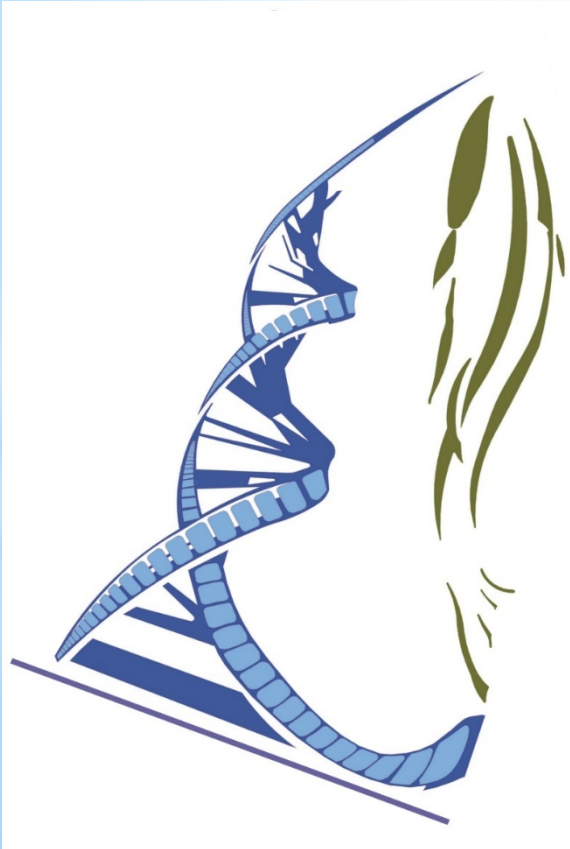


# Hagerman Genetics Lab Data Management

**Jeff Stephenson**



Columbia River Inter-Tribal  
Fish Commission

Hagerman Fish Culture  
Experiment Station



Hagerman, Idaho USA

# Presentation Subjects

- Hagerman genetics program overview
- Genetic marker types
- Genotyping methods
- Hagerman database needs: Progeny Clinical
- Shared database / FishGen
- Data use in other databases

# CRITFC Staff in Hagerman

- Supervisor/ Lead Geneticist: Shawn Narum
  - Geneticists (7)
    - Dan Hasselman
    - Ilana Koch
    - Andrew Matala
    - Steven Micheletti
    - Jeff Stephenson
    - Zhongqi Chen
    - Vacancy to fill
  - Lab Techs & Research Support(9)
    - Stephanie Harmon
    - Amanda Matala
    - Travis Jacobson
    - Lori Maxwell
    - Megan Moore
    - Vanessa Morman
    - Janae Cole
    - Rebecca Sanders
    - Mike Eastman
- Facility shared with U. of Idaho and USDA
- Facility director: Brian Small (U. of Idaho)
- USDA group leader: Ken Overturf

# Support for Specific Tribal Programs

Genetics projects with all 4 CRITFC tribes:

- Nez Perce
- Warm Springs
- Umatilla
- Yakama Nation

Research has led to >75 scientific peer-reviewed publications with tribal partners

# Primary Research Areas

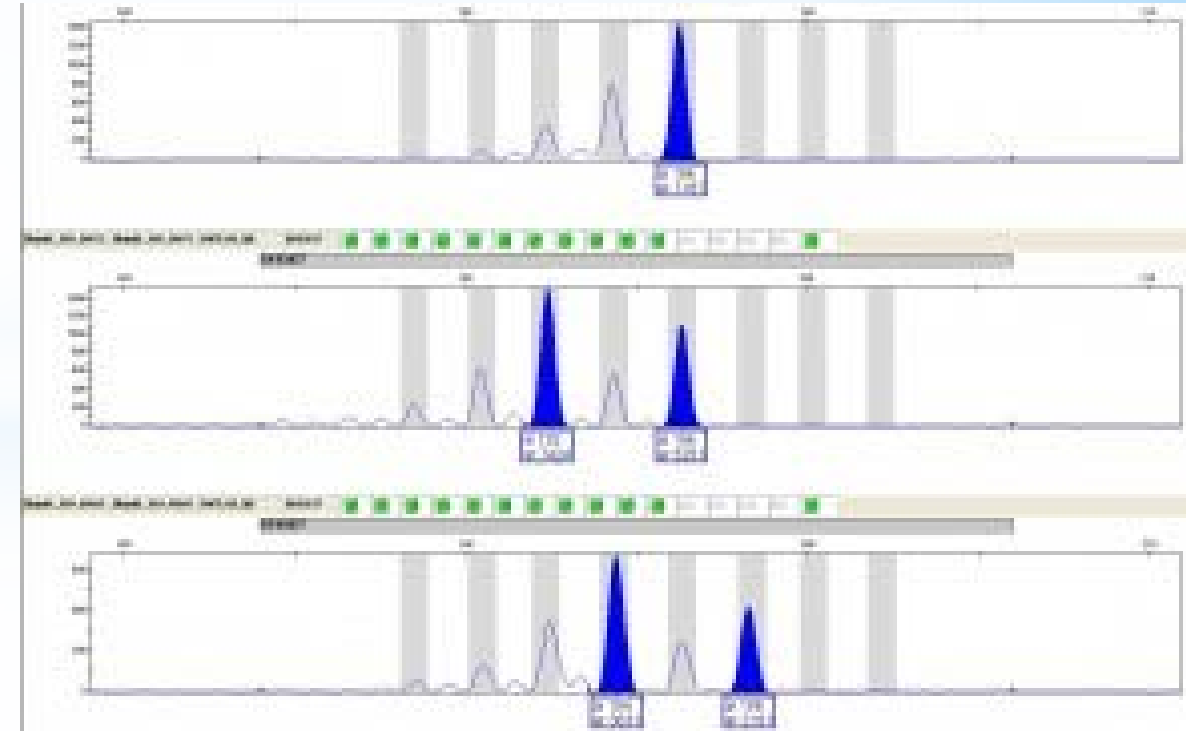
- 1) **Genetic tagging and monitoring of fisheries**
  - Stock specific abundance, harvest, and run-timing to assist fisheries management
- 2) **Genetic effects of hatchery practices**
  - Evaluate supplementation and reintroduction programs to assist with recovery of salmonids
- 3) **Genetic adaptation to local environments**
  - Investigate local adaptation and the genetic basis for traits (e.g., thermal adaptation, run-timing, maturity)



