

The feasibility of Grandparentage testing to monitor genetic risks posed by straying hatchery steelhead in the Snake River Basin



2025 Lower Snake River Compensation Plan Steelhead Symposium
January 21-23, 2025
Natural Resources Center
1387 S Vinnell Way, Boise, ID 83709



Matthew Campbell (IDFG)
Fisheries Genetics Program Coordinator



Primary EFGL Staff:
Katharine Coykendall; Audrey Harris; John Hargrove; Jesse McCane



Primary Collaborators:
Tom Delomas; USDA ARS National Cold Water Marine Aquaculture Center
Shawn Narum; Rebekah Horn; Columbia River Inter-Tribal Fish Commission

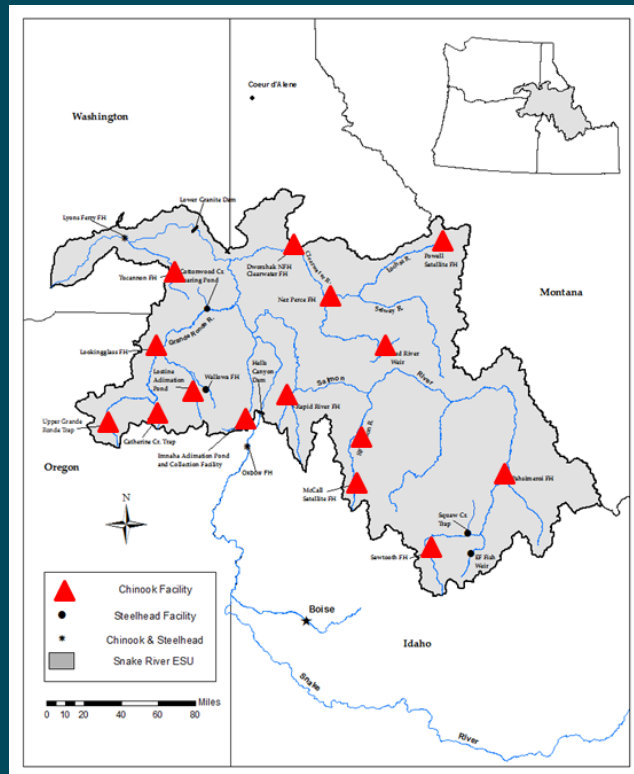


Funding:
Pacific Coast Salmon Recovery Fund

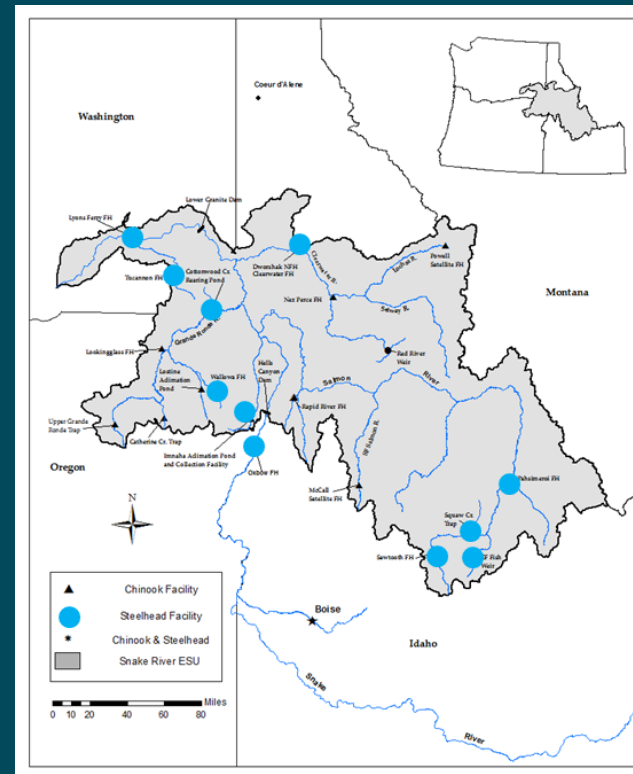
• Parentage Based Tagging (PBT)

Snake River Chinook Salmon and steelhead

- All Spring/Summer Chinook salmon broodstock sampled since **2008** (~10,000 samples annually)
- All Fall Chinook salmon (Lyons Ferry/NPT) since **2011** (~5,000 samples annually)
- Most Snake River steelhead broodstock sampled in **2008**, all since 2009 (~5,000 samples annually)



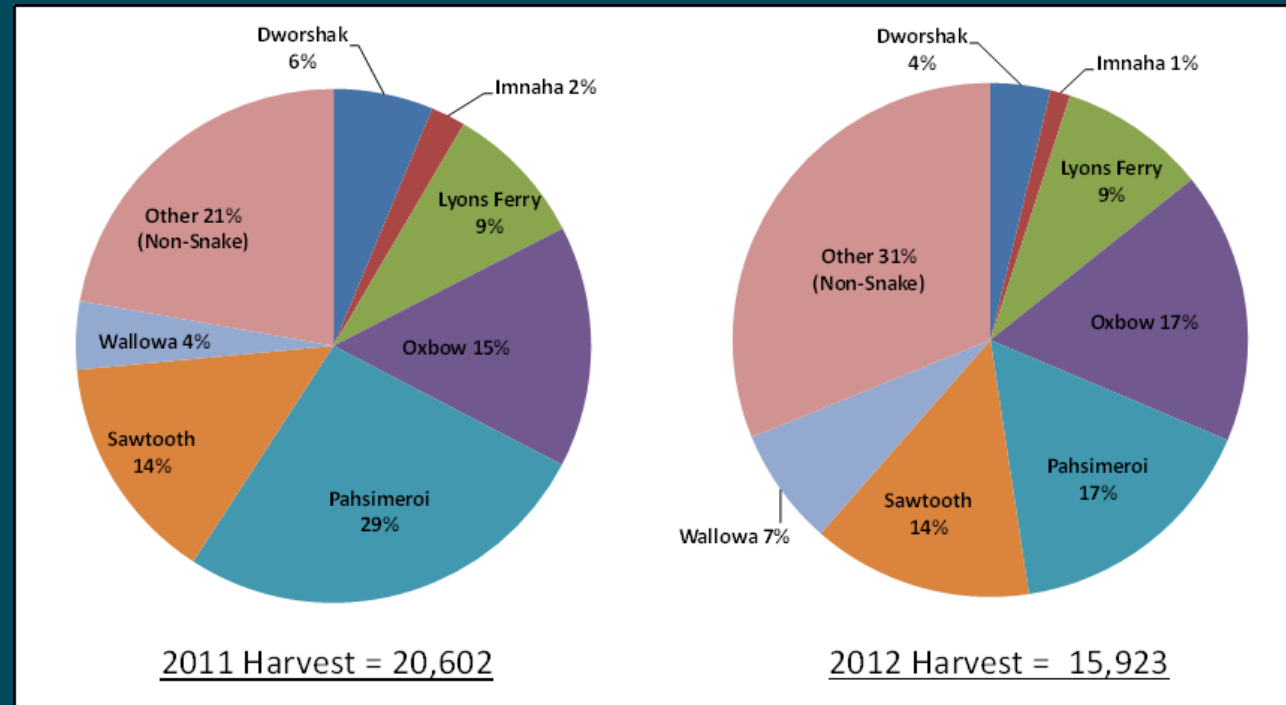
Chinook salmon



steelhead

• Benefits of PBT technology

- Better estimation of wild escapement at Lower Granite Dam (8.3% of “wild” adults are actually hatchery; Hargove et al 2021)
- Better estimation of hatchery escapement at Lower Granite Dam (SAR survival rates were 1.3–2x higher when calculated using PBT compared to PIT tags (Coykendall et al. 2022))
- Better estimation of out-of-state harvest
- Better estimation of in-state harvest



A lot of Snake River steelhead are caught in down-river fisheries!!!!

