

©Copyright 2016

Carolyn Tarpey

Contemporary population genetic structure and phylogeography of pink salmon

Carolyn Tarpey

A thesis

submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

University of Washington

2016

Committee:

Lisa Seeb

Jim Seeb

Kerry Naish

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

University of Washington

**Abstract**

Contemporary population genetic structure and phylogeography of pink salmon

Carolyn Tarpey

Chair of the Supervisory Committee:

Lisa Seeb, Research Professor

School of Aquatic and Fishery Sciences

Pink salmon *Oncorhynchus gorbuscha* are the most abundant Pacific salmon, ranging widely across drainages of the North Pacific Ocean and Bering Sea from Japan in Asia to Washington State in North America. The species are unique in that they have an obligate two-year life cycle which leads to reproductively isolated even- and odd-year lineages. Abundance of each lineage varies by location, and relative abundance can shift from one lineage to the other through time, as has been shown in the Asian portion of the range. The existence of distinct lineages of pink salmon provides a unique opportunity to address population and evolutionary genetics questions, as the lineages can serve as replicates across the entire range of the species. For instance,

population structure repeated in parallel within the lineages could help identify historical patterns of isolation due to glacial cover in the region. We used genotyping by sequencing to describe the contemporary population genetic structure of even- and odd-year lineages of pink salmon throughout their range. Our goals twofold: to conduct a population genetic characterization of the lineages to better understand their relationship, and to describe any structure within the lineages with the intent of examining the evolution of any parallel population structure identified. We genotyped 245 individuals and incorporated genotypes from 138 previously studied individuals for 16,681 SNPs; individuals originated from seven pairs of even- and odd-year populations of pink salmon spread across their range from Japan in Asia to the Pacific Northwest of North America. It has been shown that pink salmon exhibit a complex hierarchical population structure: they possess geographic structure due to their broad species range and temporal structure due to the even- and odd-year lineages. Here, hierarchical AMOVAs reveal that a significant amount of the variation in the data was explained by differences between the lineages. The lineages show differences in the distribution of diversity with the even-year lineage populations exhibiting significantly lower observed heterozygosity than those of the odd-year lineage. Within both lineages, the pattern of population structure was consistent. The Asian and northern Alaska populations show little differentiation but differ significantly from populations in southcentral Alaska and the Pacific Northwest. We tested the partitioning of variation between Asian/northern Alaska and the southcentral Alaska/Pacific Northwest within each lineage; the proportion of variation attributed to the difference between these groups was significant in both lineages, but greater in the odd-year lineage than in the even-year. Using the contemporary population structure and known paleoecological information, we inferred patterns of shared historical refugial habitat. Our data suggest that during the last glacial maximum both

lineages likely occupied a northern Beringial refugium as well as a southern North American refugium. These results highlight the influence that historical patterns of habitat availability can have on contemporary population structure, and support the hypothesis of a pre-glacial origin of the lineages.

## Table of Contents

List of Tables .....	viii
List of Figures .....	ix
Acknowledgements .....	x

### Contemporary Population Genetic Structure and Phylogeography of Pink Salmon

1. Introduction .....	1
2. Materials and Methods .....	4
2.1 Samples .....	4
2.2 RAD Sequence Data .....	4
2.3 Initial Filtering and Re-Sequencing .....	5
2.4 Genotyping .....	6
2.5 Summary Statistics and Genetic Diversity Measures .....	7
2.6 Population Structure .....	8
2.7 Neighbor-Joining Tree .....	9
2.8 <i>F</i> -Statistics .....	9
2.9 Principal Component Analysis .....	10
2.10 Analysis of Molecular Variance .....	10
2.11 Isolation by Distance .....	11
3. Results .....	12
3.1 Bioinformatics and Genotyping .....	12
3.2 Diversity within Populations and Comparison between Lineages .....	12
3.3 Diversity among Populations and Comparison between Lineages .....	14
3.3.1 <i>F</i> -Statistics .....	14
3.3.2 Neighbor-Joining Tree .....	15
3.3.3 Principal Component Analysis .....	15
3.3.4 Analysis of Molecular Variance .....	16
3.3.5 Isolation by Distance .....	17
4. Discussion .....	17
4.1 Diversity within Populations and Comparison between Lineages .....	18

4.2	Diversity among Populations and Comparison between Lineages.....	20
5.	Conclusions.....	25
6.	References.....	26
7.	Tables.....	31
8.	Figures.....	36
9.	Supplemental Files.....	43

## List of Tables

<b>Table 1.</b> Pink salmon population location and collection descriptors.....	31
<b>Table 2A.</b> Pink salmon population genetic statistics.....	32
<b>Table 2B.</b> Pink salmon population genetic statistics, continued .....	33
<b>Table 3.</b> $F_{ST}$ estimates for populations .....	34
<b>Table 4.</b> Results of three Analysis of Molecular Variance .....	35

## List of Figures

<b>Figure 1.</b> Map of sample locations with distance measures between locations.....	36
<b>Figure 2.</b> Hierarchical groupings of populations for three AMOVAs .....	37
<b>Figure 3.</b> Boxplot of individual standardized observed heterozygosity per population ...	38
<b>Figure 4.</b> Pairwise comparisons of $F_{ST}$ .....	39
<b>Figure 5.</b> Unrooted Neighbor-Joining tree.....	40
<b>Figure 6.</b> Principal Component Analysis of all populations .....	41
<b>Figure 7.</b> Map depicting glacial cover and exposed land during LGM .....	42

## **Acknowledgements**

There was an immense amount of effort that made this study possible. I would like to thank the North Pacific Anadromous Fish Commission for providing the Asian samples, along with the Alaska Department of Fish and Game, and the Washington Department of Fish and Wildlife for the North American pink salmon samples.

I would like to thank members of the Seeb lab, especially Garrett McKinney, Ryan Waples and Morten Limborg for years of enlightening conversation, advice, and help with bioinformatics, and Carita Pascal for guidance in and out of the lab. I would also like to thank my committee members Kerry Naish, Daniel Schindler and Jim Seeb who inspire me and have taught me to always ask more questions. I would especially like to thank my advisor, Lisa Seeb, who has guided and encouraged me.

Finally, I'm so lucky to have the unending love and support of my family and friends, and I thank them very much.

## 1. Introduction

An enduring goal in evolutionary biology is to describe the dynamics that result in population structure, distribution, and genetic diversity. Groups of organisms that use the same habitat sequentially, separated in time (allochronic species, Alexander & Bigelow (1960)), are an intriguing resource to investigate the drivers of genetic population structure. These completely isolated groups are subject to the same natural pressures in tandem, allowing for the study of replicated wild systems. Vertebrate allochronic species are very rare. Most known allochronic species are periodical insects such as the pine processionary moth (Santos *et al.* 2007) and aphids (Abbot & Withgott 2004).

Pink salmon *Oncorhynchus gorbuscha* are the most abundant Pacific salmon, ranging widely across drainages of the North Pacific Ocean and Bering Sea from Japan in Asia to Washington State in North America (Ruggerone *et al.* 2010). Pink salmon are anadromous and semelparous, homing to their natal streams to spawn and die. Juveniles have a very short in-stream life, migrating to the sea directly after emergence from the gravel (Ishida 1966; Quinn 2005). However, unlike other Pacific salmonids, pink salmon have an obligate two-year life cycle which leads to reproductively isolated even- and odd-year lineages (Anas 1959; Turner & Bilton 1968). These reproductively isolated lineages often spawn in the same locations simply offset by one year. However, one lineage can be much more abundant than the other in a given location, and relative abundance can shift from one lineage to the other through time (Krkosek *et al.* 2011; NPAFC 2016). In recent years, the odd-year lineage has been dominant in Asia (Gordeeva 2014; Krkosek *et al.* 2011). In North America, the pattern is more static: the even-year lineage is

dominant in northern drainages while the odd-year lineage dominates in the southern drainages; some rivers at northern and southern extremes are inhabited by only one lineage (Beacham *et al.* 2012; Krkosek *et al.* 2011).

The existence of distinct lineages combined with differing levels of abundance has precipitated a number of intriguing ecological observations (Ruggerone & Nielsen 2005). Pink salmon may exert top-down control, driving major ecological shifts between years of high and low abundances (Springer & van Vliet 2014). Their abundance in the Pacific Ocean, driven by the dominance of Asian stocks, has been highest in odd -years during recent decades. Seabird diet, body mass, and reproductive success near the Aleutian Islands are reduced in odd-numbered years when pink salmon abundance has been exceptionally high (Springer & van Vliet 2014; Toge *et al.* 2011). Similarly, the abundance of pink salmon affects other Pacific salmon species: Ruggerone *et al.* (2015) found that pink salmon abundance in the North Pacific Ocean in odd years was linked to a decline in productivity of sockeye salmon (*O. nerka*) during their second year of life at sea.

The existence of distinct lineages has also provided a unique opportunity to address population and evolutionary genetics questions as the lineages can serve as replicates across the entire range. The earliest studies used evidence of divergent frequencies at allozyme loci between the two lineages from similar locations as support for the hypothesis that most polymorphisms are selectively neutral. (Aspinwall 1974). Subsequent studies have consistently shown that the strongest signal of divergence occurs between the two lineages; even- and odd-year lineages of a single river system are more genetically diverged than populations of the same lineage from adjacent rivers (Aspinwall 1974; Churikov & Gharrett 2002; Olsen & Seeb 1998). In addition,

studies have shown that there is more genetic differentiation among the North American populations of the odd-year lineage than among the even-year lineage populations (Beacham *et al.* 2012).

Two recent genomic studies in North American populations provide insights into the relationship between environmental factors and the role of genetic background in influencing genetic structure. Seeb *et al.* (2014) identified parallel signatures of selection in populations of two lineages inhabiting the same environments using over 6,000 SNPs derived from restriction-site-associated DNA (RAD) sequencing. With the same set of loci and a haploid-assisted linkage map, Limborg *et al.* (2014) identified not only genomic regions of parallel selection, but also regions of divergent selection between lineages, suggesting that adaptation in the two lineages may have arisen from different pools of standing genetic variation.

Almost all previous studies of genetic structure of pink salmon have been limited to a single continent or single lineage. By examining populations from both lineages across the entire range of the species, we can address lingering broad-scale questions. Our first question was how do the lineages compare in their distribution of diversity and genetic structure across the range? Second, is there any evidence of population structure repeated in parallel within the lineages that could help illuminate the evolutionary history of the species? To answer these questions, we built upon the genomic studies in North American populations by Seeb *et al.* (2014) and Limborg *et al.* (2014) and examined a similar number of paired replicates along the western Pacific Ocean. This provided consistent data for a comparison of paired populations from throughout Asia and North America to explore the effects of demographic and evolutionary factors influencing genetic population structure in each lineage.