# Marker Types

## Microsatellites

- Variable number repeats
- 10-20 loci (markers) per project
- 10-50 alleles (alternative form) per locus
- Assumed to be neutral
- No longer commonly genotyped
- Keep old data in the database



# Marker Types

## Sex Marker

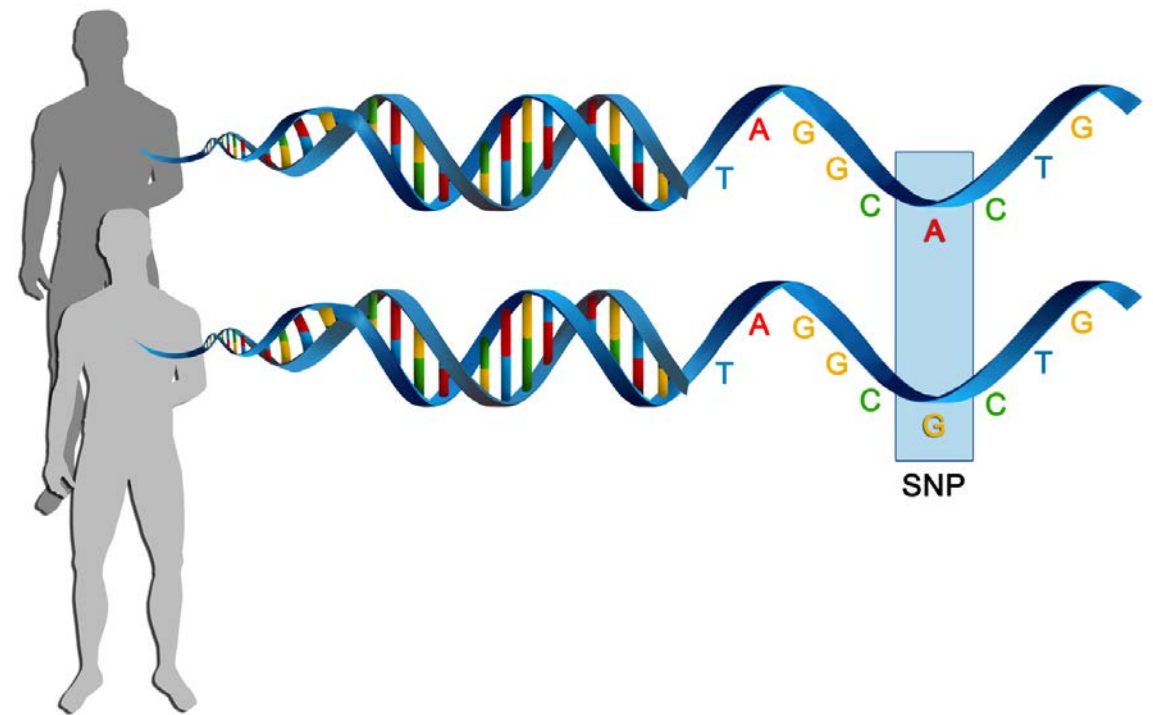
- Presence absence of the Y chromosome
- 1 locus
- 2 alleles
- Reported as X and Y
- Doesn't work for sockeye



# Marker Types

## SNP

- Single Nucleotide Polymorphism
- 96-500 loci (markers) per project
- Panels are species specific
- 2-4 alleles per locus
- Usually neutral