• Benefits of PBT technology

- PBT also can be used to identify the origin of straying hatchery fish



Transactions of the American Fisheries Society 145:671–686, 2016
© American Fisheries Society 2016
ISSN: 0002-8487 print / 1548-8659 online
DOI: 10.1080/00028487.2016.1146163

ARTICLE

Maximum Likelihood Estimation of the Proportion of Hatchery-Origin Fish on Spawning Grounds Using Coded Wire Tagging and Parentage-Based Tagging

Richard A. Hinrichsen*

Hinrichsen Environmental Services, 9034 45th Avenue Northeast, Seattle, Washington 98115, USA

Craig A. Steele and Michael W. Ackerman

Pacific States Marine Fisheries Commission, Eagle Fish Genetics Laboratory, 1800 Trout Road, Eagle, Idaho 83616, USA

Matthew R. Campbell

Idaho Department of Fish and Game, Eagle Fish Genetics Laboratory, 1800 Trout Road, Eagle, Idaho 83616, USA

Shawn R. Narum and Maureen A. Hess¹

Columbia River Inter-Tribal Fish Commission, Hagerman Fish Culture Experiment Station, 3059-F National Fish Hatchery Road, Hagerman, Idaho 83332, USA

William P. Young

Nez Perce Tribe, Department of Fisheries Resources Management, 14054 Burr Road, McCall, Idaho 83638, USA

Barbara A. Shields

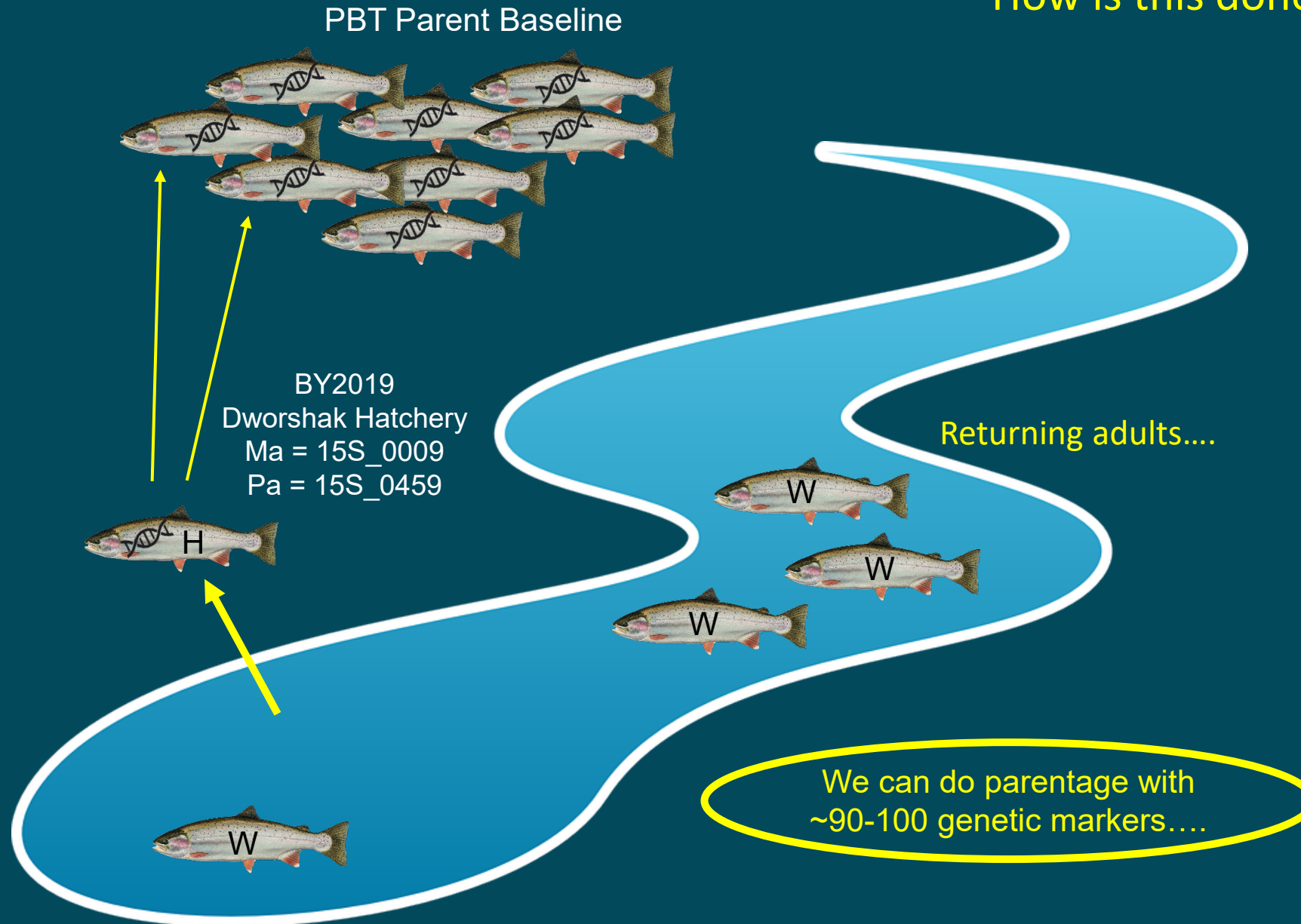
Bonneville Power Administration, 905 Northeast 11th Avenue, Portland, Oregon 97232, USA

