# Y-chromosome haplotypes are associated with variation in size and age at maturity in male Chinook salmon

Garrett J. McKinney[1a*], James E Seeb[1b], Carita E. Pascal[1c], Daniel E. Schindler[1d], Sara E. Gilk-Baumer[2e], Lisa W. Seeb[1f]

[1] School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle WA 98195-5020, USA.

[2] Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, 99518, USA

[*]**Corresponding author:** Garrett J. McKinney (email: gjmckinn@uw.edu, phone: 1-765-430-3272). Current address: NOAA National Marine Fisheries Service, Northwest Fisheries Science Center, 2725 Montlake Blvd E, Seattle, WA 98112

Emails: [a] gjmckinn@uw.edu; [b]jseeb@uw.edu; [c]cpascal@uw.edu; [d] deschind@uw.edu; [e]sara.gilk@alaska.gov ; [f]lseeb@uw.edu

# ABSTRACT

Variation in size and age at maturity is an important component of life history that is influenced both by environmental and genetic factors. In salmonids, large size confers a direct reproductive advantage through increased fecundity and egg quality in females, while larger males gain a reproductive advantage by monopolizing access to females. In addition, variation in size and age at maturity in males can be associated with different reproductive strategies; younger smaller males may gain reproductive success by sneaking in among mating pairs. In both sexes there is a trade-off between older age and increased reproductive success and increased risk of mortality by delaying reproduction. We identified four Y-chromosome haplogroups that showed regional and population-specific variation in frequency using RADseq data for 21 populations of Alaska Chinook salmon. We then characterized the range-wide distribution of these haplogroups using GT-seq assays. These haplogroups exhibited associations with size at maturity in multiple populations suggesting that the lack of recombination between X and Y-chromosomes has allowed Y-chromosome haplogroups to capture different alleles that influence size at maturity. Ultimately, conservation of life history diversity in Chinook salmon may require conservation of Y-chromosome haplotype diversity.


# Introduction

Variation in life history within populations is common across taxa and is often associated with alternative strategies for increasing fitness. This includes partial migration, where some individuals of a population migrate while others remain resident (Chapman et al. 2011), reproductive morphs that exhibit different mating strategies or sexually selected traits (Shuster 1989, Johnston et al. 2013, Küpper et al. 2015), age and size at maturity (Gibbons et al. 1981), and even length of life-span such as annual versus perennial plants (Hall et al. 2006). While life history variation is often assumed to be under the influence of many genes of small effect, examples of a large-effect genes and supergenes (sets of linked genes) influencing life history variation are increasingly being found. These include single genes that influence age at maturity (Barson et al. 2015) and sexually selected traits (Johnston et al. 2013); these also include chromosome inversions contributing to annual versus perennial life history (Twyford and Friedman 2015) and variation in migration versus residency (Pearse et al. 2014). The mechanisms underlying variation in life history have important implications for how this important diversity is maintained under different selective regimes.

50 Variation in life history strategies is exhibited by many salmon species with size and age at
51 maturity being an important component of this variation. In females, large body size confers a direct
52 reproductive advantage through increased fecundity and egg quality (Healey and Heard 1984). In
53 contrast, variation in size and age at maturity in males can be associated with different reproductive
54 strategies that exhibit frequency dependent fitness. For example, older larger males gain reproductive
55 success by monopolizing access to females while younger smaller males gain reproductive success by
56 sneaking in among mating pairs (Healey and Heard 1984, Berejikian et al. 2010). For both sexes, there is
57 a trade-off to delayed maturation where increased reproductive success is tempered by an increased
58 risk of mortality before reproduction. The optimal age at maturity for a population should represent the
59 balance between reproductive benefits and mortality costs of delayed maturation (Healey 1986), and
60 forces that change these costs or benefits could result in shifts in age composition. Nonetheless,
61 considerable diversity in age at maturation is maintained within many populations, presumably as a bet-
62 hedging strategy to spread risks over the life-cycle of these fish.

63 Age at maturity in salmon is generally thought to be a threshold trait that is dependent either
64 upon reaching a minimum size at age or upon growth rate at key periods (Healey 1991, Thorpe 2007).
65 Environmental factors that influence growth rate have been shown to influence age at maturity in many
66 species. In the wild, studies show correlations between ocean conditions such as temperature and
67 productivity and patterns of age at maturity (Otero et al. 2012, Siegel et al. 2017). In experimental
68 settings, age at maturity was manipulated through changing temperature (Heath et al. 1994, Harstad et
69 al. 2018) or food ration (Rowe and Thorpe 1990, Larsen et al. 2006).

70 In addition to environmental effects, multiple lines of evidence show a genetic component to
71 age at maturity. High heritability values suggest considerable genetic variation for age at maturity in
72 several salmon species (Gjerde 1984, Gall et al. 1988, Hankin et al. 1993, Heath et al. 1994), and
73 Quantitative Trait Locus (QTL) and Genome-Wide Association (GWAS) studies identified genomic regions
74 associated with variation in age at maturity (Barson et al. 2015, Kodama et al. 2018, Micheletti and
75 Narum 2018, Waters et al. 2018). Studies also demonstrated that offspring of alternative male
76 phenotypes exhibit different growth rates (Garant et al. 2002, Berejikian et al. 2011) with offspring of
77 early maturation phenotypes (grilse and jacks) exhibiting high growth rates. In Atlantic salmon (*Salmo*
78 *salar*), individuals with different life histories exhibit differing maturation thresholds that are genetically
79 based (Aubin Horth and Dodson 2004). Despite these findings, the genetic basis of maturation age in
80 most salmonids remains poorly understood.

81 One complicating factor is that the genetic architecture underlying variation in age at maturity
82 appears to vary among salmonid species and even among populations within a species. In Atlantic
83 salmon, age at maturity is strongly influenced by a single gene (VGLL3) (Ayllon et al. 2015, Barson et al.
84 2015). This gene exhibits sex-dependent dominance which facilitates sexually antagonistic selection
85 (Barson et al. 2015). While this gene explained 39% of the phenotypic variability in European Atlantic
86 salmon, studies in North American Atlantic salmon have shown that this association varies by population
87 (Kusche et al. 2017, Boulding et al. 2019) and this gene has not been shown to have an effect in Pacific
88 salmon (genus *Oncorhynchus)* (Micheletti and Narum 2018).

89 Chinook salmon (*O. tshawytscha*) are the largest of the Pacific salmon and follow various life
90 history strategies, spending 0 to 2 years in fresh water and 1 to 4 or more years in the ocean (Riddell et
91 al. 2018). Male Chinook salmon exhibit significant variation in size and age at maturity (Healey 1991)
92 that is linked to differential reproductive tactics and is likely controlled by both environmental and
93 genetic components (Berejikian et al. 2010, Young et al. 2013). Many populations throughout North
94 America recently experienced marked declines in size and age at maturity which may erode life history
95 variation (Lewis et al. 2015, Ohlberger et al. 2018). Explanations for decreased age at maturity have
96 focused on the impacts of fisheries induced evolution (Hard et al. 2008, Kendall et al. 2014) or changing
97 environmental conditions (Siegel et al. 2017); however, these factors alone are not consistent or
98 sufficient to explain current declines and it is likely that these declines are driven by multiple complex
99 factors (Ohlberger et al. 2018). While the genetic control of age at maturity in Chinook salmon is still
100 poorly understood, past studies offer clues to genomic regions that may be associated with maturation
101 age. In particular, Heath et al. (2002) identified a strong sex-linked component to age at maturity in
102 Chinook salmon, suggesting the influence of genes on the Y-chromosome.

103 The X and Y chromosomes in most salmonid species are morphologically undifferentiated
104 (Davidson et al. 2009). Available sequence data suggests that the primary difference between sex
105 chromosomes is an insertion containing the sex-determining gene (SDY, Yano et al. 2013), and in
106 Chinook salmon a 2.4 Mb male-specific repetitive sequence has been identified (Devlin et al. 1998).
107 While the sex-determining gene has been assigned to chromosome 17 (Ots17) in Chinook salmon
108 (Phillips et al. 2013), this region is not in either of the current genome assemblies (Christensen et al.
109 2018, Narum et al. 2018) so the exact location of *sdY* is unknown.

110 Despite a lack of large-scale differentiation, the X- and Y-chromosomes could show sequence
111 divergence due to sex-specific patterns of recombination. Recombination in females takes place along

the full length of the chromosome while recombination in males is strongly localized to telomeric regions (Lien et al. 2011), restricting recombination between the X- and Y-chromosomes. Reduced recombination between sex chromosomes is supported by a 33Mb signal of sex association observed in Atlantic salmon (Kijas et al. 2018) and a lack of recombination between the sex-determining region and an allozyme locus on the sex-chromosome in Chinook salmon (Marshall et al. 2004). In addition to facilitating divergence between sex chromosomes, sex-limited recombination could lead to the formation of different Y-chromosome haplotypes and the capture of adaptive genetic variants. If Y-chromosome haplotypes have sufficiently diverged from the X-chromosome, they may be identified through patterns of extended linkage disequilibrium.

We examined patterns of linkage disequilibrium on the sex chromosome of Chinook salmon to determine if male-specific haplotype blocks (Y-chromosome haplotypes) existed, and if so, are these haplotypes associated with variation size and age at maturity which commonly differ between sexes in Chinook salmon. We identified four Y-chromosome haplogroups in Chinook salmon from Alaska that showed regional and population-specific variation in frequency. These haplogroups showed associations with size at maturity in multiple populations suggesting that the lack of recombination between X and Y-chromosomes has allowed Y-chromosome haplogroups to capture different genetic variants influencing size and age at maturity.

# Materials and Methods

## RAD Y-chromosome Haplotypes

We used existing RADseq data to examine patterns of genetic variation in the sex chromosome. RADseq data for 21 populations of Alaska Chinook salmon were obtained from: NCBI SRA accessions SRP034950 (Larson et al. 2014) and SRP129894 (McKinney et al. 2018), bioproject PRJNA560365 (McKinney et al. 2020b), and raw data used in (Dann et al. 2018). Raw data from (Dann et al. 2018) are available from those authors upon request. Locations of populations ranged from Cook Inlet to the Upper Yukon River and include a total of 1,082 samples (Figure 1, Table 1). RADseq data were processed with Stacks V1.7 (Catchen et al. 2011, Catchen et al. 2013) using default settings with the following exceptions: process_radtags (-c -r -q -filter_illumina -t 94), ustacks (-m 2 -M 2, --model_type bounded --bound_high 0.05), cstacks (-n 2). The catalog of variation from McKinney et al. (2020b) was used for

141 consistent RADtag names among this and previous studies. A total of six individuals per population from
142 Cook Inlet was added to this catalog to allow for additional allelic variation.

