**Title**:  No Detectable Evidence of Genetic Associations with Anadromous and Resident Masu Salmon Sampled within Hokkaido

**Running Title**: No Genetic Basis of Anadromy in Masu Salmon

**Authors**: Audrey T. Dang1, Itsuro Koizumi2, Ismail K. Saglam3, Michael R. Miller4, William B. Hemstrom4

**Author Affiliations**:

1. Department of Plant Pathology, Global Disease Biology, University of California, Davis

2.

3.

4. Department of Animal Science, University of California, Davis

**Corresponding Author Information:**

William Hemstrom

whemstrom@ucdavis.edu

# Abstract:

Within Masu salmon (Oncorhynchus masou), individuals can exhibit one of two life histories: resident individuals which stay in streams for their entire life span or anadromous individuals which move to the ocean to mature before returning to their natal streams to breed. Current research has suggested that the expression of either life history is mainly influenced by environmental factors such as temperature, which influence critical traits like salmon’s body size, growth rate, lipid stores. Here, we use high-resolution reduced-representation genome sequencing to show that there are no significant genetic associations with migratory decision making in masu salmon in two streams in Hokkaido, Japan. Further research with a larger sample size from more parts of the masu range is needed in order to fully assess potential genetic influences on anadromy and residency in the species.

**Keywords:**

Migration, Anadromy, Masu Salmon, Genomics

# Introduction:

Due to the history of hydrologic modifications such as the development of reservoirs, dams, and irrigation systems, there has been an apparent decrease in the number of migratory fishes in recent years (Pringle *et al.* 2000). In North America, for example,there have been significant decreases in diversity within pacific salmonid species due to hydrologic and dam alterations, and atlantic salmon and anadromous species of sturgeons are also impacted negatively because of a loss of portions to their habitats upstream of dams (Pringle *et al.* 2000). Overall, the diversity within migratory fish species have largely declined by an average of 70% between 1970 and 2016 due to dams and other factors (Deinet *et al.* 2020).

Masu salmon (Oncorhynchus masou, otherwise known as cherry salmon) have the most limited distribution among the Pacific salmon (Augerot *et al.* 2005). Since they spend a greater portion of their life in freshwater rivers or streams than most other migratory salmon, they are particularly exposed to fishing and sportfishing pressures (Miyakoshi *et al.* 2009). Their increased vulnerability makes it critical to understand the details that influence and control migratory life history variance in the species. Masu in particular (like steelhead salmon/rainbow trout Oncorhynchus mykiss and other species to a lesser degree) are partially migratory, with large resident and anadromous populations (Morita *et al.* 2014). As is typical, anadromous masu are initially born in the rivers then migrate to the seas where they develop larger body sizes, then eventually migrate back to their natal rivers in order to breed (Dodson *et al.* 2013). Resident masu in contrast sexually mature at a younger age and stay within the river systems throughout their entire lives (Morita & Nagasawa 2010).

The decision to migrate in masu salmon is thought to be primarily influenced by environmental factors such as temperature alongside other factors such as the salmon’s body size, growth rate, lipid stores, and is heavily sex influenced (Morita *et al.* 2014; Arostegui & Quinn 2019). Migration and residency in the closely related steelhead salmon (Oncorhynchus mykiss) are controlled by similar factors. For example, resident steelhead tend to be larger within the first few months of life when compared to anadromous individuals (Morita *et al.* 2014). Of course, differences in environmental factors can influence growth rates and therefore also migration in steelhead. For example, during poor nutrient conditions, a higher proportion of individuals than expected from genetics may be anadromous; whereas during flooding or nutrient rich periods, a higher proportion of resident individuals may be observed (Kelson *et al.* 2020). Regardless of environmental conditions, however, those individuals within a population which grow faster as juveniles are more likely to stay as residents than otherwise, even under similar conditions (Kendall *et al.* 2015).

