Spring-run Workshop Fact Sheet

Identifying Spring-run

August 31st 2020

Brett Harvey[1], Melinda Baerwald[1], Daphne Gille[1]
[1]California Department of Water Resources

Introduction

Migrating juvenile spring-run Chinook Salmon occur within a mixed population of the four salmon runs in the Central Valley, and are morphologically indistinguishable from these other runs to the naked eye. However, some means of identifying juvenile spring-run salmon from the other runs migrating through the Delta will likely be critical to the development of a robust juvenile production estimate (JPE). Several approaches exist or are in development that may be applied. A single approach may not be optimal for all possible JPE approaches, and different approaches may need to be applied in combination.

Deterministic Length at Date: Historical and Current Use

The Length at Date approach assigns a run identification to a juvenile salmon based on capture date and fork length (Figure 1). The approach was devised in the 1970s after the federal listing of winter run under the Endangered Species Act as a tool to assess take of juvenile winter run salmon by the state and federal water projects (Harvey 2011). Currently there are two different Length-at-Date criteria applied in the Central Valley. The Delta Model is used for fish sampled at the state and federal water project salvage facilities and in the Yolo Bypass Fish Monitoring Program; the River Model is used everywhere else.

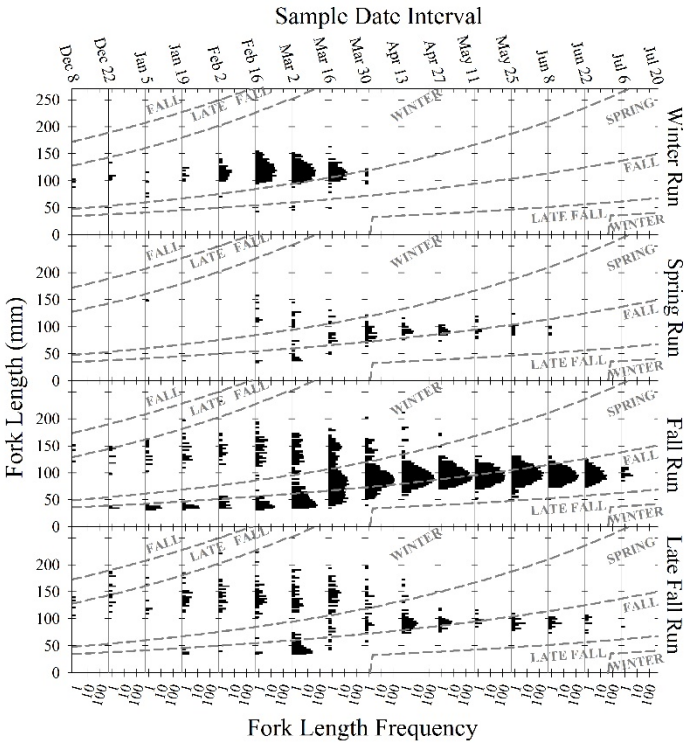

Figure 1. Fork length distribution of genetically assigned Chinook salmon (horizontal bars) overlaid on Delta Model length-at-date size ranges (gray dash lines) for fish sampled at salvage facilities from 2004 to 2010. Note: genetic tests differentiated spring run from only the Butte Creek and Mill-Deer Creeks populations (Harvey 2013).

The approach relies on two major assumptions: juvenile salmon of different runs hatch during different segregated periods of the calendar year, and all juvenile salmon grow at a constant rate. Genetic analyses show that neither of these assumptions are true, and there is large overlap in size distributions between runs (Harvey 2013, 2015). Fall run in particular have considerable size overlap with spring run (Figure 1, third panel). Due to the large abundance of fall run relative to spring run, this overlap can lead to a high number of false positive spring run assignments by Length-at-Date (Figure 2, green slice of upper left pie). Nonetheless, the approach continues to be used for run assignment in many if not most Central Valley monitoring programs, ostensibly because its speed and simplicity is useful for "real-time" management, and because its application has minimal cost.

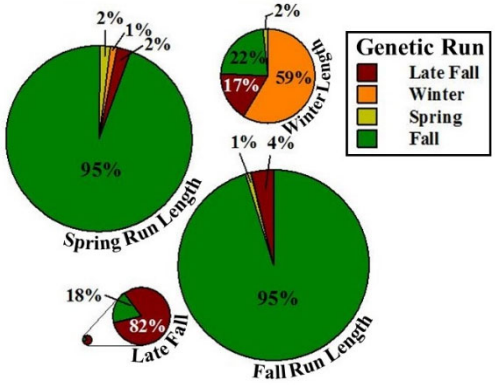

Figure 2. Proportion of genetic run sampled within each Delta Model length-at-date range at the state and federal salvage facilities, 2004-2010. Pie size represents relative number of sampled fish, N = 11,609 (Harvey 2013).