143 RADseq results were filtered to remove SNPs that were likely to genotype poorly or be
144 uninformative including collapsed paralogs, those with high missing data, and SNPs with low minor allele
145 frequency. Paralogs are common in salmonid genomes due to an ancestral whole-genome duplication
146 (Allendorf and Thorgaard 1984). With short read sequence the two copies of paralog are often
147 collapsed into a single locus that appears polyploid; these cannot be reliably genotyped in typical
148 RADseq studies because of insufficient read depth (McKinney et al. 2018). Collapsed paralogs were
149 identified using *HDplot* (McKinney et al. 2017) and removed from further analysis. Loci with more than
150 10% missing data or a minor allele frequency (MAF) < 0.01 were also removed from analysis. Finally, loci
151 were aligned to the Chinook salmon genome (Christensen et al. 2018) using Bowtie2 (Langmead and
152 Salzberg 2012) to determine genomic position; only loci that aligned to the sex chromosome (Ots17,
153 Phillips et al. 2013) with fewer than four mismatches were retained for analysis.

154 We identified putative Y-chromosome haplotypes by examining patterns of linkage
155 disequilibrium on Ots17 using network analysis. We then phased genotypes for sets of high LD loci into
156 haplotypes representing the alleles on each chromosome pair within an individual and clustered these
157 haplotypes based on similarity into haplogroups which are groups of similar haplotypes. Network
158 analysis of linkage disequilibrium is an effective approach for identifying genomic structures that exhibit
159 high LD (Kemppainen et al. 2015) and had been used to identify distinct but overlapping haplotype
160 blocks in chum salmon (McKinney et al. 2020a). Pairwise linkage disequilibrium was calculated using the
161 $r^2$ method in *Plink* (V1.9) (Purcell et al. 2007, Chang et al. 2015) and SNP pairs with $r^2 \geq 0.3$ were retained
162 for network analysis. Network analysis was conducted using the igraph package in R
163 (https://igraph.org/r) to identify sets of loci with high LD. Sets of loci that contained at least five SNPs
164 and spanned at least 1 Mb were retained for further analysis. Genotypes for retained SNPs were then
165 phased into haplotypes using fastPhase (Scheet and Stephens 2006). Putative Y-chromosome
166 haplogroups were identified by clustering haplotypes using heatmap2 in R (Warnes et al. 2020) with the
167 Ward.D clustering algorithm to minimize within group variance. Haplogroups that appear to be male-
168 specific are hereafter referred to as Y-chromosome haplogroups and assigned names based on the
169 following convention: chromosome number, MH for male haplogroup, followed by a sequential number,
170 e.g. Y-chromosome haplogroups on chromosome 17 would be Ots17-MH1, Ots17-MH2, and so on.

Allele frequencies within each haplogroup were visualized using logo plots with the ggseqlogo package in R (Wagih 2017).

Assays for SNPs that were diagnostic for Y-chromosome haplotypes were assembled into an amplicon panel (GT-seq , Campbell et al. 2015) for expanded genotyping (see methods in McKinney et al. 2020b). Primers were designed using batch primer3 (You et al. 2008). Primer design used the consensus RAD sequence for each RADtag unless SNPs occurred within 20 bp of the end of the RADtag. In these cases the consensus sequences were aligned to the genome using bowtie2 (Langmead and Salzberg 2012), and genomic sequence flanking the RADtag was added to the consensus sequence for primer design. Panel optimization followed the methods of McKinney et al. (2020b); one round of sequencing using 80 individuals was conducted to identify loci that over-amplified or produced unreliable genotypes. Amplification levels and genotyping accuracy were assessed using GT-score (McKinney et al. 2020b).

## Expanded genotyping

Additional samples ranging from Alaska to California were genotyped using the GT-seq panel to establish the geographic distribution of the haplogroups (Table 1). In addition to the Y-haplogroup markers, this panel included the sex identification marker Ots_sexy3-1 (Hess et al. 2016a) to confirm that Y-chromosome haplotypes were present only in male fish. A total of 1,341 samples from 27 populations were genotyped (Table 1); phenotypic sex was known for 193 of these. A total of 94 RADseq samples were included in the GT-seq genotyping to examine concordance between the two datasets. Samples were sequenced on a HiSeq 4000, and data were processed and genotyped using GT-score (McKinney et al. 2020b) available at https://github.com/gjmckinney/GTscore.

Two methods were used to assign GT-seq samples to Y-chromosome haplogroups. First, samples were assigned Y-chromosome haplogroups using the same methods as the RADseq samples: phasing genotypes into haplotypes using fastPhase followed by haplotype clustering using heatmap2. Genotypes from RADseq samples were included in the GT-seq haplotype assignment to assess concordance between the discriminatory ability of the full RADseq marker set and the subset that that were successfully developed into GTseq markers. Second, the expected genotype patterns for males with each haplogroup were constructed assuming fixation of alleles on the X-chromosome and the Y-chromosome. The observed genotypes were then compared to the expected genotypes. Samples were assigned to a haplogroup if the observed genotypes had less than two mismatches to the expected genotypes.

## Y-chromosome haplotype analyses

Assignments were combined for GT-seq and RADseq samples to characterize the distribution and frequency of Y-chromosome haplogroups. Data for length and age at maturity were obtained from Alaska Department of Fish and Game (ADFG) for a subset of populations in Alaska and compared with Y-chromosome haplogroup data to determine if there were relationships between Y-chromosome haplogroup and length and age at maturity. Analysis of variance (ANOVA) was used to test the significance of associations between Y-chromosome haplogroups and length; population of origin was included as a covariate. A post-hoc Tukey test was then performed to determine if differences in size distribution were significant between individual haplogroups. The significance of associations between Y-chromosome haplogroups and age at maturity in the Lower Yukon Test Fishery were assessed using an ANOVA with population of origin added as a covariate. There are multiple methods of reporting age in salmon (Koo 1962); we report freshwater and ocean age for each individual using European notation, so an individual with an age of 1.3 would have spent 1 year growing in freshwater after emergence followed by 3 years in the ocean for a total age of 5 years.

# Results

A total of 448 SNPs remained after filtering SNPs with more than 10% missing data or a MAF ≤ 0.01, removing paralogs identified by HDplot, and retaining only SNPs that aligned to chromosome 17. Retained SNPs had an average coverage of 24x.

## RADseq Y-haplotype discovery

Network analysis of linkage disequilibrium identified three sets of linked SNPs exhibiting long-distance LD (4.5-9.5 Mb each). There were 35 SNPs in these high LD sets, representing 7.8% of the SNPs on Ots17 retained after filters. Individual genotypes for these SNPs were then phased, resulting in two haplotypes per individual. Clustering of haplotypes resulted in five major groupings that exhibited extended LD up to 17.4 Mb (Figure 2). Although phenotypic sex was available from only a subset of individuals, phased haplotypes from sexed individuals were present in each of these groupings (Table 2). One of the groupings contained phased haplotypes from both males and females suggesting that this represents the X-chromosome (gray cluster, Figure 2). Four of the haplogroups contained phased haplotypes predominantly from male samples (85%-95%), suggesting that these are from the Y-chromosome (Table 2, Figure 2). In addition, all individuals with a haplotype in one of the Y-chromosome haplogroups had their second haplotype in one of the X-chromosome groupings. All Y-

chromosome haplogroups were present, at varying frequency, in each geographic region, suggesting that these haplogroups are conserved throughout this broad geographic region (Figure 1, Table 1, Table 3). Within Alaska, Y-chromosome haplogroups showed regional variation in frequency with the Ots17-MH1 haplotype exhibiting the high frequencies in most parts of Western Alaska and the Ots17-MH4 haplotype the most frequent in Cook Inlet (Figure 1). While some females had haplotypes within the Y-chromosome haplogroups, individuals had been visually sexed based on external features which is known to have variable accuracy (Lozorie and McIntosh 2014).

Allele frequencies within each haplogroup were visualized using logo plots (Figure 3). A total of 35 SNPs was found to have fixed or nearly-fixed differences in allele frequencies between Y-chromosome haplogroups and the X-chromosome. The Ots17-MH1, Ots17-MH2, and Ots17-MH3 haplogroups shared seven of the SNPs that differentiated these haplogroups from the X-chromosome (27492_30, 42530_30, 57320_53,64766_37,67724_90,5884_57,94509_60), suggesting a common evolutionary origin. The Ots17-MH4 haplogroup shared no diagnostic SNPs with the other Y-chromosome haplogroups. Several of the SNPs had allelic variation within the X-chromosome, (e.g., SNPs 57320_53 and 64766_37, Figure 3) suggesting that in some cases the Y-chromosome haplogroups have partitioned existing variation on the X-chromosome. Alternatively, some SNPs exhibited no allelic variation within the X-chromosome suggesting that alternative alleles observed on the Y-chromosome haplotypes may be novel genetic variants (e.g. SNP 71572_67, Figure 3).

## GT-seq Y-chromosome expanded sampling

A GT-seq panel was developed to genotype Y-chromosome haplotype markers for a set of samples representing the North American range of Chinook salmon. A total of 23 of the 35 RAD SNPs passed filtering criteria prior to primer design; of these 16 RAD SNPs were successfully converted to GT-seq assays (Figure 3, Table S1). Samples genotyped with GT-seq were assigned haplogroups using two methods, clustering of haplotypes using the Ward.D algorithm and assignment based on genotypes. Haplogroups Ots17-MH1 and Ots17-MH2 differed by several fixed SNPs in the RADseq dataset but only two fixed SNPs in the GT-seq panel (Figure 3). For these haplogroups an additional requirement was set that individuals must have the appropriate alleles for each of these fixed SNPs. For example, individuals in the Ots17-MH2 haplogroup must have a T allele at position 6491379 and a G allele at position 9698184.

Samples that were genotyped with both RADseq and GT-seq showed high concordance between RADseq and both GT-seq haplogroup assignment methods except for Ots17-MH3 (Table S2). When

RADseq haplogroup assignment was compared with GT-seq haplogroup assignment using Ward.D clustering, 10 of the 12 samples assigned to RADseq Ots17-MH3 were grouped with females. This is likely due to the fact that only three of the RADseq SNPs that differentiated this haplotype from the X-chromosome were successfully developed into GT-seq (Figure 3). In contrast, when GTseq samples were assigned to haplogroups based on comparing observed to expected genotypes for each haplogroup, 10 of the 12 samples correctly assigned to Ots17-MH3 while two were not assigned to any haplogroup. For samples genotyped only with GT-seq, there were no discrepancies in haplogroup assignment between the clustering and genotype-based methods except that Ots17-MH3. Ots17-MH3 had no individuals assigned using haplotype clustering but had 23 samples assigned based on genotype matching. Genotype-based matching assigned approximately 12% fewer individuals to haplotypes overall than the phased haplotype clustering (Table S2C) but was better able to assign individuals to the Ots17-MH3 haplogroup. All GT-seq samples assigned to a Y-chromosome haplogroup were genetic males based on the Ots-SEXY-3-1 sex-identification assay.