## 2. Materials and Methods

### 2.1 Samples

Tissue samples from 32 individuals per lineage from paired populations (even- and odd-year lineage populations from the same river) originated from four regions of Asia: Hokkaido, Amur, Magadan and Kamchatka (Figure 1). The samples were collected by North Pacific Anadromous Fish Commission collaborators (<http://www.npafc.org/new/index.html>) and chosen for inclusion in this study based on the presence of both even- and odd-year lineages at the site, the number of available samples, and the geographical spread of locations (Table 1). Fin clips were collected from spawning adult pink salmon and stored in 95% ethanol at room temperature. In addition, raw data from three paired North American populations: Norton Sound, Prince William Sound and Puget Sound (24 samples from each lineage per river, 20 samples for Norton Sound even-year lineage; same collection regime) from Seeb *et al.* (2014) were incorporated into the study (Table 1). Detailed information on the 396 samples is included in Supplementary file 1.

### 2.2 RAD Sequence Data

DNA extraction and library preparation for RAD sequencing of the Asian samples followed that of the North American samples and are based on the protocol in Everett *et al.* (2012). Asian samples were sequenced using a bulk pooling and re-sequencing method, whereby the RAD libraries included many individually barcoded samples with the expectation that those that had poor sequence coverage would be re-sequenced. Because the individuals with the lowest sequence coverage were re-sequenced, this process should result in more even coverage of retained sequence reads across individuals. Genomic DNA was digested using the restriction site enzyme *SbfI*, and P1 adaptors were ligated. Each individual was barcoded using 6bp long

adaptors that differed by at least 2 nucleotides. The DNA was pooled into three libraries of no more than 86 individuals each and purified. The individuals were assigned to the three libraries randomly to avoid any lane effect.

The pooled libraries were separated evenly into two subsets and sheared at the same time. The ends of the DNA were repaired, and P2 adaptors were ligated to the sheared DNA. A final PCR with conditions from Everett *et al.* (2012) amplified the P1 and P2 adaptor-ligated DNA fragments, and the product was verified on a 1% agarose gel. Each library was assessed for quality and concentration of Genomic DNA using a Bioanalyzer DNA 1000 kit (Agilent Technologies, Santa Clara, CA), and the final concentration of each library was determined using the accompanying Bioanalyzer software. The prepared libraries were assessed for DNA quality and sequenced on an Illumina HiSeq2000 sequencer (Illumina, San Diego, CA) at the University of Oregon High-Throughput Sequencing facility.

Raw sequences for North American populations were from Seeb *et al.* (2014).

### *2.3 Initial Filtering and Re-sequencing*

The raw sequence data for all fourteen populations were filtered, and polymorphic loci were identified using the software package Stacks v.1.31 (Catchen *et al.* 2011). For each lane of raw sequence, the program *process\_radtags* was used to trim the terminal nucleotide from the 101 bp raw reads, de-multiplex the individuals, and filter the low quality reads. The barcodes used as sample identifiers were replaced with a unique sample name, and all reads from multiple lanes of sequencing were combined for ease and coherence of downstream analysis.

After this initial filtering, samples with fewer than 1.5 million retained reads were re-sequenced using the above protocol and their original barcodes. The resulting raw data from the re-sequenced samples was again analyzed using the Stacks program *process\_radtags*, and all sequences were combined for downstream analysis.

## 2.4 Genotyping

The filtered raw sequences from the North American populations and the Asian populations were jointly genotyped for this study.

A Stacks catalog was built using the program *cstacks*. The catalog is a representative set of loci against which the potential loci found in the populations are compared to identify variation. For this reason, it is important that it be built with a comprehensive set of individuals that represent the potential variation found in all populations to minimize ascertainment bias. Each population was represented in the catalog by four individuals. The individuals with the highest number of sequence reads were chosen for each population, as they would likely also have the highest quality sequence as measured by depth of coverage. We allowed two mismatches between sample tags when generating the catalog (flag:  $-n\ 2$ ). The catalog also includes four female parents of the crosses used to make a linkage map in an alternate study; two females from Washington (Limborg *et al.* 2014) and two from southeastern Alaska were included (unpublished data). Locus names in the catalog were standardized for consistency with the earlier study by Limborg *et al.* (2014).

The combined sequence reads were assembled into matching stacks, and polymorphic loci (SNPs) were identified using the program *ustacks*, following the previously described methods in

Limborg *et al.* (2014) (flags: -r --model\_type bounded --bound\_low 0 --bound\_high 0.05 and -M 2). The stacks of reads for each individual were compared to the catalog using the Stacks program *sstacks*, and loci identified as matching a locus in the catalog were retained. The genotypes of loci that were present in at least 80% of the individuals were exported with the Stacks program *populations*.

## 2.5 Summary Statistics and Genetic Diversity Measures

The program GENEPOP v4.3 (Rousset 2008) was used to calculate basic diversity indices for each locus per population. SNPs were retained in the final data set based on the following criteria: 1) if there were multiple putative SNPs per locus, then we retained the single SNP with the highest minor allele frequency (MAF); 2) SNPs that had a MAF > 0.05 in at least one population, and 3) SNPs that conformed to Hardy-Weinberg equilibrium expectations ( $P$ -value > 0.05) in at least seven populations (50% of tests). To filter paralogous loci, we calculated the  $F_{IS}$  per locus (Robertson & Hill 1984), and retained those SNPs that had a global  $F_{IS} > -0.5$  (Hohenlohe *et al.* 2013). We did not discern whether SNPs were neutral or non-neutral. Though not all SNPs in this data set are neutral, we assume that the non-neutral SNPs are a very small percentage of the total number of SNPs, and their signal would be swamped by the neutral SNPs. Analysis of these North American populations by Seeb *et al.* (2014) found that 2% of the 8,036 loci they used were non-neutral. Using different methods but the same loci, Limborg *et al.* (2014) found a comparable 2.7% were non-neutral. Finally, we excluded individuals that genotyped at less than 80% of the final set of loci.

Several diversity indices were calculated. The inbreeding coefficient ( $F_{IS}$ ) was calculated per population using the program GENEPOP. We calculated the mean observed heterozygosity

within each population and compared the values for the populations within the even lineage to those within the odd lineage using the nonparametric Mann-Whitney test (Mann & Whitney 1947) in R (R Core Team 2015). We also calculated the standardized observed heterozygosity per individual ( $H_{S\_obs}$ ) within each population with the R function *genhet* (Coulon 2010). The measure is a conservative estimate of the observed heterozygosity of an individual averaged over all the markers used, standardized by the mean observed heterozygosity of the genotyped loci. It is standardized to account for individuals that are not genotyped at the same loci, to ensure that the heterozygosity is measured on the same scale (Amos 2005; Coltman *et al.* 1999). The mean values for the populations within the even-year lineage were compared to those within the odd-year lineage using the nonparametric Mann-Whitney test in R. We used the package *hierfstat* (Goudet & Jombart 2015) in R to determine the rarefied allelic richness per locus by population as another measure of the genetic diversity found in the populations, and found average values for each population and lineage. The mean values for the populations within the even-year lineage were compared to those within the odd-year lineage using the nonparametric Mann-Whitney test in R.

Finally, we estimated the contemporary genetic effective population size ( $N_e$ ) for each of the 14 populations using the linkage disequilibrium method with the program NeEstimator v.2.01 which has a bias-correction for LD (Ahfock *et al.* 2014). We used a model of random mating and excluded loci that had an overall minor allele frequency less than 0.05.

## 2.6 Population Structure

We employed several methods to evaluate population structure within and between lineages.

## 2.7 Neighbor-Joining Tree

First, we constructed a Neighbor-Joining tree (Saitou & Nei 1987) using the  $D_A$  distance (Nei 1983) with 10,000 bootstraps, using the program POPTREE2 (Takezaki *et al.* 2010) to get a broader view of the relationship of the 14 populations.

## 2.8 $F$ -Statistics

We estimated pairwise and per locus global  $F_{ST}$  based on Weir and Cockerham (1984) with GENEPOP v4.3 (Excoffier & Lischer 2010), an efficient program to generate estimates for a large number of tests.

To better discern the hierarchy of differentiation in pink salmon, we used the program GENETIX v.4.05.2 to calculate the  $F_{ST}$  based on Weir and Cockerham (1984) of six groups of populations. GENETIX is able to generate confidence intervals around the  $F_{ST}$  estimates. The estimates of  $F_{ST}$  were calculated by comparing populations within groups to each other, subdivided as follows: 1) overall  $F_{ST}$  of all 14 populations, and 2) pairwise  $F_{ST}$  by location between populations of different lineages. To characterize the magnitude of differentiation between the populations within each lineage we estimated  $F_{ST}$  values separately for each lineage: 3) all even-lineage populations and 4) all odd populations. Finally, we estimated the  $F_{ST}$  of two groups of populations within each lineage to compare the magnitude of the differentiation. The N-J tree informed the grouping of the populations tested within each lineage, as a clear parallel population structure was evident in the tree. The populations were grouped within each lineage as follows: 5) even-lineage populations Prince William Sound and Puget Sound, called the North America populations, and Hokkaido, Amur, Magadan, Kamchatka and Norton Sound

populations, which we called Beringia and 6) odd-year lineage population grouped into the same North America and the Beringia populations. Confidence intervals (95%) were estimated using 1,000 bootstraps

### *2.9 Principal Component Analysis*

To explore the relationships among populations we employed an individual-based analysis of principal components (PCA) that included all 14 populations. The eigenvalues were calculated in PLINK v1.90 (Chang *et al.* 2015) which uses a variance-standardized relationship matrix; the input was pairwise relatedness between individuals. We extracted the top 20 principal components using the default settings (flags --allow-extra-chr, and --allow-no-sex) and the resulting eigenvectors were visualized in R using the package *ggplot2* (Wickham 2009).

### *2.10 Analysis of Molecular Variance*

To test for hierarchical population structure in the data, we conducted three analysis of molecular variance (AMOVA) in Arlequin v.3.5.1.2 (Excoffier & Lischer 2010), using 10,000 permutations. The first AMOVA (A. Figure 2) evaluated the distribution of genetic diversity within and between the two lineages. The second hierarchy was tested with two AMOVAs (B. and C. Figure 2), each evaluated the distribution of diversity within each lineage; populations within lineage were grouped as follows: i) the North America group: Puget Sound and Prince William Sound, and ii) the Beringia group Hokkaido, Amur, Magadan, Kamchatka and Norton Sound.

### *2.11 Isolation by Distance*

In cases where there is limited dispersal, for instance due to natal homing, a positive relationship between geographical distance and genetic distance can occur. This is commonly referred to as isolation by distance (IBD) (Wright 1943). We limited the test of IBD to the five Beringia populations to avoid bias due to clustering of the populations, as the Mantel Test does not differentiate patterns of clustering from IBD (Merimans 2012). We were unable to test for IBD in the North America group because it only included two populations. For each lineage, two Mantel Tests (Mantel 1967) with 1,000 permutations each were conducted in R to examine whether the Beringia populations within each lineage follow the pattern of IBD. The null hypothesis was that there was no relationship between the pairwise genetic distance of the populations within lineages, measured by a standardized  $F_{ST}/(1-F_{ST})$  (Rousset 1997), and the pairwise geographical distance of the populations.