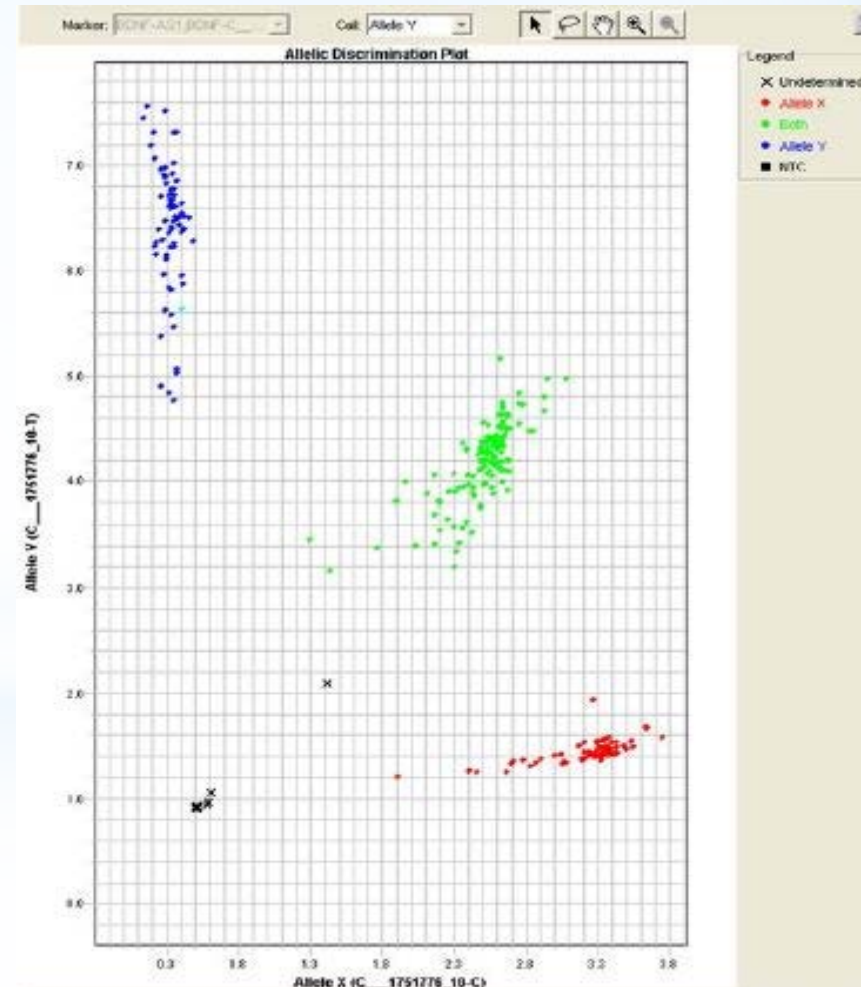




# Genotyping Methods-SNP

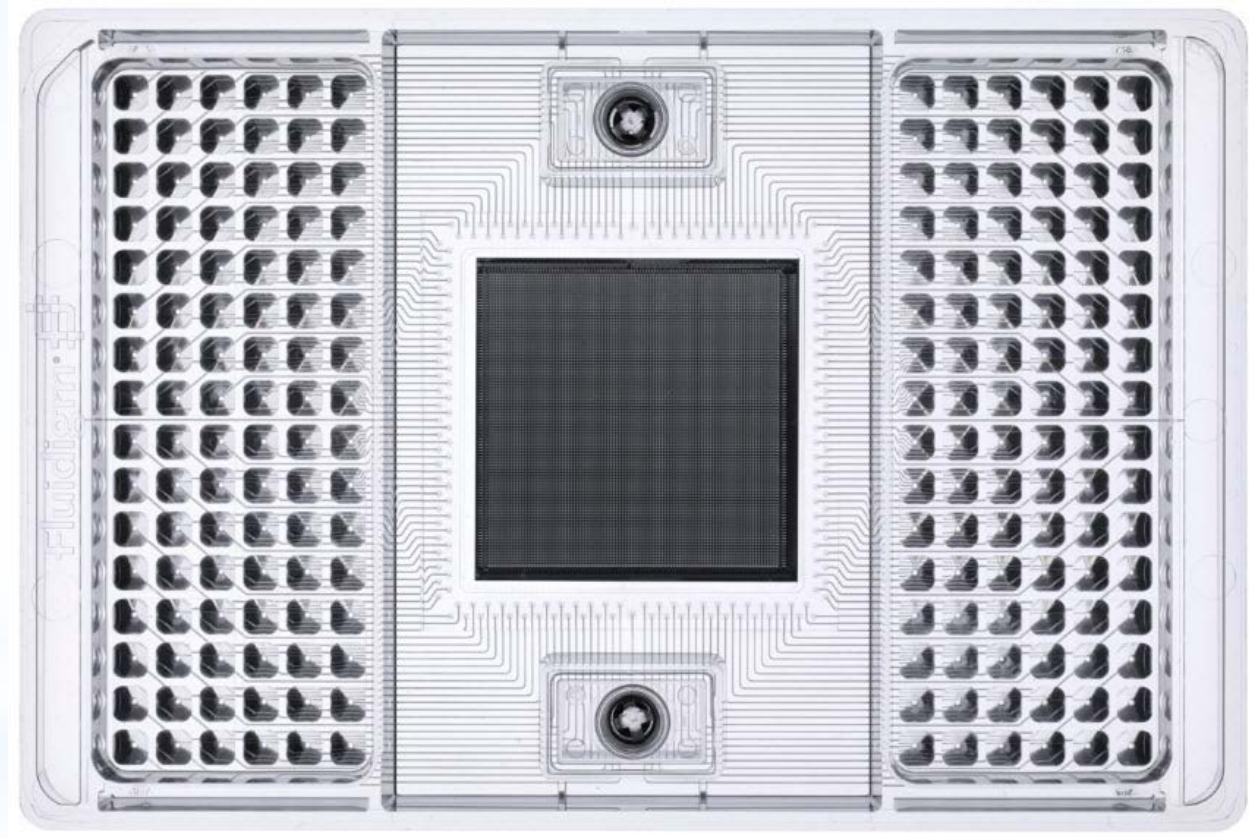
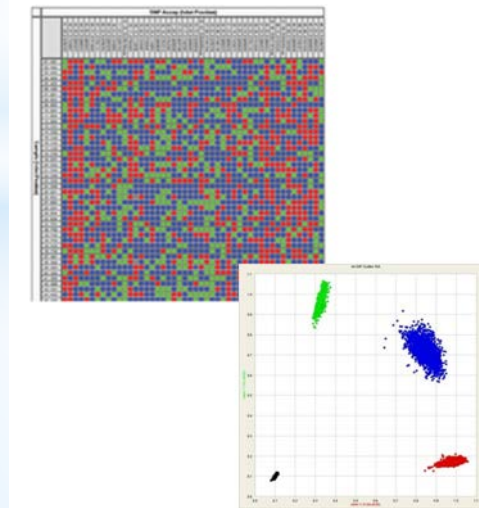
## Taqman Assays