Migration comes with a tradeoff. In steelhead, migratory juveniles tend to have low survival rates to adulthood, which is compensated for by the high fecundity of anadromous females that can significantly exceed that of residents (Pearse *et al.* 2019). This pattern likely holds true for masu as well (Morita *et al.* 2014). This trade-off is not as pronounced in male steelhead, which can mature as early as freshwater residents, avoiding the high mortality associated with migration (Pearse *et al.* 2019). This results in greater frequency of anadromy in females; a trend also seen in masu salmon (Morita *et al.* 2014). Consequently, changes in environmental factors, such as those caused by climate change, have been known to impact the decision of the masu to be either resident or anadromous (Morita *et al.* 2014). For example, Morita et al (2014) has shown that the number of resident males have increased while the proportion of anadromous males and the proportion of delayed migrants within both females and males have decreased as temperatures have increased over time, leading to a significant decrease in the overall number of anadromous individuals.

In contrast to environmental factors, the genetic factors that contribute to migration in masu salmon have not been well studied. Despite the prevalence of environmental control of migration in masu, it is not unreasonable to expect genetic elements to play some role in governing migratory decisions in the species. For example, within steelhead, there is strong evidence suggesting both an environmental and genetic basis for migratory life history. In 1944, Naeve conducted common garden and transplant experiments with sympatric rainbow trout and steelhead and found that migratory status was heritable after observing that rainbow trout offspring were considerably more likely to be resident themselves, whereas crosses between anadromous individuals resulted in the most anadromous smolts [double check paper that this was in]. Indeed, recent studies have found that the decision to migrate in steelhead is strongly linked to a large genomic inversion on chromosome Omy5 throughout their geographic range which seems to maintain a set of linked genes that controls migratory behavior (Pearse *et al.* 2014; Arostegui & Quinn 2019; Kelson *et al.* 2020). The genetic basis for migration is also likely indirectly mediated by physiological characteristics such as growth rate, which are also known to be influenced by the same large genetic elements that differ between anadromous and resident individuals (Miller *et al.* 2012). Taken together, it seems clear that migration in steelhead is jointly controlled by the interplay of genetic and environmental factors.

The degree of genetic control of migration in masu, however, is still not well known. To improve our understanding of the mechanisms behind residency and anadromy in the species, we sequenced a set of anadromous and resident masu individuals using high-resolution Restriction Associated Digest (RAD) sequencing. However, we did not find any clear indicator of the genetic difference between our anadromous and resident masu salmon. While this may suggest that there is not that there is a strong genetic basis for masu salmon migratory life history variation, further studies using a bigger sample size from a larger geographic range would need to be conducted in order to fully investigate the influences of genetics on Masu salmon migratory life cycles.

# Materials and Methods:

## Genotyping

We extracted DNA from our samples using the magnetic beads and prepared Restriction Associated Digest (RAD) libraries using the Pst1 restriction enzyme according to Ali et al (2016), then sequenced the results on an Illumina Hi-Seq 4000 using 150bp paired-end sequencing. We aligned the data to the *O. mykiss* v6 reference genome (Pearse *et al.* 2019) using Burrows-Wheeler Alignment algorithm (Li *et al.* 2009). We then filtered the data for PCR duplicates and improper pairs using SAMtools (Li *et al.* 2009). Next, we called genotypes by using the SAMtools genotype likelihood models (Li *et al.* 2009) used in the ANGSD software package with a minimum mapping and base call quality score of 20, a SNP p-value of 1e-8, and a minimum acceptable minor allele frequency of 0.05 (Korneliussen *et al.* 2014). We then removed loci out of Hardy-Weinberg Equilibrium () according to Wigginton et al.’s (Wigginton *et al.* 2005) method using the snpR R package (Hemstrom & Jones 2021).

## Diversity Stats

We calculated the observed heterozygosity (), the average number of pairwise differences (), number of private alleles, and fixation index () between populations using the snpR R package (Weir & Cockerham 1984; Hohenlohe *et al.* 2010; Hemstrom & Jones 2021).

## Association Testing

We performed a basic association test against migratory phenotype using ANGSD using the first 20 Principal Component Analysis components as covariates to correct for the related among individuals and population structure (Korneliussen *et al.* 2014). To increase our association resolution, we used used genotype likelihoods instead of called gneotypes to conduct our association test, generated in ANGSD using the same parameters described above. In order to see if any more complex genetic relationships could predict migratory phenotype, we then performed a random forest analysis in R-studio using the ranger R package (Wright & Ziegler 2017). We used 100,000 trees and used mtry of half the number of called SNPs.