Probabilistic Length at Date: Future Applications

A Probabilistic Length-at-Date approach is currently under development (Noble Hendrix, Queda Consulting, pers comm). The Probabilistic approach has a similar basic construct to the original Deterministic approach in that it relies on the fork length and sample date of a juvenile salmon to assign a run. Unlike the Deterministic approach, the Probabilistic approach may assign more than one run for a given juvenile salmon, along with a probability for each assigned run (Figure 3). The assignment probability for a given fork length and date will be based on genetic identification of catch from the preceding years of various monitoring programs, and would be updated regularly throughout a migration season as ongoing genetic identifications become available. In addition to genetic information, variables such as geographic area, flow, and temperature may be incorporated into the Probabilistic assignment model. The updated model predictions would be posted for real-time use on an internet platform such as SacPASS, possibly available as an R application.

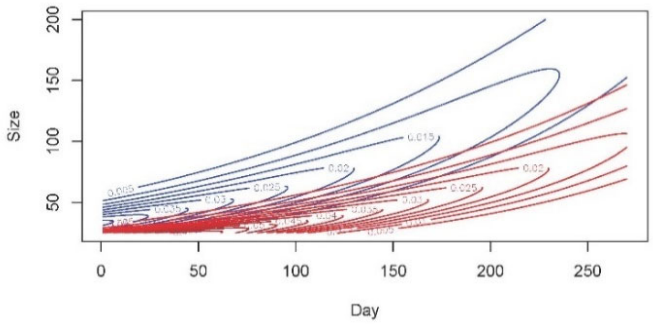

Figure 3. Conceptual depiction of Probabilistic Length-at-Date size ranges for two runs (Noble Hendrix, 2019).

Genetic Identification: The Basics

All salmon runs were originally and primarily defined by phenotypic differences among adult Chinook run-timing and spawning periods, not by differences in genetic composition or morphology. *So how do genetic tests currently in use differentiate among Central Valley salmon runs?*

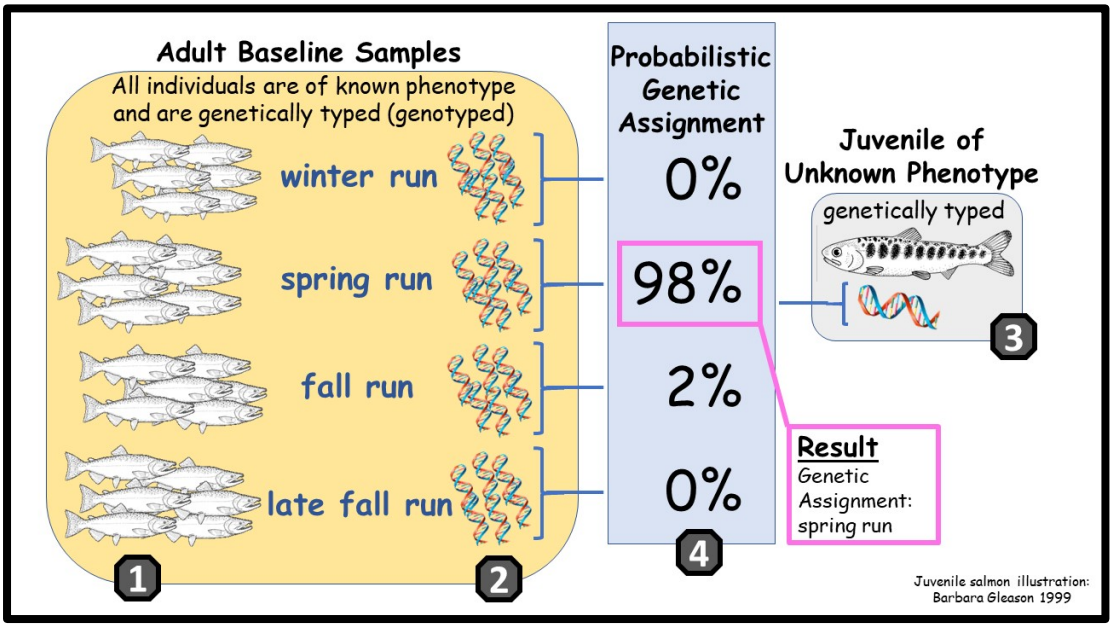

Figure 4. Basic steps in the process of genetic identification of a salmon of unknown origin include 1) collecting tissue samples from a larger number of adult salmon of known run type, called the baseline samples, 2) analyzing the genetic composition of the baseline samples at specific locations in the salmon genome, called genetic markers, 3) analyzing the genetic composition of the unknown salmon at those same genetic markers, and 4) comparing the unknown salmon genetic composition to the baseline samples to derive an probabilistic assignment.