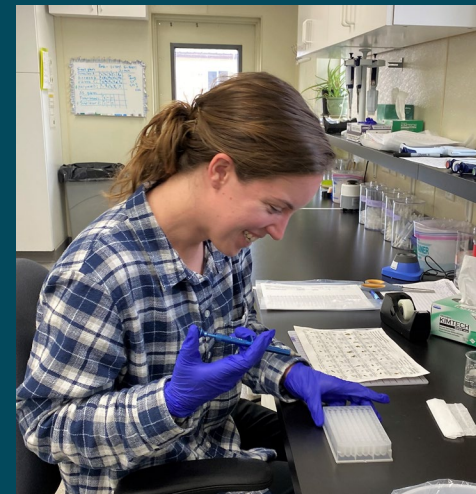
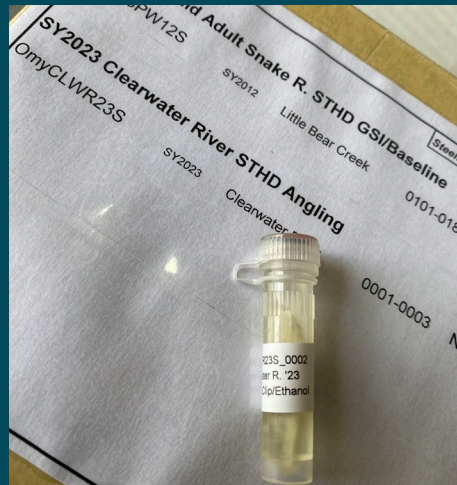
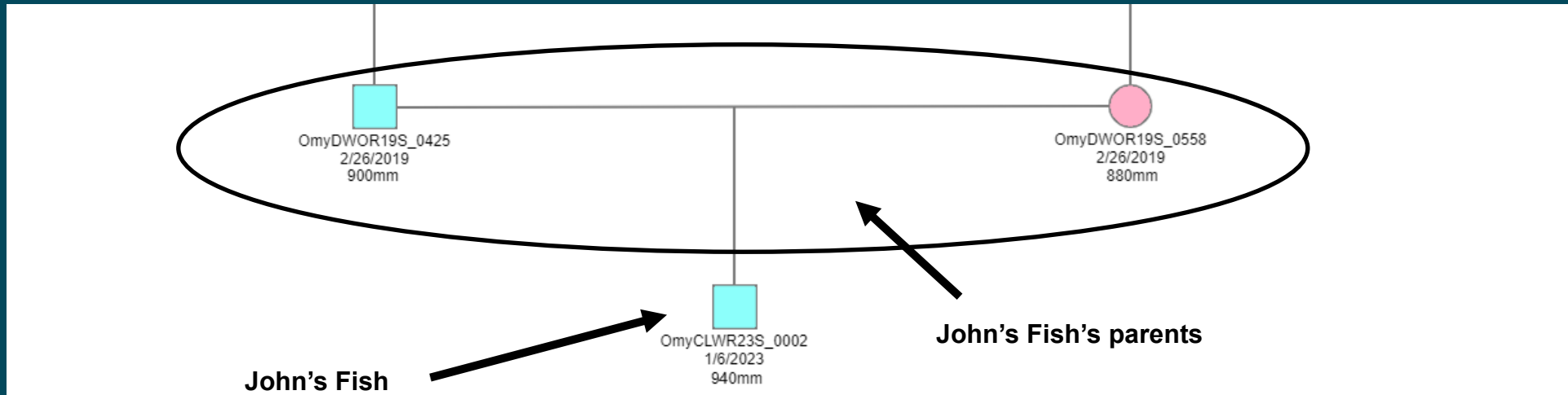
Brian L. Maschhoff

Salmonetics, 7813 1st Avenue Northwest, Seattle, Washington 98117, USA

Hinrichsen et al (2016)-"In the South Fork Salmon River application, there were 340% more PBT recoveries than CWT recoveries, leading to greater precision in release-specific values of p from maximum likelihood estimation."

How is this done???

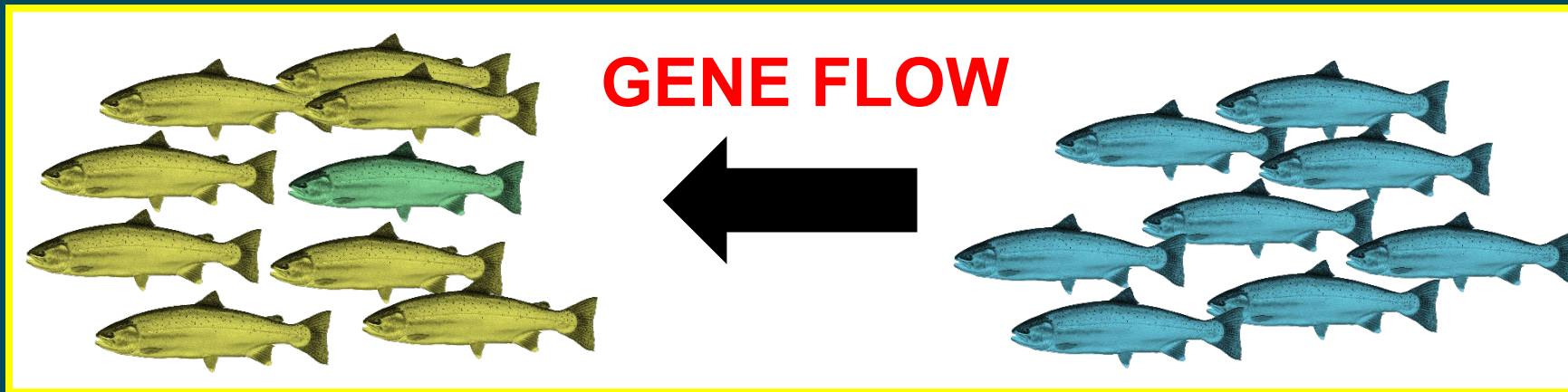




We can do parentage with
~90-100 genetic markers....

- **NOAA wants this information**

- Status assessments for ESA-listed salmonid populations in the Snake River and Columbia River basins, require reliable estimates of the proportion of hatchery-origin spawners on the spawning grounds, or pHOS (McClure et al. 2003)
- pHOS is an important measure, but another thing that geneticists and managers would like to monitor:

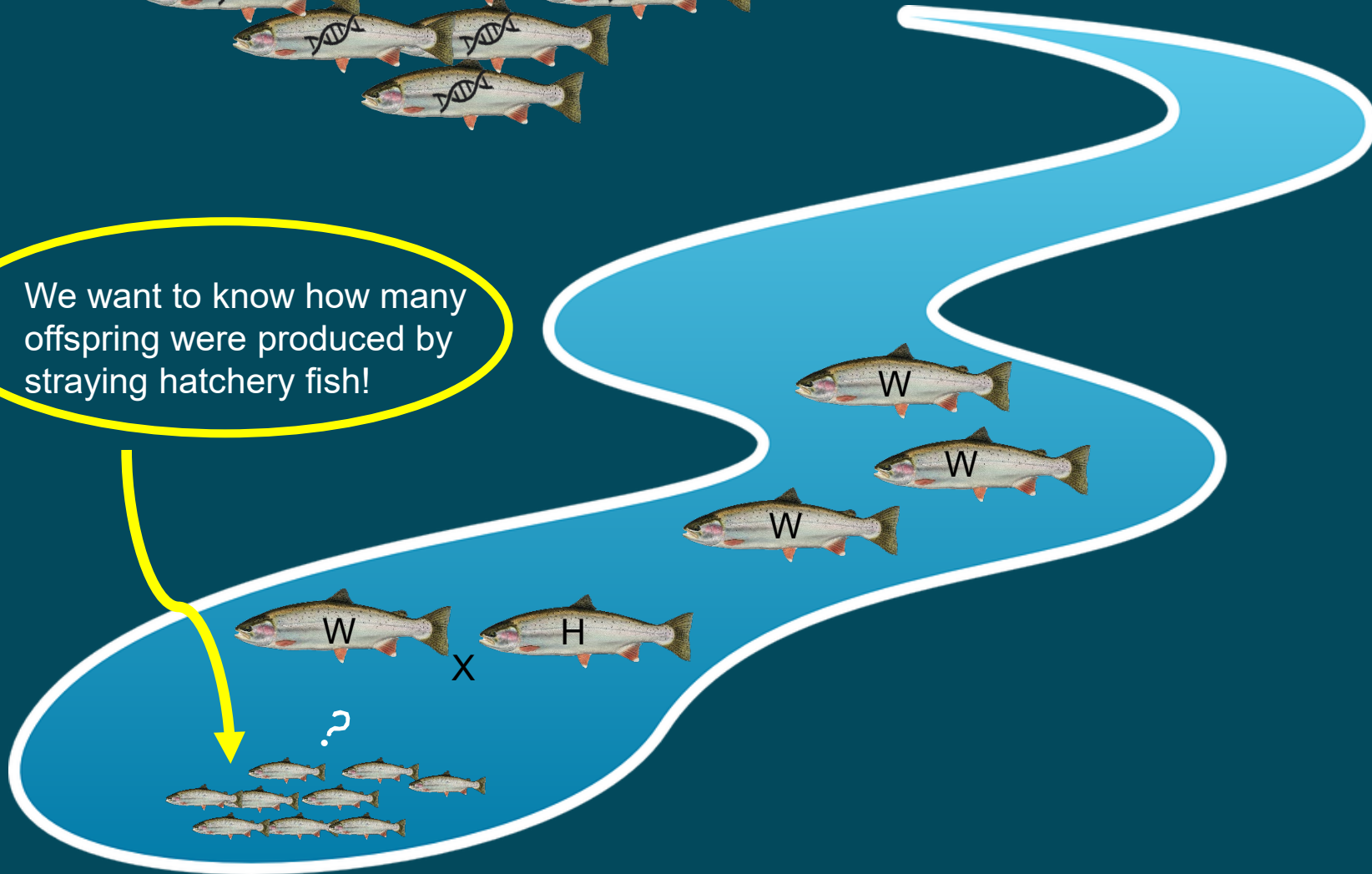


- Gene flow only occurs if hatchery fish successfully mate with wild fish and produce offspring!

PBT Parent Baseline



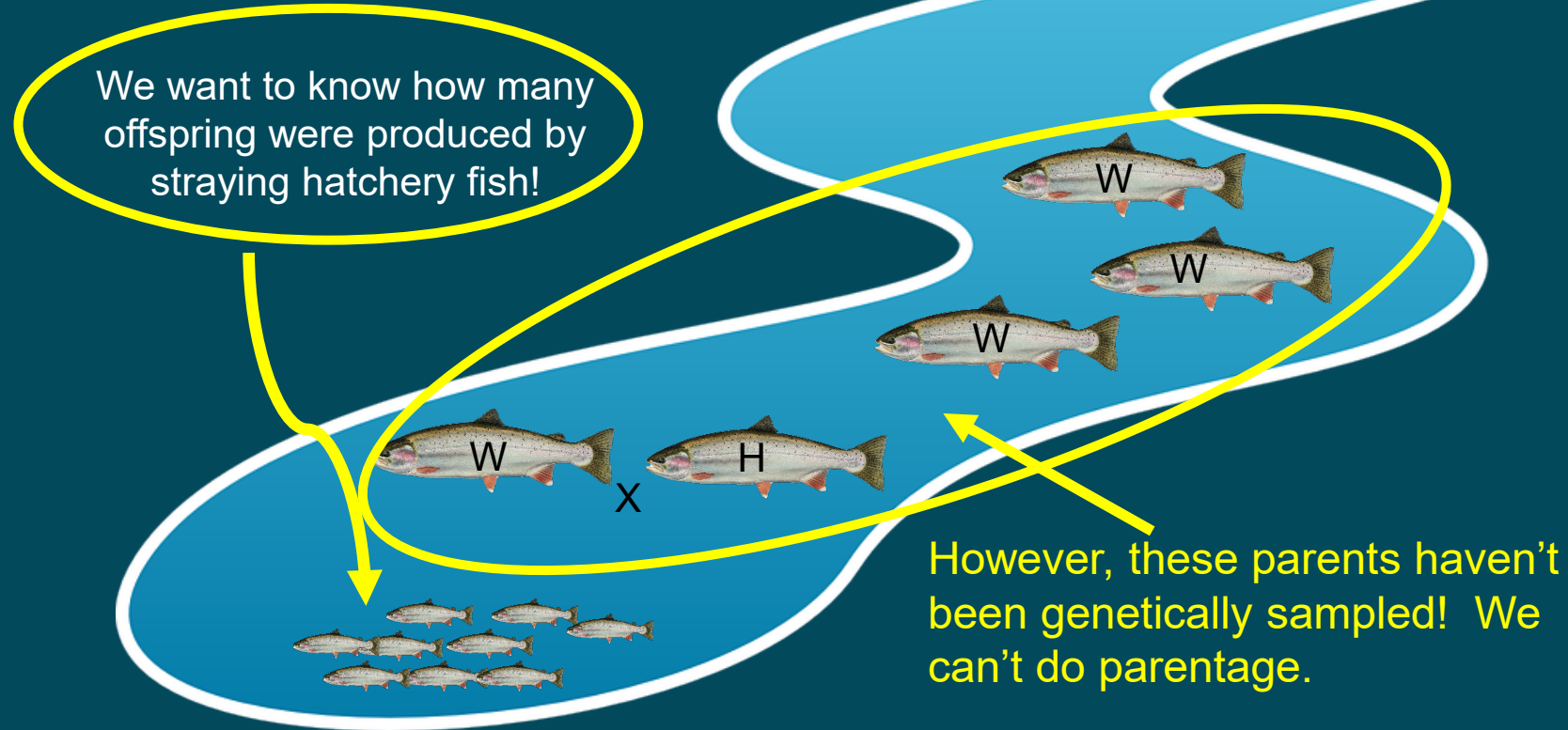
We want to know how many offspring were produced by straying hatchery fish!