Geographic distance between the populations was calculated in two ways reflecting two different hypotheses of straying and gene flow. The first pairwise distance measures were based on an abbreviated waterway distance between populations. They measured a distance between populations that imitates straying and gene flow directly from population to population across waterways (orange lines, Figure 1). The second pairwise distance measured the length of coastline between populations and reflects a pattern of straying from population to neighboring population along the coast (purple lines, Figure 1). Both measures of distance between populations were calculated in ArcGIS v.10.1 (ESRI 2012).

### 3. Results

#### *3.1 Bioinformatics and Genotyping*

Combining the new sequences for the Asian populations with the sequences for the North American populations resulted in a total of 903.5 million retained reads for 396 individuals (Supplementary file 2). On average, each individual from Asia had two million reads and each individual from North America had 2.8 million reads. From these retained reads, the Stacks pipeline identified 666,351 loci, of which 16,681 putative SNPs remained after filtering (Supplementary file 2). Thirteen individuals were genotyped at fewer than 80% of the putative 16,681 SNPs and were dropped from the analysis (Table 2A). These included five individuals from Magadan even-year and four from Magadan odd-year populations, likely a reflection of tissue quality. Two individuals from Norton Sound even-year and one each from the Kamchatka even- and odd-year lineage populations were also dropped.

#### *3.2 Diversity within populations and comparison between lineages*

We estimated diversity within each of the 14 populations and compared the results across the lineages (Table 2A). Average observed heterozygosity was 0.158 in the even-year lineage and 0.167 in the odd-year lineage; the distributions of the average observed heterozygosity ( $H_o$ ) per population in the two groups differed significantly ( $P$ -value = 0.0055, one-tailed). Additionally, for each comparison of the paired populations, the population within the even-year lineage had lower average observed and expected heterozygosities (Table 2A). In both lineages, the lowest average observed and expected heterozygosities ( $H_e$ ) were seen in the most southern populations at the end of the ranges in Asia and North America (Puget Sound and Hokkaido populations).

The individual standardized observed heterozygosity ( $H_{s\_obs}$ ) showed a particularly striking pattern (Table 2A, Figure 3). In every location, the populations of the even-year lineage had lower mean values of  $H_{s\_obs}$  than those in the odd-year lineage. The  $H_{s\_obs}$  was 0.965 in the even-year and 1.024 in the odd-year lineage, and the distributions in the two groups differed significantly ( $P$ -value = 0.0055, one-tailed). A boxplot of these values illustrates the pattern succinctly, showing the distribution of the values in the even-year lineage populations is lower than the odd-year lineage populations in all but the two southern North American populations (Figure 3).

We compared allelic richness between the lineages and across the range within the lineages. Similar to estimates based on heterozygosity, we found lower diversity in the even-year lineage in the estimation of allelic richness. The mean allelic richness was 1.60 for the even-year lineage and 1.65 for the odd-year populations; distributions in the two groups differed significantly ( $P$ -value = 0.0189, one-tailed) (Table 2A). We also found that the lowest values for both lineages were observed at the southern extent of the range; this trend was replicated on both continents.

In each lineage, the effective population size was larger for the populations at the center of the range than for those on the southern edges of the range; this mirrored trends in diversity (Table 2B). The estimates of  $N_e$  ranged from a low of 965 in the even-year Hokkaido population, to a high of 56,866 in the even-year Kamchatka population. The confidence intervals for many of the populations at the center of the range included a limit of infinity.

### 3.3 Diversity among populations and comparison between lineages

#### 3.3.1 *F*-Statistics

Non-hierarchical diversity analyses were conducted to quantify the distribution of variation. We estimated an  $F_{ST}$  value of 0.060 over all populations (Table 3A). Next, within each lineage, we calculated separate  $F_{ST}$  values: the odd-year lineage had a significantly higher  $F_{ST}$  (0.033) than the even-year lineage (0.026) (Table 3B). We then subdivided each lineage into the five Beringia populations and the two North America populations and calculated  $F_{ST}$  values. For both lineages, the estimated  $F_{ST}$  among the Beringia populations was significantly less than that estimated for the North America group (Table 3B). The  $F_{ST}$  for the even-year Beringia populations (0.011) and the North America populations (0.030) differed significantly, but were of a lesser magnitude than those observed in the odd-year lineage. In the odd-year lineage, the  $F_{ST}$  of the Beringia (0.008) and the North America (0.052) populations were strikingly different.

Additional non-hierarchical analyses were conducted in a pairwise manner. Pairwise comparisons of  $F_{ST}$  for all 14 populations range from 0.001 to 0.148 (Figure 4). Similar to the N-J tree results and PCAs, the largest values generally resulted from comparisons between lineages. Within lineages, the pattern of differentiation also followed the geographical spread of the populations, with comparisons of geographically proximal populations resulting in smaller  $F_{ST}$  values. The pairs of populations sharing the same location had higher  $F_{ST}$  than the groups of populations within the same lineage (Figure 4 and Table 3C). These patterns were also reflected in the results of the AMOVAs (see below). Looking at these comparisons by continent, we observed larger values for pairwise  $F_{ST}$  between the odd-year lineage comparisons in North America, which had a range from 0.038 to 0.081, than within the even-year lineage in North

America, which ranged from 0.028 to 0.054. Focusing on the populations on the Asian continent we observed the opposite trend, with greater differentiation in the even-year lineage than in the odd-year lineage, though the trend was much weaker. The comparisons of pairwise  $F_{ST}$  between Asian populations ranged from 0.001 to 0.023 in the even-year lineage compared to a range of 0.001 to 0.018 in the odd-year lineage.

### *3.3.2 Neighbor-Joining Tree*

We clustered populations using a Neighbor-Joining (N-J) tree. The tree shows a very distinct division between the lineages, with all branches having strong bootstrap support. All but one node was supported by 100% of the bootstrap samplings (Figure 5). Within both lineages, the most divergent branch of the tree separated the North American populations of Prince William Sound and Puget Sound (North America) from a group including all Asian populations plus the North Sound populations from Northwestern Alaska (Beringia). The branch lengths for the North America group are longer in the odd-year lineage than the even-year lineage, suggesting greater differentiation.

### *3.3.3 Principal Component Analysis*

The principal component analysis explored the relationships among all 14 populations. Echoing the N-J tree, this analysis showed that the greatest amount of variation in the data was explained by the division between the lineages (percent of variation explained by PCA axis 1: 30.24%) (Figure 6A, B). The second greatest amount of variation in the data was explained by differentiation between the populations in the odd-year lineage (PCA axis 2: 13.83%, Figure 6A, C). This appears to be dominated by the differences between the North America populations

within each lineage, as the Beringia populations cluster more closely together within lineages. Comparing the first dimension to the third shows that the differentiation between populations in the even-year lineage explains the third greatest amount of variation in the data (PCA axis 3: 8.73%, Figure 6B, C). Comparing the second and third dimensions more clearly illustrates the relationship of the populations within each lineage by removing the influence of the first dimension on the data (Figure 6C).

### 3.3.4 Analysis of Molecular Variance

We then conducted hierarchical analyses to partition the genetic variation and estimate the distribution of variation. The first AMOVA (Table 4A), evaluated the partitioning of variation between the even- and odd-lineages. A total of 5.72% of the variation was explained by differences between the lineages with 2.87% attributed to differences among populations within the lineages; both estimates were significantly different from zero ( $P < 0.0001$ ). This reflects the patterns seen in the pairwise comparisons of  $F_{ST}$  where the population comparisons with the greatest  $F_{ST}$  were those of different lineages.

Separate AMOVAs were conducted for each lineage to evaluate the partitioning of genetic variation between North America and Beringia groups within lineage. In the AMOVA for the even-year lineage (Table 4B), 2.82% of variation was explained by differences among the groupings (statistically significant at  $P < 0.05$ ) with 1.40% of variation attributed to differences among populations within the groups (statistically significant at  $P < 0.0001$ ). The AMOVA between the groups in the odd lineage was similar (Table 4C), but a greater proportion of the variation was attributed to among groups than in the even-year lineage AMOVA. For the odd-year lineage, 4.13% of variation was explained by differences among the groups (statistically

significant at  $P < 0.05$ ) with 1.46% of variation attributed to differences among populations within the groups (statistically significant at  $P < 0.0001$ ).

### *3.3.5 Isolation by Distance*

We tested for a relationship of isolation by distance within each lineage of the Beringia groups using two measures of geographic distance. The test of IBD only included the five Beringia populations to avoid bias from the strong population structure within each lineage due to clustering of the populations: the AMOVA indicated strong population differentiation between the groups Beringia and North America within each lineage. Patterns of clustering can confound tests for IBD (Merimans 2012). Within the even-year lineage, the waterway distance had a  $P$ -value of 0.510, and the coastline distance a  $P$ -value of 0.185. Within the odd-year lineage, the waterway distance had a  $P$ -value of 0.526, and the coastline distance a  $P$ -value of 0.345. No  $P$ -values were statistically significant at  $\alpha = 0.05$ . The results suggest that the Beringia populations within each lineage do not conform to the pattern of isolation by distance using either measurement of distance.

## **4. Discussion**

How do the lineages compare in their distribution of diversity and genetic structure across the range? The large number of SNPs revealed through RAD sequencing provided a detailed evaluation of the distribution of diversity among populations and between the lineages. In general, population structure was similar between lineages, likely a reflection of similar demographic and phylogeographic histories. However, the extent and distribution of individual

and population diversity varied significantly suggesting differences in the origin and age of the lineages.

#### *4.1 Diversity within populations and comparison between lineages*

We found significant differences between the lineages in the standardized observed individual heterozygosities; individuals of the odd-year lineage have significantly greater observed heterozygosity than individuals of the even-year lineage. A similar trend was observed for mean observed and expected population heterozygosity. Previously, differences in heterozygosity have only been noted among a very select subsample of loci. Limborg *et al.* (2014) found significantly higher observed heterozygosity in the odd-year lineage when compared to the even-year for a subset of loci considered candidates for selection between the lineages. The authors did not identify a difference in observed heterozygosity between the lineages when they used all loci (8,036).

There are several possible explanations for the differences between our results and those of Limborg *et al.* (2014). In general, differences in heterozygosity between groups can be a signal of ascertainment bias. This seems unlikely to be the case here, since loci were ascertained using an equal representation from each population and lineage. We think it most likely that the differences in the observed population and individual heterozygosities between the studies is a reflection of increased statistical power. We used twice the number of SNPs as Limborg *et al.* (2014), and the four new population pairs in Asia had larger sample size with a target of 32 rather than 24. Similar to Limborg *et al.* (2014), the North American populations have the smallest signal of differentiation in observed heterozygosities in our data, again likely a reflection of the smaller sample sizes from North America.