Baselines and Genetic Markers

Thousands of genetic samples are collected from adult salmon displaying the different run-timing phenotypes, and from different regions (Figure 4). These adults form a ***baseline***. The genetic composition of a set of ***genetic markers*** are characterized for each adult fish in the baseline. Most genetic tests these days use genetic markers called Single Nucleotide Polymorphisms (SNPs). For each SNP, the type of nucleotides (the building blocks of DNA) are identified for a single nucleotide pair at a specific location on the genome (Figure 5). SNPs are grouped into panels tailored for specific applications and objectives. Various lab techniques may be employed for genetic typing a SNPs panel, which differ in the number of SNPs for a given panel (96 to over 10,000), sample processing cost, turnaround time, throughput, and other factors that may influence the approach used (Figure 6, Meek & Larson 2019). Genetics labs constantly harness new technologies and develop new SNPs to update their panels and improve identification accuracy and certainty.

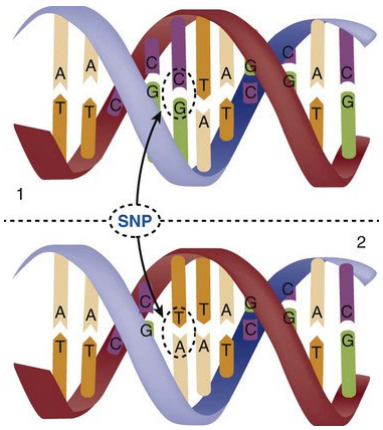

Figure 5. Graphic representation of a single nucleotide polymorphism (SNP) illustrating CG base pair on one DNA segment has changed to a TA base pair on the other. SNPs are the most common type of genetic variation between individuals of the same species (clinicalgate.com).

Probablistic Assignment

The SNPs for a specific genetic test are selected based on their collective ability to differentiate among populations of interest. In general, increasing the number of genetic markers improves the ability to differentiate among populations with increasing certainty, but the degree of improvement is highly case specific. Once a genetic test is developed, salmon of unknown origin are genetically typed at the same set of genetic markers as the baseline adults, and are assigned a probability of belonging to each run based on similarity/dissimilarity with the adult fish representing the different phenotypes in the baseline.

Genetic Test Considerations

The appropriate genetics test for differentiating among Central Valley salmon populations varies dependent on the needs and conditions of a specific application. These needs include both the biological question at hand, and the logistical requirements and constraints. Primary among biological questions is the type and level of population differentiation required, such as spring run versus not spring run, or spring run tributary of origin. Logistical considerations and constraints include cost, turn-around time, sample throughput, and the number of genetic markers needed to achieve a given level of identification accuracy (Figure 6). In general, finer scale population resolution will have higher costs. Once these parameters are defined, a geneticist can determine the most

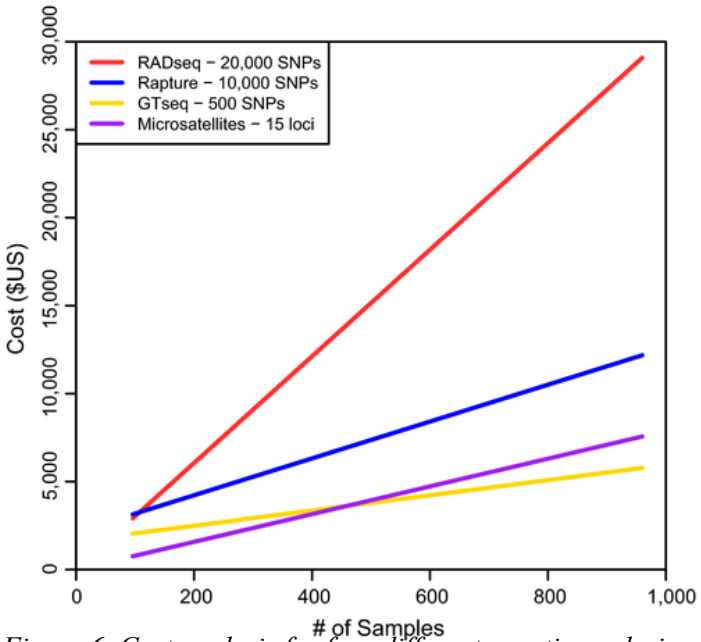

Figure 6. Cost analysis for four different genetic analysis approaches employing different numbers of SNPs. Note: results may vary for other labs dependent on lab equipment and capabilities (Meek and Larson 2019).

