**Alaska Hatchery Research Group** **Technical Document:**[i] 11

**Title:** Prioritization of pink salmon samples and analyses 2015/2016 **Version:** 1.0
**Authors:** K. Shedd, T. Dann, S. Moffitt and C. Habicht
**Date:** August 8, 2015

## Abstract

With the end of the 2015 field season, samples are available from a complete generation of odd year pink salmon samples for five streams. This will allow the first set of adult-to-adult parentage analyses for the fitness component of the research plan. However, increased sampling effort combined with unusually high-abundances in 2013 and 2015 resulted in a larger than anticipated number of samples for potential analyses (~60,000 samples). It is necessary to prioritize analyses in consideration of project objectives because there are more samples available for analysis than current funding and laboratory capacity (both otolith and genetics) can support. This document reviews project analysis components, proposes a priority order for genetic analysis of samples, and communicates the factors considered in this prioritization.

## Background of AHRP

Extensive ocean-ranching salmon aquaculture is practiced in Alaska by private non-profit corporations (PNP) to enhance common property fisheries. Most of the approximately 1.7B juvenile salmon that PNP hatcheries release annually are pink salmon in Prince William Sound (PWS) and chum salmon in Southeast Alaska (SEAK; Vercessi 2014). The large scale of these hatchery programs has raised concerns among some that hatchery fish may have a detrimental impact on the productivity and sustainability of natural stocks. Others maintain that the potential for positive effects exists. To address these concerns ADF&G convened a Science Panel for the Alaska Hatchery Research Program (AHRP) whose members have broad experience in salmon enhancement, management, and natural and hatchery fish interactions. The AHRP was tasked with answering three priority questions:

I. *What is the genetic stock structure of pink and chum salmon in each region (PWS and SEAK)*?;
II. *What is the extent and annual variability in straying of hatchery pink salmon in PWS and chum salmon in PWS and SEAK*?; and
III. *What is the impact on fitness (productivity) of natural pink and chum salmon stocks due to straying of hatchery pink and chum salmon*?

[i] This document serves as a record of communication between the Alaska Department of Fish and Game Commercial Fisheries Division and other members of the Science Panel of the Alaska Hatchery Research Program. As such, these documents serve diverse ad hoc information purposes and may contain basic, uninterpreted data. The contents of this document have not been subjected to review and should not be cited or distributed without the permission of the authors or the Commercial Fisheries Division

## Introduction

The Alaska Hatchery Research Program (AHRP) has sampled sufficient years to allow the first set of adult-to-adult parentage analyses for the fitness component of the research plan. With the end of the 2015 field season we now have collections from a complete generation of odd year pink salmon samples for five streams in PWS. The original plan called for sampling 500 adult fish per year per stream. However, subsequent power analyses indicated that larger sample sizes are required to afford reasonable chances of detecting effects, if they exist. As a result, in 2014 the science panel asked the contractor to increase sampling efforts in the pedigree streams. Increased sampling by the contractor combined with unusually high-abundances in 2013 and 2015 resulted in much larger number of samples (by an order of magnitude) for potential analyses. This number of samples outstrips available funding for analyses of otoliths and genetic samples.

## Goals of Technical Document

The purpose of this document is to propose a priority order for genetic analysis of samples, and communicate the factors considered in this prioritization. We request science panel input regarding component prioritization. Our priority list accounted for anticipated power to investigate relative reproductive success (RRS), observed stray rates from 2013, sample sizes, and stream size (i.e., depth of sampling) to determine which streams will likely provide the most valuable information. The tables below detail rough estimates of sample sizes and laboratory costs for each project component.

## Proposed analysis prioritization and plan

*Samples available through September 2015*

During the 2013 sampling season approximately 4,000 samples of adult pink salmon were collected from PWS streams identified for the pedigree analysis. Given the large return that year (leading to low proportions of potential parents being sampled) and results of simulations, it was decided that more intensive sampling was necessary. Subsequently, in 2014 approximately 8,000 samples were collected, representing larger proportions of potential parents given the estimated number of pink salmon in each stream. In 2015, the combination of another exceptional return of pink salmon to PWS streams and the increased sampling effort on pedigree streams resulted in the collection of ~59,000 samples from the pedigree streams alone (~52,000 excluding Gilmour Creek, which was not sampled in 2013).