Heterozygosity can differ between populations for many reasons. Isolation, population size and genetic drift can be major influencers of the difference in mean heterozygosity between populations, with smaller and more isolated populations possessing lower heterozygosity (Kimura 1963). Isolation and genetic drift likely play a large part in the patterns we see as the lineages have been reproductively isolated and independently influenced by random genetic drift for thousands of generations. Differences in connectivity, and demographic processes such as population bottlenecks have also likely shaped the lineages independently since their origin.

These measurable differences in individual observed heterozygosity between the lineages may offer clues to differences in the pools of standing genetic variation and genetic backgrounds between the lineages. In many species, individuals within populations that have higher heterozygosity have been shown to have greater fitness (Britten 1996; Hansson & Westerberg 2002) and increased pathogen resistance (Luikart *et al.* 2008). Interestingly, the odd-year lineage has been shown to adapt to new habitats better than the even-year lineage. In a study of the successful transplantation of both lineages from the Kamchatka region of Russia to northern Europe, habitat outside their native range, the odd-year lineage was able to readily establish self-sustaining populations, while the even-year lineage was not (Gordeeva & Salmenkova 2011). By observing the pattern of the routine failure of even-year pink salmon transplants, Gordeeva and Salmenkova (2011), propose that the lineages may differ in adaptive plasticity and that the even-year lineage may have arisen from the odd-year lineage. Further, they suggest the even-year lineage may have lost a part of its genetic variation with which adaptive potential is associated.

The differences we see in rarefied allelic richness between the lineages also support the hypothesis of different pools of standing genetic variation in the lineages. Similar to both measures of heterozygosity, we found that the odd-year lineage has greater rarefied allelic richness than the even-year lineage. Though it is less commonly used than heterozygosity as a measure of genetic diversity because of the difficulty of resolving the influence of gene flow and genetic drift, allelic richness has been used as a measure of a population's potential for adaptability and persistence (Greenbaum *et al.* 2014).

#### *4.2 Diversity among populations and comparison between lineages*

We initially examined the pattern of differentiation of the populations grouped by continent. We observed greater genetic differentiation among the populations in the odd-year lineage in North America than within the even-year lineage in North America, mirroring previous findings (Seeb *et al.* 2014). On the Asian continent we observed the opposite trend with greater differentiation in the even-year lineage than in the odd-year lineage, though the trend is much weaker.

Similar to earlier studies, the greatest signal of divergence in these 14 populations is between the lineages. This agrees with the current calculations of the origin and evolutionary timeline of pink salmon, which estimates that the species of Pacific salmon diverged between 6 and 11 million years ago (Devlin 1993). Within pink salmon, the origin of the lineages has consistently been dated to before the Last Glacial Maximum but estimates vary widely from approximately 26 thousand years ago and up to one million years ago (Bryokov *et al.* 1996; Churikov & Gharrett 2002).

One proposed theory to explain the differences in genetic variation in the lineages is that each lineage was adapted to the environment of a distinct historical refugium (Aspinwall 1974). However, in both lineages, Asian and northern Alaska populations show little differentiation to each other but differ significantly from populations in southcentral Alaska and the Pacific Northwest. This parallel population structure within each lineage suggests that both pink salmon lineages were established around the Pacific Rim before variance due to glaciation isolated populations within each lineage from the rest of the range. Our data support the theory that the lineages shared two main refugia during the Last Glacial Maximum, not lineage-specific refugia as has been suggested (Aspinwall 1974). The contemporary structure is likely a reflection of that historical reproductive isolation and adaptation to distinct habitats caused by the separation of the Beringia populations from the North American populations in both lineages, as well as differences caused by subsequent post-glacial colonization.

The habitat along the coastal North Pacific Ocean has changed drastically with the oscillating climate and resulting cycles of glaciation in the last million years (Brigham-Grette *et al.* 2003a). The Last Glacial Maximum, (LGM) was a period during the most recent glacial cycle with the highest global ice volume (Ehlers & Gibbard 2007). The LGM is thought to have reached its peak approximately 26.5 thousand years ago (Clark *et al.* 2009). The ice cover drastically altered the landscape. Most of eastern Asia and western Alaska were ice-free, whereas extensive ice sheets covered the northern latitudes of North America down to 40°N in some parts of the continent (Brigham-Grette *et al.* 2003a; Gualtieri *et al.* 2005; Hewitt 2000). These glaciers reduced the volume of water in the oceans, resulting in sea levels between 120 to 130m lower than present (Hopkins 1973; Rohling *et al.* 1998). The resulting exposed continental shelf

connected the Asian and North American continents (Bering land bridge) and provided vast, ice-free habitat in what is referred to as Beringia (Hultén 1937) (Figure 7).

The dynamic habitat of the North Pacific, especially the emergence and sinking of the Bering land bridge, has influenced the contemporary population structure of many species of varying taxa (Pielou 1991). Because of this transient connection between the Asian and North American continents, most northern Alaskan populations share affinities with Asian populations (Pielou 1991). Molecular studies on mammals, such as reindeer (Flagstad & Røed 2003), bears (Hassanin 2015; Puckett *et al.* 2015), and wolverines (Zigouris *et al.* 2013) as well as flora, (Abbott & Brochmann 2003; Eidesen *et al.* 2013; Hultén 1937), birds (Kidd & Friesen 1998; Wenink *et al.* 1996) and freshwater fish (Campbell *et al.* 2015), show that a wide variety of taxa relied on the region as an important refugium during the LGM.

In addition to Beringia, there is evidence of a second major glacial refugium in the Pacific Northwest (Pielou 1991; Shafer *et al.* 2010) (Figure 7). Retreating from the North American ice sheets during the LGM, terrestrial species either moved northwest of the ice and utilized the Beringia as refuge, or moved south and utilized terrestrial southern refugium south of the ice (Pielou 1991). Northern latitude ocean dependent species have been shown to have found refuge in the Pacific Northwest. Species such as cod (Canino *et al.* 2010), herring (Grant *et al.* 2012), and Arctic char (Moore *et al.* 2015) bear signatures of the fragmentation of habitat from the LGM in their population genetic structure.

We believe that our results suggest that the Beringia pink salmon populations persisted in Asia and the Bering land bridge. These populations may not have been as restricted in the amount of coastline habitat during the LGM as those populations that were isolated in the Pacific

Northwest. Instead Beringia pink salmon may have gained habitat with the exposure of the Bering land bridge, depending on whether sea ice cover reduced access to freshwater habitats (Noll *et al.* 2001). The changes in sea level also might have impacted connectivity between regions in Asia, as previously free flowing waterways between islands dried (Noll *et al.* 2001). Separation of populations by vicariance could have resulted in a pattern of isolation by distance. Our tests for an association between the genetic and geographic distance within each lineage did not reveal any statistical support for a pattern of isolation by distance between the Beringia populations within each lineage. However, IBD has previously been shown in Asian populations. In a study using both allozyme and microsatellite loci, Salmenkova *et al.* (2006), found a signal of IBD only in the even lineage, though the samples did not span the range of Beringia populations we represent here. Instead they only included a set of populations from the northern Sea of Okhotsk, akin to our Magadan population, and the southern tip of Sakhalin Island, near Hokkaido, Japan. Hawkins *et al.* (2002) also found IBD in Asia using populations from Japan to the southern coast of the Bering Sea, though the signal was stronger in the even-year lineage than the odd.

That we find no signal of IBD in either lineage in Beringia might be because of the large number of loci employed (16,681), which categorized very low differentiation between the populations, and the large range of populations, which span a considerable geographic distance. Possibly, the lack of IBD reflects a history of gene flow between the populations in Beringia. Of the Pacific salmon, there is some evidence that pink salmon have the highest straying rate (Gordeeva 2014). Pink salmon juveniles migrate to saltwater immediately after hatching, and therefore may not strongly imprint on natal streams which may explain their higher straying rates as returning adults (Quinn 2005). An increase in the available habitat during the LGM coupled with the

higher straying rate may have resulted in reduced population structure among the Beringia populations.

In contrast to the homogeneity within the lineages in Beringia, the North America populations show a greater genetic differentiation within lineages. The odd-year lineage within North America is more differentiated than the even. This could be due to the inherent differences in the standing genetic variation present in the lineages (Limborg *et al.* 2014); a factor that could be independent, and predate, the pattern of historical refugia, or the influence of random genetic drift and demographic processes such as bottlenecks in the post-glacial colonization of populations.

Another explanation for the population differentiation between the two southern North American populations is a phenomena referred to as *refugia within refugia*, (Gómez & Lunt 2007). The two populations from North America were isolated from the Beringia populations by the ice sheets in North America, finding refuge in the Pacific Northwest. But within that North American refugium, populations may have experienced further isolation from one another. For instance, a segment of populations may have used a separate, more isolated portion of the refugium, and been exposed to population demographic processes and random genetic drift independent of other populations. They subsequently would have also experienced separate post-glacial recolonization dynamics. There is evidence of a cryptic refugia in the Alexander Archipelago and Haida Gwaii, off the coast of British Columbia, that was used as a *refugia within a refugia* for some species of flora, birds and small mammals during the LGM (Shafer *et al.* 2010). These two factors are not mutually exclusive; both standing genetic variation and a

pattern of *refugia within refugia* could account for the differing population structure within the lineages in North America.

Signals of population fragmentation are also seen in other species of Pacific salmon. A contact zone for chum salmon (*O. keta*) exists on the North Alaska Peninsula, a likely consequence of fragmentation during the last ice age followed by recolonization (Petrou *et al.* 2013; Seeb *et al.* 1999). Similar divergence among population groups along the North Alaska Peninsula has been observed in Chinook (*O. tshawytscha*), coho (*O. kisutch*) and sockeye salmon (Habicht *et al.* 2010; Smith *et al.* 2001; Templin *et al.* 2011). With the limited density of populations, we do not have enough resolution along the northern North American coast to estimate a location of post glacial secondary contact, nor can we fully understand the consequences of colonization from multiple refugia.

## 5. Conclusions

Utilizing paired populations from the extent of the range allowed a more complete view of the population genetic structure of the species. Distinct differences in standing pools of genetic variation in the lineages were identified, as measured by individual observed heterozygosity, observed heterozygosity and allelic richness by population. The hierarchical nature of the lineages was also leveraged to identify parallel population structure within the lineages. The differentiation between the North America and Beringia populations could be explained by a history of glacial refugia that divided the Beringia populations from the southern North American populations but was shared by the lineages.

