyses (Supplementary Table 2.1). We found that the average r^2 over all populations was 0.014 (Table 2.2). However, the Plenty Bear Creek and Ilnik River samples had an r^2 that was approximately twice that of other populations. The median estimate of N_e was 822, and values ranged from 333 to 7,287.

Locus-specific F_{ST} ranged from 0.0004 at *Oke_zn593-152* to 0.1482 at *Oke_FANK1-96* (Supplementary Table 2.1). When allele frequencies were plotted against sampling location, 42 loci showed gradients in allele frequency across the Alaska Peninsula. Several loci (*Oke_FANK1-96, Oke_ROA1-209,* and *Oke_U509-219*) were characterized by abrupt and coincident changes in allele frequency (Figure 2.2) between the Meshik River and Ilnik River populations, a region separated by approximately 100 waterway km (Figure 2.1).

Global F_{ST} was 0.036 and pairwise population F_{ST} values ranged from 0.006 to 0.067 (Table 2.3). All pairwise comparisons of populations were statistically significant ($\alpha = 0.05$). In PCA plots, considerable variation occurred among individuals within each population but recognizable groups of individuals were evident (Figure 2.3). Individuals from Whale Mountain Creek, Wiggly Creek, Plenty Bear River, and Meshik River (locations 1-4 in Table 2.1) grouped together, and were clearly distinct from individuals from Joshua Green and Frosty Creek (locations 10 and 11). These findings indicate that hierarchical population structure exists and that individuals sampled from the eastern end of the Alaska Peninsula are genetically distinct from individuals sampled from the western tip of the Alaska Peninsula.

Genetic differentiation increased with geographic distance, and Mantel tests found significant isolation by distance (maximum distance ~ 550 km; $p < 0.0001$, $r^2 = 0.53$) over all of the populations sampled (Figure 2.4). However, the IBD slope was non-linear over increasing spatial scales, indicating that populations on the north Alaska Peninsula are not in migration-drift equilibrium (Figure 2.5A). The smallest IBD slope occurred at the smallest spatial scale (< 50 km).

2. Simulated data

Secondary contact simulations

When secondary contact was simulated for 100 generations, populations were highly differentiated. Genetic differentiation was strongly correlated to step distance between populations (Figure 2.6A) but that relationship was non-linear (Figure 2.5B). When IBD slope was examined at increasing spatial scales, the smallest slopes were found at small distances.

Genetic differentiation decreased after more than 250 generations of secondary contact because gene flow homogenized differences in allele frequency across the two lineages (Figure 2.6B). At 1,000 generations of secondary contact, populations exhibited a strongly linear IBD relationship ($r^2 = 0.97$, $p < 0.0001$) and approximately linear IBD slope (Figure 2.5B), indicating that they were approaching migration-drift equilibrium. When lineages were connected by gene flow for that length of time, it was impossible to detect the signal of secondary contact by examining the IBD slope (Figure 2.5B).

Single lineage simulations

We found statistically significant correlations between increasing geographic and genetic distances for all simulations, regardless of number of generations (100, 250, 500, and 1,000) spent in a one-dimensional stepping stone model of migration (Figure 2.7A and B). However, IBD slope was strongly non-linear for populations simulated with 100 and 250 generations in a one-dimensional stepping stone model (Figure 2.5C). For these simulations, IBD slope was greatest at small spatial scales. IBD slope became more constant over distance as the number of generations in one-dimensional stepping stone model increased, indicating that populations approached migration-drift equilibrium.

DISCUSSION

A central purpose of population genetics is to understand the contribution of different evolutionary processes on population structure. However, it can be very difficult to discern between historical and contemporary factors. We explored the different factors shaping population structure of chum salmon by using a combined experimental and simulation approach.

Population structure in empirical data

We found that significant structure exists among chum salmon populations spawning in different rivers on the northern Alaska Peninsula. Genetic differentiation was greatest among populations sampled from the northeastern and northwestern ends of the Alaska Peninsula. Although sampling locations spanned a linear distance of only ~500 km, the overall F_{ST} of these populations was 0.036. In comparison, summer spawning chum salmon in western Alaska rivers that are thousands of km apart have a regional F_{ST} of only 0.004, as estimated using a subset of SNPs used in this study (Seeb et al. 2011). These findings suggest that the Alaska Peninsula is a relative hotspot of genetic differentiation for chum salmon, and the results agree with previous studies using various genetic markers to describe the population structure of the species in Alaska (Seeb and Crane 1999; Smith and Seeb 2008; Seeb et al. 2011).

Several behavioral and geographic factors might explain this striking pattern of genetic differentiation. The strong homing behavior of salmon to their natal stream has been well documented (reviewed in Quinn 2005). Homing restricts gene flow between populations and contributes to genetic divergence, as the stochastic process of genetic drift causes populations to have different allele frequencies at neutral DNA markers. Homing and straying patterns are likely influenced by geographical features of the surrounding landscape because physical barriers (dams, waterfalls, etc.) may contribute to restricted gene flow and increased differentiation.

There are no obvious physical barriers to straying in our study system, and populations do not have notable differences in spawn timing. Rivers on the northern Alaska Peninsula run roughly parallel to each other and flow north to the Bering Sea along relatively flat terrain. All except two of the waterways sampled in our study (Plenty Bear Creek and Meshik River) are hydrologically separated (Figure 2.1). It is possible that straying between spawning sites on the northern Alaska Peninsula is less than other parts of the species range where rivers share a common outlet or confluence. If parallel and separate rivers influence straying rates and gene flow between spawning populations, then one might expect to see the similarly high patterns of genetic differentiation in other regions where chum salmon spawn in separate watersheds. Previous research (Olsen et al. 2008) examining the population structure of chum salmon spawning in parallel waterways near Norton Sound found very low levels of genetic differentiation. Similarly, chum salmon populations spawning on the southern Alaska Peninsula show little differentiation (see chapter 1, this thesis). This suggests that hydrographic structure might not be the most important driver of genetic diversity in our study region.

Other researchers (Seeb and Crane 1999) reasoned that strong genetic differentiation between spawning populations on the northern Alaska Peninsula is driven by secondary contact between two genetic lineages, originating from Cascadian and Beringian refugia, respectively. Their conclusions were supported by multivariate analyses of allozyme genotypes demonstrating that populations from the northeastern Alaska Peninsula were similar to populations from western Alaska, while populations from the northwestern Alaska Peninsula were similar to populations from the southern range of the species (British Columbia and Washington State). Secondary contact zones are often characterized by steep gradient in genetic or morphological characteristics (Barton and Hewitt 1985) that degrades over time when the differences are not maintained by selection, leaving behind a smooth allele frequency gradient (Endler 1977). We found evidence of allele frequency gradients at numerous loci in our empirical data. In addition, several markers (*Oke_FANK1-96, Oke_ROA1-209,* and *Oke_U509-219*) showed abrupt and coincident changes in allele frequency. These loci had a maximum difference of 0.40 in allele frequency between chum salmon populations in the Meshik and Ilnik rivers, spawning locations separated by approximately 100 waterway km (Figure 2.1). These abrupt differences in allele frequency are a curious pattern.

In addition to secondary contact, there are several other processes that produce allele frequency gradients and hierarchical population structure, including incomplete sampling of populations across the study area. A previous study has shown that patchy sampling of individuals across a study region can lead to signals of population clustering and hierarchy that do not necessarily reflect hierarchical clustering (Schwartz and McKelvey 2009). This problem is exacerbated when there is a strong correlation between genetic and geographic distance. The authors suggest sampling on a fine-scale relative to species life history, and examining data for patterns of spatial autocorrelation to avoid finding false signals of population structure. We reason that because our sampling targeted spawning aggregations that were tens to hundreds of km apart, it is unlikely that we failed to sample any important intermediate populations and detected a false signal of hierarchical population structure.

Comparison of IBD slope in empirical and simulated data sets

We investigated whether secondary contact is the most likely driver of genetic differentiation, as suggested by Seeb and Crane (1999). Empirical data were compared to output from two different models, the first representing colonization of the Alaska Peninsula by secondary contact, and the second representing colonization by a single genetic lineage. Our simulations indicated that if the northern Alaska Peninsula were recolonized by chum salmon more than 4,000 ybp (~1000 chum salmon generation ago), then neutrally evolving populations should be approaching migration-drift equilibrium (Figure 2.6 and 2.7). In that case, the signal of secondary contact would be eroded and a strong IBD relationship would be present.

However, if colonization of the Alaska Peninsula were more recent (100, 250, or 500 generations), simulations suggest that we should be able to detect secondary contact by examining IBD slope at different spatial scales. Data simulated by the secondary contact and single lineage models had characteristic and non-linear IBD slope patterns at these time scales. Simulated populations had smallest IBD slopes at small spatial distances when simulated with 100-500 generations of secondary contact (Figure 2.5B). In contrast, simulated populations had largest IBD slopes at small spatial scales when simulated by the single lineage colonization model (Figure 2.5C). Empirical data had non-linear IBD slope, but it is unclear if its pattern most closely matched colonization simulated by recent secondary contact or by a single lineage (Figure 2.5A). Of course several assumptions were made in the simulations that may not reflect the biological reality of our study system. The most important of these is that all simulated populations have the same effective population size (N_e = 1,000) and an equal migration rate (m) connects populations. This could be problematic since genetic divergence among populations at neutral markers is controlled by the number of migrants per generation (N_em). Given the remote location of our sampling sites, there are scarce estimates of numbers of spawning adults in the sampled populations that can be used as a conceptual guide for models. The only available data on this subject are reports of escapement across river weirs (Murphy and Hartill 2009) indicating that chum salmon populations range from several hundreds to thousands of individuals in regional spawning streams (with yearly variation). These observations are similar to genetic estimates of N_e (Table 2.2), and are consistent with our model parameter N_e.

Geological evidence consistent with recent secondary contact

Considering that the Cordilleran Ice Sheet receded from the Alaska Peninsula approximately 14,000 ybp, it may seem unlikely that secondary contact occurred only several hundred generations ago. However, the dynamic geological history of the region might have mediated timing of colonization of the Alaska Peninsula after the last ice age. Isostatic rebound of land after glacial retreat caused relative sea level to slowly drop over time. As a result, present-day coastal spawning habitats on the north Alaska Peninsula were covered by sea water as late as 7,500 ybp (Jordan 2001). The Alaska Peninsula has considerable volcanic and seismic activity due to subduction of the Pacific tectonic plate below the North American plate. There is evidence that an earthquake occurring ~2,100 ybp was accompanied by subsidence of land. As a result, sea level rose by 2-3 meters, altering coastal geography and ecology of the north Alaska Peninsula dramatically (Jordan 2001). The region is a broad coastal plain whose elevation ranges from 0 to 150 m above sea level, and a rise of 2 meters in sea level likely inundated coastal salmon spawning habitats and rearing lakes with salt water.