The majority of male (phenotypic or genetic) Chinook salmon in Alaska were assigned to Y-chromosome haplogroups with regional variation in assignment rate. Overall, 90% of phenotypic males (93/103) and 87% of genetic males (232/266) were assigned to a haplogroup. Genetic males that were not assigned Y-chromosome haplotypes were concentrated in Southeast Alaska; 30 of the 40 unassigned males were from the Little Port Walter and Pullen Creek populations. Only 58% of males from Little Port Walter and a single male from Pullen Creek could be assigned to a haplogroup. These samples had low missing data suggesting that the low assignment rate is due to low prevalence of these haplogroups in this region. Excluding these two populations results in 98% haplogroup assignment of male Chinook salmon in Alaska. Distribution of Y-chromosome haplotypes varied regionally: Y-chromosome haplotype blocks identified in Alaska were present in nearly all genetic males within Alaska but had rare occurrence outside of Alaska (Table 1). In total, eight males assigned to Ots17-MH1 were found in British Columbia, and 24 males assigned to Ots17-MH4 were found in British Columbia, Washington, and Idaho.

## Size and Age at Maturity

Size at maturity data were available for nine populations in this study; five populations were genotyped using RADseq data, three were genotyped using GT-seq data, and one was genotyped using both RADseq and GT-seq. Populations were grouped by region (Yukon River, Western Alaska, and Cook Inlet) for visualization. Boxplots of size at maturity for each Y-chromosome haplotype showed a consistent relationship throughout these regions (Figure 4). The Ots17-MH1 haplogroup had the

smallest individuals, the OTS17-MH2 haplogroup was associated with intermediate sized fish, and the Ots17-MH4 haplogroup was associated with the largest individuals in each region. The Ots17-MH3 haplogroup showed inconsistent results with small fish in some regions and large fish in others. ANOVA results showed that both haplogroup and population were significantly associated with variation in size at maturity. Post-hoc Tukey tests were conducted to determine which haplotypes had statistically significant differences in size. In the Yukon River, individuals with the Ots17-MH4 haplotype were significantly larger ($p<0.05$) than fish with the Ots17-MH1 or Ots17-MH2 haplotype. While this pattern was repeated in Western Alaska, the relationship was not statistically significant, likely due to low sample size of the Ots17-MH4 haplogroup (Figure 4). In Cook Inlet, the Ots17-MH1, Ots17-MH2, and Ots17-MH3 all had significant differences in size.

Age at maturity data were available for 177 males from six of the populations in this study (Table 1). The number of males with age data in each population ranged from 19 (Coal Creek) to 57 (Lower Yukon Test Fishery). There was a significant association between Y-chromosome haplogroup and age at maturity ($p=0.047$). The distribution of age at maturity by haplogroup was plotted for the Lower Yukon Test Fishery as this collection had the most samples with age data. Histograms of age at maturity for each Y-chromosome haplotype revealed that the Ots17-MH2 haplotype had approximately three times as many 1.3 fish as 1.4 fish while the Ots17-MH4 haplotype had nearly even proportions of 1.3 and 1.4 fish (Figure 5). Boxplots of size at age showed that fish with the Ots17-MH4 haplotype were larger for age 1.3 and 1.4 than fish with the Ots17-MH2 haplotype. Results were statistically significant for age 1.3 fish but not age 1.4, possibly due to low sample size. The smallest fish had the Ots17-MH1 haplotype; however, only two fish in the Lower Yukon Test Fishery had both size and age data for this haplotype.

# Discussion

Variation in life history within populations is common across taxa and is often assumed to be under the influence of many genes of small effect; however, examples of a genes and regions of large effect on life history variation, including single genes, small genomic regions, or chromosome inversions, are increasingly being found (Johnston et al. 2013, Pearse et al. 2014, Barson et al. 2015, Twyford and Friedman 2015). The genetic mechanism underlying variation in life history has important implications for how life history diversity is maintained under different selective regimes (Hess et al. 2016b, Prince et al. 2017), particularly in the case of sexually antagonistic selection where males and females have different fitness optima (Barson et al. 2015, Pearse et al. 2019).

Size and age at maturity are ecologically and evolutionarily important traits in Chinook salmon. Numerous studies have examined ongoing declines in age at maturity; however, it has been difficult to disentangle the interactions of environmental and genetic causes of this decline. Size and age associated markers and genes have previously been identified in genetic studies of Chinook salmon (Micheletti and Narum 2018, Waters et al. 2018); however, results were not consistent across populations, and no markers were located on the sex chromosome. We show that a conserved set of Y-chromosome haplotypes is associated with variation in size and age at maturity in Chinook salmon across the Yukon River, Western Alaska, and Cook Inlet. These observations open a new line of research into the genetic basis of age at maturity in salmonids.

## Range-wide distribution of haplotypes

Chinook salmon are represented by multiple genetically distinct lineages throughout the species range (Waples et al. 2004, Beacham et al. 2006, Moran et al. 2012). These lineages often show little gene flow among them due to differences in geographic range or spawn timing, and this isolation may result in different sets of Y-chromosome haplogroups and patterns of recombination across lineages. The Y-chromosome haplogroups that we identified through extended linkage disequilibrium were consistently observed throughout Chinook salmon populations from the Upper Yukon River south to Little Port Walter in Southeast Alaska. The occurrences become rarer in Southeast Alaska, and few individuals south of Southeast Alaska could be assigned haplotypes suggesting that the haplotypes identified within Alaska are regionally restricted. This corresponds with observed breakpoints between Chinook salmon lineages near Cape Fairweather (Templin et al. 2011) which is approximately 320 km northwest of Little Port Walter and 160 km west of Pullen Creek. While rare, the Ots17-MH4 haplogroup was found in some individuals as far south as Idaho in the northwestern continental United States, suggesting that this may be broadly distributed even if at low frequency. It is likely that other Y-chromosome haplogroups exist outside of Alaska, but we were unable to identify them with this dataset. The expanded survey used only a subset of SNPs, rather than the full RADseq dataset, that characterized the Y-chromosome haplogroups in Alaska. If different SNPs are associated with Y-chromosome haplogroups in other regions of the species range, we would not be able to identify them in this study. The presence of additional haplogroups outside Alaska could be determined by examining reduced-representation or whole-genome sequence data from additional populations.

Populations throughout Alaska showed variation in haplotype frequency which may be ecologically significant given the association between haplotypes and size and age at maturity.

Populations from Western Alaska and the Anvik River in the Lower Yukon River had a greater proportion of the Ots17-MH1 haplogroup which was associated with smaller fish. Populations in the Middle and Upper Yukon River and Cook Inlet were primarily composed of Ots17-MH2 and Ots17-MH4 haplogroup which were associated with larger fish. While we did not have enough samples with size data to characterize size distributions within regions, this finding is consistent with a long-term analysis of Chinook salmon returns by Lewis et al. (2015) who reported smaller fish on average in Kuskokwim and Nushagak river populations from Western Alaska relative to Yukon River and Cook Inlet populations. In addition, the Cook Inlet populations sampled in this study are near and share a common migration pathway with the Kenai River which has historically produced large Chinook salmon (Lewis et al. 2015, Schoen et al. 2017). The relationship between size and haplotype also varied by region, with the Yukon River having the smallest difference in sizes between haplotypes and Cook Inlet having the largest difference in sizes (Figure 4). While the magnitude of difference appears to be largest in Cook Inlet, it is difficult to accurately assess statistical significance due to the low sample size when splitting samples among regions. Taken together, these results suggest that differing frequencies of Y-chromosome haplotypes may contribute to regional variation in size of Chinook salmon and that the effect of haplotype on size can vary between regions, potentially due to other genetic influence or different environmental conditions. The Ots17-MH3 haplogroup was unusual in that it showed no consistent pattern, with large fish in some regions and small fish in other regions (Figure 4). This haplogroup also showed the least differentiation from the X-chromosome based on RADseq data (Figure 3) and may not contain adaptive variants influencing size or age at maturity.

## Resolving sexual conflict

Different fitness optima for males and females are common across species and are difficult to resolve when adaptive loci are on autosomes. One mechanism is for the same alleles to exhibit sex-specific dominance, such as the VGLL3 gene which influences age at maturity in Atlantic salmon (Barson et al. 2015). Another mechanism is to partition adaptive variants between non-recombining regions of sex chromosomes, such as genes governing coloration and fin morphology in Poecilids; these genes are attractive in males but would increase predation risk in females (Anna Lindholm and Felix Breden 2002). The existence of Y-chromosome haplotypes demonstrates not only that genetic variation is partitioned between the X- and Y-chromosomes, but that Y-chromosome variants have partitioned different genetic variants (Figure 3). While it is unlikely that the specific SNPs we observed are adaptive, these haplotypes are associated with variation in adaptive traits. This suggests that different adaptive variants

have been captured by Y-chromosome haplotypes. Y-chromosome haplotypes also varied in the chromosomal regions where the X- and Y-chromosome were differentiated. SNPs that differentiated the Ots17-MH1, Ots17-MH2, and Ots17-MH3 haplogroups from the X-chromosome were generally concentrated from approximately 4 Mb up to 22 Mb (Fig. 2, 3). One exception was a single SNP at ~600 Kb. SNPs that characterize the Ots17-MH4 haplogroup were concentrated from ~600 Kb to 6Mb along chromosome 17. The haplogroups may have captured different adaptive variants in each of these regions. This partitioning of genetic variation can resolve sexual conflict in age at maturity and provide a mechanism for the evolution of life history diversity in males.

## Importance of Y-haplotypes for life history diversity

Male Chinook salmon exhibit life history diversity related to maturation age. Older, larger males are believed to have greater reproductive success through their ability to monopolize access to females. Males that mature younger and smaller as jacks are generally believed to have reduced reproductive success but have greater survival to maturation due to reduced risk of ocean mortality (Berejikian et al. 2010). These alternate life histories of male Chinook salmon may exhibit frequency dependent fitness which in theory should exhibit stable proportions; however, this assumes populations are at equilibrium. Male sockeye salmon (*O. nerka*) exhibit similar frequency dependent life history variation, but persistent demographic shifts towards an abundance of jacks (males that mature one year earlier than the earliest maturing females) has occurred in some populations as a result of strong selection events coupled with variation in recruitment (DeFilippo et al. 2019). If maturation size and age in males are strongly influenced by an individual's Y-chromosome haplotype, then size-selective fishing practices or even size selective predation by marine predators (Ohlberger et al. 2019, Seitz et al. 2019) may result in shifts in haplogroup frequency, demographic changes, and loss of age diversity that are difficult to recover. The markers we developed can be used to characterize historic (i.e., from archived samples) and current Y-chromosome haplogroup diversity in Alaska Chinook salmon to determine if demographic shifts correspond with shifts in frequencies of Y-chromosome haplogroups. Ultimately, conservation of life history diversity in Chinook salmon may require conservation of Y-chromosome haplogroup diversity.