## 6. References

- Abbot P, Withgott JH (2004) Phylogenetic and molecular evidence for allochronic speciation in gall-forming aphids (*Pemphigus*). *Evolution* **58**, 539-553.
- Abbott RJ, Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hulten. *Molecular Ecology* **12**, 299-313.
- Ahfock D, Wood I, Stephen S, Cavanagh CR, Huang BE (2014) Characterizing uncertainty in high-density maps from multiparental populations. *Genetics* **198**, 117-128.
- Alexander RD, Bigelow RS (1960) Allochronic speciation in field crickets, and a new species, *Acheta veletis*. *Evolution* **14**, 334-346.
- Amos W (2005) *IRmacroN3* documentation. URL: [www.zoo.cam.ac.uk/directory/william-amos/](http://www.zoo.cam.ac.uk/directory/william-amos/). Accessed June, 2016.
- Anas RE (1959) Three-year-old pink salmon. *Journal of Fisheries Research Board of Canada* **16**, 91-94.
- Aspinwall N (1974) Genetic analysis of North American populations of the pink salmon, *Oncorhynchus gorbuscha*, possible evidence for the neutral mutation-random drift hypothesis. *Evolution* **28**, 295-305.
- Beacham TD, McIntosh B, MacConnachie C, Spilsted B, White BA (2012) Population structure of pink salmon (*Oncorhynchus gorbuscha*) in British Columbia and Washington, determined with microsatellites. *Fisheries Bulletin* **110**, 242-256.
- Brigham-Grette J, Gualtieri LM, Glushkova OY, *et al.* (2003a) Chlorine-36 and C-14 chronology support a limited Last Glacial Maximum across central Chukotka, northeastern Siberia, and no Beringian ice sheet. *Quaternary Research* **59**, 386-398.
- Brigham-Grette J, Keigwin L, Driscoll N (2003b) Investigating the paleoclimate of an Arctic gateway. *Geotimes* **48**, 20-23.
- Britten HB (1996) Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* **50**, 2158-2164.
- Bryokov VA, Polyakova N, Skurikhina LA, Kukhlevsky AD (1996) Geographical and temporal mitochondrial DNA variability in populations of pink salmon. *Journal of Fish Biology* **48**, 899-909.
- Campbell MA, Takebayashi N, Lopez JA (2015) Beringian sub-refugia revealed in blackfish (*Dallia*): implications for understanding the effects of Pleistocene glaciations on Beringian taxa and other Arctic aquatic fauna. *BMC Evol Biol* **15**, 144.
- Canino MF, Spies IB, Cunningham KM, Hauser L, Grant WS (2010) Multiple Ice-Age refugia in Pacific cod, *Gadus macrocephalus*. *Mol Ecol* **19**, 4339-4351.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *Genetics* **1**, 171-182.
- Chang CC, Chow CC, Tellier LC, *et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7.
- Churikov D, Gharrett AJ (2002) Comparative phylogeography of the two pink salmon broodlines: an analysis based on a mitochondrial DNA genealogy. *Molecular Ecology* **11**, 1077-1101.
- Clark PU, Dyke AS, Shakun JD, *et al.* (2009) The Last Glacial Maximum. *Science* **325**, 710-714.

- Coltman DW, Pilkington JG, Smith JA, M. J, Pemberton JM (1999) Parasite-Mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53**, 1259-1267.
- Coulon A (2010) *genhet*: an easy-to-use R function to estimate individual heterozygosity. URL: <http://www.aureliecoulon.net/research/ac-computer-programs.html>. *Mol Ecol Resour* **10**, 167-169.
- Ehlers J, Gibbard PL (2007) The extent and chronology of Cenozoic Global Glaciation. *Quaternary International* **164-165**, 6-20.
- Eidesen PB, Ehrich D, Bakkestuen V, *et al.* (2013) Genetic roadmap of the Arctic: plant dispersal highways, traffic barriers and capitals of diversity. *New Phytol* **200**, 898-910.
- ESRI (2012) ArcGIS Desktop: Release 10.1. Environmental Systems Research Institute, Redlands, CA.
- Everett MV, Miller MR, Seeb JE (2012) Meiotic maps of sockeye salmon derived from massively parallel DNA sequencing. *BMC Genomics* **13**, 521.
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564-567.
- Flagstad Ø, Røed KH (2003) Refugial origins of reindeer (*Rangifer tarandus L.*) inferred from mitochondrial DNA sequences. *Evolution* **57**, 658-670.
- Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of Southern European Refugia* (eds. Weiss S, Ferrand N), pp. 155-188.
- Gordeeva NV (2014) Estimates of population differentiation in pink salmon *Oncorhynchus gorbuscha* with using of the microsatellite markers may be underestimated due to high population sizes. *Journal of Ichthyology* **54**, 347-358.
- Gordeeva NV, Salmenkova EA (2011) Experimental microevolution: transplantation of pink salmon into the European North. *Evolutionary Ecology* **25**, 657-679.
- Goudet J, Jombart T (2015) hierfstat: estimation and tests of hierarchical *F*-statistics. URL: <https://cran.r-project.org/web/packages/hierfstat/hierfstat.pdf>.
- Grant WS, Liu M, Gao T, Yanagimoto T (2012) Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographics in Pacific herring (and other species). *Mol Phylogenet Evol* **65**, 203-212.
- Greenbaum G, Templeton AR, Zarmi Y, Bar-David S (2014) Allelic richness following population founding events – a stochastic modeling framework incorporating gene flow and genetic drift. *PloS one*, 1-23.
- Gualtieri L, Vartanyan S, Brigham-Grette J, Anderson P (2005) Evidence for an ice-free Wrangel Island, northeast Siberia during the Last Glacial Maximum. *Boreas* **34**, 264-273.
- Habicht C, Seeb LW, Myers KW, Farley EV, Seeb JE (2010) Summer–fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. *Transactions of the American Fisheries Society* **139**, 1171-1191.
- Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* **11**, 2467-2474.
- Hassanin A (2015) The role of Pleistocene glaciations in shaping the evolution of polar and brown bears. Evidence from a critical review of mitochondrial and nuclear genome analyses. *C R Biol* **338**, 494-501.
- Hawkins S, Varnavskaya NV, A. ME, *et al.* (2002) Population structure of odd-broodline Asian pink salmon and its contrast to the even-broodline structure. *Journal of Fish Biology* **60**, 370-388.

- Hewitt G (2000) The genetic legacy of the Quaternary Ice Ages. *Nature* **405**, 907-913.
- Hohenlohe PA, Day MD, Amish SJ, *et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology* **22**, 3002-3013.
- Hopkins DM (1973) Sea level history in Beringia during the past 250,000 years. *Quaternary Research* **3**, 520-540.
- Hultén E (1937) *Outline of the history of Arctic and Boreal biota during the Quaternary period*, Stockholm.
- Ishida T (1966) Salmon of the North Pacific Ocean- Part III: A review of the life history of the North Pacific salmon 2, pink salmon of the far east (ed. Research CoBa), pp. 29-39.
- Kidd MG, Friesen VL (1998) Analysis of mechanisms of microevolutionary change in *Cepphus guillemots* using patterns of control region variation. *Evolution* **52**, 1158-1168.
- Kimura M (1963) The mutation load in small populations. *Genetics* **48**, 1303-1312.
- Krkosek M, Hilborn R, Peterman RM, Quinn TP (2011) Cycles, stochasticity and density dependence in pink salmon population dynamics. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2060-2068.
- Limborg MT, Waples RK, Seeb JE, Seeb LW (2014) Temporally isolated lineages of pink salmon reveal unique signatures of selection on distinct pools of standing genetic variation. *J Hered* **105**, 741-751.
- Luikart G, Pilgrim K, Vistry J, Ezenwa VO, Schwartz MK (2008) Candidate gene microsatellite variation is associated with parasitism in wild bighorn sheep. *Biol Lett* **4**, 228-231.
- Mann HB, Whitney DR (1947) On a test of whether one of two random variables in stochastically larger than the other. *The Annals of Mathematical Statistics* **18**, 50-60.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.
- Merimans PG (2012) The trouble with isolation by distance. *Molecular Ecology* **21**, 2839-2846.
- Moore J-S, Bajno R, Reist JD, Taylor EB (2015) Post-glacial recolonization of the North American Arctic by Arctic char (*Salvelinus alpinus*): genetic evidence of multiple northern refugia and hybridization between glacial lineages. *Journal of Biogeography* **42**, 2089-2100.
- Nei M (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of molecular evolution* **19**, 153-170.
- Noll C, Kinas NM, Varnavskaya NV, *et al.* (2001) Analysis of contemporary genetic structure of even-broodyear populations of Asian and western Alaskan pink salmon, *Oncorhynchus gorbuscha*. *Fishery Bulletin* **99**, 123-138.
- NPAFC (2016) NPAFC Pacific salmonid catch statistics (updated 29 March 2016) (ed. North Pacific Anadromous Fish Commission V).
- Olsen JB, Seeb LW (1998) Genetic interpretation of broad-scale microsatellite polymorphism in odd-year pink salmon. *Transactions of the American Fisheries Society* **127**, 535-550.
- Petrou EL, Hauser L, Waples RS, *et al.* (2013) Secondary contact and changes in coastal habitat availability influence the nonequilibrium population structure of a salmonid (*Oncorhynchus keta*). *Molecular Ecology* **22**, 5848-5860.
- Pielou EC (1991) *After the Ice Age: The return of life to glaciated North America* The University of Chicago Press.

- Puckett EE, Etter PD, Johnson EA, Eggert LS (2015) Phylogeographic analyses of American black bears (*Ursus americanus*) suggest four glacial refugia and complex patterns of postglacial admixture. *Mol Biol Evol* **32**, 2338-2350.
- Quinn TP (2005) *The behavior and ecology of Pacific salmon and trout* University of Washington Press.
- R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Robertson A, Hill WG (1984) Deviations from Hardy-Weinberg proportions: Sampling variances and use in estimation of inbreeding coefficients. *Genetics* **107**, 703-718.
- Rohling EJ, Fenton M, Jorissen FJ, *et al.* (1998) Magnitudes of sea-level lowstands of the past 500,000 years. *Nature* **394**, 162-165.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* **145**, 1219-1228.
- Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* **8**, 103-106.
- Ruggerone GT, Connors BM, Ii TQ (2015) Productivity and life history of sockeye salmon in relation to competition with pink and sockeye salmon in the North Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* **72**, 818-833.
- Ruggerone GT, Nielsen JL (2005) Evidence for competitive dominance of pink salmon (*Oncorhynchus gorbuscha*) over other salmonids in the North Pacific Ocean. *Reviews in Fish Biology and Fisheries* **14**, 371-390.
- Ruggerone GT, Peterman RM, Dorner B, Myers KW (2010) Magnitude and trends in abundance of hatchery and wild pink salmon, chum salmon, and sockeye salmon in the North Pacific Ocean. *Marine and Coastal Fisheries* **2**, 306-328.
- Saitou N, Nei M (1987) The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.
- Salmenkova EA, Gordeeva NV, Omel'chenko VT, *et al.* (2006) Genetic differentiation of pink salmon *Oncorhynchus gorbuscha* Walbaum in the Asian part of the range. *Russian Journal of Genetics* **42**, 1148-1163.
- Santos H, Rousset J, Magnoux E, *et al.* (2007) Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings. Biological sciences / The Royal Society* **274**, 935-941.
- Seeb JE, Habicht C, Templin WD, *et al.* (1999) Allozyme and mitochondrial DNA variation describe ecologically important genetic structure of even-year pink salmon inhabiting Prince William Sound, Alaska. *Ecology of Freshwater Fish*.
- Seeb LW, Waples RK, Limborg MT, *et al.* (2014) Parallel signatures of selection in temporally isolated lineages of pink salmon. *Molecular Ecology* **23**, 2473-2485.
- Shafer AB, Cullingham CI, Cote SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Mol Ecol* **19**, 4589-4621.
- Smith CT, Nelson RJ, Wood CC, Koop BF (2001) Glacial biogeography of North American coho salmon (*Oncorhynchus kisutch*). *Molecular Ecology* **10**, 2775-2785.
- Springer AM, van Vliet GB (2014) Climate change, pink salmon, and the nexus between bottom-up and top-down forcing in the subarctic Pacific Ocean and Bering Sea. *Proc Natl Acad Sci U S A*.