Interestingly, there is also a gap in the archaeological record during this period. Although there is evidence of human settlement on the Alaska Peninsula beginning 5,000 ybp, no human artifacts are dated from 2,500 – 2,200 ybp due to the destruction or abandonment of villages (Maschner 1999). This gap in the archaeological record has been attributed to the sudden increase in sea level and degradation of coastal resources that may have forced human emigration. It was not until 1,250 ybp that settlements were constructed next to modern pink (*O. gorbuscha*) and chum salmon streams, and salmon procurement intensified again (Maschner 1999). This finding suggests that contemporary ecology of coastal habitats was established at that time (~ 300 chum salmon generations).

The geological and archaeological evidence, in conjunction with our results and previous phylogeographic research conducted on the species (Seeb and Crane 1999), suggests that the most likely explanation for striking genetic differentiation on the north Alaska Peninsula is recent colonization. We were not able to distinguish through the simulations whether colonization occurred by a single lineage or by two lineages that met in a secondary contact zone. However, it is clear that the empirical populations are not in migration-drift equilibrium. Further research in this region could also explore contemporary landscape characteristics that might influence gene flow, and test whether correlations exist between abrupt changes in allele frequency and differences in environmental parameters such as temperature or flow rate in salmon spawning streams. By using this approach it may be possible to differentiate between the effects of different evolutionary processes on the genome.

TABLES

Table 2. 1. Sampling information associated with collection sites, including geographical location, collection dates, and number of individuals sampled.

Location name	Location number	Latitude	Longitude	Year(s)	Month	Number of individuals
Whale Mountain Creek	1	58.2549	-156.5817	2010	August	95
Wiggly Creek	2	56.9846	-157.6637	2009	August	95
Plenty Bear River	3	56.7070	-158.2994	2009	August	50
Meshik River	4	56.6100	-158.5014	2009	August	95
Ilnik River	5	56.4390	-160.0142	2002	July	50
Cape Seniavin	6a	56.4381	-160.0143	2001	August	55
	6b	56.4402	-160.0109	2009	August	21
	6c	56.4334	-160.0250	2010	August	30
Frank's Lagoon	7	56.0449	-160.5066	2010	July	95
Moller Bay	8	55.7779	-160.3461	2009	August	95
Deer Valley	9	55.7160	-160.7860	2008	August	95
Joshua Green	10	55.3717	-162.4746	2009	August	95
Frosty Creek	11	55.1947	-162.8583	2009	August	95

Table 2. 2. Estimates of the effective population size (N_e), 95% confidence intervals of N_e, and overall correlation coefficient (r^2) of linkage disequilibrium for empirical populations.

Location name	**Location number**	**Estimated N_e**	**95% CI of N_e**	**Overall r^2**
Whale Mountain Creek	1	822	471- 2879	0.0115
Wiggly Creek	2	868	452- 7384	0.0134
Plenty Bear River	3	7286	605- infinite	0.0225
Meshik River	4	457	318- 786	0.0112
Ilnik River	5	333	202- 868	0.0236
Cape Seniavin	6	437	290- 842	0.0131
Frank's Lagoon	7	339	254- 499	0.0121
Moller Bay	8	617	386- 1438	0.0111
Deer Valley	9	3324	798 - infinite	0.0116
Joshua Green	10	3211	775- infinite	0.0116
Frosty Creek	11	1360	584- infinite	0.0113

Table 2.1. Pairwise population comparisons in the empirical data. Pairwise distance between populations (km) is shown above the diagonal. Linearized F_{ST} between populations is shown below the diagonal. All pairwise comparisons were statistically significant ($p < 0.0001$).

Population	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	*	225	278	273	296	314	372	409	411	522	551
2	0.018	*	163	158	188	214	275	312	314	434	464
3	0.010	0.007	*	35	107	132	194	231	233	352	382
4	0.014	0.012	0.009	*	102	128	189	226	228	348	377
5	0.034	0.039	0.030	0.038	*	34	95	133	134	254	284
6	0.037	0.034	0.031	0.040	0.002	*	61	99	100	220	250
7	0.041	0.044	0.042	0.043	0.020	0.027	*	37	40	166	195
8	0.038	0.048	0.045	0.044	0.027	0.033	0.024	*	50	187	217
9	0.036	0.046	0.040	0.044	0.023	0.030	0.025	0.003	*	175	205
10	0.072	0.081	0.080	0.073	0.047	0.058	0.045	0.026	0.027	*	41
11	0.066	0.070	0.069	0.066	0.039	0.047	0.036	0.020	0.020	0.006	*

FIGURES

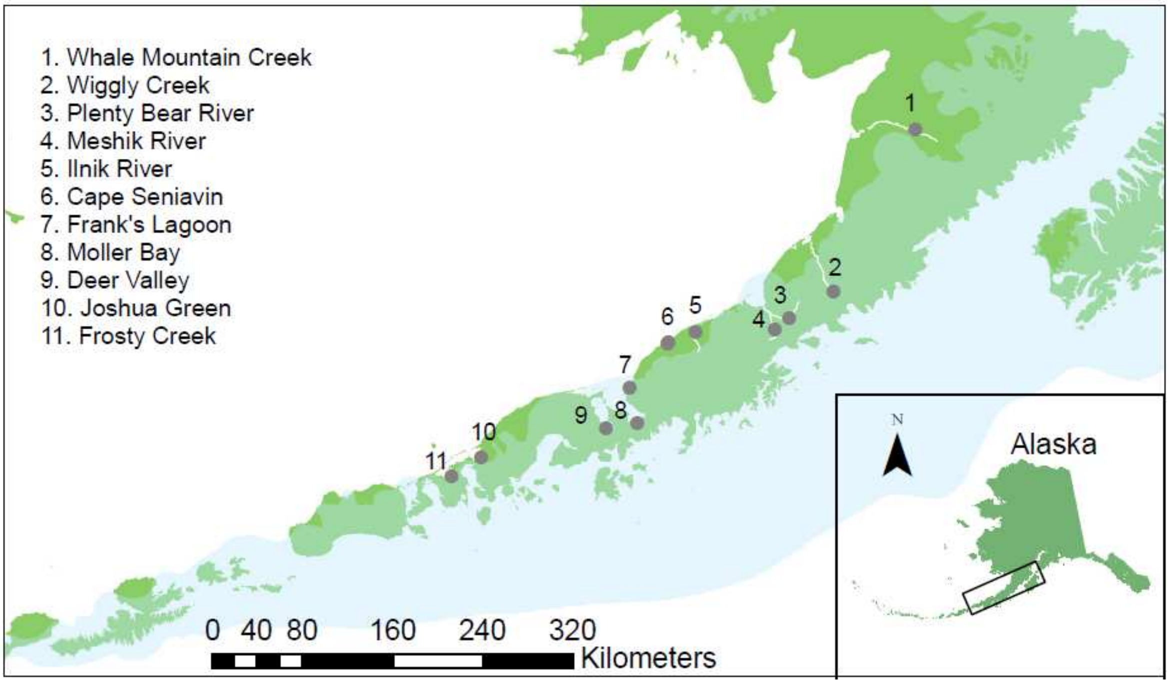


Figure 2.1. Map of the Alaska Peninsula, showing the sampling locations of spawning chum salmon. Landmasses are depicted in green, and the translucent overlay represents the Cordilleran ice sheet circa 25,000 years before present.

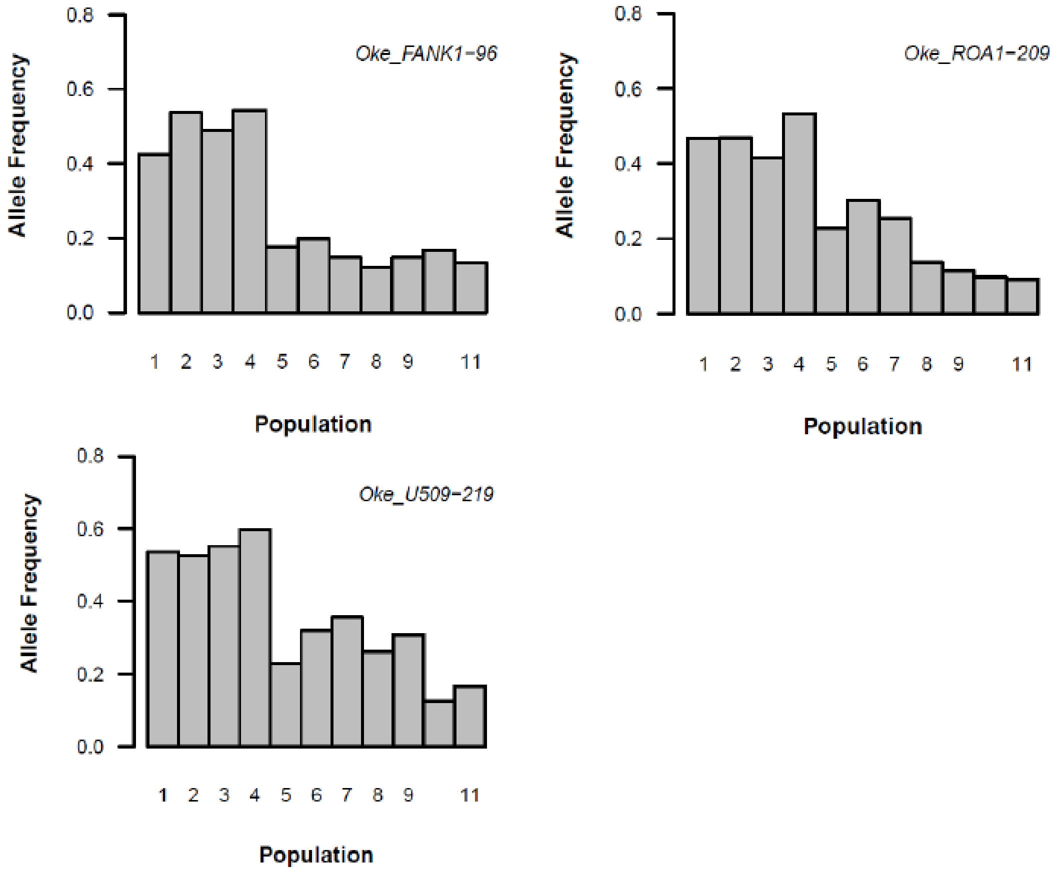


Figure 2. 2. Loci with large allele frequency differences between populations of the Meshik and Ilnik rivers (#4 and #5). Populations are ordered from east to west on the Alaska Peninsula.