- Single locus
- 96-384 samples at a time



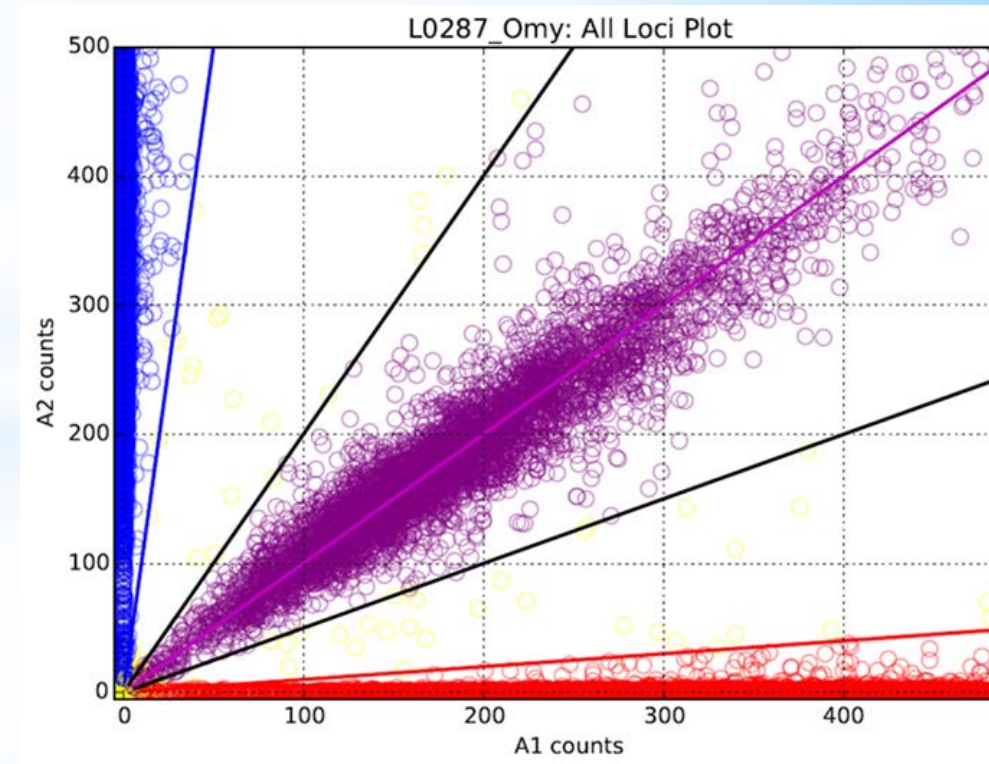
# Genotyping Methods-SNP Fluidigm Array

- 96 Loci at a time
- 96 Samples at a time
- 18,000 Allelic genotypes



# Genotyping Methods-SNP GT-seq

- Genotyping in thousands by sequencing
- Developed at Hagerman Genetics Lab
- Targeted amplicon sequencing
- ~300 Loci at a time
- Typically 1000 Samples at a time
- 600,000 Allelic genotypes
- ~125,000 samples in 2017
- 75,000,000 alleles (database needs)





# Genotyping Methods-SNP PoolSeq

- Genotypes at the population level
- Can produce 5-15 million of SNPs
- Good at finding selective loci
  - Run timing
  - Growth rate/size
  - Thermal tolerance
  - Age at maturity
- Data does not go into a database, but can be added to existing GT-seq panels. More data into the database



# Hagerman Database Needs

- Lots of individual samples
- Lots of field collection data
- Lots of genotypes
- Our first database was called Vanessa
- Data integrity failure with less than 20,000 samples

# Hagerman Database Needs

- Lots of individual samples
- Lots of field collection data
- Lots of genotypes
- Our first database was complicated
- Data integrity failure around 2005





# The Most Powerful Risk Modeling and Pedigree Software for Clinicians

[DISCOVER PROGENY CLINICAL](#)

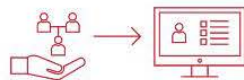
## New High Risk Triage Screening Tools

See our quick screening tools to identify high risk patients for breast, colorectal and other cancers...

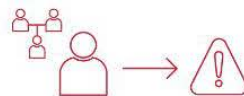
[See Features](#)

## Top Features

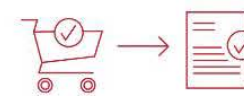
Progeny Clinical simplifies the process of managing family history, assessing risk and determining treatment options for your patients.