- Takezaki N, Nei M, Tamura K (2010) POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* **27**, 747-752.
- Templin WD, Seeb JE, Jasper JR, Barclay AW, Seeb LW (2011) Genetic differentiation of Alaska Chinook salmon: the missing link for migratory studies. *Mol Ecol Resour* **11 Suppl 1**, 226-246.
- Toge K, Yamashita R, Kazama K, *et al.* (2011) The relationship between pink salmon biomass and the body condition of short-tailed shearwaters in the Bering Sea: can fish compete with seabirds? *Proceedings of the Royal Society B: Biological Sciences* **278**, 2584-2590.
- Turner CE, Bilton HT (1968) Another pink salmon (*Oncorhynchus gorbuscha*) in its third year. *Journal of Fisheries Research Board of Canada* **25**, 1993-1996.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
- Wenink PW, Baker AJ, Rosner H-U, Tilanus MGJ (1996) Global mitochondrial DNA phylogeography of Holarctic breeding dunlins (*Calidris alpina*). *Evolution* **50**, 318-330.
- Wickham H (2009) *ggplot2: elegant graphics for data analysis*. URL : <http://ggplot2.org/> Springer-Verlag, New York.
- Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.
- Zigouris J, Schaefer JA, Fortin C, Kyle CJ (2013) Phylogeography and post-glacial recolonization in wolverines (*Gulo gulo*) from across their circumpolar distribution. *PloS one* **8**, e83837.

## 7. Tables

**Table 1.** Location and collection year information for even- and odd-year populations of pink salmon.

<b>Location</b>	<b>River</b>	<b>Collection Year</b>		<b>Lat.</b>	<b>Long.</b>
		<b>even</b>	<b>odd</b>		
<b>Hokkaido</b>	Kushiro River	2006	2007	42.980	144.378
<b>Amur</b>	Amur River	2010	2011	52.938	139.727
<b>Magadan</b>	Tauy River	2012	2009	59.714	148.929
<b>Kamchatka</b>	Haylylulya River	2010	2009	57.769	162.485
<b>Norton Sound<sup>1</sup></b>	Nome River	1994	1991	64.483	-165.301
<b>Prince William Sound<sup>1</sup></b>	Koppen Creek	1996	1991	60.706	-145.898
<b>Puget Sound<sup>1</sup></b>	Snohomish River	1996	2003	48.014	-122.181

<sup>1</sup> From Seeb *et al.* (2014)

**Table 2A.** Population genetic statistics for samples of even- and odd-year populations of pink salmon, including final sample number, rarefied allelic richness, mean observed heterozygosity ( $H_o$ ), mean expected heterozygosity ( $H_e$ ), and mean individual standardized heterozygosity ( $H_{S\_obs}$ ).

Location	Final Sample Number		Allelic Richness		Mean $H_o$		Mean $H_e$		Mean $H_{S\_obs}$	
	even	odd	even	odd	even	odd	even	odd	even	odd
<b>Hokkaido</b>	32	32	1.595	1.645	0.159	0.170	0.160	0.170	0.968	1.037
<b>Amur</b>	32	32	1.625	1.677	0.161	0.171	0.163	0.172	0.988	1.051
<b>Magadan</b>	27	28	1.623	1.677	0.161	0.171	0.161	0.171	0.986	1.048
<b>Kamchatka</b>	31	31	1.619	1.679	0.160	0.172	0.161	0.171	0.983	1.056
<b>Norton Sound</b>	18	24	1.617	1.678	0.153	0.170	0.160	0.172	0.937	1.045
<b>Prince William Sound</b>	24	24	1.593	1.623	0.161	0.162	0.160	0.165	0.988	0.996
<b>Puget Sound</b>	24	24	1.558	1.555	0.149	0.154	0.154	0.156	0.910	0.939
<b>Overall Mean</b>	27	28	1.604	1.648	0.158	0.167	0.160	0.168	0.966	1.024

**Table 2B.** Population genetic statistics for samples of even- and odd-year lineage populations of pink salmon, including inbreeding coefficient ( $F_{IS}$ ) and estimates of effective population size ( $N_e$ ) with a 95% confidence interval.

Location	Mean $F_{IS}$		Estimated $N_e$ (95% CI of $N_e$ )	
	even	odd	even	odd
<b>Hokkaido</b>	0.011	0.006	965 (917 – 10,167)	1,699 (1,571 – 1,850)
<b>Amur</b>	0.011	0.001	20,731 (10,088 – Infinite)	7,051 (52,545 – 10,709)
<b>Magadan</b>	0.003	-0.001	6,673 (4,660 – 11,736)	20,321 (9,348 – Infinite)
<b>Kamchatka</b>	0.004	-0.004	56,866 (13,291 – Infinite)	18,847 (94,523 – 2,505,396)
<b>Norton Sound</b>	0.042	0.009	16,395 (6,094 – Infinite)	5,143 (3,777 – 8,050)
<b>Prince William Sound</b>	-0.007	0.015	3,990 (3,064 – 5,714)	2,565 (2,153 – 3,171)
<b>Puget Sound</b>	0.032	0.017	1,110 (1,010 – 1,232)	5,916 (3,992 – 11,412)

**Table 3.** Estimates of  $F_{ST}$  for: **A.** all populations, **B.** by lineage: the even- and odd-year lineages and groups of populations within each lineage, and **C.** pairwise by location. Confidence intervals are based on 1,000 bootstraps.

<b>Comparison</b>	<b>No. Populations</b>	<b><math>F_{ST}</math></b>	<b>95% CI</b>
<b>A. Over all populations</b>	14	0.060	(0.059 - 0.062)
<b>B. By lineage</b>			
Odd	7	0.033	(0.032 - 0.035)
Odd Beringia	5	0.008	(0.007 - 0.009)
Odd North America	2	0.052	(0.049 - 0.053)
Even	7	0.026	(0.025 - 0.027)
Even Beringia	5	0.011	(0.010 - 0.012)
Even North America	2	0.030	(0.029 - 0.033)
<b>C. Pairwise by location</b>			
Hokkaido	2	0.088	(0.084 - 0.089)
Amur	2	0.063	(0.062 - 0.066)
Magadan	2	0.060	(0.058 - 0.062)
Kamchatka	2	0.060	(0.059 - 0.063)
Norton Sound	2	0.062	(0.059 - 0.064)
Prince William Sound	2	0.094	(0.089 - 0.095)
Puget Sound	2	0.148	(0.143 - 0.145)

**Table 4.** Results of AMOVA comparing the amount variation between lineages and between groups of populations.

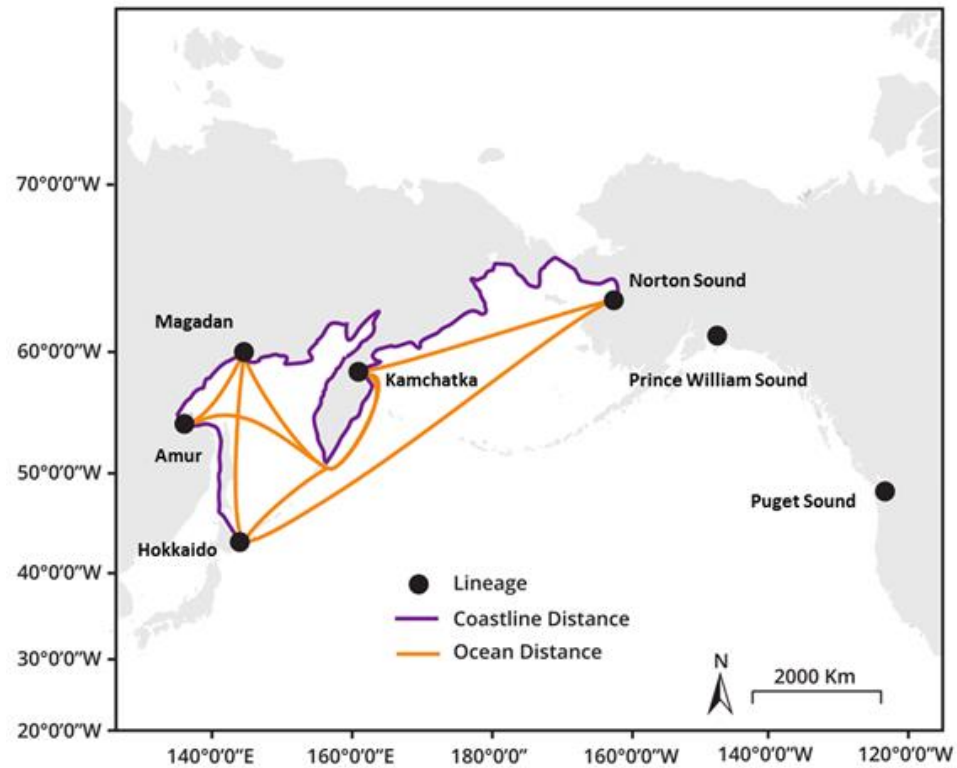
AMOVA	Source of Variation	Amount of Variation (%)	Fixation Indices
<b>A. Between lineages</b>	Between lineages	<b>5.72</b>	$F_{ST}$ 0.086**
	Among populations within lineages	<b>2.87</b>	$F_{SC}$ 0.030**
	Within populations	91.41	$F_{CT}$ 0.057**
<b>B. Between even lineage groups</b>	Between groups	<b>2.82</b>	$F_{ST}$ 0.042**
	Among populations within groups	<b>1.40</b>	$F_{SC}$ 0.014*
	Within populations	95.78	$F_{CT}$ 0.028**
<b>C. Between odd lineage groups</b>	Between groups	<b>4.13</b>	$F_{ST}$ 0.056**
	Among populations within groups	<b>1.46</b>	$F_{SC}$ 0.015*
	Within populations	94.40	$F_{CT}$ 0.041**