*ADF&G Cordova otolith lab capacity*

The ADF&G Cordova otolith lab estimates they will be able to read a maximum of 35,000 otoliths from 2015 samples given the current level of funding and staffing available. Assuming the ocean test fishery, stock structure, and stream straying samples take precedence, there will be capacity to read between 5,000 and 10,000 otoliths for the pedigree analysis (Tables 1 &2). An

additional 1,500 pedigree samples may be made available by reducing the number of collections included in the stock structure analysis.

Table 1. Timeline and estimated sample sizes for otolith reading by the ADF&G Cordova laboratory. Samples are organized by project for pink and chum salmon from Prince William Sound.

| Priority | Project | Sample Estimate | Cumulative Estimate | Expected Completion Date |
|---|---|---|---|---|
| 1 | Ocean test fishery | 3,564 | 3,564 | Oct 13, 2015 |
| 2 | Straying streams | ~18,200 | 21,764 | Jan 12, 2016 |
| 3 | Pink salmon stock structure | ~2,500 | 24,264 | Jan 25, 2016 |
| 4 | Pink salmon pedigree | ~59,000 | 83,264 | Feb 29, 2016 [a] |
| | Total samples | ~83,300 | | |

[a] Available funds ($96,700) will be depleted.

Table 2. Cordova otolith laboratory budget outline based on anticipated otolith personnel read rates and lab supplies for 35,000 otoliths.

*Personnel wages and benefits*

| Position | No. Positions | Monthly cost | Months each | Total Cost |
|---|---|---|---|---|
| Fish/Wildlife Technician II | 3 | $ 5,911 | 5 | $ 88,662 |

*Commodities*

| | |
|---|---|
| Lab supplies (sand paper, glue, slides, and labels) | $ 12,632 |
| All lines total | $ 101,294 |
| Total budget allocation | $96,700 |
| Remaining balance | ($4,594) |

*Gene Conservation Lab capacity*

The GCL timeline is dictated by the reading of matched otoliths, completion of the sequencing contract to discover new SNPs, and capacity in the lab for analysis with current technology. Recent power analyses were performed to explore the implications of the many single-parent families in the parentage analysis. The results indicate that twice as many markers (192 instead of 96) will be required to achieve the precision and accuracies necessary. While this does not

double the laboratory cost, it does increase the per-fish cost from $25 to $32. Under the current project plan, genotyping of all 2013 potential parents and 2015 potential offspring (~47K individuals not including expected number of hatchery strays in 2015) would not be complete until spring of 2017 and would cost ~$1,500,000.

Screening for this many markers can be most cost-effectively accomplished with recently developed technology. However, purchasing, installing, and implementing this new technology will add an uncertain amount of time (see "Mitigating circumstances" below). Therefore, the timeline and cost estimates provided below are based on using the current technology.

*Proposed priorities*

Given throughput and funding limitations, we recommend prioritizing sample analysis by focusing on depth of analysis at the expense of breadth of analysis (Table 3). We think that focusing on one or two streams and maximizing statistical power is the best approach to successfully accomplish some of the program objectives and will provide information for subsequent decisions. We also recommend genetic analysis of only natural-origin fish for the $F_1$ (offspring) collection(s) in 2015. Reductions in cost and necessary lab capacity achieved by excluding hatchery-origin $F_1$'s will save funds in the current context, however, the hatchery-origin $F_1$'s will need to be genotyped at a later date if fitness analyses continue for a second generation (2017 return). Findings from this work will provide the most solid initial evidence to evaluate the null hypothesis (that hatchery-origin fish spawning in the wild do not impact the fitness of wild fish) for one or two creeks in a single generation. Increasing breadth at the expense of depth is more likely to yield equivocal results.