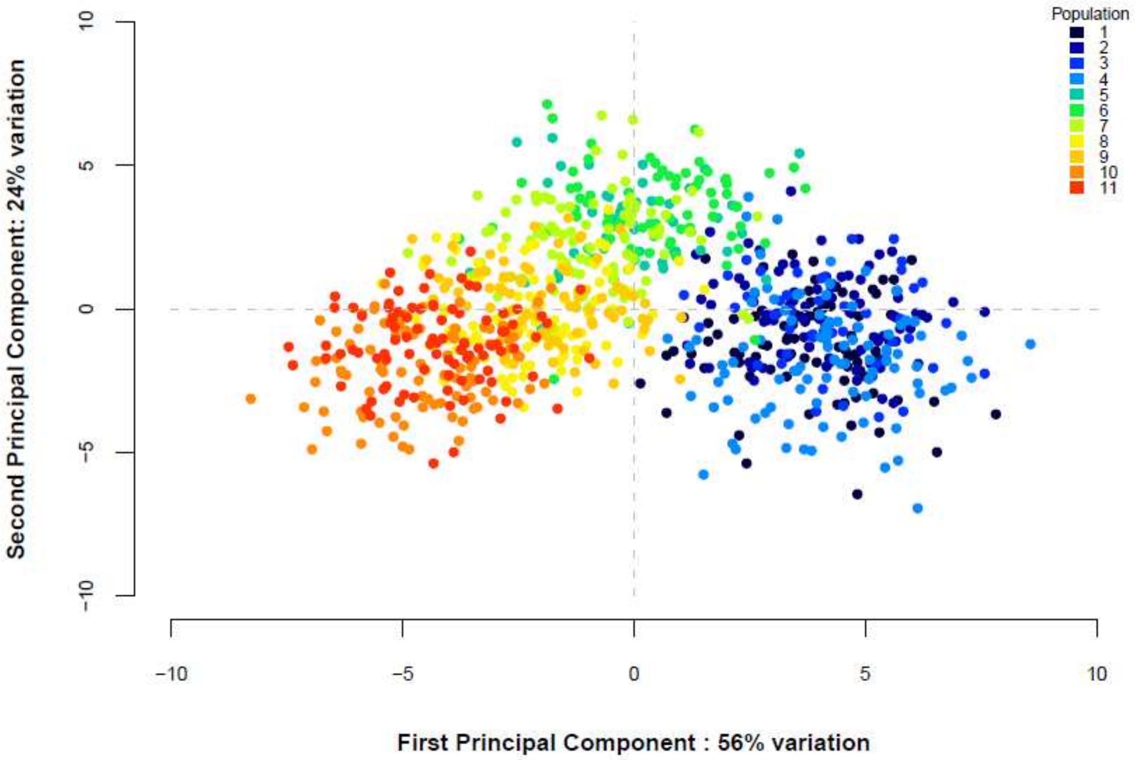


Figure 2. 3. PCA of individuals based on allele frequencies at all loci. Each dot is an individual and colors represent the different populations sampled. Populations are numbered 1(Whale Mountain Creek) to 11 (Frosty Creek), from east to west along the Alaska Peninsula.

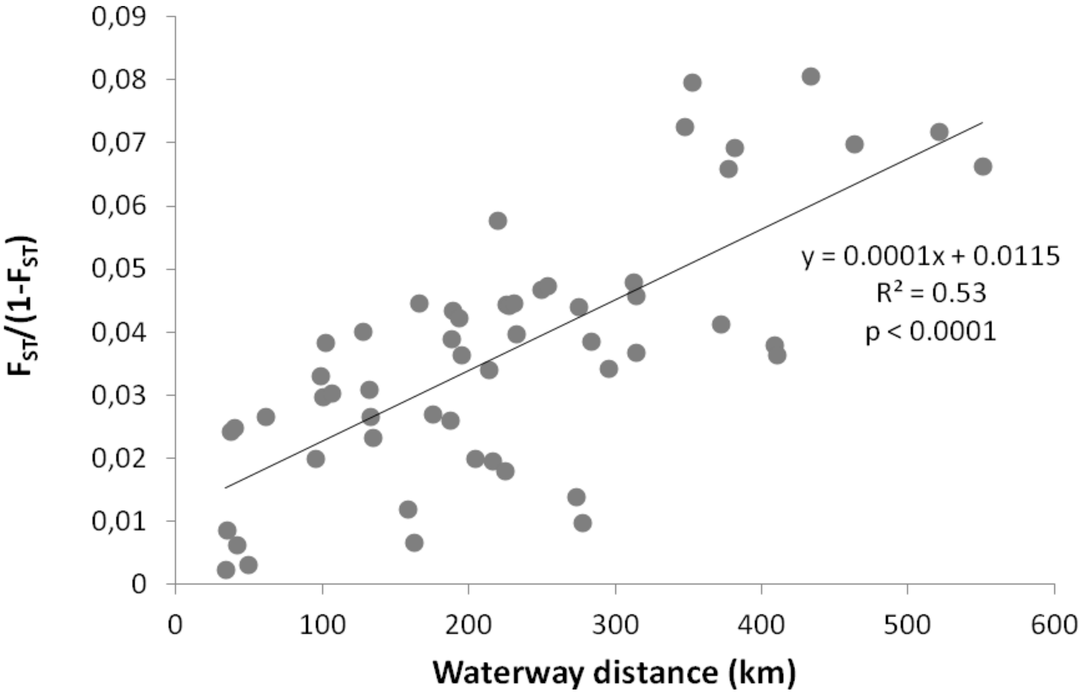


Figure 2. 4. Regression of waterway distance to F_{ST}/ (1-F_{ST}) in the empirical data set.

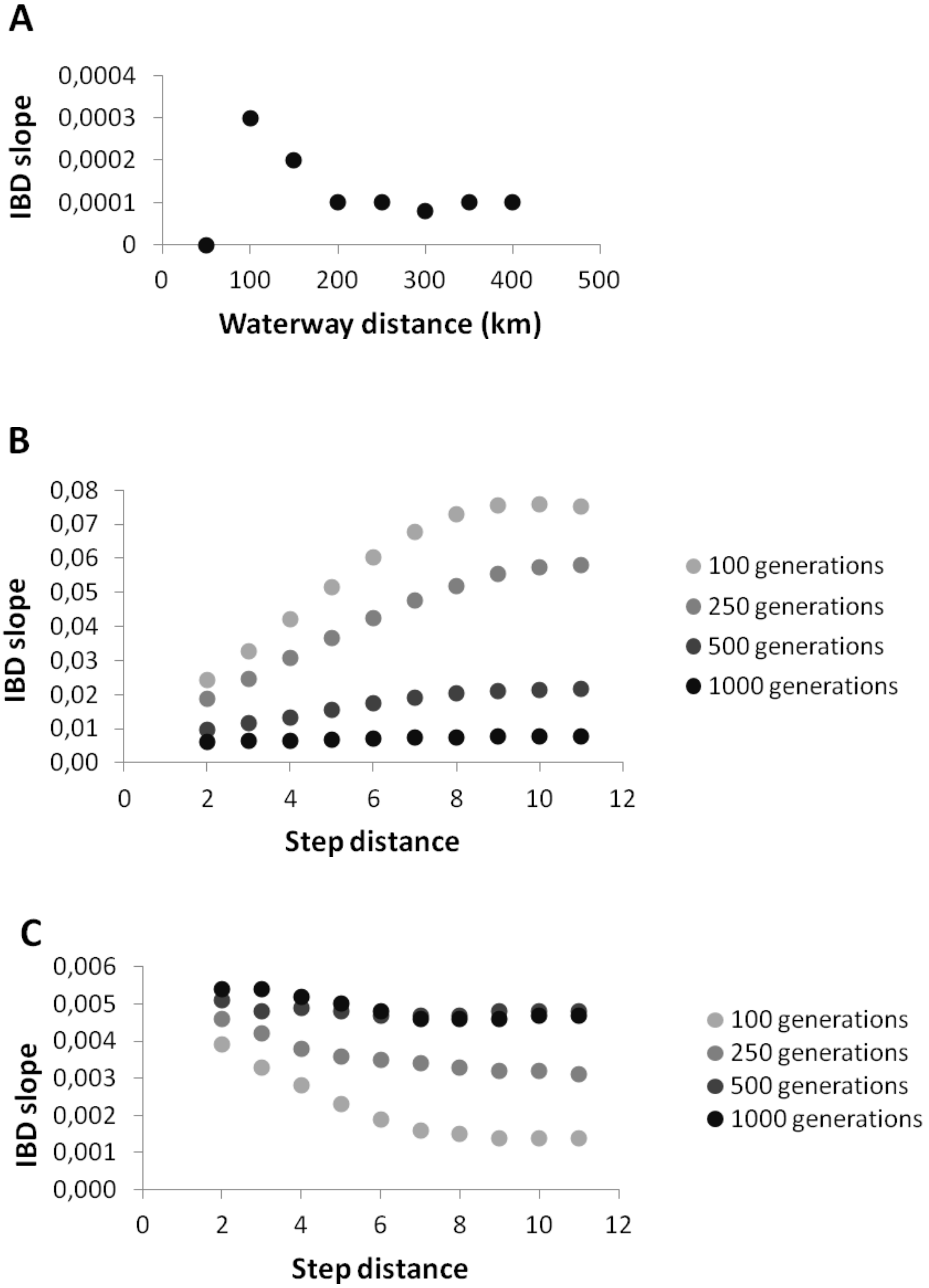


Figure 2. 5. IBD slopes at increasing spatial scales for A) empirical data, B) secondary contact simulations (100, 250, 500, and 1,000 generations) and C) single lineage simulations (100, 250, 500, and 1,000 generations).