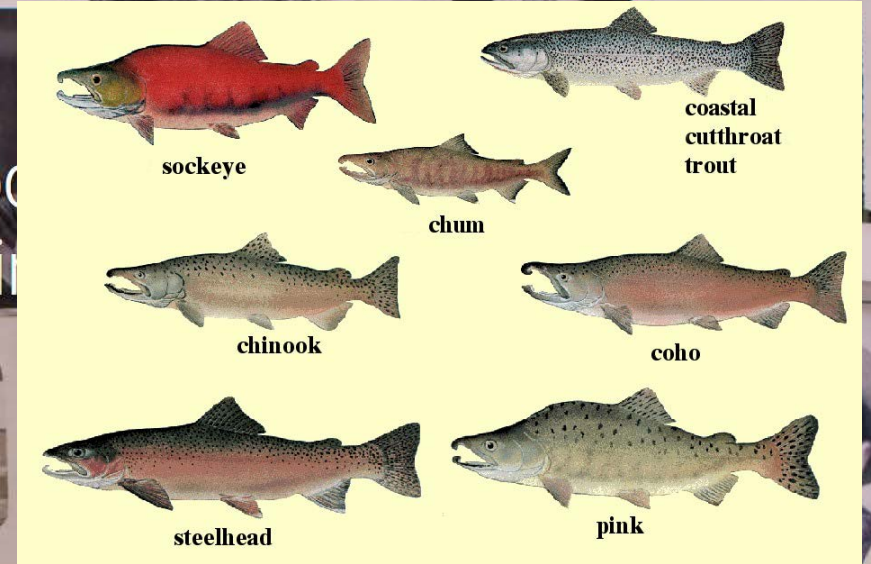
Collect Family History



Assess Risk



Order and Track Genetic Testing



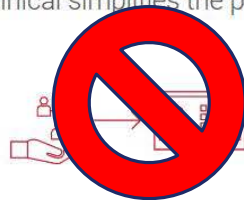
## New High Risk Family Screening Tools

See our quick screening tools to identify high risk patients for breast, colorectal and other cancers...

[See Features](#)

## Top Features

Progeny Clinical simplifies the process of managing family history, assessing risk and determining treatment options for your patients.



Collect Family History



Assess Risk

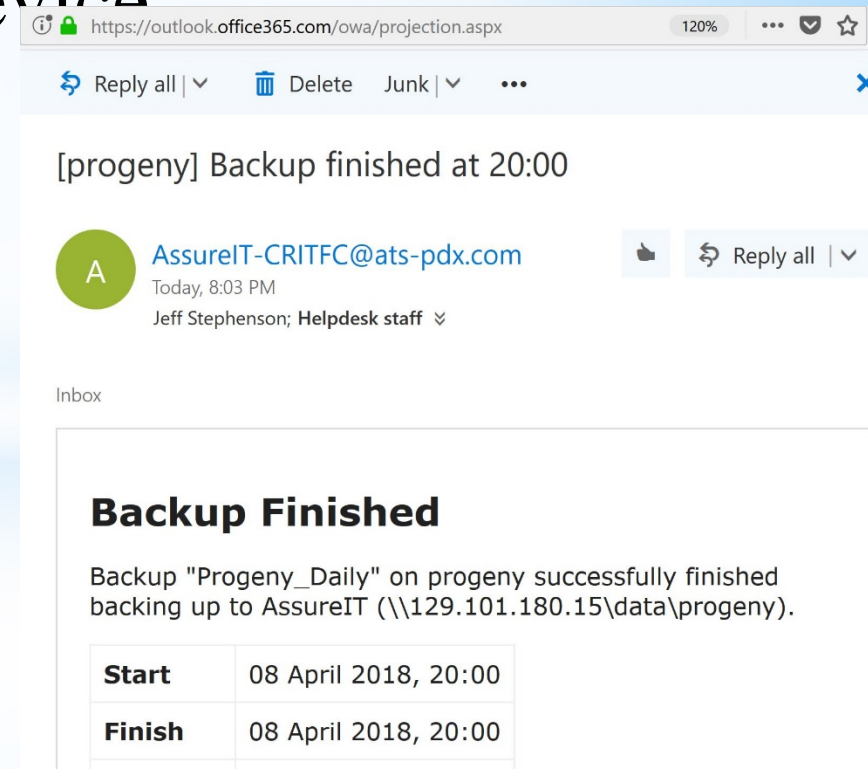


Order and Track Genetic Testing



# Where is the database?

- Physical server in my office
- Accessible only on our internal network
- Hourly backups of the database
- Nightly backups to an external device device in a data closet
- Nightly offsite backups



# What is actually in the Progeny database?

- 920,000 Individual IDs (samples)
- Field collection data
- Sample location within the lab
- Processing records
- Genotype data

# What is actually in the Progeny database?

- 920,000 Individual IDs (samples)
  - Chinook 570,000
  - Steelhead 213,000
  - Sockeye 50,000
  - Sturgeon 10,000
  - ~50 Misc. species

# What is actually in the Progeny database?

- Field collection data
  - Sample Name-Usually modified
  - Species
  - Location
  - Date-Year only works
  - ~10 more required fields
  - ~200 Misc. fields, often project specific



# What is actually in the Pro

- Sample location within the lab
  - Pedigree = Container
  - Containers stored in numerical order



# What is actually in the Progeny database?

- Processing records
  - Sample confirmed
  - Sample preservation type
  - Extract plate
  - Genotyping Information