\*significant at  $P < 0.05$

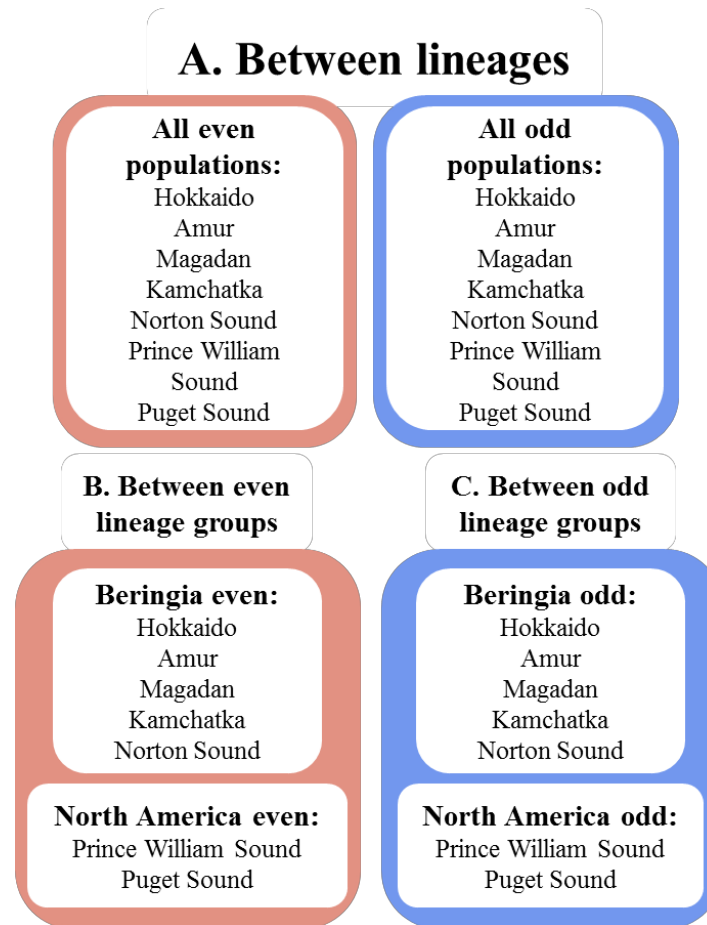
\*\*significant at  $P < 0.001$

## 8. Figures

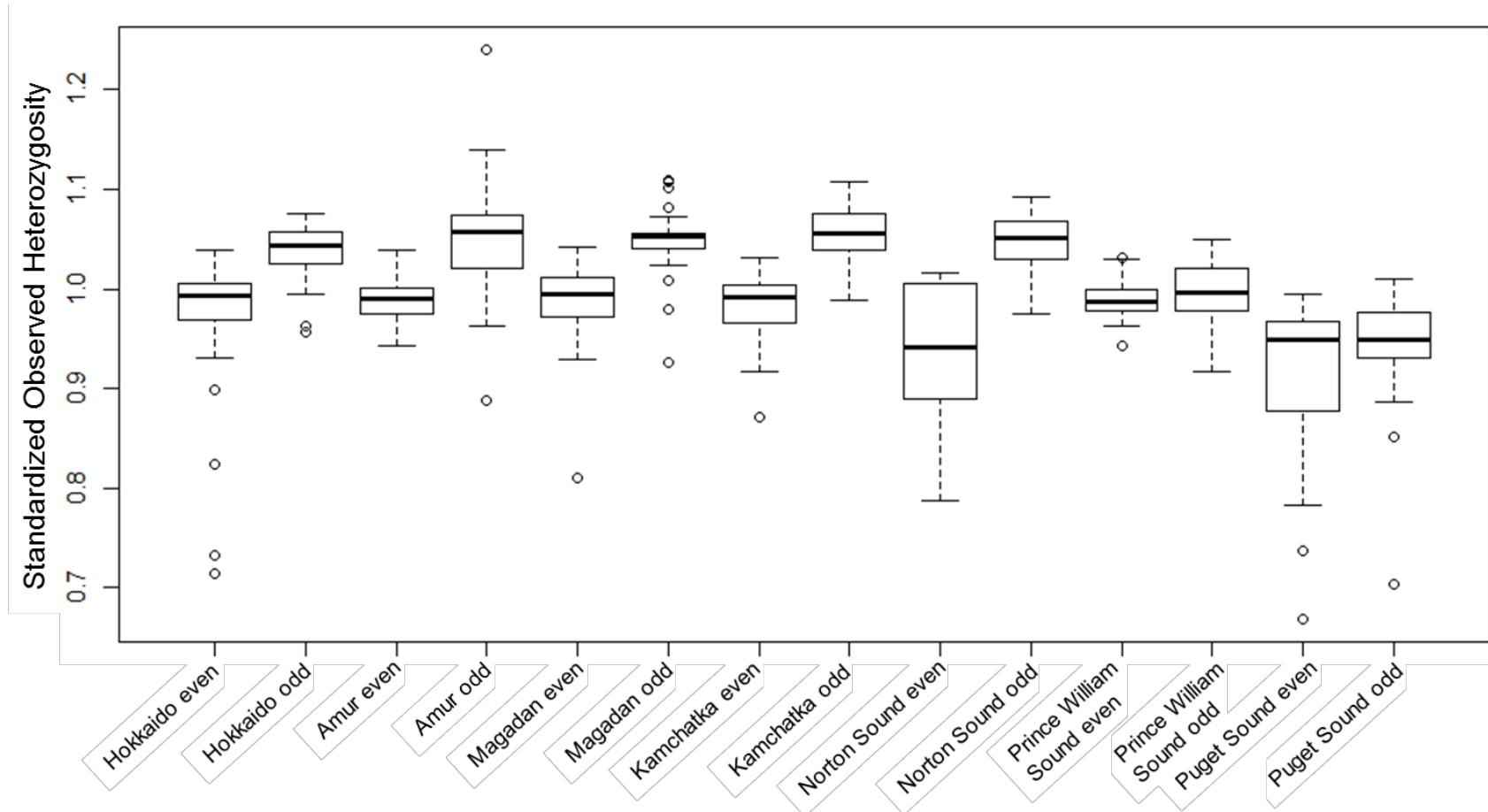
**Figure 1.** Map depicting the two methods of measuring the distance between locations. The coastline distance between each location is shown in purple; the distance between locations over water is shown in orange.



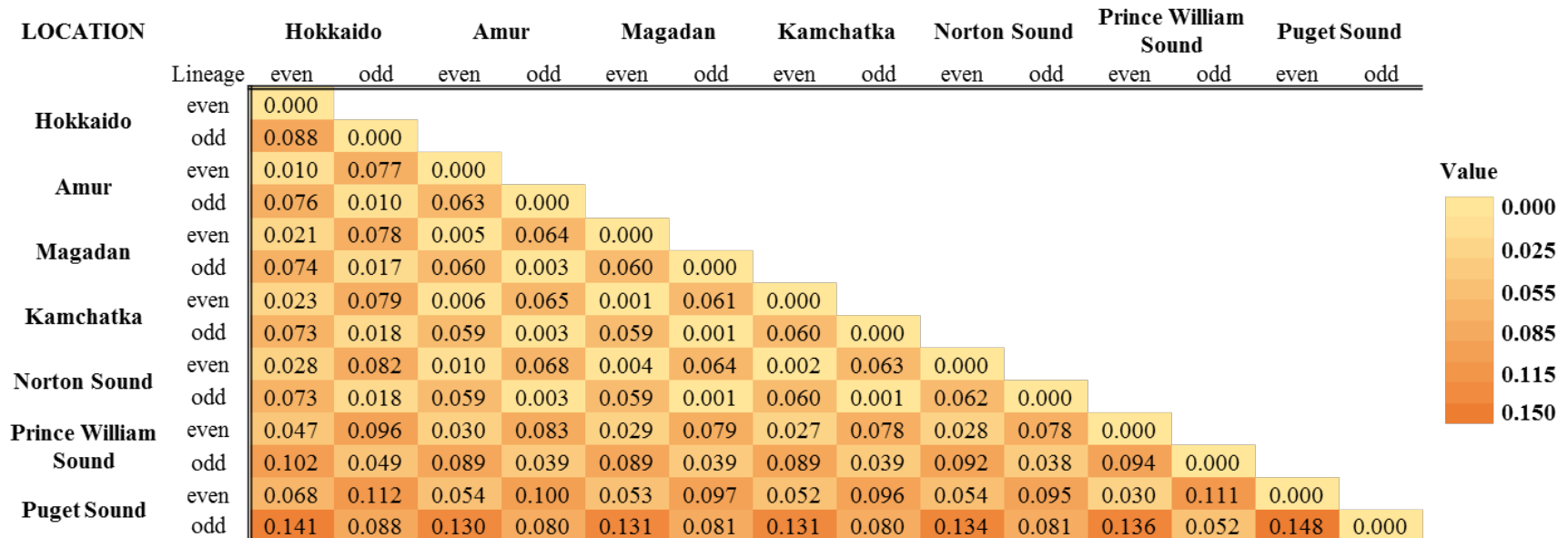
**Figure 2.** The hierarchical grouping of populations for three AMOVAs to test for variation between lineages and between groups of populations.