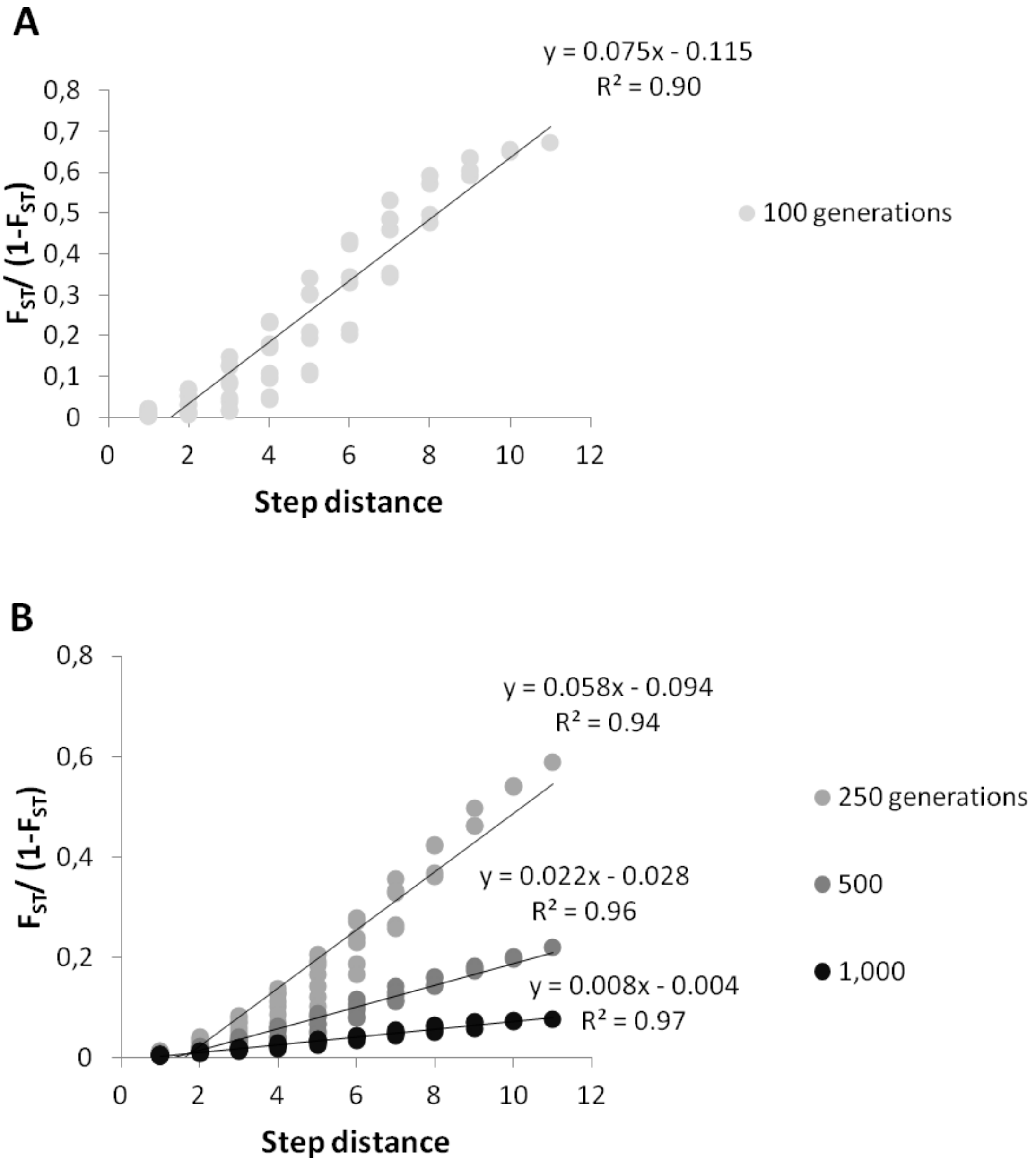
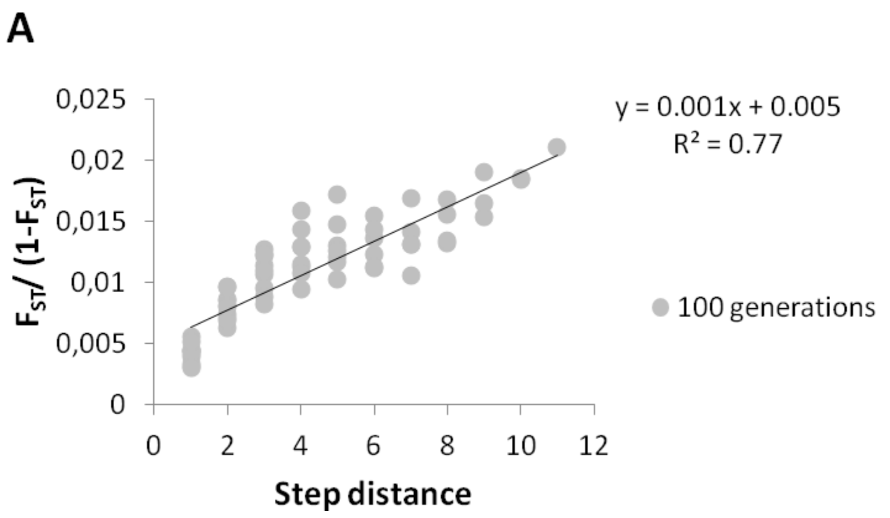


Figure 2. 6. Regression of waterway distance to $F_{ST}/(1-F_{ST})$ for data simulated by modeling A) 100 generations and B) 250, 500, and 1,000 generations of secondary contact.



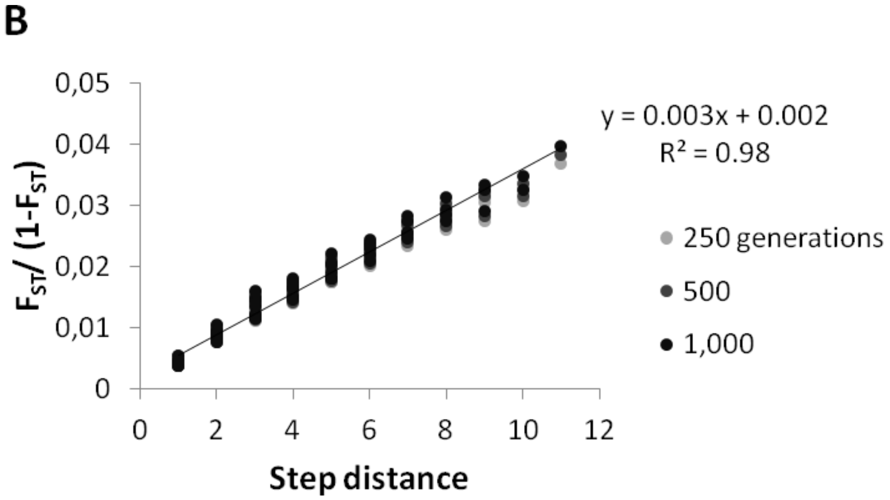


Figure 2. 7. Regression of waterway distance to $F_{ST}/(1-F_{ST})$ for data simulated by modeling A) 100 generations and B) 250, 500, and 1,000 generations of one-dimensional stepping stone migration by a single lineage.

REFERENCES

Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N (2008) Genetic effects of harvest on wild animal populations. Trends in Ecology & Evolution 23 (6):327-337.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nature Reviews Genetics 11 (10):697-709.

Balloux F (2001) Easypop (Version 1.7): A computer program for population genetics simulations. Journal of Heredity 92 (3):301-302.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annual Review of Ecology and Systematics 16:113-148.

Beacham TD, Candy JR, Wallace C, Urawa S, Sato S, Varnavskaya NV, Le KD, Wetklo M (2009) Microsatellite stock identification of chum salmon on a Pacific Rim basis. North American Journal of Fisheries Management 29 (6):1757-1776.

Beacham TD, Murray CB (1987) Adaptive variation in body size, age, morphology, egg size, and developmental biology of chum salmon (*Oncorhynchus keta*) in British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 44 (2):244-261.

Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. Evolution 45 (4):1016-1035.

Bradbury IR, Bentzen P (2007) Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. Marine Ecology Progress Series 340:245-257.

Brigham-Grette J (2001) New perspectives on Beringian Quaternary paleogeography, stratigraphy, and glacial history. Quaternary Science Reviews 20 (1-3):15-24.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18 (5):249-256.

Burnham KP, Anderson DR (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2 edn. Springer, New York.

Cadrin SX, Friedland KD, Waldman JR (eds) (2005) Stock Identification Methods: Applications in Fishery Science. Elsevier Academic Press, Burlington.

Carmelli D, Cavalli-Sforza LL (1976) Some models of population structure and evolution. Theoretical Population Biology 9 (3):329-359.

Castric V, Bernatchez L (2003) The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchill). Genetics 163 (3):983-996.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM (2009) The Last Glacial Maximum. Science 325 (5941):710-714.

Cunningham KM, Canino MF, Spies IB, Hauser L (2009) Genetic isolation by distance and localized fjord population structure in Pacific cod (*Gadus macrocephalus*): limited effective dispersal in the northeastern Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences 66 (1):153-166.

Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J, Veillette JJ (2002) The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. Quaternary Science Reviews 21 (1-3):9-31.

Elfstrom CM, Smith CT, Seeb LW (2007) Thirty-eight single nucleotide polymorphism markers for high-throughput genotyping of chum salmon. Molecular Ecology Notes 7 (6):1211-1215.

Endler JA (1977) Geographic variation, speciation, and clines Princeton University Press, Princeton.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10 (3):564-567.

Fall JA, Brown C, Turek MF, Braem N, Simon JJ (2007) Alaska subsistence salmon fisheries 2007 annual report. Alaska Department of Fish and Game, http://www.subsistence.adfg.state.ak.us/techpap/TP346.pdf

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. Molecular Ecology 19 (17):3773-3788.

Gisclair BR (2009) Salmon bycatch management in the Bering Sea walleye pollock fishery: threats and opportunities for Western Alaska. In: Krueger CC, Zimmerman CE (eds) Pacific Salmon: Ecology and Management of Western Alaska's Populations. American Fisheries Society, Bethesda, pp 799-816.

Haflinger K, Gruver J, (2009) Rolling hot spot closure areas in the Bering Sea walleye pollock fishery: estimated reduction of salmon bycatch during the 2006 season. In: Krueger CC, Zimmerman CE (eds) Pacific Salmon: Ecology and Management of Western Alaska's Populations. American Fisheries Society, Bethesda, pp 817-826.

Hellberg ME (1994) Relationships between inferred levels of gene flow and geographic distance in a philopatric coral, *Balanophyllia elegans*. Evolution 48 (6):1829-1854.

Hellberg ME (1995) Stepping-stone gene flow in the solitary coral *Balanophyllia elegans*: equilibrium and nonequilibrium at different spatial scales. Marine Biology 123 (3):573-581.

Hendry AP, Day T (2005) Population structure attributable to reproductive time: isolation by time and adaptation by time. Molecular Ecology 14 (4):901-916.

Hewitt G (2011) Quaternary phylogeography: the roots of hybrid zones. Genetica 139 (5):617-638.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 359 (1442):183-195.

Hoffman JI, Dasmahapatra KK, Amos W, Phillips CD, Gelatt TS, Bickham JW (2009) Contrasting patterns of genetic diversity at three different genetic markers in a marine mammal metapopulation. Molecular Ecology 18 (14):2961-2978.

Humphries EM, Winker K (2011) Discord reigns among nuclear, mitochondrial and phenotypic estimates of divergence in nine lineages of trans-Beringian birds. Molecular Ecology 20 (3):573-583.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53 (6):1898-1914.

Johnson OW, Grant WS, Kope RG, Neely K, Waknitz FW, Waples RS (1997) Status review of chum salmon from Washington, Oregon, and California. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, http://www.nwfsc.noaa.gov/assets/25/4235_06172004_120642_chum.pdf

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24 (11):1403-1405.

Jordan J (2001) Late Quaternary sea level change in Southern Beringia: postglacial emergence of the Western Alaska Peninsula. Quaternary Science Reviews 20 (1–3):509-523.

Karlstrom TNV, Ball G (1969) The Kodiak Island Refugium. Ryerson Press, Boreal Institute, University of Alberta, Calgary.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49 (4):561-576.