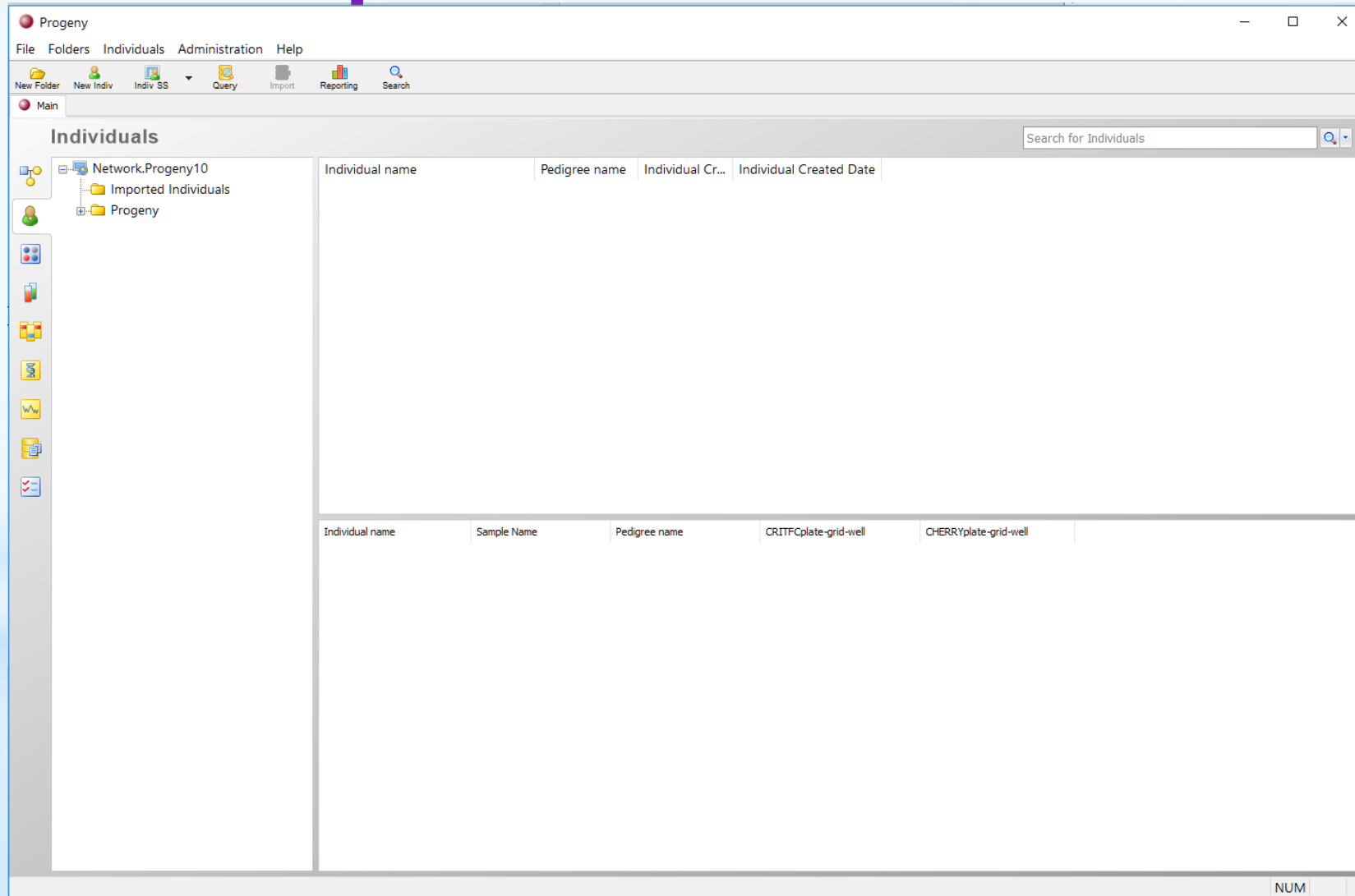
# What is actually in the Progeny database?

- Genotype data with basic QA/QC
  - No template controls
  - Positive controls on every plate
  - Minimum genotyping percentage

	Individual name		OmyY1_2SEXY-a1	OmyY1_2SEXY-a2	Omy_114315-438-a1	Omy_114315-438-a2	Omy_130524-160-a1	Omy_130524-160-a2	Omy_97077-73-a1	Omy_97077-73-a2	Omy_97660-230-a1	Omy_97660-230-a2	Omy_97865-196-a1	Omy_97865-196-a2	Omy_97954-618-a1	Omy_97954-618-a2	Omy_aldB-165-a1	Omy_aldB-165-a2	Omy_aromat-280-a1	Omy_aromat-280-a2	Omy_arp-630-a1	Omy_arp-630-a2	Omy_aspAT-123-a1
OmyCEC17r-Roza02	X	Y	A	A	B	B	B	B	B	A	A	A	A	A	B	A	B	A	A	A	B	A	
OmyCEC17r-Roza03	X	Y	A	B	A	A	B	B	B	A	A	A	A	A	B	A	A	A	A	A	B	A	
OmyCEC17r-Roza04	X	Y	A	A	B	B	B	B	B	A	A	A	A	A	B	A	A			A	B	A	
OmyCER15r-15BK049	X	Y	A	B	A	A	B	B	B	A	B	A	A	A	A	B	B	A	A	B	B	A	
OmyCER15r-15BS055	X	Y	A	A	A	B	B	B	B	A	B	A	A	A	A	A	B	A	A	A	B	A	
OmyCER15r-15BS057	X	Y	A	A	A	B	A	B	B	A	A	A	A	B	B	A	B	A	A	A	B	A	
OmyCER15r-15BS059	X	Y	A	A	A	B	A	B	B	A	A	A	A	A	A	A	A	A	A	A	B	A	
OmyCER16r-16DG265	X	Y	B	B	A	B	B	B	B	B	B	A	A	A	A	B	B	B	B	A	B	A	
OmyCER16r-16DG266	X	X	B	B	B	B	B	B	B	A	A	A	A	A	B	A	B	A	A	A	A	A	
OmyCER16r-16DG268	X	X	A	A	A	B	B	B	B	A	B	A	A	B	B	A	A	A	A	A	B	A	
OmyCER16r-1J1	X	X	A	A	B	B	B	B	B	A	B	A	A	B	B	A	A	A	A	A	A	A	
OmyCER16r-1J2	X	X	A	B	B	B	B	B	B	A	A	A	A	A	B	B	B	A	B	B	B	A	
OmyCER16r-1J3	X	Y	A	A	B	B	B	B	B	B	B	A	A	B	B	A	B	A	B	A	A	A	
OmyCER16r-1J4	X	Y	A	A	B	B	B	B	B	B	B	A	B	B	B	B	B	A	A	A	A	A	