Hypotheses of population structure and delineation of management units using genetic data are typically based on genome-wide analyses consistent with the assumption that major life history traits are controlled by many genes with small effects. Waples and Lindley (2018) recently commented on the new challenges facing existing conservation frameworks when associations are identified between one or a very few genes and key life history traits. Their comment was prompted by the recent

identification of SNPs from a GREB1L gene that explain a large proportion of the variation associated with seasonal timing of adults returning to spawn steelhead (*O. mykiss*) and Chinook salmon (Hess et al. 2016b, Prince et al. 2017). Conservation of the Y-chromosome haplotype shares similar challenges to the GREB1L situation. Waples and Lindley (2018) pose a series of key questions to help provide an informed basis for decisions or management actions. Among other questions, they argue that a full understanding of the distribution of the variation in space and time is needed and that investigations into the genes and mechanisms responsible for the life history variation should be initiated. In the case of the Y-haplotypes, additional questions exist such as: What are the causal variants within these haplotype blocks? Do haplotype blocks exhibit consistent phenotypes in different environments and with different genetic backgrounds? What proportion of variance in size and age at maturity are explained by these haplotype blocks relative to other regions of the genome? If haplotype blocks are found in other regions of the Chinook salmon range, do they function in a similar manner to those suggested by the results of this study?

# Conclusion

Variation in size and age at maturity is common across taxa and is often associated with alternative strategies for increasing fitness; however, the genetic basis of this variation is largely unknown. We identified Y-chromosome haplogroups that are associated with size, and likely age, at maturity in Chinook salmon throughout Alaska. These haplogroups were primarily restricted to western and southcentral Alaska Chinook salmon where the most diversity in age-at-maturity exists, and likely represent a subset of the total diversity across the species range. Interestingly, this region holds the most stable populations of Chinook salmon in the Northeast Pacific (Griffiths et al. 2014), possibly due to the bet-hedging benefits of diversity in age at maturity. It is possible that each Chinook salmon lineage has a specific set of haplogroups and relationships between haplotypes and size/age at maturity may differ by lineage. The discovery of Y-chromosome haplotypes and their potential effect on life history variation in Chinook salmon may help explain the causes and consequences of the recent declines in size and age of adult Chinook salmon, trends that are most pronounced in the region with the highest haplotype diversity. Ongoing efforts to understand the causes of these declines point to size-specific mortality of maturing fish but also require an unknown evolutionary basis (Ohlberger et al. 2019). The results presented here point to such a mechanism for the genetic control of changes in size-at-age and age-at-maturity in Chinook salmon, as future changes in environmental conditions and selective fishing may lead to further demographic responses in this economically and ecologically important species.

# Acknowledgements

We would like to thank Chris Habicht, Bill Templin, Wes Larson, and Todd Seamons for helpful discussions and support for this research. This research was partially funded by North Pacific Research Board grant 1724-000, AKSSF grants 44812, 44913 and 44914, and Pacific States Marine Fisheries Commission grant # 16-108G to Alaska Department of Fish and Game. GJM was partially supported by a National Research Council postdoctoral fellowship. The statements, findings, and conclusions are those of the authors and do not necessarily reflect the views of the NOAA, the U.S. Department of Commerce, or the ADFG.

# References


Allendorf, F. W., and G. H. Thorgaard. 1984. Tetraploidy and the evolution of salmonid fishes. Pages 1-53 *in* B. Turner, editor. Evolutionary Genetics of Fishes. Plenum Publishing Corporation.

Anna Lindholm, and Felix Breden. 2002. Sex chromosomes and sexual selection in poeciliid fishes. The American Naturalist **160**:S214-S224. doi:10.1086/342898

Aubin Horth, N., and J. J. Dodson. 2004. Influence of Individual Body Size and Variable Thresholds on the Incidence of a Sneaker Male Reproductive Tactic in Atlantic Salmon. Evolution **58**:136-144. doi:10.1111/j.0014-3820.2004.tb01580.x

Ayllon, F., E. Kjaerner-Semb, T. Furmanek, V. Wennevik, M. F. Solberg, G. Dahle, G. L. Taranger, K. A. Glover, M. S. Almen, C. J. Rubin, R. B. Edvardsen, and A. Wargelius. 2015. The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (*salmo salar l.*) males. PLoS Genet **11**:e1005628. doi:10.1371/journal.pgen.1005628

Barson, N. J., T. Aykanat, K. Hindar, M. Baranski, G. H. Bolstad, P. Fiske, C. Jacq, A. J. Jensen, S. E. Johnston, S. Karlsson, M. Kent, T. Moen, E. Niemela, T. Nome, T. F. Naesje, P. Orell, A. Romakkaniemi, H. Saegrov, K. Urdal, J. Erkinaro, S. Lien, and C. R. Primmer. 2015. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature **528**:405-408. doi:10.1038/nature16062

Beacham, T. D., K. L. Jonsen, J. Supernault, M. Wetklo, L. Deng, and N. Varnavskaya. 2006. Pacific rim population structure of Chinook salmon as determined from microsatellite analysis. Transactions of the American Fisheries Society **135**:1604-1621. doi:10.1577/T06-071.1

Berejikian, B. A., D. M. Van Doornik, and J. J. Atkins. 2011. Alternative male reproductive phenotypes affect offspring growth rates in Chinook salmon. Transactions of the American Fisheries Society **140**:1206-1212. doi:10.1080/00028487.2011.618351

Berejikian, B. A., D. M. Van Doornik, R. C. Endicott, T. L. Hoffnagle, E. P. Tezak, M. E. Moore, and J. Atkins. 2010. Mating success of alternative male phenotypes and evidence for frequency-dependent selection in Chinook salmon, *Oncorhynchus tshawytscha*. Canadian Journal of Fisheries and Aquatic Sciences **67**:1933-1941. doi:10.1139/f10-112

Boulding, E. G., K. P. Ang, J. A. K. Elliott, F. Powell, and L. R. Schaeffer. 2019. Differences in genetic architecture between continents at a major locus previously associated with sea age at sexual maturity in European Atlantic salmon. Aquaculture **500**:670-678. doi:https://doi.org/10.1016/j.aquaculture.2018.09.025

Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Mol Ecol Resour **15**:855-867. doi:10.1111/1755-0998.12357

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Mol Ecol **22**:3124-3140. doi:10.1111/mec.12354

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. G3 (Bethesda) **1**:171-182. doi:10.1534/g3.111.000240

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience **4**. doi:10.1186/s13742-015-0047-8

Chapman, B. B., C. Brönmark, J.-Å. Nilsson, and L.-A. Hansson. 2011. The ecology and evolution of partial migration. Oikos **120**:1764-1775. doi:10.1111/j.1600-0706.2011.20131.x

Christensen, K. A., J. S. Leong, D. Sakhrani, C. A. Biagi, D. R. Minkley, R. E. Withler, E. B. Rondeau, B. F. Koop, and R. H. Devlin. 2018. Chinook salmon (*Oncorhynchus tshawytscha*) genome and transcriptome. PLoS One **13**:e0195461. doi:10.1371/journal.pone.0195461

Dann, T. H., C. Habicht, W. D. Templin, L. W. Seeb, G. J. McKinney, and J. S. Seeb. 2018. Identification of genetic markers useful for mixed stock analysis of Chinook salmon in Cook Inlet, Alaska. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J18-04, Anchorage. Available at: http://www.adfg.alaska.gov/FedAidPDFs/RIR.5J.2018.04.pdf.

Davidson, W. S., T. K. Huang, K. Fujiki, K. R. von Schalburg, and B. F. Koop. 2009. The sex determining loci and sex chromosomes in the family salmonidae. Sexual Development **3**:78-87. doi:10.1159/000223073

DeFilippo, L. B., D. E. Schindler, J. Ohlberger, K. L. Schaberg, M. B. Foster, D. Ruhl, and A. E. Punt. 2019. Recruitment variation disrupts the stability of alternative life histories in an exploited salmon population. Evol Appl **12**:214-229. doi:10.1111/eva.12709

Devlin, R. H., G. W. Stone, and D. E. Smailus. 1998. Extensive direct-tandem organization of a long repeat DNA sequence on the y chromosome of Chinook salmon (*Oncorhynchus tshawytscha*). Journal of Molecular Evolution **46**:277-287. doi:10.1007/PL00006304

Gall, G. A. E., J. Baltodano, and N. Huang. 1988. Heritability of age at spawning for rainbow trout. Aquaculture **68**:93-120.

Garant, D., P.-M. Fontaine, S. P. Good, J. J. Dodson, and L. Bernatchez. 2002. The influence of male parental identity on growth and survival of offspring in Atlantic salmon (*Salmo salar*). Evolutionary Ecology Research **4**:537-549.

Gibbons, J. W., R. D. Semlitsch, J. L. Greene, and J. P. Schubauer. 1981. Variation in Age and Size at Maturity of the Slider Turtle (*Pseudemys scripta*). The American Naturalist **117**:841-845. doi:10.1086/283774

Gjerde, B. 1984. Response to individual selection for age at sexual maturity in Atlantic salmon. Aquaculture **38**:229-240. doi:https://doi.org/10.1016/0044-8486(84)90147-9

Griffiths, J. R., D. E. Schindler, J. B. Armstrong, M. D. Scheuerell, D. C. Whited, R. A. Clark, R. Hilborn, C. A. Holt, S. T. Lindley, J. A. Stanford, and E. C. Volk. 2014. Performance of salmon fishery portfolios across western North America. Journal of Applied Ecology **51**:1554-1563.

Hall, M. C., C. J. Basten, and J. H. Willis. 2006. Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in Mimulus guttatus. Genetics **172**:1829-1844. doi:10.1534/genetics.105.051227

Hankin, D. G., J. W. Nicholas, and T. W. Downey. 1993. Evidence for inheritance of age of maturity in Chinook salmon (*Oncorhynchus tshawytscha*). Canadian Journal of Fisheries and Aquatic Sciences **50**:347-358. doi:doi.org/10.1139/f93-040

Hard, J. J., M. R. Gross, M. Heino, R. Hilborn, R. G. Kope, R. Law, and J. D. Reynolds. 2008. Evolutionary consequences of fishing and their implications for salmon. Evolutionary Applications **1**:388-408. doi:10.1111/j.1752-4571.2008.00020.x

Harstad, D. L., D. A. Larsen, J. Miller, I. Adams, D. K. Spangenberg, S. Nance, L. Rohrbach, J. G. Murauskas, and B. R. Beckman. 2018. Winter-rearing temperature affects growth profiles, age of maturation, and smolt-to-adult returns for yearling summer Chinook salmon in the Upper Columbia River basin. North American Journal of Fisheries Management **38**:867-885. doi:doi:10.1002/nafm.10186

Healey, M. C. 1986. Optimum size and age at maturity in pacific salmon and effects of size-selective fisheries. Can. Pec. Publ. Fish. Aquat. Sci. **89**:39-52.