If this approach is taken, and analyses are limited to one stream, we recommend analyzing samples from Stockdale Creek. Stockdale Creek offers the best combination of 1) adequate sampling of 2013 parents, 2) intermediate stray rate (10.2% in 2013), and 3) an intermediate population size, resulting in a good depth of sampling coverage that will likely provide the most statistical power of all the streams.

Stockdale Creek is also the only pink salmon stream in PWS where alevin were sampled so we will already have all the 2013 parents genotyped if parentage analysis of alevin becomes a priority in the future. By starting with Stockdale Creek, we will be able to fine-tune our genotyping capabilities with the pink salmon SNPs under development and see what information is provided by parentage analysis from the stream with the highest power to detect a difference in reproductive success. These results can then be used to inform future analysis decisions based on the utility of the data for a given level of funding and staffing.

Table 3. Approximate sample sizes available and proposed priority for the six streams in the pedigree analysis for the odd-year run of pink salmon in Prince William Sound. Sample sizes include the parents from 2013 and the potential offspring from 2015.

| GCL Priority | Project Component | Samples available | | Laboratory Genotyping Cost | 2013 Stray rate | Likely Statistical Power | Rationale |
|---|---|---|---|---|---|---|---|
| | | Otolith | Genotype | | | | |
| 1 | Stockdale Creek Adult | 8,602 | ~9,000 | $288,000 | 10.2% | High | Intermediate stray rate and high power |
| 2 | Hogan Bay Adult | 9,441 | ~5,000 | $160,000 | 56.4% | High | High stray rate and high power |
| 3 | Erb Creek Adult | 13,039 | ~12,000 | $384,000 | 10.8% | Medium | Intermediate stray rate and medium power |
| 4 | Spring Creek Adult | 12,469 | ~13,500 | $432,000 | 1.5% | Low | Low stray rate but low power |
| 5 | Stockdale Creek 2014 Alevin | - | 2,728 | $87,300 | 10.2% | Likely Low | Only alevin stream |
| 6 | Paddy Creek Adult | 8,710 | ~7,500 | $240,000 | 15.3% | Very Low | Intermediate stray rate and very low power |

Note: These numbers assume genotyping all 2013 adults regardless of origin (potential parents), but only natural-origin adults for 2015 (potential offspring). Numbers of natural-origin adults for 2015 were estimated assuming the same stream-specific stray rates as 2013. Laboratory genotyping costs with GCL's current genotyping technology are estimated at $32/fish.

If funding is available for additional sample analyses, we recommend adding Hogan Bay. This addition will increase breadth by including another location and a different (higher) stray rate. This set of samples is also the only other stream which is likely to yield high statistical power based on the sample sizes and escapement sizes.

*Proposed timeline (Stockdale only)*

Below is a brief timeline for the analysis of the Stockdale Creek pink salmon samples.

| Component | Start date | End date |
|---|---|---|
| Receive all samples from PWSSC | September 2015 | October 2015 |
| Separate heart from otoliths for Stockdale samples | October 2015 | November 2015 |
| New SNP markers available | | February 2016 |
| Read otoliths from 2015 Stockdale samples | November 2015 | March 2016 |
| Genotype 2013 & 2015 Stockdale samples | April 2016 | May 2016 |
| Parentage analysis on Stockdale samples | May 2016 | June 2016 |
| Report results of parentage analysis and RRS | | July 2016 |

## Mitigating circumstances

The GCL's current Fluidigm® genotyping platform costs ~$32/fish for the anticipated 192 SNP markers that will be analyzed for parentage analysis. While the Fluidigm® platform is highly cost effective for genotyping fish for 24, 48 or 96 SNPs, other next-generation sequencing technologies offer reduced costs and increased efficiency for genotyping 192 SNPs (Campbell et al. 2014). The GCL is currently exploring these recently developed technologies to bring down the laboratory genotyping costs for this project from ~$32/fish to ~$24/fish, resulting in a 25% savings. Given the number of samples expected to be genotyped for this program, the savings will be large.