# What does Progeny look like?

## Graphical User Interface





Progeny

File Folders Individuals Administration Help

New Folder New Indiv Indiv SS Query Import Reporting Search

Main

### Individuals

Search for Individuals

Network.Progeny10

Imported Individuals

Progeny

- Pedigree (sample container)
- Individuals
- Samples
- Markers
- Genotypes
- Analysis
- Fields

Individual name	Pedigree name	Individual Cr...	Individual Created Date
-----------------	---------------	------------------	-------------------------

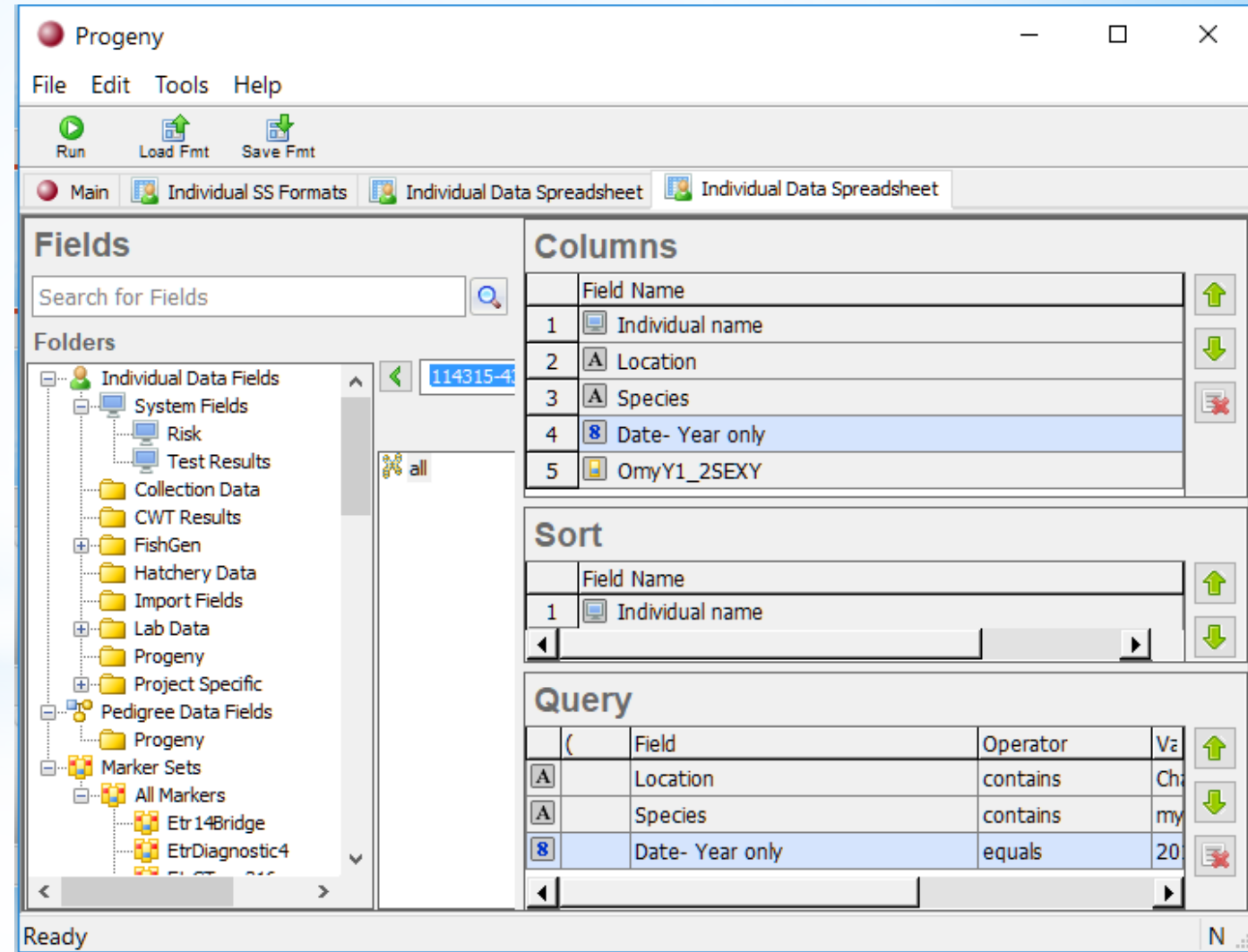
Individual name	Sample Name	Pedigree name	CRITFCplate-grid-well	CHERRYplate-grid-well
-----------------	-------------	---------------	-----------------------	-----------------------






NUM

# How to get data out of Progeny?


## Query by any field of interest

- What data is needed
- Sort order
- Which Individuals






	Field Name
1	 Individual name
2	 Location
3	 Species
4	 Date- Year only
5	 OmyY1_2SEXY

## Sort

	Field Name
1	 Individual name
<div> <div>◀</div> <div></div> <div>▶</div> </div>	

## Query

	(	Field	Operator	Value
		Location	contains	Ch
		Species	contains	my
		Date- Year only	equals	20