Healey, M. C. 1991. Life history of Chinook salmon (*Oncorhynchus tshawytscha*). Pages 311-393 *in* L. Margolis and C. Groot, editors. Pacific salmon life histories. UBC Press, Vancouver, B.C.

Healey, M. C., and W. R. Heard. 1984. Inter- and intra-population variation in the fecundity of chinook salmon (*Oncorhynchus tshawytscha*) and its relevance to life history theory. Canadian Journal of Fisheries and Aquatic Sciences **41**:476-483. doi:10.1139/f84-057

552 Heath, D. D., R. H. Devlin, J. W. Heath, and G. K. Iwama. 1994. Genetic, environmental and interaction
553 effects on the incidence of jacking in *Oncorhynchus tshawytscha* (Chinook salmon). Heredity
554 (Edinb) **72**:146-154.
555 Heath, D. D., L. Rankin, C. A. Bryden, J. W. Heath, and J. M. Shrimpton. 2002. Heritability and Y-
556 chromosome influence in the jack male life history of Chinook salmon (*Oncorhynchus*
557 *tshawytscha*). Heredity (Edinb) **89**:311. doi:10.1038/sj.hdy.6800141
558 Hess, J. E., N. R. Campbell, A. P. Matala, D. J. Hasselman, and S. P. Narum. 2016a. Genetic assessment of
559 Columbia River stocks, 4/1/2014-3/31/2105 annual report, 2008-907-00. CRITFC. Available from
560 http://www.critfc.org/wp-content/uploads/2016/04/16-03.pdf.
561 Hess, J. E., J. S. Zendt, A. R. Matala, and S. R. Narum. 2016b. Genetic basis of adult migration timing in
562 anadromous steelhead discovered through multivariate association testing. Proc Biol Sci **283**.
563 doi:10.1098/rspb.2015.3064
564 Johnston, S. E., J. Gratten, C. Berenos, J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, and J. Slate.
565 2013. Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature
566 **502**:93-95. doi:10.1038/nature12489
567 Kemppainen, P., C. G. Knight, D. K. Sarma, T. Hlaing, A. Prakash, Y. N. Maung Maung, P. Somboon, J.
568 Mahanta, and C. Walton. 2015. Linkage disequilibrium network analysis (LDna) gives a global
569 view of chromosomal inversions, local adaptation and geographic structure. Mol Ecol Resour
570 **15**:1031-1045. doi:10.1111/1755-0998.12369
571 Kendall, N. W., U. Dieckmann, M. Heino, A. E. Punt, and T. P. Quinn. 2014. Evolution of age and length at
572 maturation of Alaskan salmon under size-selective harvest. Evolutionary Applications **7**:313-322.
573 doi:doi:10.1111/eva.12123
574 Kijas, J., S. McWilliam, M. Naval Sanchez, P. Kube, H. King, B. Evans, T. Nome, S. Lien, and K. Verbyla.
575 2018. Evolution of sex determination loci in Atlantic salmon. Scientific Reports **8**:5664.
576 doi:10.1038/s41598-018-23984-1
577 Kodama, M., J. J. Hard, and K. A. Naish. 2018. Mapping of quantitative trait loci for temporal growth and
578 age at maturity in coho salmon: Evidence for genotype-by-sex interactions. Marine Genomics
579 **38**:33-44. doi:https://doi.org/10.1016/j.margen.2017.07.004
580 Koo, T. S. Y. 1962. Age determination in salmon. Pages 37-48 *in* T. S. Y. Koo, editor. Studies of Alaska red
581 salmon. University of Washington Press,, Seattle.
582 Küpper, C., M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell, S. B. McRae, T. C. Morgan, N. Karlionova,
583 P. Pinchuk, Y. I. Verkuil, A. S. Kitaysky, J. C. Wingfield, T. Piersma, K. Zeng, J. Slate, M. Blaxter, D.
584 B. Lank, and T. Burke. 2015. A supergene determines highly divergent male reproductive morphs
585 in the ruff. Nature Genetics **48**:79. doi:10.1038/ng.3443

586 https://www.nature.com/articles/ng.3443#supplementary-information
587 Kusche, H., G. Cote, C. Hernandez, E. Normandeau, D. Boivin-Delisle, and L. Bernatchez. 2017.
588 Characterization of natural variation in North American Atlantic Salmon populations
589 (Salmonidae: *Salmo salar*) at a locus with a major effect on sea age. Ecol Evol **7**:5797-5807.
590 doi:10.1002/ece3.3132
591 Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357-
592 359. doi:10.1038/nmeth.1923
593 Larsen, D. A., B. R. Beckman, C. R. Strom, P. J. Parkins, K. A. Cooper, D. E. Fast, and W. W. Dickhoff. 2006.
594 Growth modulation alters the incidence of early male maturation and physiological
595 development of hatchery-reared spring Chinook salmon: A comparison with wild fish.
596 Transactions of the American Fisheries Society **135**:1017-1032. doi:10.1577/t05-200.1
597 Larson, W. A., J. E. Seeb, C. E. Pascal, W. D. Templin, and L. W. Seeb. 2014. Single-nucleotide
598 polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock

identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. Canadian Journal of Fisheries and Aquatic Sciences **71**:698-708. doi:10.1139/cjfas-2013-0502

Lewis, B., W. S. Grant, R. E. Brenner, and T. Hamazaki. 2015. Changes in size and age of Chinook salmon *Oncorhynchus tshawytscha* returning to Alaska. PLoS One **10**:e0130184-e0130184. doi:10.1371/journal.pone.0130184

Lien, S., L. Gidskehaug, T. Moen, B. J. Hayes, P. R. Berg, W. S. Davidson, S. W. Omholt, and M. P. Kent. 2011. A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. BMC Genomics **12**:615. doi:10.1186/1471-2164-12-615

Lozorie, J. D., and B. C. McIntosh. 2014. Sonar esimation of salmon passage in the Yukon River near Pilot Station, 2012. Alaska Department of Fish and Game, Anchorage, AK.

Marshall, A. R., K. L. Knudsen, and F. W. Allendorf. 2004. Linkage disequilibrium between the pseudoautosomal PEPB-1 locus and the sex-determining region of Chinook salmon. Heredity (Edinb) **93**:85. doi:10.1038/sj.hdy.6800483

McKinney, G. J., M. V. McPhee, C. E. Pascal, J. E. Seeb, and L. W. Seeb. 2020a. Network analysis of linkage disequilibrium reveals genome architecture in chum salmon. G3: Genes|Genomes|Genetics. doi:10.1534/g3.119.400972

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

McKinney, G. J., R. K. Waples, C. E. Pascal, L. W. Seeb, and J. E. Seeb. 2018. Resolving allele dosage in duplicated loci using genotyping-by-sequencing data: A path forward for population genetic analysis. Mol Ecol Resour **18**:570-579. doi:10.1111/1755-0998.12763

McKinney, G. J., R. K. Waples, L. W. Seeb, and J. E. Seeb. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Mol Ecol Resour **17**:656-669. doi:10.1111/1755-0998.12613

Micheletti, S. J., and S. R. Narum. 2018. Utility of pooled sequencing for association mapping in nonmodel organisms. Mol Ecol Resour **18**:825-837. doi:10.1111/1755-0998.12784

Moran, P., D. J. Teel, M. A. Banks, T. D. Beacham, M. R. Bellinger, S. M. Blankenship, J. R. Candy, J. C. Garza, J. E. Hess, S. R. Narum, L. W. Seeb, W. D. Templin, C. G. Wallace, and C. T. Smith. 2012. Divergent life-history races do not represent Chinook salmon coast-wide: the importance of scale in Quaternary biogeography. Canadian Journal of Fisheries and Aquatic Sciences **70**:415-435. doi:10.1139/cjfas-2012-0135

Narum, S. R., A. D. Genova, S. J. Micheletti, and A. Maass. 2018. Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. Proceedings of the Royal Society B: Biological Sciences **285**:20180935. doi:doi:10.1098/rspb.2018.0935

Ohlberger, J., D. E. Schindler, E. J. Ward, T. E. Walsworth, and T. E. Essington. 2019. Resurgence of an apex marine predator and the decline in prey body size. Proc Natl Acad Sci U S A **116**:26682-26689. doi:10.1073/pnas.1910930116

Ohlberger, J., E. J. Ward, D. E. Schindler, and B. Lewis. 2018. Demographic changes in Chinook salmon across the Northeast Pacific Ocean. Fish and Fisheries **19**:533-546. doi:doi:10.1111/faf.12272

Otero, J., A. J. Jensen, J. H. L'Abee-Lund, N. C. Stenseth, G. O. Storvik, and L. A. Vollestad. 2012. Contemporary ocean warming and freshwater conditions are related to later sea age at maturity in Atlantic salmon spawning in Norwegian rivers. Ecol Evol **2**:2192-2203. doi:10.1002/ece3.337

Pearse, D. E., N. J. Barson, T. Nome, G. Gao, M. A. Campbell, A. Abadia-Cardoso, E. C. Anderson, D. E. Rundio, T. H. Williams, K. A. Naish, T. Moen, S. Liu, M. Kent, M. Moser, D. R. Minkley, E. B. Rondeau, M. S. O. Brieuc, S. R. Sandve, M. R. Miller, L. Cedillo, K. Baruch, A. G. Hernandez, G. Ben-Zvi, D. Shem-Tov, O. Barad, K. Kuzishchin, J. C. Garza, S. T. Lindley, B. F. Koop, G. H.

Thorgaard, Y. Palti, and S. Lien. 2019. Sex-dependent dominance maintains migration supergene in rainbow trout. Nat Ecol Evol **3**:1731-1742. doi:10.1038/s41559-019-1044-6

Pearse, D. E., M. R. Miller, A. Abadia-Cardoso, and J. C. Garza. 2014. Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. Proc Biol Sci **281**:20140012. doi:10.1098/rspb.2014.0012

Phillips, R. B., L. K. Park, and K. A. Naish. 2013. Assignment of Chinook salmon (*Oncorhynchus tshawytscha*) linkage groups to specific chromosomes reveals a karyotype with multiple rearrangements of the chromosome arms of rainbow trout (*Oncorhynchus mykiss*). Genes Genomes Genetics **3**:2289-2295. doi:10.1534/g3.113.008078

Prince, D. J., S. M. O'Rourke, T. Q. Thompson, O. A. Ali, H. S. Lyman, I. K. Saglam, T. J. Hotaling, A. P. Spidle, and M. R. Miller. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Sci Adv **3**:e1603198. doi:10.1126/sciadv.1603198

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet **81**:559-575. doi:10.1086/519795

Riddell, B. R., R. D. Brodeur, A. V. Bugaev, P. Moran, J. M. Murphy, J. A. Orsi, M. trudel, L. A. Weitkamp, B. K. Wells, and A. C. Wertheimer. 2018. Ocean ecology of Chinook Salmon. Pages 55-696 *in* R. J. Beamish, editor. The ocean ecology of Pacific salmon and trout. American Fisheries Society, Bethesda, Maryland.