appropriate available techniques, and develop an approach, or explain the tradeoffs of different potential approaches.

Special Challenge: Feather River Spring Run

When Oroville Dam was constructed, spring run access to historical holding and spawning habitat in the upper watershed was eliminated. Feather River spring run have continued to migrate early and hold in the dams cold tailwaters, an area overlapping the historical spawning region of Feather River fall run. The subsequent interbreeding between these once reproductively isolated populations has made genetic differentiation difficult, giving rise to the term Feather River "sprall" run.

Recent Advancements in Chinook Genetics

Recently, the phenotypic trait that defines spring-run salmon, early migration with immature eggs, was found to be strongly associated with SNPs located around and within an adjacent pair of genes, Greb1L and Rock1 (Prince et al. 2017, Narum et al. 2018). Genetic tests on Rogue River and Klamath River Chinook salmon, using only two SNPs in the region of Greb1L, provided strong evidence that salmon with spring run genotypes at both chromosomes (i.e. homozygous spring run genotypes) almost universally express the early-migrating spring run phenotype in these rivers, and homozygous fall run genotypes express a late-migrating fall run phenotype, while salmon with a spring run allele on one chromosome and a fall run

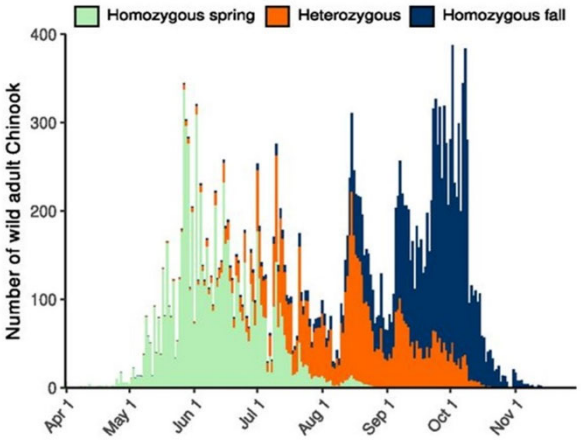

Figure 7. Run timing distribution of Rogue River Chinook salmon genotyped as homozygous spring run, homozygous fall run, or heterozygous (fall and spring run), at two SNPs in the region of the Greb1L gene (Thompson et al. 2019).

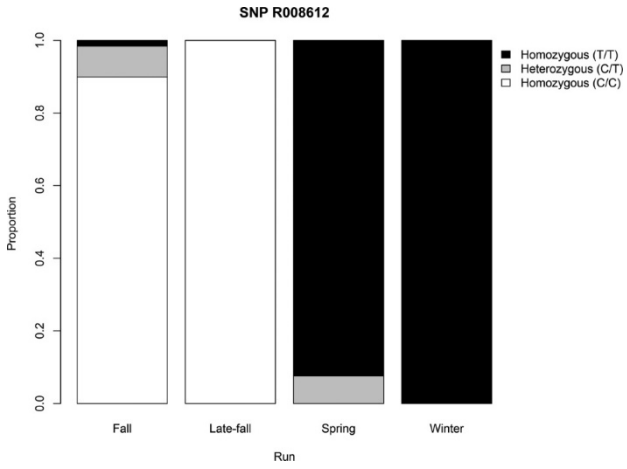

Figure 9. Proportion of individuals among the four Central Valley Chinook salmon runs that had each of three possible genotypes at a single GREB1L associated SNP (Meek et al. 2020).