Koizumi I, Yamamoto S, Maekawa K (2006) Decomposed pairwise regression analysis of genetic and geographic distances reveals a metapopulation structure of stream-dwelling Dolly Varden charr. Molecular Ecology 15 (11):3175-3189.

Laurent L, Casale P, Bradai MN, Godley BJ, Broderick GG, A C, Schroth W, Schierwater B, Levy AM, Freggi D, Abd El-Mawla EM, Hadoud DA, Gomati HE, Domingo M, Hadjichristophorou M, Kornaraky L, Demirayak F, Gautier CH (1998) Molecular resolution of marine turtle stock composition in fishery bycatch: a case study in the Mediterranean. Molecular Ecology 7 (11):1529-1542.

Lin J, Quinn TP, Hilborn R, Hauser L (2008) Fine-scale differentiation between sockeye salmon ecotypes and the effect of phenotype on straying. Heredity 101 (4):341-350.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18 (4):189-197.

Mann DH, Hamilton TD (1995) Late Pleistocene and Holocene paleoenvironments of the North Pacific coast. Quaternary Science Reviews 14 (5):449-471.

Mann DH, Peteet DM (1994) Extent and timing of the Last Glacial Maximum in southwestern Alaska. Quaternary Research 42 (2):136-148.

Maschner HDG (1999) Prologue to the prehistory of the lower Alaska Peninsula. Arctic Anthropology 36 (1/2):84-102.

McGlauflin MT, Schindler DE, Seeb LW, Smith CT, Habicht C, Seeb JE (2011) Spawning habitat and geography influence population structure and juvenile migration timing of sockeye salmon in the Wood River lakes, Alaska. Transactions of the American Fisheries Society 140 (3):763-782.

Murphy RL, Hartill TG (2009) The north Alaska Peninsula salmon report to the Alaska Board of Fisheries, 2010. Alaska Department of Fish and Game, http://www.adfg.alaska.gov/FedAidPDFs/FMR09-53.pdf

Murray CB, McPhail JD (1988) Effect of incubation temperature on the development of five species of Pacific salmon (*Oncorhynchus*) embryos and alevins. Canadian Journal of Zoology 66 (1):266-273.

Olsen JB, Flannery BG, Beacham TD, Bromaghin JF, Crane PA, Lean CF, Dunmall KM, Wenburg JK (2008) The influence of hydrographic structure and seasonal run timing on genetic diversity and isolation-by-distance in chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 65 (9):2026-2042.

Palsbøll PJ, Bérubé M, Allendorf FW (2007) Identification of management units using population genetic data. Trends in Ecology & Evolution 22 (1):11-16.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications 13 (1):S146-S158.

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

Pella JJ, Milner GB (1987) Use of genetic marks in stock composition analysis. In: Ryman N, Utter FM (eds) Population Genetics and Fishery Management. University of Washington Press, Seattle, pp 247-276.

Pewé TL (1975) Quaternary Geology of Alaska. U.S. Geological Survey Professional Paper, 835. U.S. Government Printing Office, Washington, D.C.

Quinn TP (2005) The Ecology and Behavior of Pacific Salmon and Trout. University of Washington Press, Seattle.

Ramstad KM, Woody CA, Sage GK, Allendorf FW (2004) Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. Molecular Ecology 13 (2):277-290.

Raymond M, Rousset F (1995) Genepop version 1.2: population genetics software for exact tests and ecumenicism Journal of Heredity 86: 248-249.

Redenbach Z, Taylor EB (2002) Evidence for historical introgression along a contact zone between two species of char (PISCES:SALMONIDAE) in northwestern North America Evolution 56 (5):1021-1035.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-Statistics under isolation by distance. Genetics 145 (4):1219-1228.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources 8:103-106.

Ryman N (2006) Chifish: a computer program testing for genetic heterogeneity at multiple loci using chi-square and Fisher's exact test. Molecular Ecology Notes 6 (1):285-287.

Salo EO (1991) Life history of chum salmon (*Oncorhynchus keta*). In: Groot C, Margolis L (eds) Pacific Salmon Life Histories. University of British Columbia Press, Vancouver, pp 231–309.

Sato S, Kojima H, Ando J, Ando H, Wilmot R, Seeb L, Efremov V, LeClair L, Buchholz W, Jin D-H, Urawa S, Kaeriyama M, Urano A, Abe S (2004) Genetic population structure of chum salmon in the Pacific rim inferred from mitochondrial DNA sequence variation. Environmental Biology of Fishes 69 (1):37-50.

Sawyer S, Felsenstein J (1983) Isolation by distance in a hierarchically clustered population. Journal of Applied Probability 20 (1):1-10.

Schwartz M, McKelvey K (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conservation Genetics 10 (2):441-452.

Seeb JE, Pascal CE, Ramakrishnan R, Seeb LW (2009) SNP genotyping by the 5'-nuclease reaction: advances in high throughput genotyping with non-model organisms. In: Komar A (ed) Methods in Molecular Biology, Single Nucleotide Polymorphisms. 2 edn. Humana Press, New York, pp 277-292.

Seeb LW, Crane PA (1999) High genetic heterogeneity in chum salmon in Western Alaska, the contact zone between northern and southern lineages. Transactions of the American Fisheries Society 128 (1):58-87.

Seeb LW, Templin WD, Sato S, Abe S, Warheit K, Park JY, Seeb JE (2011) Single nucleotide polymorphisms across a species’ range: implications for conservation studies of Pacific salmon. Molecular Ecology Resources 11:195-217.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47 (1):264-279.

Smith CT, Baker J, Park L, Seeb LW, Elfstrom C, Abe S, Seeb JE (2005a) Characterization of 13 single nucleotide polymorphism markers for chum salmon. Molecular Ecology Notes 5 (2):259-262.

Smith CT, Elfstrom CM, Seeb LW, Seeb JE (2005b) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14 (13):4193-4203.

Smith CT, Nelson RJ, Wood CC, Koop BF (2001) Glacial biogeography of North American coho salmon (*Oncorhynchus kisutch*). Molecular Ecology 10 (12):2775-2785.

Smith CT, Seeb LW (2008) Number of alleles as a predictor of the relative assignment accuracy of short tandem repeat (STR) and single-nucleotide-polymorphism (SNP) baselines for chum salmon. Transactions of the American Fisheries Society 137 (3):751-762.

Swenson Nathan G, Howard Daniel J (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. The American Naturalist 166 (5):581-591.

Tallman RF, Healey MC (1994) Homing, straying, and gene flow among seasonally separated populations of chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 51 (3):577-588.

Tamkee P, Parkinson E, Taylor EB (2010) The influence of Wisconsinan glaciation and contemporary stream hydrology on microsatellite DNA variation in rainbow trout (*Oncorhynchus mykiss*). Canadian Journal of Fisheries and Aquatic Sciences 67 (6):919-935.

Taylor BL, Dizon AE (1999) First policy then science: why a management unit based solely on genetic criteria cannot work. Molecular Ecology 8:S11-S16.

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

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology 21 (16):3907-3930.

Utter F, Ryman N (1993) Genetic markers and mixed stock fisheries. Fisheries 18 (8):11-21.

Waples RS (1990) Temporal changes of allele frequency in Pacific salmon: implications for mixed-stock fishery analysis. Canadian Journal of Fisheries and Aquatic Sciences 47 (5):968-976.

Waples RS, Do C (2008) Ldne: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8 (4):753-756.

Weir B, Cockerham C (1984) Estimating F-Statistics for the analysis of population structure. Evolution 38 (6):1358-1370.

Wilmot RL, Everett RJ, Spearman WJ, Baccus R, Varnavskaya NV, Putivkin SV (1994) Genetic stock structure of western Alaska chum salmon and a comparison with Russian Far East stocks. Canadian Journal of Fisheries and Aquatic Sciences 51 (S1):84-94.

Witt JDS, Zemlak RJ, Taylor EB (2011) Phylogeography and the origins of range disjunctions in a north temperate fish, the pygmy whitefish (*Prosopium coulterii*), inferred from mitochondrial and nuclear DNA sequence analysis. Journal of Biogeography 38 (8):1557-1569.

Wolfe RJ (2004) Local traditions and subsistence: a synopsis from twenty-five years of research by the state of Alaska. Alaska Department of Fish and Game, http://www.adfg.alaska.gov/techpap/Tp284.pdf

Wright S (1943) Isolation by distance. Genetics 28 (2):114-138.

Zimmerman CE, Finn JE (2012) A simple method for in situ monitoring of water temperature in substrates used by spawning salmonids. Journal of Fish and Wildlife Management 3 (2):1-8.

APPENDICES

Supplementary Table 1.1. SNP loci used in the first chapter. Includes information on F_{ST}, observed heterozygosity (H_o) over all populations sampled, and the citation in which the SNP was originally described. SNPs occur at codominant loci, unless marked with the subscript $_{mtDNA}$, indicating that they are found in mitochondrial DNA.

Locus	F_{ST}	H_o	Citation Key
Oke_ACOT-100	0.0339	0.460	5
Oke_AhR1-78	0.0507	0.468	3
Oke_arf-319	0.0339	0.446	2
Oke_ATP5L-105	0.0349	0.466	5
Oke_azin1-90	0.0297	0.367	5
Oke_brd2-118	0.0384	0.361	4
Oke_brp16-65	0.0253	0.465	4
Oke_CATB-60	0.0666	0.117	5
Oke_ccd16-77	0.0475	0.477	4
Oke_CD81-108	0.0448	0.239	5
Oke_CD81-173	0.0842	0.393	5
Oke_CKS1-94	0.0235	0.385	5
Oke_CKS-389	0.0664	0.431	1
Oke_Cr30$_{mtDNA}$			1
Oke_Cr386$_{mtDNA}$			1
Oke_ctgf-105	0.0198	0.146	3
Oke_DCXR-87	0.0339	0.235	5
Oke_e2ig5-50	0.0301	0.469	5
Oke_eif4g1-43	0.1025	0.377	5
Oke_f5-71	0.0485	0.429	5
Oke_FANK1-166	0.0349	0.391	5
Oke_FBXL5-61	0.0369	0.327	5
Oke_gdh1-191	0.0506	0.481	5
Oke_gdh1-62	0.0626	0.441	5
Oke_GHII-3129	0.0216	0.133	3
Oke_glrx1-78	0.019	0.402	5
Oke_GPDH-191	0.0207	0.399	2
Oke_GPH-105	0.0431	0.473	3
Oke_HP-182	0.0338	0.410	3
Oke_il-1racp-67	0.0231	0.222	2
Oke_IL8r2-406	0.0254	0.330	1
Oke_KPNA2-87	0.0358	0.234	3
Oke_LAMP2-186	0.0473	0.471	5
Oke_mgll-49	0.0375	0.466	5
Oke_MLRN-63	0.0261	0.472	5
Oke_Moesin-160	0.0126	0.206	2
Oke_nc2b-148	0.0356	0.385	5
Oke_ND3-69$_{mtDNA}$			1