Rowe, D., and J. E. Thorpe. 1990. Suppression of maturation in male Atlantic salmon parr (*Salmon salar L.*) by reduction in feeding and growth during spring months. Aquaculture **86**:291-313.

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. American journal of human genetics **78**:629-644. doi:10.1086/502802

Schoen, E. R., M. S. Wipfli, E. J. Trammell, D. J. Rinella, A. L. Floyd, J. Grunblatt, M. D. McCarthy, B. E. Meyer, J. M. Morton, J. E. Powell, A. Prakash, M. N. Reimer, S. L. Stuefer, H. Toniolo, B. M. Wells, and F. D. W. Witmer. 2017. Future of Pacific salmon in the face of environmental change: Lessons from one of the world's remaining productive salmon regions. Fisheries **42**:538-553. doi:10.1080/03632415.2017.1374251

Seitz, A. C., M. B. Courtney, M. D. Evans, and K. Manishin. 2019. Pop-up satellite archival tags reveal evidence of intense predation on large immature Chinook salmon (*Oncorhynchus tshawytscha*) in the North Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences **76**:1608-1615. doi:10.1139/cjfas-2018-0490

Shuster, S. M. 1989. Male alternative reproductive strategies in a marine isopod crustacean (*Paracerceis sculpta*): The use of genetic markers to measure differences in fertilization success among alpha-, beta-, and gamma-males. Evolution **43**:1683-1698. doi:10.1111/j.1558-5646.1989.tb02618.x

Siegel, J. E., M. V. McPhee, and M. D. Adkison. 2017. Evidence that marine temperatures influence growth and maturation of Western Alaskan Chinook salmon. Marine and Coastal Fisheries **9**:441-456. doi:10.1080/19425120.2017.1353563

Templin, W. D., J. E. Seeb, J. R. Jasper, A. W. Barclay, and L. W. Seeb. 2011. Genetic differentiation of Alaska Chinook salmon: the missing link for migratory studies. Mol Ecol Resour **11 Suppl 1**:226-246. doi:10.1111/j.1755-0998.2010.02968.x

Thorpe, J. E. 2007. Maturation responses of salmonids to changing developmental opportunities. Marine Ecology Progress Series **335**:285-288.

Twyford, A. D., and J. Friedman. 2015. Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. Evolution **69**:1476-1486. doi:10.1111/evo.12663

Wagih, O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics **33**:3645-3647. doi:10.1093/bioinformatics/btx469

Waples, R. S., and S. T. Lindley. 2018. Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. Evolutionary Applications **11**:1518-1526. doi:10.1111/eva.12687

Waples, R. S., D. J. Teel, J. M. Myers, and A. R. Marshall. 2004. Life-history divergence in Chinook salmon: Historic contingency and parallel evolution. Evolution **58**:386-403. doi:doi:10.1111/j.0014-3820.2004.tb01654.x

Warnes, G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber, and A. Liaw. 2020. gplots: Various R programming tools for plotting data.

Waters, C. D., J. J. Hard, M. S. O. Brieuc, D. E. Fast, K. I. Warheit, C. M. Knudsen, W. J. Bosch, and K. A. Naish. 2018. Genomewide association analyses of fitness traits in captive-reared Chinook salmon: Applications in evaluating conservation strategies. Evol Appl **11**:853-868. doi:10.1111/eva.12599

Yano, A., B. Nicol, E. Jouanno, E. Quillet, A. Fostier, R. Guyomard, and Y. Guiguen. 2013. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. Evolutionary applications **6**:486-496. doi:10.1111/eva.12032

You, F. M., N. Huo, Y. Q. Gu, M. C. Luo, Y. Ma, D. Hane, G. R. Lazo, J. Dvorak, and O. D. Anderson. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics **9**:253. doi:10.1186/1471-2105-9-253

Young, B., D. V. Conti, and M. D. Dean. 2013. Sneaker "jack" males outcompete dominant "hooknose" males under sperm competition in Chinook salmon (*Oncorhynchus tshawytscha*). Ecology and Evolution **3**:4987-4997. doi:doi:10.1002/ece3.869

# Data Accessibility

GT-seq data will be deposited to NCBI SRA upon acceptance.

# Supporting Information

**Table S1:** Primer and probe sequence for GT-seq loci that were used for expanded genotyping of Y-chromosome haplotypes.

**Table S2:** Concordance between methods of Y-haplogroup assignment. Samples genotyped with both RADseq and GT-seq are shown in A and B. GT-seq samples were assigned to haplogroups using a clustering algorithm in A and by matching observed genotypes to expected reference genotypes in B. Samples genotyped only with GT-seq are shown in C.

# Tables and Figures

732 Table 1. Populations used in this study. RADseq data was used for haplogroup discovery while GT-seq data was used to confirm that
733 haplogroups were male-specific and to expand the geographic distribution of haplogroups. For populations with both RADseq and GT-seq
734 samples, samples sizes for RADseq are given first. Frequency of Y-chromosome haplogroups are given for populations in Alaska in A. Due to the
735 low proportion of males assigned to Y-chromosome haplogroups outside of Alaska, the number of individuals assigned to each Y-chromosome
736 haplogroup for these populations are given in B.

A

| Map Number | Population | Data Type | Region | Latitude | Longitude | N | Ots17-MH1 | Ots17-MH2 | Ots17-MH3 | Ots17-MH4 | N Aged Males |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Big Salmon River | RAD | Yukon River* | 61.87 | -134.92 | 49 | 0.00 | 0.78 | 0.00 | 0.22 | 0 |
| 2 | Kantishna River | GT-seq | Yukon River | 64.76 | -149.97 | 48 | 0.09 | 0.91 | 0.00 | 0.00 | 0 |
| 3 | Lower Yukon Test Fishery | GT-seq | Yukon River | 62.79 | -164.81 | 95 | 0.05 | 0.70 | 0.07 | 0.18 | 57 |
| 4 | Anvik River | RAD | Yukon River | 62.68 | -160.21 | 56 | 0.47 | 0.47 | 0.02 | 0.04 | 0 |
| 5 | Tubutulik River | RAD | Western Alaska | 64.74 | -161.89 | 56 | 0.82 | 0.09 | 0.06 | 0.03 | 0 |
| 6 | Necons River | RAD | Western Alaska | 61.10 | -153.85 | 47 | 0.14 | 0.14 | 0.18 | 0.55 | 0 |
| 7 | George River | RAD/GT-seq | Western Alaska | 61.90 | -157.71 | 48/47 | 0.21 | 0.32 | 0.13 | 0.34 | 0 |
| 8 | Kogrukluk River | RAD | Western Alaska | 60.84 | -157.85 | 59 | 0.63 | 0.09 | 0.22 | 0.06 | 0 |
| 9 | Aniak River | RAD | Western Alaska | 61.06 | -159.18 | 47 | 0.33 | 0.11 | 0.53 | 0.03 | 0 |
| 10 | Kisaralik River | RAD | Western Alaska | 60.86 | -161.24 | 48 | 0.68 | 0.06 | 0.26 | 0.00 | 0 |
| 11 | Kwethluk River | RAD | Western Alaska | 60.81 | -161.45 | 47 | 0.50 | 0.25 | 0.22 | 0.03 | 0 |
| 12 | Kanektok River | RAD/GT-seq | Western Alaska | 59.75 | -161.93 | 48/47 | 0.44 | 0.36 | 0.19 | 0.00 | 0 |
| 13 | Arolik River | RAD | Western Alaska | 59.69 | -161.88 | 48 | 0.40 | 0.40 | 0.14 | 0.06 | 0 |
| 14 | Goodnews River | RAD | Western Alaska | 59.25 | -161.36 | 47 | 0.24 | 0.43 | 0.00 | 0.33 | 0 |
| 15 | Togiak River | RAD | Western Alaska | 59.09 | -160.37 | 48 | 0.24 | 0.57 | 0.10 | 0.10 | 21 |
| 16 | Iowithla River | RAD | Western Alaska | 59.18 | -158.06 | 48 | 0.22 | 0.57 | 0.14 | 0.08 | 0 |
| 17 | Stuyahok River | RAD | Western Alaska | 59.68 | -156.17 | 48 | 0.35 | 0.32 | 0.13 | 0.19 | 0 |
| 18 | Koktuli River | RAD | Western Alaska | 59.94 | -156.43 | 56 | 0.39 | 0.39 | 0.14 | 0.09 | 0 |
| 19 | Karluk River | GT-seq | Western Alaska | 57.57 | -154.38 | 48 | 0.00 | 1.00 | 0.00 | 0.00 | 25 |
| 20 | Montana Creek | GT-seq | Cook Inlet | 62.18 | -149.95 | 48 | 0.22 | 0.30 | 0.04 | 0.44 | 0 |
| 21 | Chuitna River | RAD | Cook Inlet | 61.20 | -151.66 | 57 | 0.16 | 0.75 | 0.00 | 0.09 | 39 |
| 22 | Sucker Creek | RAD | Cook Inlet | 61.51 | -150.83 | 57 | 0.06 | 0.88 | 0.00 | 0.06 | 0 |
| 23 | Talachulitna River | RAD | Cook Inlet | 61.62 | -151.15 | 57 | 0.13 | 0.65 | 0.00 | 0.22 | 23 |
| 24 | Theodore Creek | RAD | Cook Inlet | 61.49 | -151.09 | 56 | 0.26 | 0.67 | 0.05 | 0.03 | 0 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | Coal Creek | RAD | Cook Inlet | 61.62 | -151.76 | 55 | 0.00 | 0.68 | 0.00 | 0.32 | 19 |
| 26 | Pullen Creek** | GT-seq | Southeast Alaska | 59.45 | -135.32 | 48 | 0.00 | 1.00 | 0.00 | 0.00 | 0 |
| 27 | Little Port Walter-Unuk River stock*** | GT-seq | Southeast Alaska | 56.09 | -131.06 | 48 | 0.71 | 0.00 | 0.00 | 0.29 | 0 |
| B | | | | | | | | | | | |
| - | Harrison River | GT-seq | British Columbia | 49.28 | -121.92 | 48 | NA | NA | NA | NA | 0 |
| - | Big Qualicum Hatchery | GT-seq | British Columbia | 49.40 | -124.62 | 48 | NA | NA | NA | NA | 0 |
| - | Kitwanga River | GT-seq | British Columbia | 55.10 | -128.09 | 48 | 4 | NA | NA | 1 | 0 |
| - | Kitsumkalum River | GT-seq | British Columbia | 54.52 | -128.66 | 48 | 3 | NA | NA | 3 | 0 |
| - | Morice River | GT-seq | British Columbia | 54.41 | -126.75 | 48 | 1 | NA | NA | 6 | 0 |
| - | Rapid River Hatchery | GT-seq | Idaho | 45.35 | -116.40 | 48 | NA | NA | NA | NA | 0 |
| - | McCall Fish Hatchery South Fork Salmon River | GT-seq | Idaho | 44.67 | -115.71 | 48 | NA | NA | NA | 5 | 0 |
| - | Marblemount Fish Hatchery | GT-seq | Washington | 48.52 | -121.42 | 48 | NA | NA | NA | NA | 0 |
| - | Soos Creek Hatchery | GT-seq | Washington | 47.31 | -122.16 | 48 | NA | NA | NA | NA | 0 |
| - | Columbia River at Wells Hachery | GT-seq | Washington | 47.95 | -119.87 | 48 | NA | NA | NA | NA | 0 |
| - | Quinault Lake Pens | GT-seq | Washington | 47.47 | -123.89 | 48 | NA | NA | NA | NA | 0 |
| - | Lyons Ferry Hatchery | GT-seq | Washington | 46.59 | -118.22 | 48 | NA | NA | NA | 2 | 0 |
| - | Wenatchee River at Tumwater Dam | GT-seq | Washington | 47.61 | -120.72 | 48 | NA | NA | NA | 7 | 0 |
| - | Spring Creek Hatchery | GT-seq | Washington | 45.73 | -121.55 | 48 | NA | NA | NA | NA | 0 |
| - | Rock Creek, Umpqua River | GT-seq | Oregon | 43.34 | -123.00 | 48 | NA | NA | NA | NA | 0 |
| - | Cole River Hatchery, Rogue River | GT-seq | Oregon | 42.66 | -122.69 | 48 | NA | NA | NA | NA | 0 |
| - | Cedar Creek Hatchery | GT-seq | Oregon | 45.22 | -123.84 | 48 | NA | NA | NA | NA | 0 |
| - | McKenzie Hatchery | GT-seq | Oregon | 44.11 | -122.68 | 48 | NA | NA | NA | NA | 0 |
| - | Coleman National Fish Hatchery | GT-seq | California | 40.40 | -122.14 | 48 | NA | NA | NA | NA | 0 |