## Population Structure and Migratory Relatedness

We also conducted a Principle Component Analysis chart (PCA) using the ANGSD-IBS method (Korneliussen *et al.* 2014). To look for patterns in relatedness between resident and anadromous individuals, we calculated the mean PC1 and PC2 coordinates for each stream, then calculated the distance between each individual sample the mean for the other population. We fit a simple linear regression with the distance to the other population cluster as the response variable and the migratory phenotype as the explanatory variable.

# Results:

## Genotyping

The number of unfiltered reads are 163,200,129. After filtering, the number of reads were 90,449,126. We called a total of 71,175 SNPs and 67,256 IBS loci.

## Diversity Stats

The mean values are relatively similar between the anadromous and residents individuals and between the CH and TK populations (see Table 1). The values between the two populations and between anadromous and resident individuals were very low, which indicated that there is not much genetic diversity between the two populations and two phenotypes. There are also no private alleles between the two populations, which also indicated low genetic differentiation between the two populations of salmon.

## Association

We did not find any strong associations with migratory life history status at any loci (Figure 1), and the random forest only successfully predicted 25 out of 52 anadromous individuals correctly and 17 out of 44 resident individuals correctly for a total prediction accuracy of 38.6%.

## Influence of Anadromy on Population Structure

We found that the anadromous individuals clustered significantly closer to the cluster center for individuals from the other population (one-way ANOVA, p-value = 0.0419, see Figure 2). This indicates that the anadromous individuals within one population were more closely related to the individuals in the other population than were residents.

# Discussion:

## Genetic Associations with Migration

We did not find any significant genetic associations with migratory phenotype (see Figure 1). This indicates that in the streams we examined, there may not be any loci in the genome that strongly contribute to migratory life history decisions. It is not therefore surprising that the random forest which we conducted also did not have strong predictive power for anadromy and residency, since it would not be able to accurately group the individuals to their actual life history if there are no genetic elements that contribute to migratory phenotype.

However, our study included samples from only a small region at the center of the masu range. It is possible that if we sampled from other areas, we could uncover loci that influence migratory life history phenotypes under environmental conditions that differ from those in our study systems. This could occur if the effect of migratory genes depends on environmental conditions (GxE). This phenomena is seen in steelhead, where variance in environmental conditions across the species’ range can determine whether life histories are more genetically or environmentally based (Kendall *et al.* 2015).

## Patterns of Relatedness Between Anadromous and Resident Individuals

We found that although our two populations (TK and CH) were genetically distinct from one another, there was substantial overlap between the anadromous and resident individuals within each population, indicating that resident and anadromous individuals freely interbreed within populations. Interestingly, we did find that anadromous individuals were significantly more closely related to individuals from the stream from which they did not originate. This could potentially be attributed to where the samples were collected. For example, if the anadromous individuals from one stream were collected closer to the river mouth than residents, and thus geographically closer to the other stream, this could explain why we observed that they were more genetically similar to individuals from that stream. Alternatively, if anadromous individuals are more likely to stray than residents (since they may return to the wrong river system), and their offspring are more likely to be anadromous themselves, then this could also produce the pattern we observed.

However, this hypothesis could only be true if there are transgenerational environmental or genetic correlations with anadromy. Given that we did not find any obvious genetic associations with anadromy, the latter seems more likely. For example, transgenerational environmental correlations anadromy could exist if anadromous individuals tend to breed in areas where there are less nutrients available, since smaller body sizes induce the decision of an anadromous life cycle. Under these circumstances, the offspring of migrants would be more likely to be migrants themselves, allowing for the localized build up of alleles brought in by strays.