Locus	F_{ST}	H_o	Citation Key
Oke_NUPR1-70	0.0272	0.304	5
Oke_pgap-111	0.0558	0.443	5
Oke_pgap-92	0.0853	0.338	5
Oke_PPA2-635	0.0709	0.283	3
Oke_psmd9-57	0.0208	0.195	5
Oke_rab5a-117	0.0714	0.416	5
Oke_ras1-249	0.0486	0.478	3
Oke_RFC2-618	0.0711	0.029	2
Oke_RH1op-245	0.0275	0.292	2
Oke_RS27-81	0.0142	0.258	5
Oke_RSPRY1-106	0.0576	0.147	4
Oke_serpin-140	0.034	0.410	2
Oke_slc1a3a-86	0.0545	0.469	5
Oke_sylc-90	0.0452	0.448	5
Oke_TCP1-78	0.0241	0.152	3
Oke_Tf-278	0.0371	0.428	3
Oke_thic-84	0.0507	0.478	5
Oke_U200-385	0.0183	0.484	2
Oke_U212-87	0.0541	0.130	2
Oke_u217-172	0.0184	0.452	2
Oke_U302-195	0.0757	0.174	3
Oke_U502-241	0.0162	0.197	3
Oke_U504-228	0.0212	0.331	3
Oke_U506-110	0.0671	0.464	3
Oke_U507-286	0.023	0.490	3
Oke_U509-219	0.0569	0.404	3
Oke_U1002-262	0.0403	0.472	4
Oke_U1008-83	0.0271	0.158	4
Oke_U1010-251	0.0459	0.313	4
Oke_U1012-241	0.0304	0.485	4
Oke_U1015-255	0.0328	0.435	4
Oke_U1016-154	0.0286	0.446	4
Oke_U1017-52	0.014	0.488	4
Oke_U1018-50	0.0113	0.064	4
Oke_U1021-102	0.0288	0.305	4
Oke_U1022-139	0.0316	0.419	4
Oke_U1023-147	0.0338	0.474	4
Oke_U1024-113	0.0268	0.213	4
Oke_U1025-135	0.0403	0.047	4
Oke_U2006-109	0.017	0.451	5
Oke_U2007-190	0.0443	0.478	5
Oke_U2011-107	0.024	0.153	5
Oke_U2015-151	0.0248	0.078	5
Oke_U2025-86	0.0237	0.480	5
Oke_U2029-79	0.0386	0.476	5
Oke_U2031-37	0.0379	0.106	5
Oke_U2032-74	0.0349	0.166	5

Locus	F_{ST}	H_o	Citation Key
Oke_U2034-55	0.0949	0.447	5
Oke_U2035-54	0.0344	0.039	5
Oke_U2037-76	0.0351	0.088	5
Oke_U2041-84	0.0177	0.428	5
Oke_U2043-51	0.025	0.274	5
Oke_U2048-91	0.031	0.441	5
Oke_U2050-101	0.0227	0.243	5
Oke_U2053-60	0.0274	0.488	5
Oke_U2054-58	0.0368	0.209	5
Oke_U2056-90	0.0383	0.475	5
Oke_U2057-80	0.0237	0.458	5

Supplementary Table 2.1. SNP loci used in the second chapter. Several loci were removed from analyses of the data because of linkage disequilibrium (LD), or very low minor allele frequency.

SNP	Global F_{ST}	Reason for removal from dataset	Citation Key
Oke_ACOT-100	0.0425		5
Oke_AhR1-278	0.0102		3
Oke_AhR1-78	0.0631		3
Oke_APOB-60	0.075		5
Oke_arf-319	0.053		2
Oke_ATP5L-105	0.0303		5
Oke_ATP5L-248		in LD with *Oke_ATP5L-105*	5
Oke_azin1-90	0.0311		5
Oke_brd2-118	0.0366		4
Oke_brp16-65	0.0092		4
Oke_CATB-60	0.0977		5
Oke_ccd16-77	0.1025		4
Oke_CD123-62	0.1101		5
Oke_CD81-108	0.0559		5
Oke_CD81-173	0.057		5
Oke_cjo57-86	0.0291		5
Oke_CKS1-70	0.0389		5
Oke_CKS1-94	0.0149		5
Oke_CKS-389	0.0282		1
Oke_CO1A1-72	0.006		5
Oke_CO1A1-76	0.0034		5
Oke_col1a2-62	0.0564		5
Oke_Cr30		monomorphic mitochondrial	1
Oke_Cr386		monomorphic mitochondrial	1
Oke_ctgf-105	0.0189		3
Oke_CTR2-82	0.0214		5
Oke_DBLOH-79	0.0189		5
Oke_DCXR-87	0.0294		5
Oke_DM20-548	0.024		1
Oke_e2ig5-50	0.0219		5
Oke_EF2-394	0.0608		5
Oke_eif4ebp2-64	0.0074		2
Oke_eif4g1-43	0.0343		5
Oke_f5-71	0.0264		5
Oke_FANK1-166		in LD with *Oke_FANK1-96*	5
Oke_FANK1-96	0.1411		5
Oke_FBXL5-61	0.0339		5
Oke_gdh1-191		in LD with *Oke_gdh1-234*	5
Oke_gdh1-234	0.0338		5
Oke_gdh1-62		in LD with *Oke_gdh1-234*	5
Oke_GHII-2943	0.0283		3
Oke_GHII-3129		in LD with *Oke_GHII-2943*	3

SNP	**Global F_{ST}**	**Reason for removal from dataset**	**Citation Key**
Oke_glrx1-78	0.0123		5
Oke_GNMT-100	0.0611		5
Oke_GnRH-373	0.0468		1
Oke_GnRH-527	0.0859		1
Oke_GPDH-191	0.0061		2
Oke_GPH-105	0.0191		3
Oke_GPH-78	0.0559		3
Oke_H2AX-72	0.039		5
Oke_hmgb1-66	0.0142		5
Oke_hnRNPL-239		average minor allele frequency < 0.01	3
Oke_HP-182	0.0886		3
Oke_hsc71-199	0.0682		2
Oke_HSP90BA-299		average minor allele frequency < 0.01	3
Oke_il-1racp-67	0.0119		2
Oke_IL8r2-406	0.0183		1
Oke_IL8r-272		in LD with *Oke_IL8r2-406*	1
Oke_KPNA2-87	0.0174		3
Oke_lactb2-71	0.0145		5
Oke_lamp2-138	0.0581		5
Oke_LAMP2-186	0.0255		5
Oke_mcfd2-86	0.0288		5
Oke_METK2-97	0.0335		5
Oke_mgll-49	0.0286		5
Oke_MLRN-63	0.0111		5
Oke_Moesin-160	0.0035		2
Oke_nc2b-148	0.022		5
Oke_ND3-69		monomorphic mitochondrial	1
Oke_ndub3-58	0.0031		5
Oke_NHERF-123		in LD with *Oke_NHERF-54*	5
Oke_NHERF-54	0.0706		5
Oke_NUPR1-70	0.0601		5
Oke_PDIA3-475		in LD with *Oke_PDIA3-82*	5
Oke_PDIA3-82	0.0207		5
Oke_pgap-111		in LD with *Oke_pgap-92*	5
Oke_pgap-92	0.0525		5
Oke_pnrc2-78	0.0087		5
Oke_PPA2-635	0.1354		3
Oke_psmd9-188	0.006		5
Oke_psmd9-57	0.0141		5
Oke_rab5a-117	0.0481		5
Oke_ras1-249	0.0507		3
Oke_RFC2-618	0.0632		2
Oke_RH1op-245	0.0269		2
Oke_ROA1-209	0.1361		5
Oke_RPN1-80	0.0321		5
Oke_RS27-81	0.0174		5
Oke_RS27-94	0.0325		5

SNP	Global F_{ST}	Reason for removal from dataset	Citation Key
Oke_RS9-379		average minor allele frequency < 0.01	5
Oke_RSPRY1-106	0.1217		4
Oke_serpin-140	0.0496		2
Oke_slc1a3a-86	0.0349		5
Oke_sylc-90	0.0221		5
Oke_TCP1-78	0.0222		3
Oke_TCTA-202		in LD with *Oke_TCTA-99*	5
Oke_TCTA-99	0.0526		5
Oke_Tf-278	0.0633		3
Oke_thic-84	0.0438		5
Oke_txnrd1-74	0.0181		5
Oke_u0602-244		average minor allele frequency < 0.01	4
Oke_U1001-79		in LD with *Oke_DUBLOH-79*	4
Oke_U1002-165		in LD with *Oke_U1002-262*	4
Oke_U1002-262	0.0523		4
Oke_U1008-83	0.0089		4
Oke_U1010-154		average minor allele frequency < 0.01	4
Oke_U1010-251		average minor allele frequency < 0.01	4
Oke_U1012-241	0.0089		4
Oke_U1012-60		in LD with *Oke_U1012-241*	4
Oke_U1015-255	0.0226		4
Oke_U1016-154	0.04		4
Oke_U1017-52	0.0083		4
Oke_U1018-50		average minor allele frequency < 0.01	4
Oke_U1019-218	0.0111		4
Oke_U1020-75	0.0136		4
Oke_U1021-102	0.0253		4
Oke_U1022-114		average minor allele frequency < 0.01	4
Oke_U1022-139	0.0099		4
Oke_U1023-147	0.0155		4
Oke_U1024-113	0.0186		4
Oke_U1025-135	0.0491		4
Oke_U1027-89	0.0185		4
Oke_U1028-100	0.0345		4
Oke_U1031-132	0.0154		4
Oke_U1103-150	0.0348		5
Oke_u1-519	0.0107		1
Oke_U2001-629		average minor allele frequency < 0.01	5
Oke_U2002-200	0.0328		5
Oke_U2003-142	0.0291		5
Oke_u200-385	0.0492		2
Oke_U2005-62	0.0087		5
Oke_U2006-109	0.0023		5
Oke_U2007-190	0.0614		5
Oke_U2010-94	0.0311		5
Oke_U2011-107	0.0227		5
Oke_U2015-151	0.019		5