*The Yukon River is part of Western Alaska but was analyzed separately for this study. **Only one male from Pullen Creek was assigned to a Y-chromosome haplogroup. ***Only 11 out of 20 males from Little Port Walter Unuk River Stock were assigned to a Y-chromosome haplogroup.

740 Table 2. Number of haplotypes from RADseq samples assigned to each haplogroup by phenotypic sex.
741 Each individual has two haplotypes. Males should have one haplotype assigned to a Y-chromosome
742 haplogroup and one haplotype assigned to an X-chromosome haplogroup. Females should have both
743 haplotypes assigned to an X-chromosome haplogroup.

744

| | Sex | | | | |
|---|---|---|---|---|---|
| Heatmap Group Color | F | M | Unknown | Putative Chromosome | Haplogroup Name |
| gray | 168 | 114 | 999 | X | X |
| pink | 1 | 21 | 206 | Y | Ots17-MH1 |
| green | 9 | 55 | 54 | Y | Ots17-MH2 |
| blue | 0 | 7 | 78 | Y | Ots17-MH3 |
| purple | 2 | 11 | 66 | Y | Ots17-MH4 |

745

746

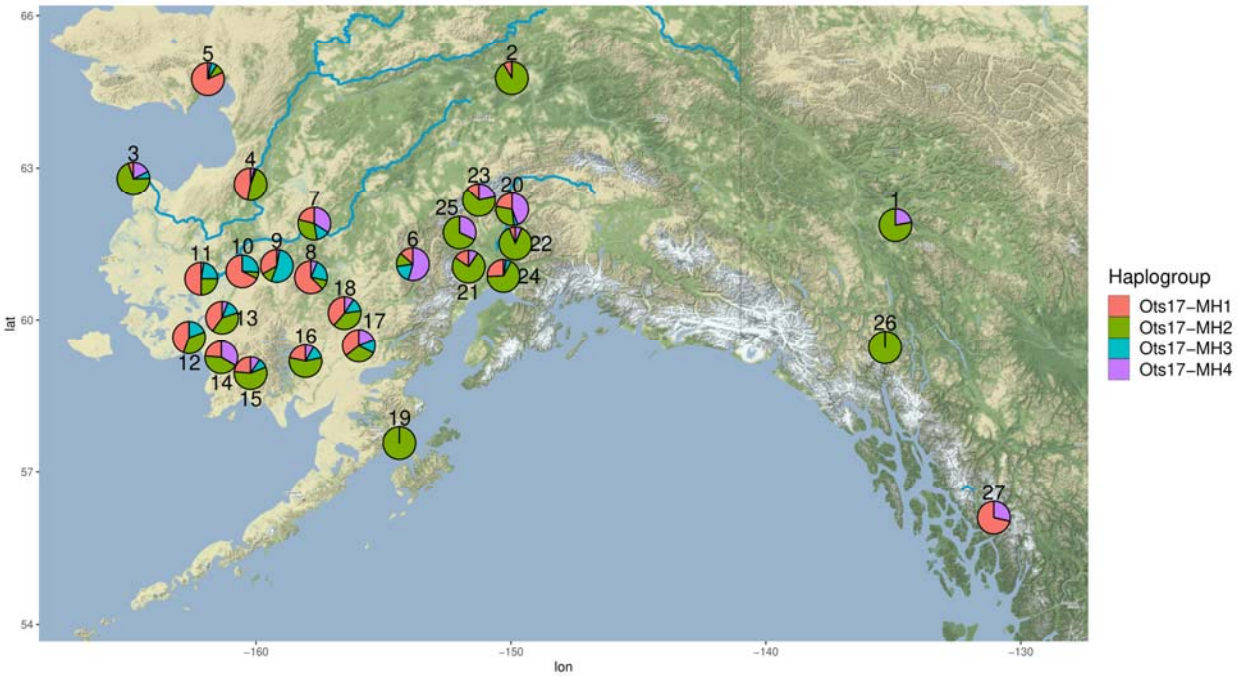
747 Table 3. Number of haplotypes from RADseq samples assigned to each haplogroup by region. Each
748 individual has two haplotypes. Males should have one haplotype assigned to a Y-chromosome
749 haplogroup and one haplotype assigned to an X-chromosome haplogroup. Females should have both
750 haplotypes assigned to an X-chromosome haplogroup.

| Heatmap Group Color | Cook Inlet | Western Alaska | Yukon River | Putative Chromosome | Haplogroup Name |
|---|---|---|---|---|---|
| gray | 410 | 942 | 139 | X | X |
| pink | 20 | 183 | 25 | Y | Ots17-MH1 |
| green | 112 | 130 | 39 | Y | Ots17-MH2 |
| blue | 2 | 82 | 1 | Y | Ots17-MH3 |
| purple | 20 | 53 | 6 | Y | Ots17-MH4 |

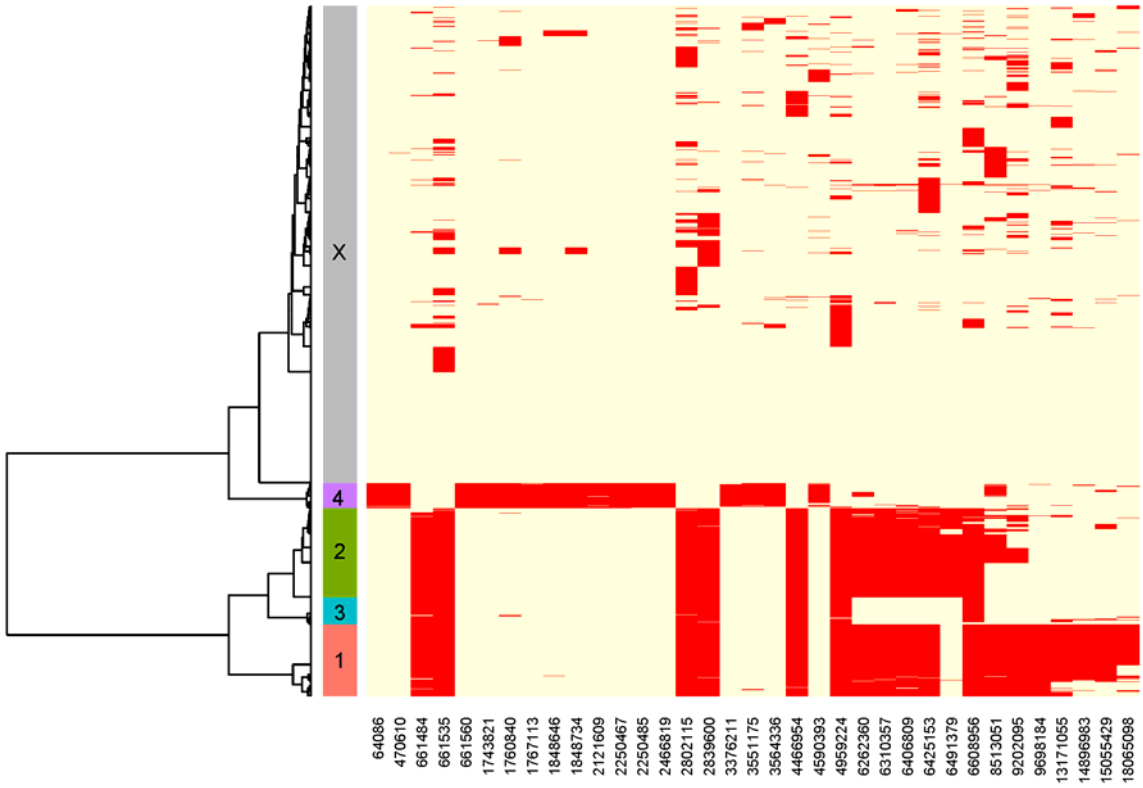
751

752

Figure 1. Frequency of Y-chromosome haplogroups throughout Alaska based on RADseq and GT-seq data. Locations of populations are approximate to prevent overlap of pie charts. Population names are given in Table 1. Note that the location for the Lower Yukon Test Fishery (population 3) indicates where fish were caught on their return to the Yukon River; these samples may represent fish from many populations that spawn throughout the Yukon River.



759 Figure 2. Plot of haplotype clusters identified by phasing high LD loci from the RADseq dataset. Each
760 individual is represented by two haplotypes corresponding to each chromosome of Ots17. For each
761 SNP, the most common allele is in yellow and the least common allele is in red. Haplotypes were
762 clustered into haplogroups, and six major haplotype clusters were identified; these are denoted by
763 different colors along the sample dendrogram (y-axis). The gray haplogroup represents X-chromosomes
764 while four haplogroups (pink=Ots17-MH1, green=Ots17-MH2, blue=Ots17-MH3, purple=Ots17-MH4)
765 represent Y-chromosomes; numeric designations only are labeled on plot. SNP position on Ots17 is
766 given on the x-axis.