PBT Parent Baseline

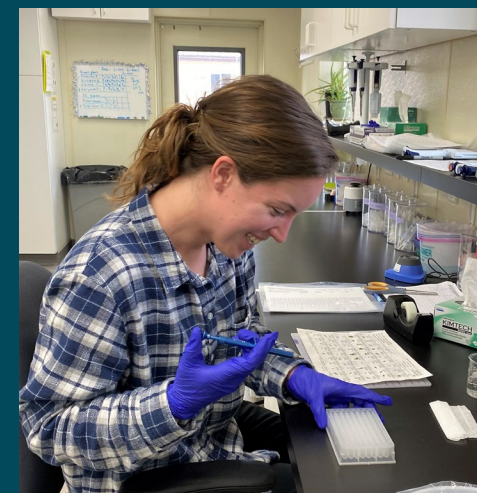
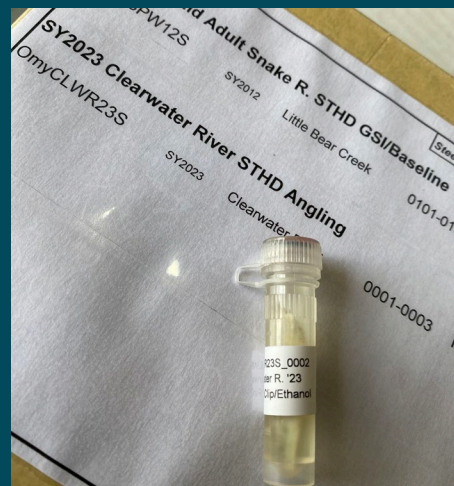
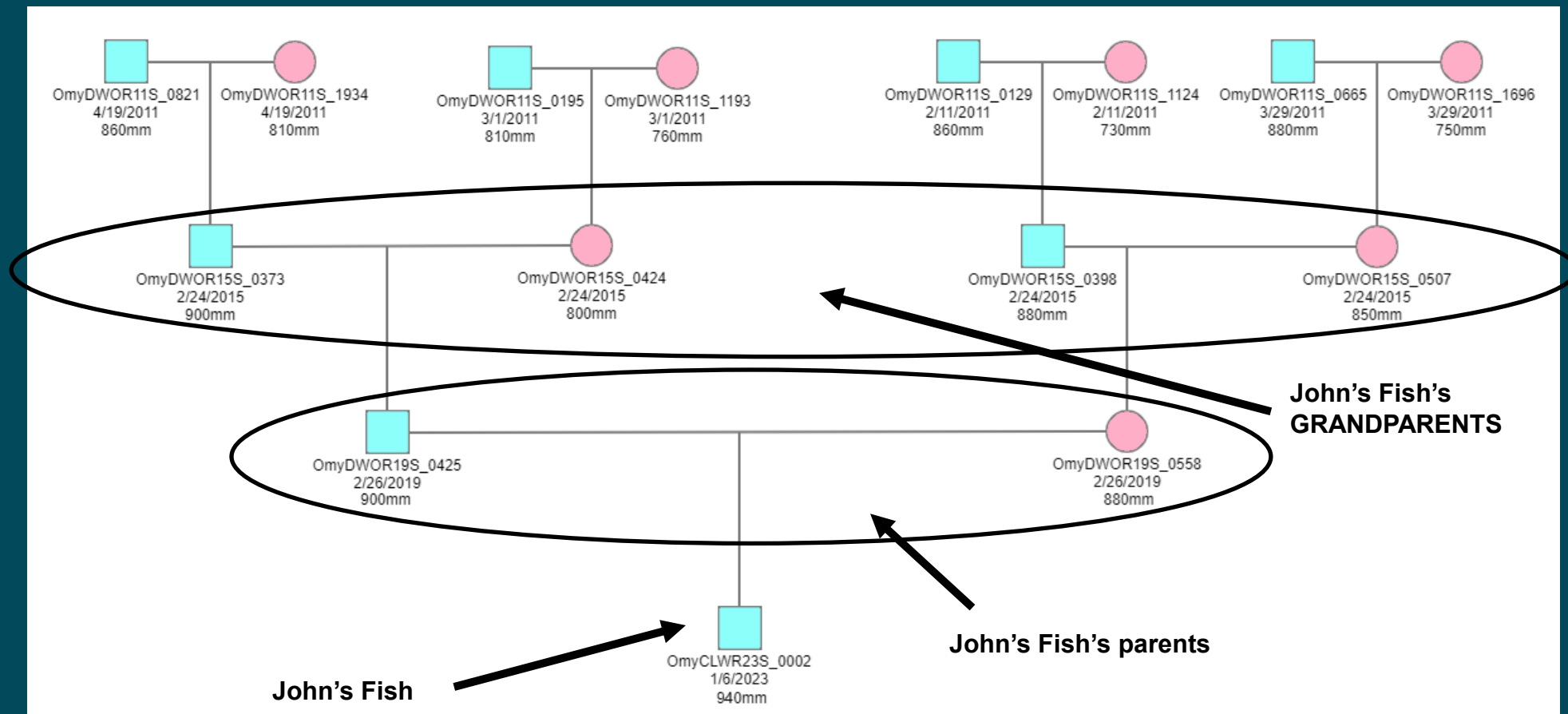


We want to know how many offspring were produced by straying hatchery fish!



However, these parents haven't been genetically sampled! We can't do parentage.