SNP	Global F_{ST}	Reason for removal from dataset	Citation Key
Oke_U2016-118	0.0139		5
Oke_U2017-87	0.0028		5
Oke_U2019-112	0.0309		5
Oke_U2020-51	0.0326		5
Oke_u202-131		average minor allele frequency < 0.01	2
Oke_U2021-86	0.0334		5
Oke_U2022-101	0.0178		5
Oke_U2023-99	0.0589		5
Oke_U2024-93	0.0072		5
Oke_U2025-86	0.0208		5
Oke_U2026-64	0.039		5
Oke_U2029-79	0.0096		5
Oke_U2031-37	0.0348		5
Oke_U2032-74	0.0152		5
Oke_U2033-122	0.0456		5
Oke_U2034-55	0.0282		5
Oke_U2035-54	0.0184		5
Oke_U2037-76	0.0169		5
Oke_U2038-32	0.0706		5
Oke_U2040-77	0.0158		5
Oke_U2041-84	0.0129		5
Oke_U2042-61	0.0467		5
Oke_U2043-51	0.0146		5
Oke_U2045-43	0.0173		5
Oke_U2047-49	0.037		5
Oke_U2048-91	0.0357		5
Oke_U2049-99	0.027		5
Oke_U2050-101	0.0173		5
Oke_U2052-56	0.0244		5
Oke_U2053-60	0.0469		5
Oke_U2054-58	0.0248		5
Oke_U2056-90	0.0533		5
Oke_U2057-80	0.0308		5
Oke_U212-87	0.0139		2
Oke_u216-222	0.0178		2
Oke_u217-172	0.0071		2
Oke_U302-195	0.0239		3
Oke_U305-130	0.0845		3
Oke_U305-307		in LD with *Oke_U305-130*	3
Oke_U401-143	0.0331		3
Oke_U401-220		in LD with *Oke_ U401-143*	3
Oke_U502-241		average minor allele frequency < 0.01	3
Oke_U503-272		average minor allele frequency < 0.01	3
Oke_U504-228	0.0463		3
Oke_U505-112	0.0102		3
Oke_U506-110	0.0034		3
Oke_U507-286	0.0253		3

SNP	**Global F_{ST}**	**Reason for removal from dataset**	**Citation Key**
Oke_U507-87		in LD with *Oke_U507-286*	3
Oke_U509-219	0.113		3
Oke_U510-204	0.0171		3
Oke_U511-271	0.0234		3
Oke_U514-150	0.0616		3
Oke_UBA3-245	0.0212		4
Oke_uqcrfs-69		average minor allele frequency < 0.01	4
Oke_XBP1-82	0.0261		5
Oke_zn593-152	0.0007		5

Citation Key

1. Smith CT, Baker J, *et al*. (2005)
2. Smith CT, Elfstrom CM, *et al*. (2005)
3. Elfstrom CM, Smith CT, Seeb LW (2007)
4. Seeb JE, Pascal CE, *et al*. (2011)
5. Chapter 2, this thesis.

Supplementary Table 2.2. Sequence information for SNP loci developed in the second chapter.

Assay Name	Gene	SNP (VIC/FAM)	Forward Primer Sequence	Reverse Primer Sequence
Oke_ACOT-100	Similar to acyl-coenzyme A thioesterase 3	C/G	TCAGGGACGATAAAGGGATCATCTT	GGGAGAGACACAGGTCTACCT
Oke_APOB-60	Apolipoprotein B	C/T	CTGTGGATGGTATTCTGGATGCA	GGCACTACAAAAGAGGGAATCTCA
Oke_ATP5L-105	ATP synthase subunit g, mitochondrial putative	C/G	GTGCACACCAATCCATTTCTGAAT	TGTTTTAAGGTGTGACTTGCTGGTA
Oke_ATP5L-248	ATP synthase subunit g, mitochondrial putative	A/T	CTAGTGGATTGTGGCTTACGTCAA	GGATTCTGACTGTGGGTGTTTAACA
Oke_azin1-90	Antizyme inhibitor 1 (azin1)	C/T	GGGAATAGTGTCATTTGGGATGCAT	GGTGAATGATATTCTGTAGTCATATTGCTTGA
Oke_CATB-60	Cathepsin B precursor putative	C/T	GCTTCTATGGGTCCTACTACCGTAT	GCACTCCTCTTACACAACTCTGA
Oke_CD123-62	Cell division cycle protein 123 homolog, putative	A/G	GAACAGCAGTGAATCGGTTACCT	TTGACGCTGTGTGTCTTCGA
Oke_CD81-108	CD81 antigen putative	G/T	CAGTATCATCATACAGCACAGATACAACA	GCCTGCTTGTTATACTGACAGTCAA
Oke_CD81-173	CD81 antigen putative	A/C	GATGACTGGAGTCAGCTTGCA	TTTCCTTGGCTCATCTTGCTGTA
Oke_cjo57-86	CJ057 protein	A/C	CAGAAGGTCTAAAGGTCCTTAACAATCA	CAGATTGAACAGTGCCCAGAGA
Oke_CKS1-70	Cyclin-dependent kinases regulatory subunit 1	G/A	GCTACCTCTATCATACCGCCAATATT	ATGTACATCTCTCCATGTGTTTTGGT
Oke_CKS1-94	Cyclin-dependent kinases regulatory subunit 1, putative	G/T	TCTTCGACATGTTTAATCGAACAGAAGT	CCAGCTTTCGCTTGTCAAAACG
Oke_CO1A1-72	Collagen alpha-1I chain precursor, putative	G/A	CTGTCAATGGAGGGTGATCGAAAT	CACCTAAGGTCCTGCAAGCA
Oke_CO1A1-76	Collagen alpha-1I chain precursor, putative	A/T	GAGGGTGATCGAAATTGTGCTACTA	ACGACTGACATCAAGATAAATCTGCTAATTT
Oke_col1a2-62	Collagen alpha-2I chain precursor, putative	G/A	GCAGGAAACCACTCTCATTCTTACT	AGACTTAGGAAATTGCACCTGCTTA
Oke_CTR2-82	Low affinity cationic amino	C/T	GCAGCAGACAC	CCATTCCCATCGGCA

Assay Name	**Gene**	**SNP (VIC/FAM)**	**Forward Primer Sequence**	**Reverse Primer Sequence**
	acid transporter 2		ACCGAAGTA	TCGT
Oke_DBLOH-79	Diablo homolog, mitochondrial precursor, putative	C/T	GCAGATATGCCTCAGGGATGT	GACAGTCAAAGGATCAAGCTACCT
Oke_DCXR-87	L-xylulose reductase, putative	A/T	GTCACCCAGAACAATAGAATGAGTCT	TCTAACACACCCACAATCTGCAAAA
Oke_e2ig5-50	E2-induced gene 5 protein homolog (e2ig5)	C/T	GCACTGCTCATTCTGTCACATG	GGGAGTTCTTAGTGTGACCATAGAG
Oke_EF2-394	Elongation factor 2	C/T	GCTTAACTGCTGTTTCTGCTATAGG	GCAGTCTCCTCCTTCTTGAAGTT
Oke_eif4g1-43	Eukaryotic translation initiation factor 4 gamma 1, putative	G/T	GCACCCAACAGTTCATCATGTAAGT	CCACCCCCAGTAGTCAATCCT
Oke_f5-71	Coagulation factor V precursor, putative	C/T	CTCAAATTTCCCTTTGACATCAATTCATCA	GATCCTCATGCACATCCAAACG
Oke_FANK1-166	Fibronectin type 3 and ankyrin repeat domains protein 1	C/T	ACTCACGTGTGGTAGAGACAGA	AGACTGAGAATCACAAGACCAACTG
Oke_FANK1-96	Fibronectin type 3 and ankyrin repeat domains protein 1	A/C	GGCTCACCTGGATGACATTATATAGG	CGTGAGTACACAACACTCTTTCAGT
Oke_FBXL5-61	F-box/LRR-repeat protein 5, putative	G/A	TGGTGTGTAACGTCAGTGACTTAAG	CACCTCTAGAAATGACATGATCAGTGT
Oke_gdh1-191	Glutamate dehydrogenase (gdh1.2 gene)	A/G	GTGGAGACCAAACCCAGTAGAAC	GGGTTAGAGGTCAGGGTTAGAG
Oke_gdh1-234	Glutamate dehydrogenase (gdh1.2 gene)	C/T	CAAACCCAGTAGAACCCTGTGT	CTGGGAATGGTGATATATGTGTTCCT
Oke_gdh1-62	Glutamate dehydrogenase (gdh1.2 gene)	C/T	CCACGTGATACAGGGAGATGTG	CACACACACTGACACGTACTGT
Oke_glrx1-78	Glutaredoxin-1	C/T	CGCTCCGTCCAGTGATGTC	GGCCAAAGAGGTATTGACAAAGTAC
Oke_GNMT-100	Glycine N-methyltransferase	C/T	GCGTCCACGCTCGTCAT	AGCGTGGACTCCATCATGTTG
Oke_H2AX-72	Histone H2A.x	C/T	AGGGCAACTACGCTCATAGAGTA	CGAGGACAGCAGCCATGTA
Oke_hmgb1-66	high-mobility group box 1 (hmgb1)	G/T	GGAAACAGAATAACTACTAAGACCCTACATTATAAC	ACGCCCCCATTGGAAACC
Oke_lactb2-71	Beta-lactamase-like protein 2	G/A	CGTCGTGAACCATGAGTGCAATA	TTCGCACAACTCTGGACGATAG

Assay Name	Gene	SNP (VIC/FAM)	Forward Primer Sequence	Reverse Primer Sequence
Oke_lamp2-138	Lysosome-associated membrane glycoprotein 2 precursor, putative	G/A	GCATGTTACAGGACGGCAAGA	GGCTGGTATCACTGTGACATTCA
Oke_LAMP2-186	Lysosome-associated membrane glycoprotein 2 precursor, putative	A/G	TTCTAGCCATGACCCAATGAAAGG	AACTGCTCCAAATGCTGGTTAGTA
Oke_mcfd2-86	Multiple coagulation factor deficiency 2 (mcfd2)	C/T	GGCTTGAGGGCCACATTG	GTCAAAACAAAATCTGTGCAACCCT
Oke_METK2-97	S-adenosylmethionine synthetase isoform type-2, putative	C/T	CCAGGACGAAGGTCAAAGTTCTT	GGCACATCCCAGAAGAGTGA
Oke_mgll-49	Monoglyceride lipase	A/T	ACATTGTAATCTGTATTAGTCCAATGCAGAC	GGTACCACCTGCAACATCAAC
Oke_MLRN-63	Myosin regulatory light chain 2, smooth muscle isoform, putative	G/A	CCATTTCAGCATTGCCAGATTTGAAA	GATGTCACAGACCAGTACCATGTTT
Oke_nc2b-148	Down-regulator of transcription 1, Dr1 (nc2b), mRNA	A/C	CCAGCCTATTTCCTTTAGTGCATATGA	GCACCCTATTCCCTACATGGT
Oke_ndub3-58	NADH dehydrogenase 1 beta subcomplex subunit 3 (ndub3)	C/T	GAGGCTTCAGTCGTCCTGTATC	CAGCGAAGCCCCATTTAAAGC
Oke_NHERF-123	Ezrin-radixin-moesin-binding phosphoprotein 50, putative	G/T	AGGAGTGAGGGCGAGAGAA	GGAACGAACTCTTCAGTAACCT
Oke_NHERF-54	Ezrin-radixin-moesin-binding phosphoprotein 50, putative	G/A	CCCTCAATTAGCACATGAAAATCACA	CTCCTTCTTTTGCTCTCTCTCTCAA
Oke_NUPR1-70	Nuclear protein 1, putative	G/T	AGACGGTGAACTCTGCTGTAGA	TCCCTTCACTGAAGCTACAGTCA
Oke_PDIA3-475	Protein disulfide-isomerase A3	A/G	CCCGGTTCCTCCAGTAGTTG	CTGGTGGCCTACTACGATGTG
Oke_PDIA3-82	Disulfide-isomerase A3 precursor, putative	A/C	TGCCTACGATGGACCCAGAA	ACTCAGCCCATCAGGACAAGA
Oke_pgap-111	Post-GPI attachment to proteins 2 (pgap2)	C/T	TGCAGATCTCAATTTGAACGACCTAT	AGACGACCATTATGGCTAACGTT
Oke_pgap-92	Post-GPI attachment to proteins 2 (pgap2)	C/G	TGCAGATCTCAATTTGAACGACCTAT	AGACGACCATTATGGCTAACGTT
Oke_pnrc2-78	Proline-rich nuclear receptor coactivator 2 (pnrc2)	G/A	CGTGACAGCAGGGAGATGA	CATCTCTAGGCATGCACCTTGA
Oke_psmd9-188	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 9 (psmd9)	C/T	ACTGAGGCAATATTCTGCAGGTT	GGGCTTGCGAATTAGTGATGAAATC
Oke_psmd9-57	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 9 (psmd9)	C/T	ACTGTAGTGACTGCATTTCATATTGCT	ACCAGTTGTATTTTATTGTGCCAAATGAC