767

768

769

Figure 3. Logo plots showing allele frequencies within each of the male haplogroups and the X-chromosome for the RADseq samples. For each SNP, the frequency of alleles within each haplotype (range = 0-1) is shown by the height of the allele. Alleles which are putatively fixed relative to the X-chromosome are bounded by boxes. SNPs identified by tag number that were successfully converted to GT-seq assays are in red on the top X-axis. SNP positions on Ots17 are on the bottom X-axis.

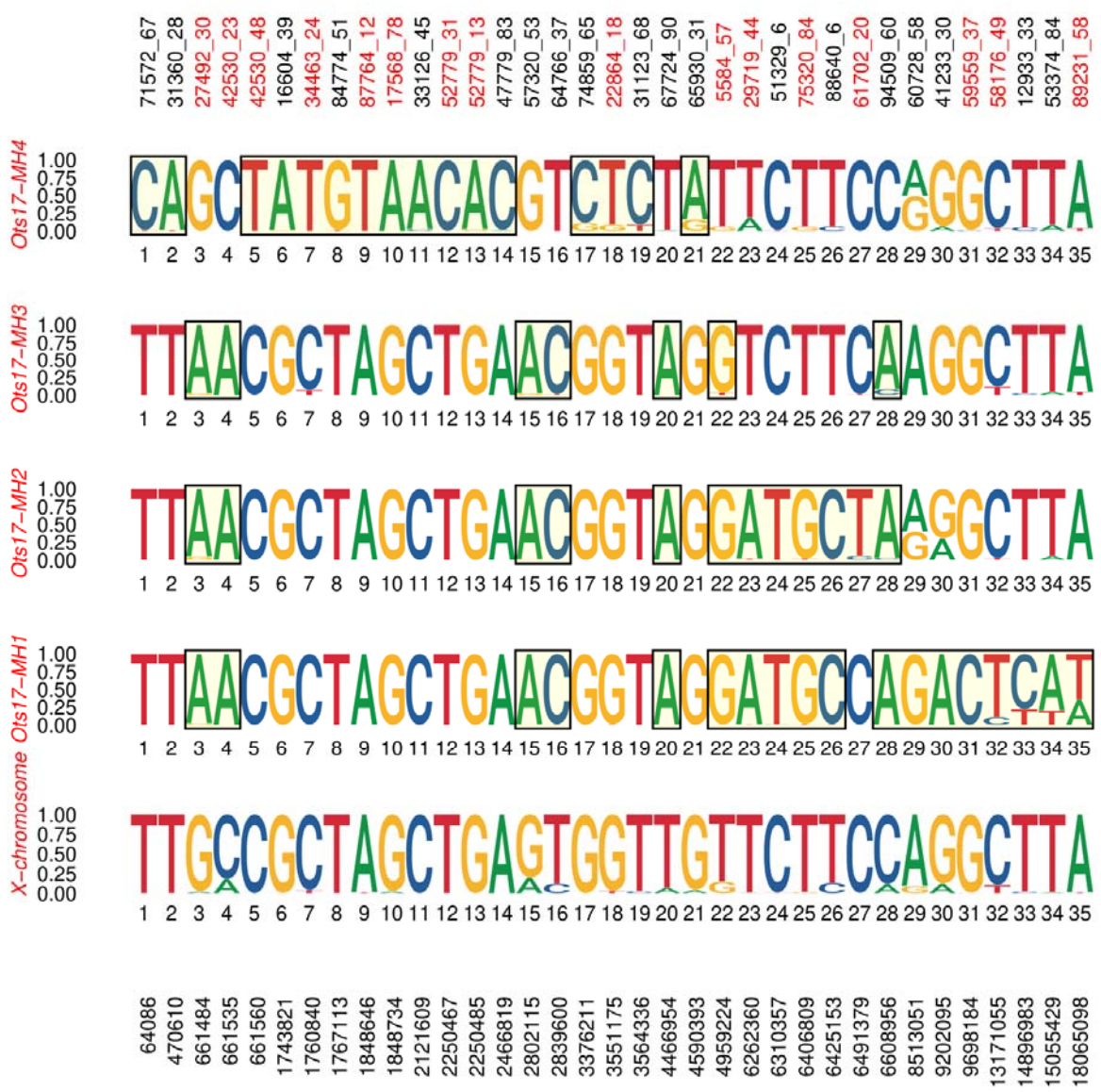
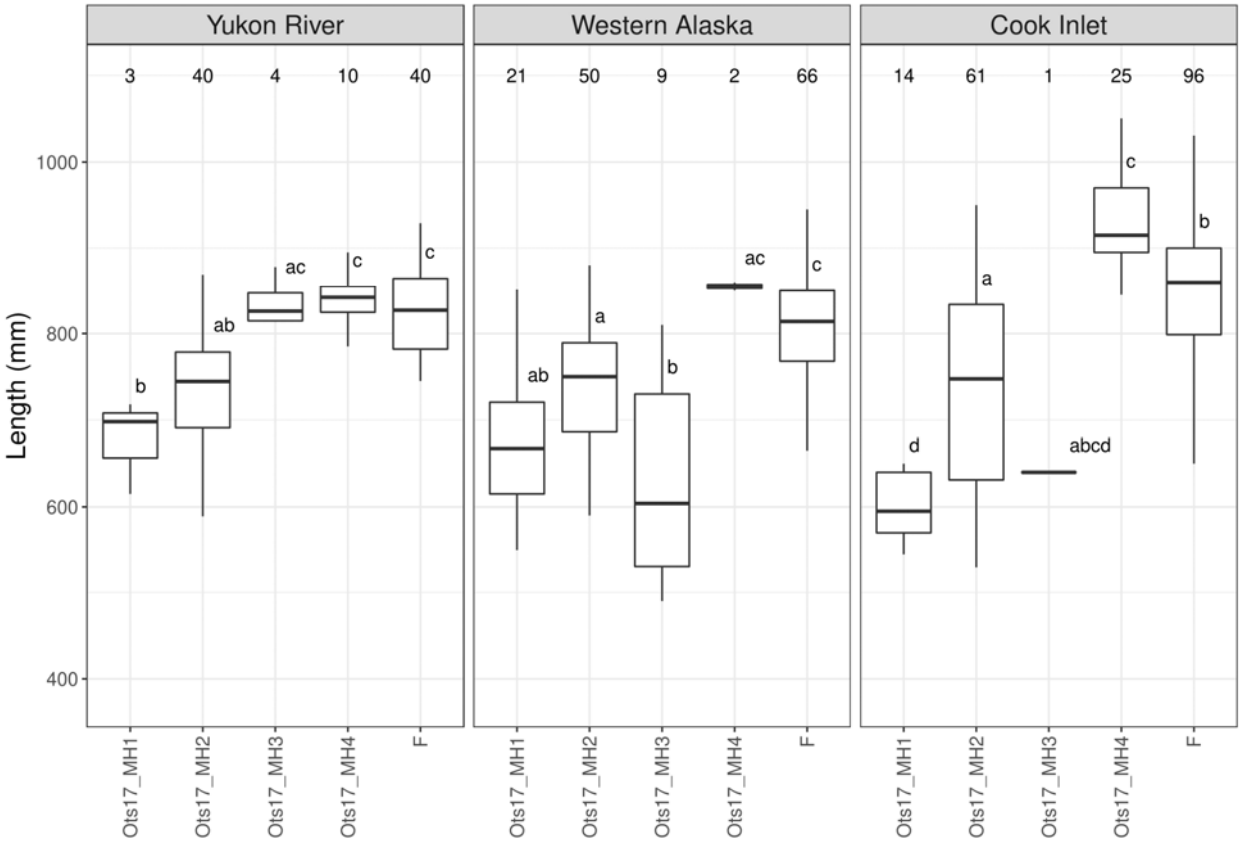
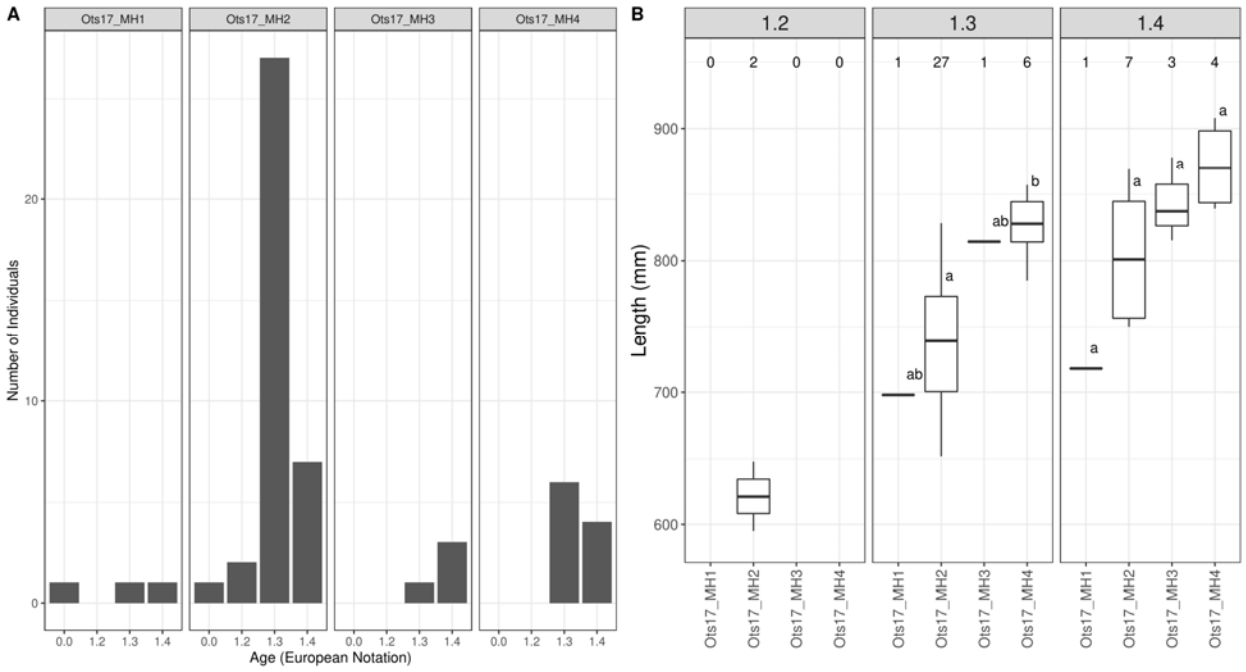


Figure 4. Distribution of size at maturity for each Y-chromosome haplogroup for Alaska Chinook salmon. Female (F) size at maturity is included for comparison. Samples sizes for each haplogroup are given above the boxplots. Within each region, Y-chromosome haplogroups with statistically different lengths are represented by different letters. Haplogroups with two letters (i.e., ab) do not have statistically different size distributions from haplogroups with a or b.



785 Figure 5. Distribution of A) age at maturity and B) size at age for each Y-chromosome haplogroup in the
786 Yukon River. Within each age class, Y-chromosome haplogroups with statistically different lengths are
787 represented by different letters. Sample sizes for each haplotype are given above the boxplots.




788