## Questions for the Science Panel

1. Is the proposed prioritized approach the best method to provide necessary information within the limitations of time and funding?
2. If not, is there information or a consideration that we have not considered?
3. Is Stockdale Creek the appropriate stream to analyze first?
4. Should the analysis be extended to include Hogan Creek or some other creek?

## Science Panel Review and Comments

*This technical document has had partial review – see comments below:*

*The recommendation is to process approximately half the collected otoliths at Hogan Bay and Stockdale Creeks and see if that number is adequate for analysis of fitness. The results of that work will guide decisions in regard to other samples collected.*

John H Clark 10/4/2015

*Thank you for the opportunity to review the document, it is both informative and helpful.*

*In reading the document, I am uncertain why you have chosen Stockdale Creek as the top priority for parental analysis. While it has the potential to assist in determining if alevin analysis might be useful, it has a lower stray rate in 2013 (10% vs 56%) and higher sample sizes (9,000 vs 5,000) and thus cost than is the case for Hogan Bay. It strikes me that if there are differences in fitness, we may be more likely to see such in the Hogan Bay analysis. Could you provide me with the rationale used to prioritize Stockdale over Hogan Bay?*

*I second question: In terms of the equipment question, how many samples would have to be run before the savings per fish would/could account for the cost of the new equipment? You have indicated a potential cost savings of $8/ per fish; not knowing the cost of the equipment, it is hard to consider whether or not it is worth investing in the equipment that might result in the cost saving. For instance, if we went with just Stockdale and Hogan, genetic cost savings would be about 14,000 x $8 or $112,000. If the equipment is less than that, then serious consideration might be given to purchase it.*

Chris Habicht 10/6/2014

*Thanks for your timely response on this important subject.*

*To address your first question, we prioritized Stockdale Creek over Hogan Bay for 2013 largely due to escapement size and stray rate.*

- *Escapement size: In 2013, Hogan Bay had an aerial survey estimated escapement of ~47K vs. Stockdale Creek's estimate of ~4K. While both streams have similar power for a given F1 sampling proportion, if Hogan Bay had a similarly high escapement compared to Stockdale Creek in 2015 (don't have aerial survey estimates yet), then it will take a lot more F1 fish to reach a similar sampling proportion and thus similar level of power (see Y-axis in attached). In addition, the X-axis of the power curves is the reproductive success of the natural-origin fish. To move to the right, the higher the natural-origin returns need to be in 2015 relative to 2013. If Hogan Creek had a large return in 2013, and it continues to have a high stray rate in 2015, it will be harder to attain high reproductive success of natural fish. The lower escapement and lower stray rates of Stockdale Creek in 2013 make it more likely that the reproductive success of natural fish returning in 2015 will be higher. Finally, the higher escapement to Hogan Bay in 2013 and thus lower sampling proportion of F0's will result in a smaller proportion of F1's that had their parents sampled. This will result in an even smaller proportion of F2's (2017 return) that had grandparents sampled in 2013.*

- *Stray rate: The ~10% stray rate for Stockdale in 2013 is more "representative" of pink salmon streams in PWS than Hogan Bay (higher escapement and a stray rate in 2013 of ~56%).*

*Additionally, if 2013 is representative of stray rates for the odd-year broodline for these two streams, the >50% stray rate of Hogan Bay is more likely to have eroded more potential adaptation to wild conditions over the past 15+ generations of hatchery influence. If there are adaptive genetic differences between natural- and hatchery-origin pink salmon in PWS that could lead to differential RRS, then Stockdale Creek may provide more contrast.*

*To address your second question, we are looking at the purchase of new equipment but need to make sure that this purchase makes both financial sense and does not delay data acquisition. As you point out, if we analyze both Stockdale and Hogan, we might realize $112K in savings for this portion of the project alone. Additional savings would be achieved with the even-year pink salmon analysis, analysis of samples from the other four pedigree streams in PWS and the chum salmon analyses in SEAK. In addition, this equipment might save funds for other genetics projects, so the department is considering potentially purchasing part or all of the equipment with other funding. However, cost savings are based on our best understanding of the technology and implementing this technology adds uncertainty to both the timeline and cost. We are continuing to assess whether and when to switch over.*