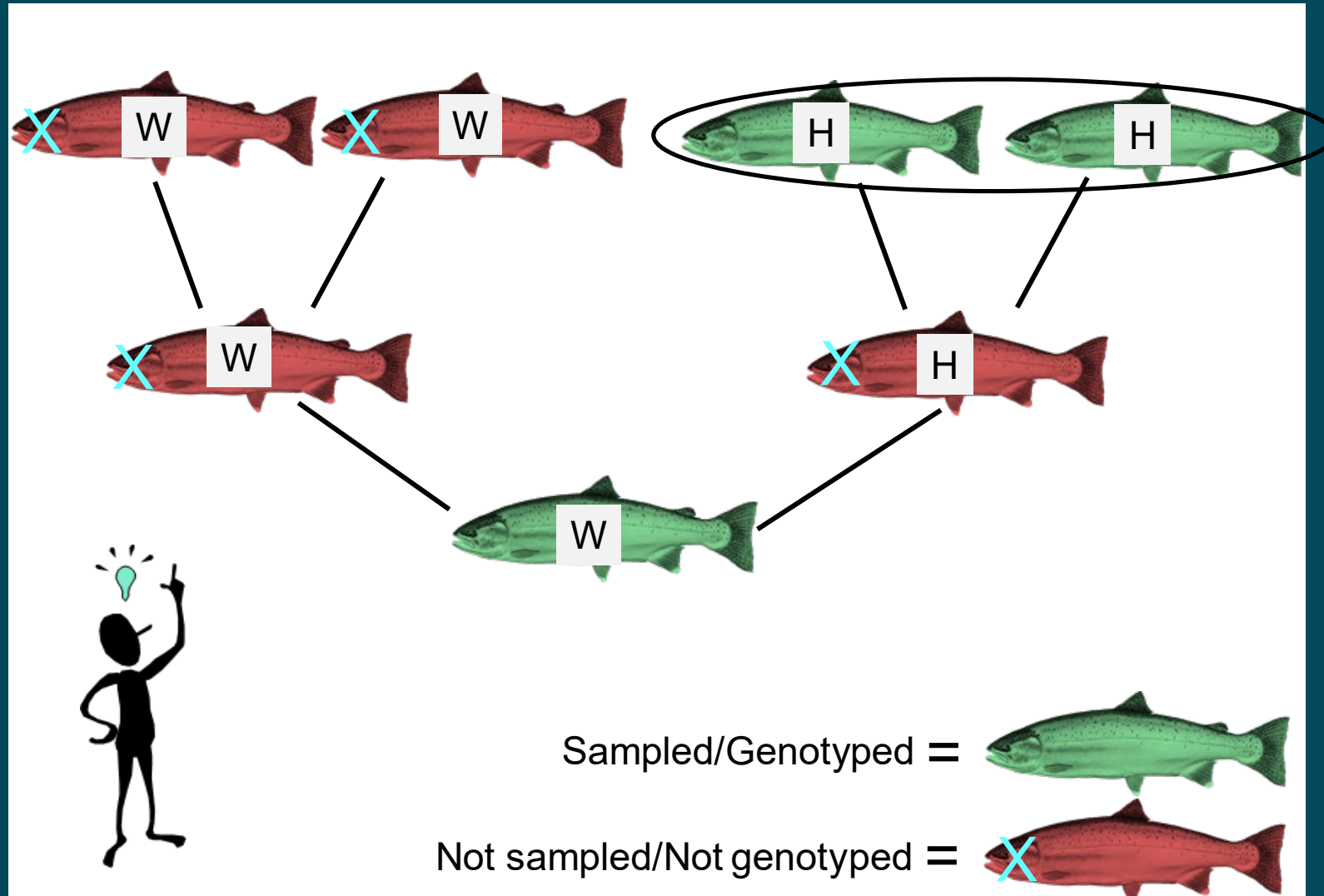




Grandparentage Testing



- **What are we proposing?**
 - With sufficient genetic markers can extend PBT to identify grandparent-grandchild relationships?

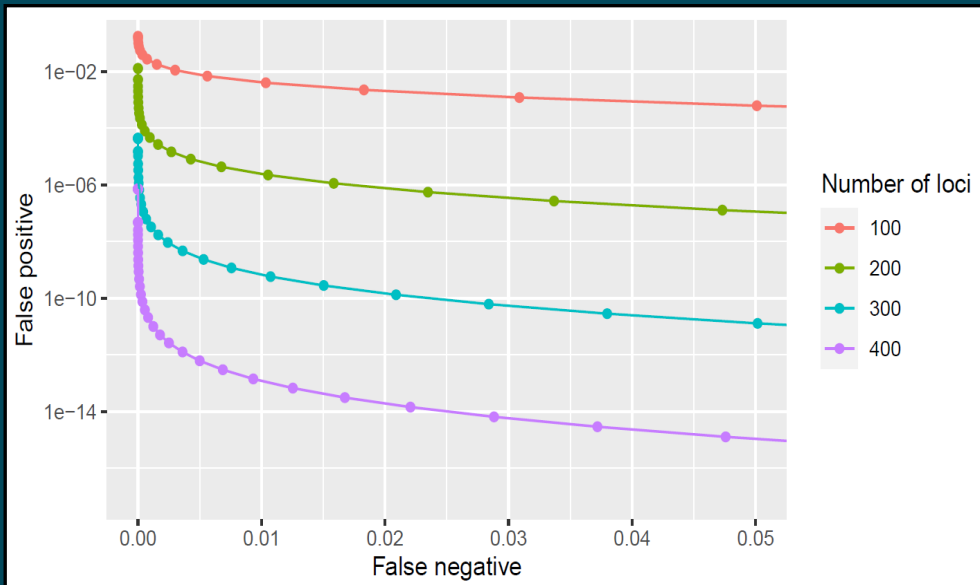


- **History**

- ✓ We developed statistical methods for assigning grandparents and estimating error rates for a genetic panel
- ✓ We have implemented these methods in a package for R statistical software at <https://github.com/delomast/gRandma>
- ✓ Preliminary analysis shows that 300 – 500 genetic markers are sufficient to accurately assign grandparents basin-wide
- ✓ We proposed an empirical validation and demonstration of this new technique



Tom Delomas



North American Journal of
Fisheries Management

ARTICLE

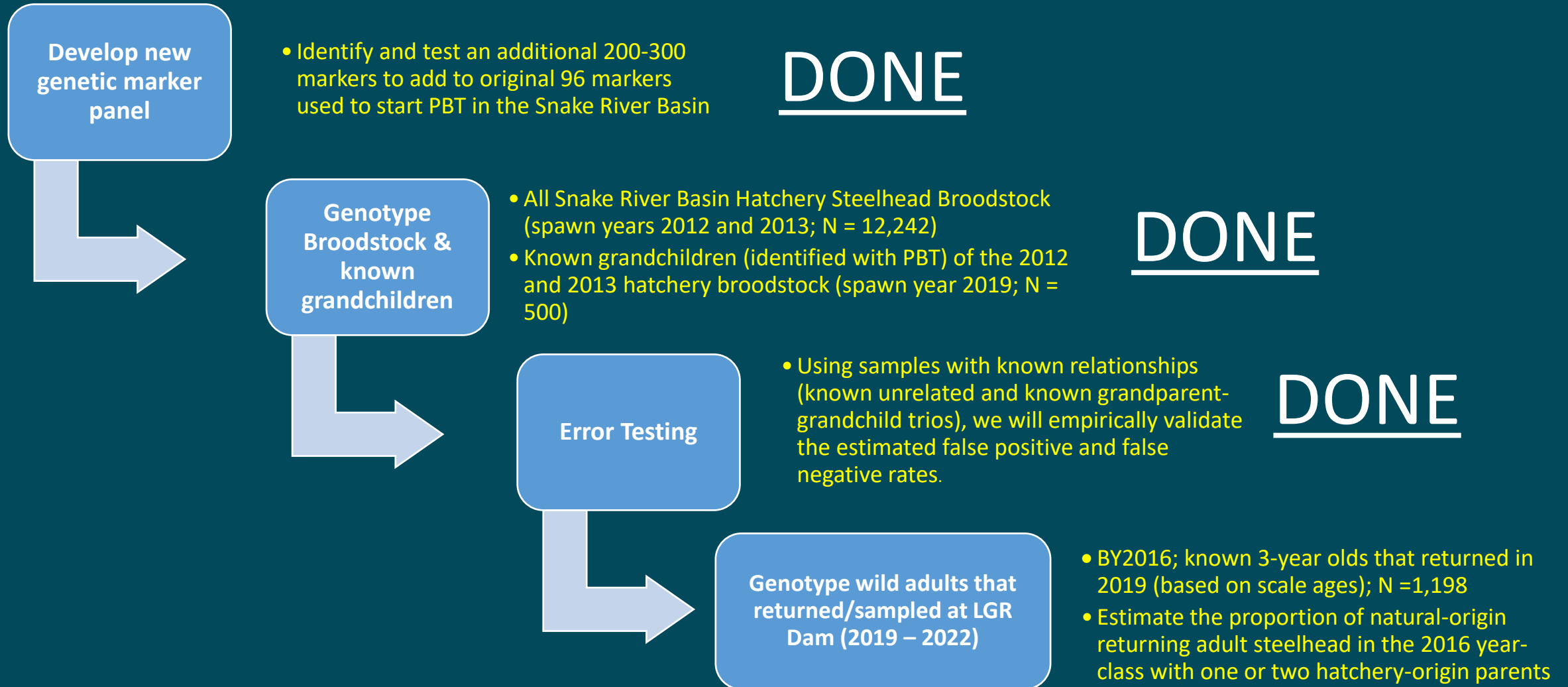
Grandparent inference from genetic data: The potential for parentage-based tagging programs to identify offspring of hatchery strays