Assay Name	**Gene**	**SNP (VIC/FAM)**	**Forward Primer Sequence**	**Reverse Primer Sequence**
Oke_rab5a-117	Ras-related protein Rab-5A	C/T	GGGAATAACAGTCATTGCAGCATTT	CCATTGTTGGAAACTGGACAGC
Oke_ROA1-209	Heterogeneous nuclear ribonucleoprotein A1, putative	G/A	CAGGGTTTGATTGGTTAACTTACATTGAAT	GCTGGATCTCTCATTACCTGTAGGT
Oke_RPN1-80	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 precursor	G/A	CACGCACCTTGCTAAGATAACAG	GGCTCTACCGCCAAGATAAAGTTAT
Oke_RS27-81	40S ribosomal protein S27	G/A	GCAACAAAGTGGACTATCACATTGAA	GCACCCAAGAAAGATTGATCCAGAA
Oke_RS27-94	40S ribosomal protein S27	G/A	CACTTCTAGATCAATCCGCTGTTTC	GCGACTCCAGCCTTGACA
Oke_RS9-379	40S ribosomal protein S9	A/T	GCAATCCTCCATACATTACCTGTCA	GTCTATAAGACTGCCCAGAACCAA
Oke_slc1a3a-86	Solute carrier family 1 (glial high affinity glutamate transporter), member 3a	C/T	TGTCTTCATCTGTGGACTCCTACA	TCACCATGACAACTCACCTAGATGA
Oke_sylc-90	Leucyl-tRNA synthetase, cytoplasmic (sylc)	A/G	TTGAGGAAACCACTGGTCTTACAAG	GCATCCTTCCTACCTTCCTTGAG
Oke_TCTA-202	T-cell leukemia translocation-altered gene protein homolog	A/C	AGTTTAGCACTTACCTTGTCGGT	CAGTCTCATTGCCATCCATTTCG
Oke_TCTA-99	T-cell leukemia translocation-altered gene protein homolog	C/G	GCTAGCACTTCATGGCAGCAT	GCCACCATCATCACGTTTAGTTCTA
Oke_thic-84	Acetyl-CoA acetyltransferase, cytosolic (thic)	C/T	GCTGCTGTCTTAAACCACATTCTACA	GCCTTCCTATTTGTCCTGTTCTCT
Oke_txnrd1-74	Thioredoxin reductase 1, putative	A/G	CACCCCATGTGCCTTTGGA	GACGTCCTGGCCCTTAACC
Oke_U1103-150		C/A	ACCTCTCGACTTATTCCAAATTTTGTTACA	GGTATTGTGTGTAGTGGTTAAAGGG
Oke_U2001-629		C/T	CCCCACTCCTTCTACTCATCCAT	TTAGTACAAATGAACGAGGGTTGGAA
Oke_U2002-200		A/C	CCAGTGTTGTAGAAAACATGTGCTCTA	GCGCTTACGCTTCAATTTGCA
Oke_U2003-142		C/T	CTCCTCACTAACAGTAGCTGCAATT	CCTTGAAAGTGAGATATCTGTAGTTCGT
Oke_U2005-62		A/G	GTACAGCAGAGACTAAAGCTAT	GAGGTCAAGGCTTCAACATCAC

Assay Name	Gene	SNP (VIC/FAM)	Forward Primer Sequence	Reverse Primer Sequence
			ACAACA	
Oke_U2006-109		G/T	CCAACACCACTTTCCATTAATAAGCA	GCACACCCTAATTGACAAACAAACC
Oke_U2007-190		C/G	ACAGGCTGTGATGAGTTAACAATGTAAA	CATGTCGTCTACTTTGATGCCAATT
Oke_U2010-94		C/T	CCGCAGACAGTGGTCAATACT	GCCCTTCTCTTTCTCCATACTTTTCT
Oke_U2011-107		G/T	CCGTTTCTGTCAGACTCTGGTAAA	CTGGAGTGACTCAGGATCATAGC
Oke_U2015-151		C/T	GCATTTTATCCTCAAACTTTTCAACTGACA	ACGAATCCACCTAAAATCCACCAAA
Oke_U2016-118		C/T	ACGTGTCCTGTTCAAATTAGCAGTA	GAGGTGCATGCTTTTGTTCCA
Oke_U2017-87		A/C	CAGGAGCCATTGGAAGAGTAGAG	CCATGATTTGAAAAGAGCTGAACCAT
Oke_U2019-112		A/C	GAATTTGACTGCCTGGCGAAAG	CGTACTGCTGATCCCAAATGATTTT
Oke_U2020-51		C/T	GCTCTGGTCCAGGTCGTT	ACGTTCTTCGCCTCATGTTACATTA
Oke_U2021-86		A/C	TGTGGCTCCAGCCAAAGTT	GCATCCTCAGTTCCAGCATAATGAT
Oke_U2022-101		G/C	TGTCCTAATGACAGGCCTTGC	GTCACTGCAGCCTAACGTTATATTG
Oke_U2023-99		C/T	CACTATTTTGACAAGTGTTAAGATCATTTTGTGT	TGTGATCAACAGTTTTACACTCAATGGA
Oke_U2024-93		C/T	CGTCTTTCCAATACCACAGAGATACA	GAAGTGTTGACCTCTGCTCCTT
Oke_U2025-86		G/A	AAATCCCCATGGAGAAACACAATGA	ATTGTCCTTCCGCAGTGGT
Oke_U2026-64		G/T	CTTCCCACGTCTTTTCTGTCTCA	GCCTCTCACGTTACACTGTCATTTT
Oke_U2029-79		C/T	GGTTTGATTTCGTCGCGATTTGA	AAATCCCAGGGAGCGAAAGTC
Oke_U2031-37		A/T	CACACTTTCAATCAAATGTTGTGCAG	CGTTTGAGACGCCCTTCACT
Oke_U2032-74		G/A	GCTATTCCAATGTAAATCCTGTACTGTGT	AACCCTATCTGCTCATTGGTCATG

Assay Name	Gene	SNP (VIC/FAM)	Forward Primer Sequence	Reverse Primer Sequence
Oke_U2033-122		G/T	ACGCCCTTCCCCGATTC	GGCCTGGGTATGACTCAACATG
Oke_U2034-55		C/T	GGGAAGAAAAGCCTACCATAAACAG	CCCAGAGCGAATGCCAACA
Oke_U2035-54		G/A	CGCCAATAACGCTCCAACAAC	CTTCACACCCTGAGAACTGGTTTTA
Oke_U2037-76		C/G	CATATCAGGTGTGTCTCAACAGTCT	GGCATTCACTACATCACATGACCTT
Oke_U2038-32		C/T	CGACTTCCTGGCGTCATCTTC	GGGTACTTTCATACAGTACAGCTCT
Oke_U2040-77		A/C	GGGCTAGAATTCTACTTGGTGACA	CCTTCCACAGTCTCATTTTGCTCTT
Oke_U2041-84		G/T	CCAGACCATGTGCTTGTTTGTCATA	GTGAATATTTTGGCAAGCCTGTACA
Oke_U2042-61		C/T	GCAATGCACATCTGAAATCTGCTAT	ACTCATTTTCAGATGTTCTTCCTTTGTCT
Oke_U2043-51		G/A	CACAAACCTACTACAGACAGCAGTT	GCCAGCTTTGTAGTCTTGTGGAAA
Oke_U2045-43		G/A	GACCCCGGGTCACCAC	CCAGGTGTGGCAGTGGAT
Oke_U2047-49		G/A	TTAGCTAGTATGTTAGCGTAGCACTT	TGCTGAATTTGAGGAACTAAACGTTAGT
Oke_U2048-91		A/C	AGTTGGGTCTTAAAGATGATCATTTGCT	GGACTCTTGACGGCCATCTTA
Oke_U2049-99		C/T	CATTGTAGCAGAGGGTCAACGATAT	ACACACGGCATTGCAAACTC
Oke_U2050-101		C/T	CTCTGAGTGTCACAATCACATATCGT	GTGTAAACGCATTCGAAGTCCTTT
Oke_U2052-56		C/T	GTGCCATGTTAGCCAAAAAGTTTCA	TCCATGTTAGCAGCGAACGTT
Oke_U2053-60		C/T	TCTGCTTTTGTCGTCTCACCAA	CACACGAGGGTGGACTTAGTT
Oke_U2054-58		C/T	CGTCTCATTCAGCTCTTTGATGTC	TTGGTTCAAGATGACACTGGTGT
Oke_U2056-90		G/T	CCATCACGTCACCATTACACTGT	GACATTAGCTGGCAGTCTGATCA
Oke_U2057-80		A/G	GCAGTTGTCAT	GCCCCTCGTTCATTTT

Assay Name	**Gene**	**SNP (VIC/FAM)**	**Forward Primer Sequence**	**Reverse Primer Sequence**
			GGCAGTAAGG	CAGATG
Oke_XBP1-82	X-box-binding protein 1, putative	C/T	TCTGCTCCGGAGTCTTCTGTAT	AAGGAGAGTGTTTAACAAAATTATACAGGATGT
Oke_zn593-152	Zinc finger protein 593 (zn593)	A/C	GTTTTGAAAAGTTATTTCTCGCGTAGATTAAGA	AACTAGCTAGTTATCTAGTAGTAGCTAAATTAGCT