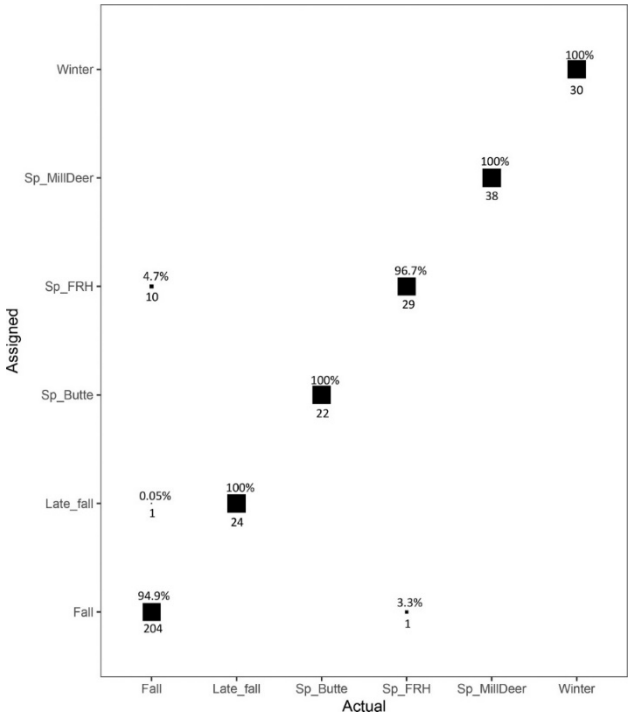

Figure 8. Matrix comparing actual phenotypic run with genetically assigned run for Central Valley salmon populations using a large number of SNPs (Meek et al. 2020).

allele on the other chromosome (i.e. heterozygous), display intermediate run-timing which overlaps with the run-timing of homozygous spring and fall run salmon (Figure 7, Thompson et al. 2019). The degree of overlap for heterozygous salmon appears to vary by watershed. Ongoing work shows a similarly strong association exists between Greb1L associated SNPs and phenotypic run-timing for Sacramento River Chinook salmon, including Feather River salmon (Figure 8: Meek et al. 2020; Carlos Garza NOAA, personal communication). As previously discussed, assignment accuracy of current genetic tests varies dependent on the level of differentiation a test was designed to resolve, which in turn depends on the purpose, cost, and other constraints that were considered in the design of the test. If the objective is to distinguish spring run from other runs, or to distinguish between the four Central Valley runs, a high degree of accuracy can be obtained (Figure 9: Meek et al. 2020, Carlos Garza pers comm), but fine-scale differentiation within a run will likely be associated with higher costs per individual sample, which may become prohibitive for large sampling programs.

CRISPR-based Genetic Assays

Rapid portable genetic testing tools, first developed in 2017 for human disease detection during outbreaks, are being developed for potential field-based fish identification (Baerwald et al. 2020). The first of these technologies was SHERLOCK (Gootenberg et al. 2017), followed by DETECTR (Chen et al. 2018), but other similar innovative tools will likely be available in the future. Essentially, these tests search for the presence of a specific DNA nucleotide sequence in a sample of genetic material, such as a swab taken from the mucus of a juvenile salmon. If the nucleotide sequence is in the sample, the detection is indicated by a line on a paper strip (like a home pregnancy test) or a fluorescent reaction (Figure 10). These tests can be carried out anywhere (field, salvage facility, lab), with minimal equipment and training, and return results in as little as 30 minutes. Once a specific test has been developed, test production is very low cost.

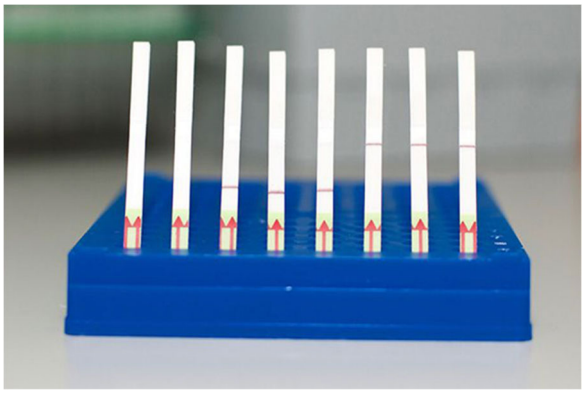

Figure 10. Photograph of SHERLOCK technology applied using test strips which visually indicate genetic identification by the presence or absence of a line on each test strip (Photo Credit: Zhang lab, Broad Institute of MIT and Harvard, Broadinstitute.org).

Questions

For JPE scenarios and priorities proposed by breakout group participants (e.g. cost, turnaround time, population resolution, and acceptable level of uncertainty):
At what life stages and geographic locations would identification be necessary?
What existing identification tools, or set of tools could be used?
What are the challenges to identification?
What new tools would and could be developed to meet those challenges?