## Conservation Ramifications

Our results suggest that, at least in some areas, residency and anadromy within masu salmon are not significantly genetically controlled and that fisheries managers and conservation groups should therefore focus more on the possible environmental correlations with migration (such as temperature and resource availability) when planning conservation efforts targeting masu salmon. Unfortunately, this also means that genetic monitoring for the anadromous potential of a stream is probably often not possible, since it is not possible to predict migratory phenotype from genotypic data. These results may not extrapolate across the entire masu range: further studies with larger sample sizes from more areas could help determine if migratory life history decisions are influenced by the additive or multiplicative effects of genetics and environment or simply a result of purely environmental effects across the full range of the species.

**Acknowledgements:**

# References:

Ali OA, O’Rourke SM, Amish SJ, *et al.* (2016) [RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping](https://doi.org/10.1534/genetics.115.183665). *Genetics*, **202**, 389–400.

Arostegui MC, Quinn TP (2019) [Reliance on lakes by salmon, trout and charr (Oncorhynchus, Salmo and Salvelinus): An evaluation of spawning habitats, rearing strategies and trophic polymorphisms](https://doi.org/10.1111/faf.12377). *Fish and Fisheries*, **20**, 775–794.

Augerot X, Foley DN, Steinback C, *et al.* (2005) *Atlas of Pacific salmon: the first map-based status assessment of salmon in the North Pacific*. University of California Press Berkeley, CA.

Deinet S, Scott-Gatty K, Rotton H, *et al.* (2020) The living planet index (LPI) for migratory freshwater fish: Technical report.

Dodson JJ, Aubin-Horth N, Thériault V, Páez DJ (2013) [The evolutionary ecology of alternative migratory tactics in salmonid fishes](https://doi.org/10.1111/brv.12019). *Biological Reviews*, **88**, 602–625.

Hemstrom W, Jones M (2021) [snpR: user friendly population genomics for SNP datasets with categorical metadata](https://doi.org/10.22541/au.161264719.94032617/v1). *Authoria*, 507615.

Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) [Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags](https://doi.org/10.1371/journal.pgen.1000862). *PLOS Genetics*, **6**, 1–23.

Kelson SJ, Carlson SM, Miller MR (2020) [Indirect genetic control of migration in a salmonid fish](https://doi.org/10.1098/rsbl.2020.0299). *Biology Letters*, **16**, 20200299.

Kendall NW, McMillan JR, Sloat MR, *et al.* (2015) [Anadromy and residency in steelhead and rainbow trout (oncorhynchus mykiss): A review of the Processes and Patterns](https://doi.org/10.1139/cjfas-2014-0192). *Canadian Journal of Fisheries and Aquatic Sciences*, **72**, 319–342.

Korneliussen TS, Albrechtsen A, Nielsen R (2014) [ANGSD: Analysis of Next Generation Sequencing Data](https://doi.org/10.1186/s12859-014-0356-4). *BMC Bioinformatics*, **15**, 356.

Li H, Handsaker B, Wysoker A, *et al.* (2009) [The Sequence Alignment/Map format and SAMtools](https://doi.org/10.1093/bioinformatics/btp352). *Bioinformatics*, **25**, 2078–2079.

Miller MR, Brunelli JP, Wheeler PA, *et al.* (2012) [A conserved haplotype controls parallel adaptation in geographically distant salmonid populations](https://doi.org/10.1111/j.1365-294X.2011.05305.x). *Molecular Ecology*, **21**, 237–249.

Miyakoshi Y, Sasaki Y, Fujiwara M, *et al.* (2009) [Implications of Recreational Fishing on Juvenile Masu Salmon Stocked in a Hokkaido River](https://doi.org/10.1577/m08-107.1). *North American Journal of Fisheries Management*, **29**, 33–39.

Morita K, Nagasawa T (2010) [Latitudinal variation in the growth and maturation of masu salmon (Oncorhynchus masou) parr](https://doi.org/10.1139/F10-028). *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 955–965.

Morita K, Tamate T, Kuroki M, Nagasawa T (2014) [Temperature-dependent variation in alternative migratory tactics and its implications for fitness and population dynamics in a salmonid fish](https://doi.org/10.1111/1365-2656.12240). *Journal of Animal Ecology*, **83**, 1268–1278.