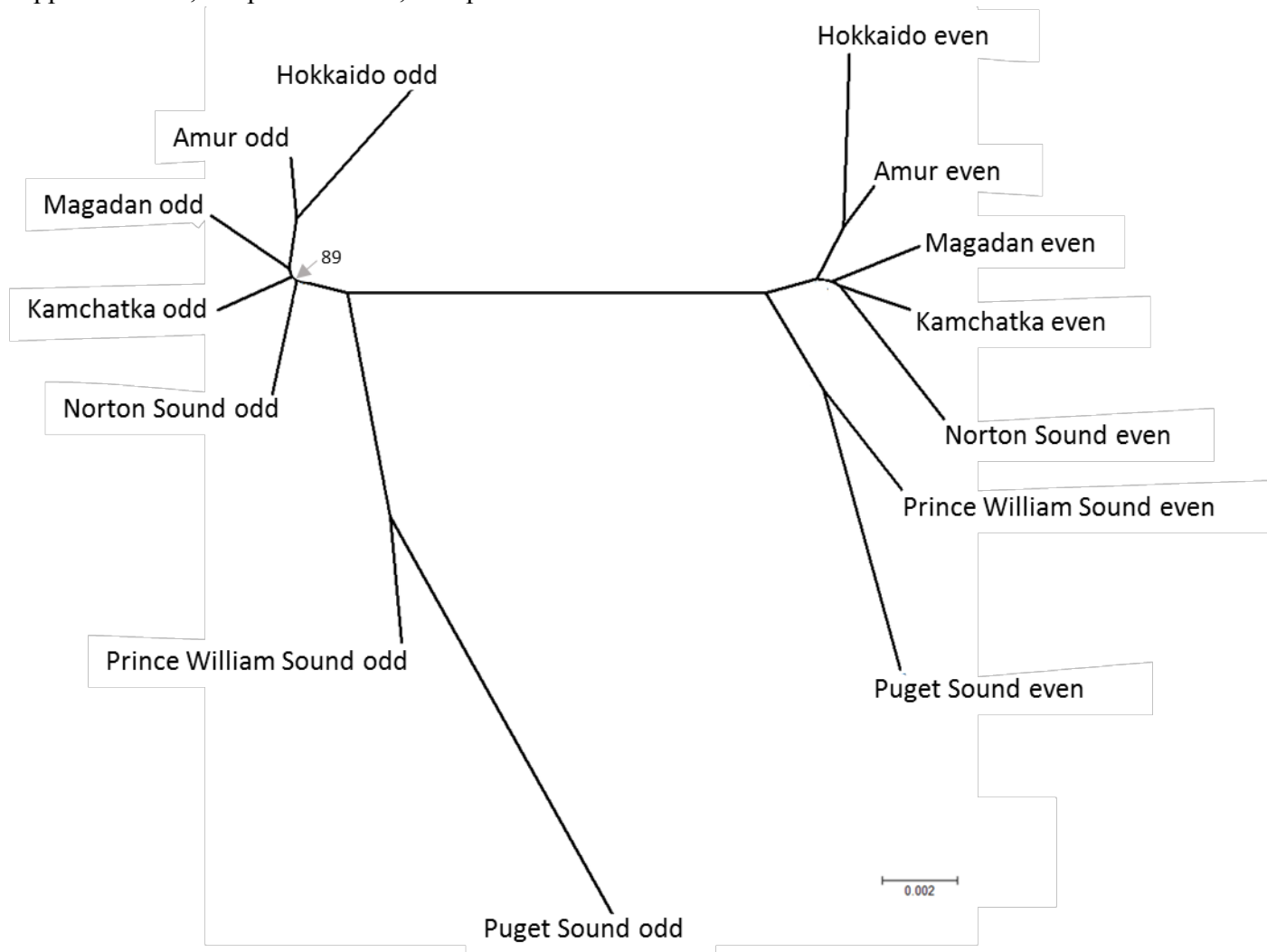
**Figure 3.** Boxplot of individual standardized observed heterozygosity ( $H_{s\_obs}$ ) across loci within each populations, for all populations.  $H_{s\_obs}$  was measured as the proportion of heterozygotes standardized by the mean observed heterozygosity of the genotyped loci. For each population, the thick black lines show the median values and the box depicts the quartiles of the individual values. Whiskers show 1.5x the interquartile range or the maximum value, with circles representing individuals with  $H_{s\_obs}$  that fall outside of that range.



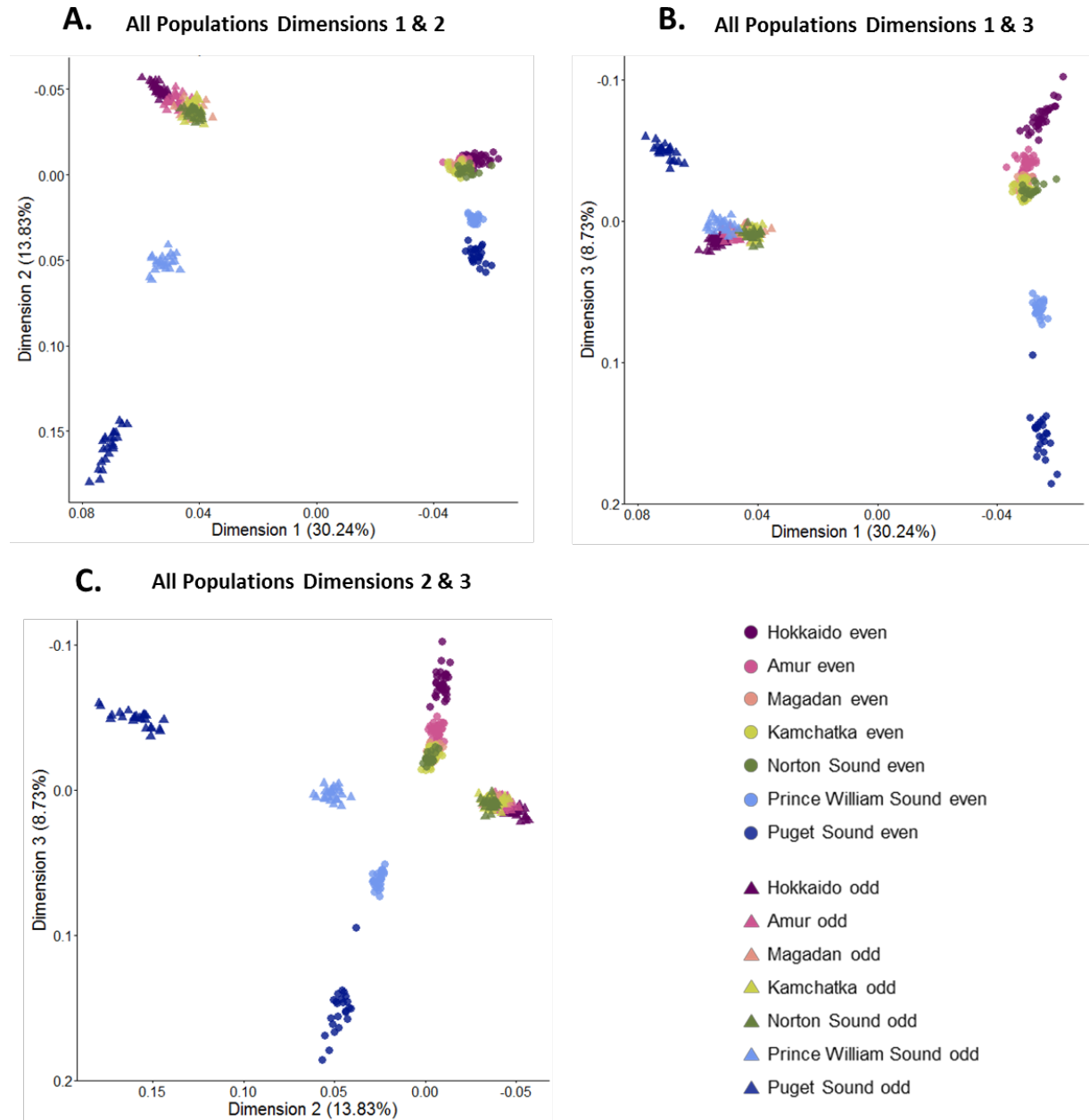
**Figure 4.** Comparisons of pairwise  $F_{ST}$  between all fourteen populations. Populations are ordered geographically, from the southern Asian populations to the Southern North American populations. Conditional formatting highlights the pairwise comparisons with the greatest  $F_{ST}$ .



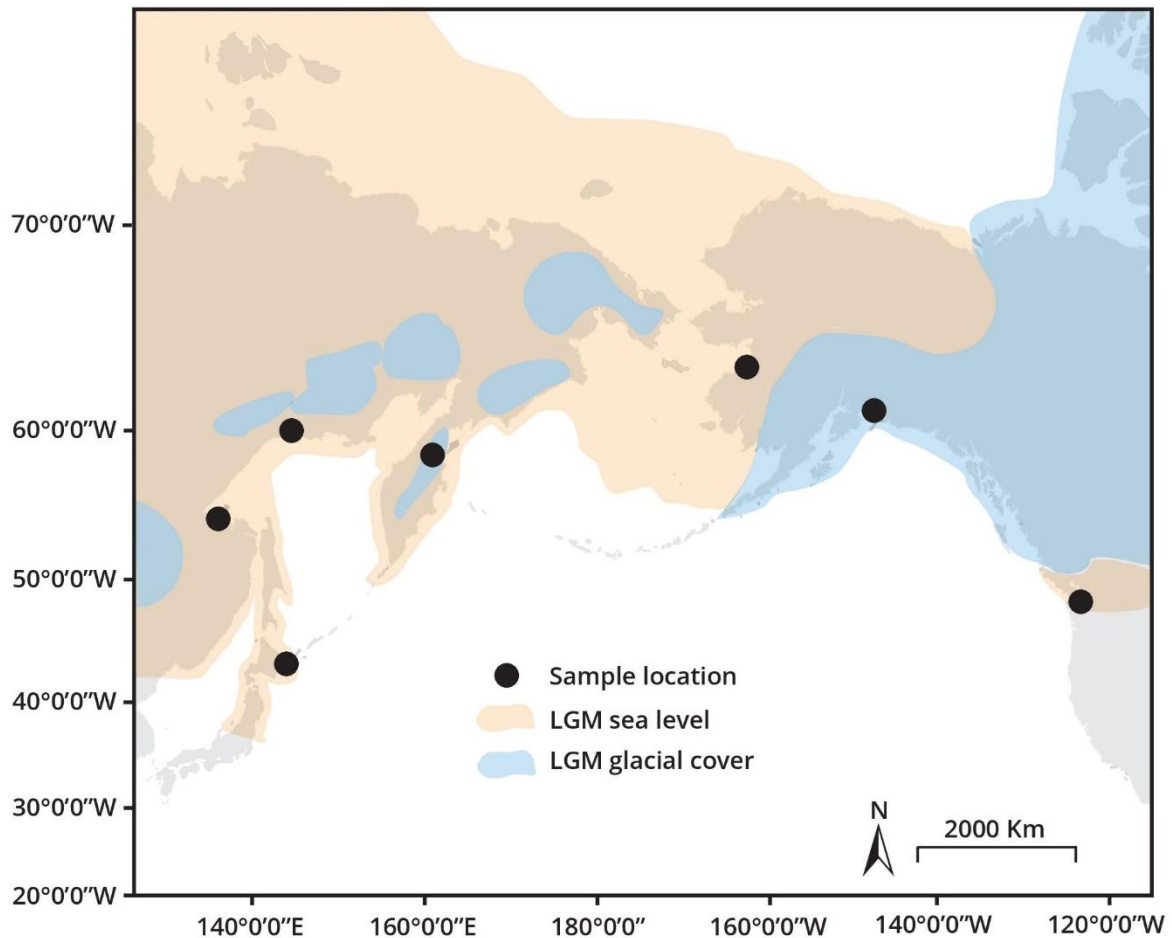
**Figure 5.** Unrooted Neighbor-Joining tree based on  $D_A$  distance (Nei 1983) using 16,681 SNPs. All branches had 100% bootstrap support with 10,000 permutations, except where noted.



**Figure 6.** Individual based PCA exploring the relationship between all 14 populations: **A)** dimensions 1 vs 2 **B)** dimensions 1 vs 3 **C)** dimensions 2 vs 3. Even-year lineage represented by the circles and the odd-year lineage represented by the triangles, with each location denoted by a different color. The percent of variation explained by each axis is in parenthesis.



**Figure 7.** Map of the present day coastline (grey) overlaid with the exposed land during lowered sea levels (-120 to -130m lower than present, orange) and glacial cover (blue) during the LGM, approximately 18-26 thousand years ago. Historical coastline adapted from Grant *et al.* (2012), Brigham-Grette *et al.* (2003b) and Pielou (1991). Glacial cover adapted from Shafer *et al.* (2010), Ehlers and Gibbard (2007), and Pielou (1991).



## Supplemental Files

**File 1.** Detailed information for 396 population samples.

**File 2.** Sequence information for 16,681 loci.