*Moving forward, the escapement numbers from this year will help determine where we stand on the power curves (attached). It makes sense to me for the science panel to have these numbers to incorporate into their decision. Is there other information/outcomes that the science panel should be consider before making a decision?*

Chris Habicht 10/7/2015 in response to JHC on costs

*This question is more difficult to answer than one might think. We will not pay list prices on these instruments. We have been meeting with sales/technical folks who represent competing technologies (we were in a meeting with one group yesterday afternoon) to figure out the price structure of the capital and the operating costs. These companies make much of their income on the consumables, so they may offer heavy discounts on the hardware so that we purchase their consumables. We will be putting together a request for bids soon and we will need to include both capital and operating costs into the bid selection. Tyler's PhD training will really come in handy in writing and evaluating these bids.*

*So, to give you a ball-park capital cost for the equipment, we are looking at somewhere between $150K and $300K.*

Alex Wertheimer 10/7/2015

*Thanks for providing Tech Document 11. The problem of too many samples and not enough funding definitely requires prioritization, and I appreciate the very clear and explicit description of how the gene lab thinks this should be done. It was always the intent of the project to examine the fitness question for both a high (50%) and low-intermediate stray rate. The original plan for PWS pink salmon was of course to have three streams in each category, which would ideally*

*provide some replication of fitness estimates for both high and low stray streams. Reality happens, and the escapement numbers, sampling rates, and processing costs have eliminated the ideal, but because of the ambitious sampling strategy, there are promising sample sets from at least one of each stream type. I understand your argument that low stray rate stream (Stockdale) is more representative of "average" stray rates. However, it will be more difficult to assess the reproductive success of hatchery-origin parents because of their lower incidence in the population. Also, your argument that it will be more difficult to find differences in a 50% stray rate system because of homogenization with hatchery and wild fish assumes that reproductive success of hatchery parents is close enough to "blend" out any differences. If reproductive success is very poor, then differences should persist. Lack of a difference in reproductive success at high straying (we accept the null hypothesis) would still provide insight into the magnitude of the introgression "problem."*

*I have no problem with Stockdale being analyzed first, but I think we need to figure out how to get the Hogan samples processed as well. If equipment efficiencies can make this feasible, great! If not, then I would prioritize the Hogan samples over 2017 sampling and sample processing from the pedigree streams.*

John Burke 10/7/2015

*Just a short comment…pretty much agree with what Alex said below, though it would serve our larger argument better to look at Stockdale since that is a more usual situation than one where half the spawners were from an enhancement program. One thing, we have always thought that even if there was some measure of loss in fitness in the F1's…that could disappear in the F2's. This has been part of the argument from the beginning. Dropping the meaningful pedigree sampling that would include these second generational outcomes, we would be missing something that could prove very important in this assessment. Of course, if no loss of fitness is found in the F1's, it may not make any sense to continue. The issue, we probably not have results in time to make that decision. The other issue that is important to us is that we do not ignore the work in SE as this is not only about pink salmon in PWS. The outcome could be different in chums whose progeny spread themselves across three return years and there is a much general greater mix of fish from different sources in any brood year.*

*Not having Bill Gates support the project, is a real issue. The "budget committee" is now functioning. We will shortly have a better understanding of the funding, at least what is in hand. We will also have a better understanding of how much will be required to go forward at different levels of intensity. Some things are obviously going to have to be sorted out and prioritized.*

*For now, I don't see a problem going forward with Stockdale, but before much else happens it is important that we sit down together and take a hard look at what funding we have and what is possible.*

Ron Josephson 11/6/2015

*Today in discussions with Eric Knudsen I mentioned the concept of sub-sampling results from one of the creeks. Eric had some good thoughts.*