Pearse DE, Barson NJ, Nome T, *et al.* (2019) [Sex-dependent dominance maintains migration supergene in rainbow trout](https://doi.org/10.1038/s41559-019-1044-6). *Nature Ecology & Evolution*, **3**, 1731–1742.

Pearse DE, Miller MR, Abadía-Cardoso A, Garza JC (2014) [Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout](https://doi.org/10.1098/rspb.2014.0012). *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20140012.

Pringle CM, Freeman MC, Freeman BJ (2000) [Regional Effects of Hydrologic Alterations on Riverine Macrobiota in the New World: Tropical-Temperate Comparisons: The massive scope of large dams and other hydrologic modifications in the temperate New World has resulted in distinct regional trends of b](https://doi.org/10.1641/0006-3568(2000)050%5b0807:REOHAO%5d2.0.CO;2). *BioScience*, **50**, 807–823.

Weir BS, Cockerham CC (1984) [Estimating F-Statistics for the Analysis of Population Structure](https://doi.org/10.1111/j.1558-5646.1984.tb05657.x). *Evolution*, **38**, 1358–1370.

Wigginton JE, Cutler DJ, Abecasis GR (2005) [A Note on Exact Tests of Hardy-Weinberg Equilibrium](https://doi.org/10.1086/429864). *The American Journal of Human Genetics*, **76**, 887–893.

Wright MN, Ziegler A (2017) [ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R](https://doi.org/10.18637/jss.v077.i01). *Journal of Statistical Software; Vol 1, Issue 1 (2017)*.

**Data Accessibility:**

Raw fastq and metadata files for the data presented here will be available from the NCBI SRA upon publication. The Github repository containing the scripts used to analyze this data are located at <https://github.com/hemstrow/masu_migration>. The script used to produce this paper specifically is located at <https://github.com/hemstrow/masu_migration/blob/master/paper/paper_draft.Rmd>.

**Benefit Statement:**

**Author Contributions:**

ATD and WBH analyzed the data. ATD prepared the manuscript. IK collected samples, designed the study, and provided feedback on the manuscript. MRM and IKS assisted with study design and provided feedback on the paper.

# Tables and Figures:

Tables go here

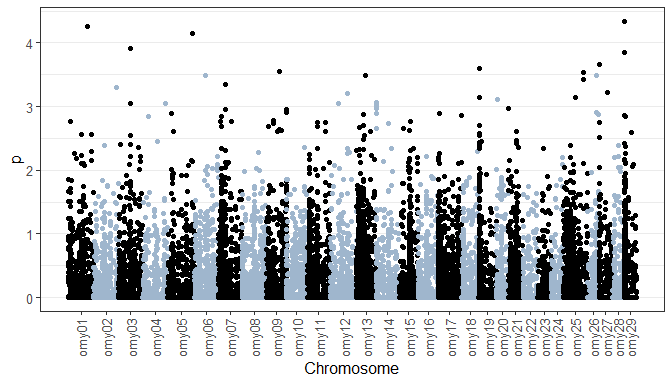


Figure 1: Association across the genome with migratory status showing that no loci were found that were significantly associated with the given traits (anadromy and residency)

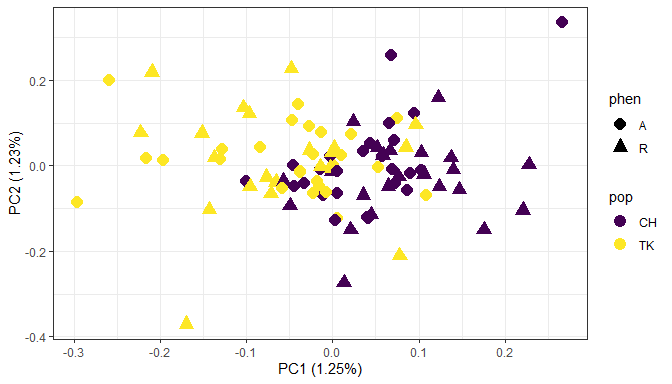


Figure 2: PCA plot showing the relatedness between anadromous and resident individuals in TK and CH populations