Thomas A. Delomas ✉ Matthew Campbell

First published: 27 October 2021 | <https://doi-org.libproxy.boisestate.edu/10.1002/nafm.10714>

- **Next Steps?**

- ✓ We proposed an empirical proof-of-concept study to validate and demonstrate this new technology in Snake River steelhead (**Funded by PCSRF in 2021: 01121SC**)



Estimating grandparentage error rates

- Using known grandparentage assignments from multi-generational pedigrees, we can estimate false positive and false negative rates

		Assignment outcome	
Known relationship		Identified as grandparents	Not identified as grandparents
	True grandparents (P)	True positive (TP)	False negative (FN) - Type II error
	Unrelated grandparents (N)	False positive (FP) - Type I error	True negative (TN)

Estimating grandparentage error rates

Known relationship	Assignment outcome	
	Identified as grandparents	Not identified as grandparents
True grandparents (P) P = 1,198	True positive (TP)	False negative (FN) - Type II error FN = 80
Unrelated grandparents (N) N = 20,107,644	False positive (FP) - Type I error FP = 2	True negative (TN)

$$\text{false negative rate (FNR)} = \frac{FN}{P} = \frac{80}{1198} = 0.067$$


related to the number of **true grandparent trios** considered

Interpretation – for every true comparison we make, there is a 0.067 probability of rejecting a correct assignment

Estimating grandparentage error rates

Known relationship	Assignment outcome	
	Identified as grandparents	Not identified as grandparents
True grandparents (P) P = 1,198	True positive (TP)	False negative (FN) - Type II error FN = 80
Unrelated grandparent trios (N) N = 20,107,644	False positive (FP) - Type I error FP = 2	True negative (TN)

$$\text{false positive rate (FPR)} = \frac{FP}{N} = \frac{2}{20,107,644} = 9.95 \times 10^{-8}$$


related to the number of **unrelated grandparent trios** considered

Interpretation – for every unrelated comparison we make, there is a 9.95×10^{-8} probability of accepting an incorrect assignment

Importance of hatchery metadata

- Data collected at hatcheries during spawning can greatly decrease the number of comparisons made during grandparentage analyses by reducing the number of crosses considered
- For the SY2012 steelhead broodstock at Oxbow hatchery, with 477 broodstock and 700 potential grandchildren:

$$\# \text{ comparisons} = \# \text{ grandparent crosses} \times \# \text{ grandchildren}$$

Metadata	# Crosses	# Grandchildren	# Comparisons	# False positives
None	113,526	700	79,486,200	7.94
Phenotypic sex	56,376	700	39,463,200	3.93
Phenotypic sex + spawn date	5,268	700	3,687,600	0.37
Cross records	246	700	172,200	0.02

Estimating expected false negatives and positives

- For an analysis with 2,009 wild-origin potential grandchildren at Lower Granite and 28,796 grandparent crosses:

Assignment outcome		
Total comparisons made	Identified as grandparents	Not identified as grandparents
57,851,164	476	57,850,688

$$\# \text{ false positives} = 9.95 \times 10^{-8} \times 57851164 = 5.75$$

$$\# \text{ false negatives} = 0.067 \times 476 = 31.89$$

Low expected numbers of false positives and false negatives show that grandparentage is possible on a basin-wide scale when hatchery metadata is available

- **Grandparentage Results:**

Potential Grandchildren	N = 2,009 samples, SY2020-2021 putatively wild fish from Lower Granite ✓ No PBT assignments (i.e. NOT a hatchery no-clip) ✓ BY2016 scale ages
Potential Grandparents	N = 10,213 samples, SY2012-2023 broodstock
Assignments	N = 374 (18.6%) individuals assigned to grandparents in the SY2012-2013 broodstocks ✓ 81.4% (1635) of individuals assigned to zero sets of hatchery grandparents ✓ 14.3% (288) of individuals assigned to one set of hatchery grandparents ✓ 4.3% (86) of individuals assigned to two sets of hatchery grandparents

- **Grandparentage Results:**
 - ✓ **All** hatcheries were identified as having some individuals that strayed and spawned successfully
 - ✓ Using PIT detections of grandchildren indicated that **most** returned to areas managed for hatchery fish: Upper Salmon, S.F. Clearwater, Grande Ronde (Cottonwood)
 - ✓ **Few** grandchildren were identified in areas managed for wild fish (Upper Clearwater, MFSR, SFSR)
 - ✓ But....low sample sizes

• Current Work?

- Focus on juvenile samples from a mix of areas, some managed for hatchery fish and some managed for wild fish.

Pedigree	Stream/Site	N
OmyBGCS18C	Big Creek (MFSR)	400
OmyFILW18C	Fish Creek (Lochsa/Upper Clearwater)	635
OmyHAYW18C	Hayden Creek (Lemhi River)	150
OmyMRST18C	Marsh Creek (MFSR)	75
OmyNFSW18C	N.F. Salmon	150
OmyPAHS18C	Pahsimeroi River	150
OmyRRHW18C	Rapid River (Little Salmon River)	300
OmySAWW18C	Sawtooth (Upper Salmon)	140
	Total Samples	2,000
	per sample cost:	\$15
	Amount	\$ 30,000.00

- Received funding from PCSRF to complete this additional genotyping (January 2025)

- # Challenges?