Example: What are the challenges related to differentiating Sacramento from San Joaquin spring run at the salvage facilities, and what existing or potential tools could address those challenges?

References Consulted for This Fact Sheet

Baerwald, Melinda R., Alisha M. Goodbla, Raman P. Nagarajan, Jonathan S. Gootenberg, Omar O. Abudayyeh, Feng Zhang, Andrea D. Schreier. 2020. Rapid and accurate species identification for ecological studies and monitoring using CRISPR-based SHERLOCK. Molecular Ecology Resources 20:4, 961-970. DOI: 10.1111/1755-0998.13186

Ford, M., K. Nichols, R. Waples, E. C. Anderson, M. Kardos, I. Koch, G. McKinney, M. R. Miller, J. Myers, K. Naish, S. Narum, K. G. O'Malley, D. Pearse, T. Seamons, A. Spidle, P. Swanson, T. Q. Thompson, K. Warheit, and S. Willis. 2020. Reviewing and Synthesizing the State of the Science Regarding Associations between Adult Run Timing and Specific Genotypes in Chinook Salmon and Steelhead: Report of a workshop held in Seattle, Washington, 27–28 February 2020. U.S. Department of Commerce, NOAA Processed Report NMFS-NWFSC-PR-2020-06. https://doi.org/10.25923/mv3n-zb79

Harvey, B. N. 2011. Length-at-date criteria to classify juvenile Chinook Salmon in the California Central Valley: development and implementation history. Interagency Ecological Program Newsletter 24(3).

Harvey, Brett N., Carol Stroble. 2013. Comparison of genetic versus Delta Model length-at-date race assignments for juvenile Chinook salmon at state and federal south delta salvage facilities. Technical Report 88 for the Interagency Ecological Program for the San Francisco Bay/Delta Estuary, March 2013.

Harvey, Brett N., David P. Jacobson & Michael A. Banks. 2014. Quantifying the Uncertainty of a Juvenile Chinook Salmon Race Identification Method for a Mixed-Race Stock. North American Journal of Fisheries Management, 34:6, 1177-1186, DOI: 10.1080/02755947.2014.951804

Hendrix, Noble, Corey Phillis, Brett Harvey, Josh Israel. Probabilistic Length-at-Date (PLAD): Technical Discussion. Presentation at the PLAD Technical Team. June 19, 2019, Sacramento CA.

Meek, Mariah H., Melinda R. Baerwald, Molly R. Stephens, Alisha Goodbla, Michael R. Miller, Katharine M. H. Tomalty, Bernie May. 2016. Sequencing improves our ability to study threatened migratory species: Genetic population assignment in California's Central Valley Chinook salmon. Ecology and Evolution, 6: 7706–7716.

Meek, MH, WA Larson. 2019. The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. - Molecular Ecology Resources, 19(4): 795-803.

Meek, Mariah H., Molly R. Stephens, Alisha Goodbla, Bernie May, and Melinda R. Baerwald. 2020. Identifying hidden biocomplexity and genomic diversity in Chinook salmon, an imperiled species with a history of anthropogenic influence. Canadian Journal of Fisheries and Aquatic Sciences. 77: 534–547 DOI: 10.1139/cjfas-2019-0171

Narum SR, Di Genova A, Micheletti SJ, Maass A. 2018 Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. Proc. R. Soc. B 285: 20180935. http://dx.doi.org/10.1098/rspb.2018.0935

Prince, Daniel J., Sean M. O'Rourke, Tasha Q. Thompson, Omar A. Ali, Hannah S. Lyman, Ismail K. Saglam, Thomas J. Hotaling, Adrian P. Spidle, Michael R. Miller. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Science Advances. 3. e1603198. DOI: 10.1126/sciadv.1603198.

Thompson, Tasha Q., M. Renee Bellinger, Sean M. O'Rourke, Daniel J. Prince, Alexander E. Stevenson, Antonia T. Rodrigues, Matthew R. Sloat, Camilla F. Speller, Dongya Y. Yang, Virginia L. Butler, Michael A. Banks, Michael R. Miller. 2019. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. Proceedings of the National Academy of Sciences 116 (1): 177-186. DOI:10.1073/pnas.1811559115

Personal Communications

Carlos Garza, NOAA Fisheries (phone conversation, July 16, 2020)
Mariah Meek, Michigan State University (phone conversation, July 21, 2020)
Jeffrey Rodzen, CA Department of Fish and Wildlife (MS Teams meeting, August 18, 2020)