*One was that sub-sampling might be best done on a day basis; e.g. include all the fish on every 4th day or someother increment. I would add it could also be by tray. But Eric and I both agree that if it were by fish the chances of matching data would be challenging. (Mostly due to otoltihs and how they process them).*

*The other thought Eric had was that we could calculate area under the curve estimates based on the PWSSC foot surveys. These are likely more frequent and consistent than aerial surveys. With an estimate of total escapement and the known number of sampled fish we would know what proportion of the escapement was sampled.*

*The Cordova staff and Xinxian have routinely done AOC estimates for PWS and might be able to come up with estimates pretty easily.*

*I also want to mention a tentative date Of March 2nd and 3rd for a meeting with the contractor in Anchorage. This will be just prior to a PWSAC board meeting and some of you will be at that meeting. It also would provide opportunity for some of PWSAC board to attend our meeting.*

*The final point is that Eric also asked if any work had been done on the 2015 aelvin sampling at Fish Creek on Douglas. He was interested, as am I. (We were talking about this in the context of a measure of their efficiency at sampling adults).*

Dave Bernard 11/9/2015

*A couple of thoughts on estimating the sampled portion of the escapement in a stream by using area-under-the curve (AUC) methods to estimate abundance. An accurate estimate from the AUC method requires knowledge of how long spawners remain to be counted (stream life) and the accuracy of the counts (observer efficiency). Without knowledge of these two variables, the AUC expansion will be a biased. The result can still be used to produce relative weights for use in studies like ours, but will probably underestimate true abundance in the stream and therefore overestimate the portion of that abundance sampled.*

*I'm not saying that the foot surveys can't be used to accurately estimate abundance, only that accuracy will depend on more than just counts from foot surveys.*

Chris Habicht 11/10/2015

*I just wanted to answer Eric Knudsen's question regarding the progress on the 2015 chum salmon alevin from Fish Creek on Douglas.*

*These fish are available for analysis, but we need to settle on a marker suite and then determine prioritization for screening in the lab.*

- *Marker suite selection: We have genotyped all ~1K 2013 Fish Creek adults and 567 Fish Creek alevin for the 188 SNPs used in the Western Alaska Salmon Stock Identification Program. These samples are currently being genotyped for the new 96 Western Alaska Salmon Coalition SNPs. We anticipate data by mid-December. Once we have the data, we will be able to 1) provide final parentage analysis results, and 2) make our marker selection for parentage analysis*

- *Prioritization for laboratory screening: We have 2,626 Fish Creek adults from 2014 and 1,985 Fish Creek alevin from 2015. Total project sample size = 4,611 fish. This project is not on the lab schedule and the science panel will need to determine the prioritization for these samples. We plan to put together a Technical Document that lays out the costs and products that would result from these analyses. This document should help with determining prioritization of these analyses.*

*We just finished a proposal to Saltonstall-Kennedy and we are currently putting together a proposal for North Pacific Research Board to fund aspects of this program, so we do not anticipate completing this TD until December.*

Ron Josephson 12/8/2015 – Today, Alex Wertheimer, John Clark, Bill Templin, Sam Rabung and I met to discuss this issue.

*The group decided that prioritizing the Stockdale and Hogan Bay samples makes the most sense at this point in time. Given the number of samples from each system, it is reasonable to only process half the samples for otolith reading this winter. The best way to do that is to process every other tray, based on sampling date and chronological numbering of the trays.*

*Bill explained that SNPs may be run this spring but more likely not till this fall. The prioritization of these systems fits well with the expected timelines and budgets for SNP analysis.*

*On a sidebar issue, the group also decided that the otoliths that have been read for Hartney (304), Fish (360), Coghill (234), Cabin (297), and Constantine (322), was adequate for estimation of hatchery origin proportions. With the exception of Cabin Creek (11%) the hatchery proportions were 2% or less for current processed otoliths.*

## References

Campbell, N.R., Harmon, S.A., and Narum, S.R. 2014. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources: 13.