- Current technology would require us to run 2 genetic marker panels rather than one. **More time and \$.**
- Large numbers of broodstock would need to be re-genotyped to allow this technology to be used starting this year. **More time and \$.**
- **Collection of accurate hatchery metadata is needed!**
 - ✓ Without hatchery data, all possible pairs of potential grandparents must be considered for every sampled offspring, resulting in an extremely large number of trio comparisons.
 - ✓ Each comparison carries a small probability of a false-positive error. As the number of comparisons increases, the cumulative likelihood of false positives also increases. Hatchery data (e.g., spawning records, sex, and timing) reduces the total number of plausible pairs by excluding combinations that are biologically or temporally impossible.
 - ✓ This is available in the Snake River Basin. Cross-record information may not currently be collected at hatcheries outside the basin.

Questions???



How are Grandparent assignments made???

Grandparent assignments in this study are made using a likelihood-based approach to compare genetic relationships among putative grandparents and grandchildren. The method calculates the likelihood of a trio (two grandparents and one grandchild) being biologically related versus being unrelated by analyzing their genotypes. Key points include:

- **Likelihood Ratios:** Log-likelihood ratios (LLRs) are computed for the relationship hypothesis (grandparent-grandchild) compared to unrelated trios.
- **Mendelian Inheritance:** Calculations incorporate allele frequencies, Mendelian inheritance laws, and genotyping error models.
- **Error Filtering:** Trios with excessive Mendelian incompatibilities (MIs) are excluded using a predetermined threshold.
- **Statistical Validation:** False-positive and false-negative error rates are assessed using Monte Carlo simulations with methods like importance sampling and stratified sampling. The process is implemented through an R package ("gRandma"), which enables efficient grandparent inference by integrating existing genetic data and analytical frameworks, particularly for monitoring hatchery programs.

The proposed approach provides a likelihood-based framework to infer grandparent-grandchild trios using genotypes from two putative grandparents and one putative grandchild, while considering genotyping errors. The method is implemented through an R package called gRandma.

Methodology for Estimating Type I and Type II Errors

False-Positive Error Rate (Type I Error):

Definition: The likelihood of incorrectly assigning a grandparent-grandchild relationship to an unrelated trio. Approaches:

Importance Sampling:

- Simulates genotypes focused on scenarios likely to yield false positives and adjusts the results using statistical weights.
- Stratified Sampling: Categorizes simulated data by the number of Mendelian incompatibilities (MIs) and focuses computational resources on groups most likely to yield false positives.

Estimation Process:

- Simulates genotypes for trios assuming various relationships (e.g., unrelated, great-aunts, cousins).
- Uses Monte Carlo techniques to refine error rate estimates for rare events, achieving sufficient precision for comparisons involving thousands to millions of trios.
- Incorporates genotyping error by modeling the distribution of observed genotypes relative to true genotypes.

Methodology for Estimating Type I and Type II Errors

False-Negative Error Rate (Type II Error):

Definition: The probability of failing to detect a true grandparent-grandchild relationship.

Estimation:

- Simulates true grandparent-grandchild trios using allele frequencies and genotyping error models.
- Tracks trios excluded due to MIs or log-likelihood ratios (LLRs) below the critical threshold.
- Uses Monte Carlo simulations to calculate the proportion of excluded true relationships.

Simulation Parameters:

Genetic Marker Panels:

SNPs (500–700 loci) or microhaplotypes (200–300 loci) are simulated to optimize power and accuracy.

Marker Characteristics:

Focuses on biallelic or multiallelic loci with predefined heterozygosity levels to assess realistic error rates.

Comparison Metrics:

Evaluates error rates for both unrelated and closely related trios under varying levels of data availability (e.g., sex, spawning records).

Parentage-based genetic tagging - PBT (Hatchery Fish)

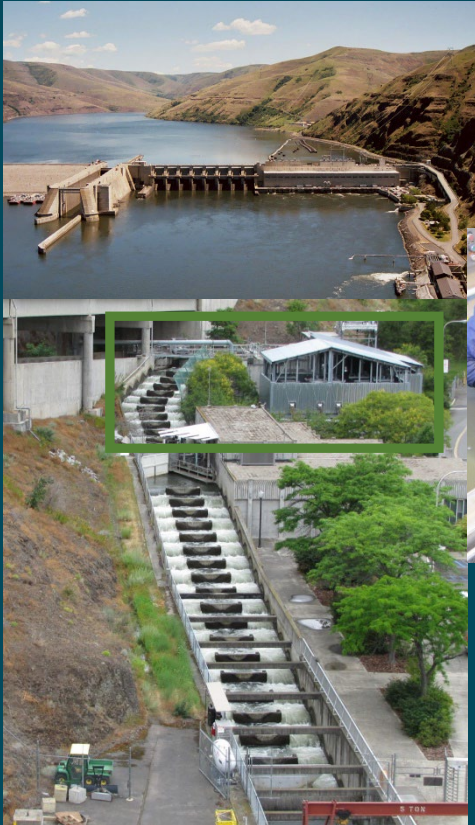
- Proposed by Eric Anderson and Carlos Garza (NOAA SWFSC) in 2005 as an alternative to CWTs
- Tested and implemented by IDFG and CRITFC in the Snake River and Columbia Rivers, respectively.
- Genetic-based fish tagging method that involves genotyping hatchery broodstock
- By genetically sampling the parents, all offspring are genetically “tagged”, because you can use dual-parentage assignments to assign any offspring to their parents
- Information obtained is the same as CWT, but improved tagging rate of hatchery fish (~95-100%)
- Juvenile handling not required prior to release
- Genetic tags can’t be lost and there is no differential mortality issues
- ‘Tag’ recovery is non-lethal, and possible at all life stages

2 = 6,000!!!!



• Error Testing Methodology

- ✓ Error Testing
- ✓ Details for testing shown in table



Test Individuals	Details
Potential Grandchildren	701 SY2020 hatchery fish from Lower Granite, each with 4 known hatchery grandparents based on pedigree data.
Potential Grandparents	10,213 SY2012-2013 broodstock individuals.
Known Trios	1,198 known grandparents based on overlap of genotyping success, crosses, spawn years, etc.

- **Error Testing Results**

Empirical Error Rates	Details
- Empirical False Positive Rate	1.00e-5% (2 false positives for unrelated individuals in 19,985,537 comparisons)
- Empirical False Negative Rate	6.7% (80 of 1,198 known assignments)

What if we don't have known pedigrees?

- When we don't have information about the numbers of true and unrelated grandparent trios considered (e.g., in wild populations), we can still estimate the number of expected false positives and false negatives using the previously estimated FPR and FNR

$$\# \text{ false positives} = FPR \times \text{total \# comparisons made}$$

$$\# \text{ false negatives} = FNR \times \text{total \# accepted assignments}$$

Since every comparison made increases the estimated number of false positives, it's imperative to reduce the number of comparisons we make*