Run



Edit



Save



Print



Load Fmt



Save Fmt



Add Row



Delete Row



Export



Import

3<sup>1</sup>2

Count



Main



Individual SS Formats



Individual Data Spreadsheet



Individual Data Spreadsheet

	Individual name ▼	Location ▼	Species ▼	Date- ▼	OmyY1_ ▼	OmyY1_2SEXY-a2
1	OmyYRC17k-0095	Yakima River Chandler	Oncorhynchus mykiss	2017	X	X
2	OmyYRC17k-0096	Yakima River Chandler	Oncorhynchus mykiss	2017	X	X
3	OmyYRC17k-0097	Yakima River Chandler	Oncorhynchus mykiss	2017		
4	OmyYRC17k-0098	Yakima River Chandler	Oncorhynchus mykiss	2017	X	X
5	OmyYRC17k-0099	Yakima River Chandler	Oncorhynchus mykiss	2017	X	X
6	OmyYRC17k-0100	Yakima R				X
7	OmyYRC17k-0101	Yakima R				X
8	OmyYRC17k-0102	Yakima R				X
9	OmyYRC17k-0103	Yakima R				X
10	OmyYRC17k-0104	Yakima R				Y
11	OmyYRC17k-0105	Yakima R				X
12	OmyYRC17k-0106	Yakima R				X
13	OmyYRC17k-0107	Yakima R				X
14	OmyYRC17k-0108	Yakima R				X
15	OmyYRC17k-0109	Yakima River Chandler	Oncorhynchus mykiss	2017	X	X

Count

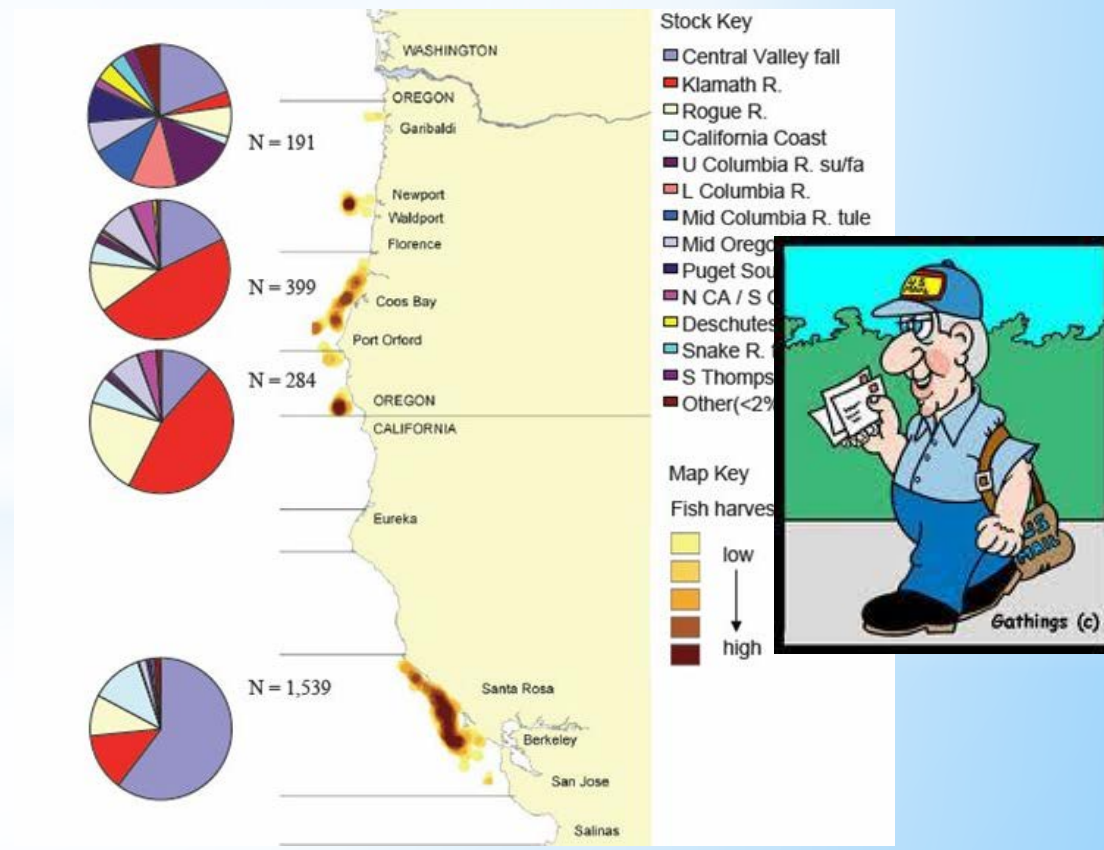
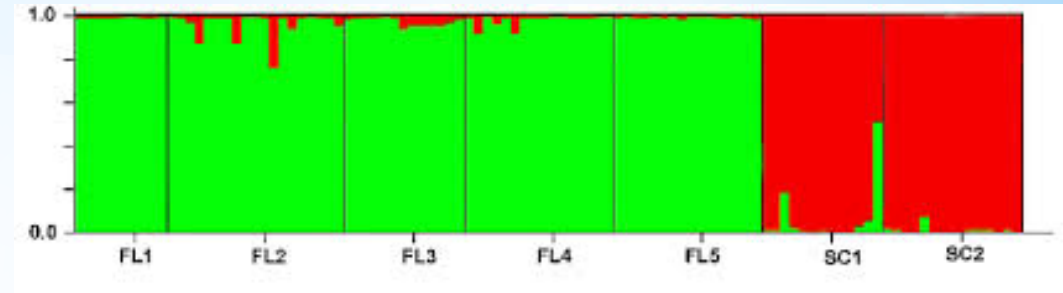


This spreadsheet contains 144 items.

OK

# What happens to data after progeny

- Project specific analysis
  - Further QA/QC work
  - Various statistical analysis
    - Structure analysis
    - Genetic Stock Identification
    - Parentage assignment
- FishGen –Shared database



# Before FishGen Remove

- Special characters e.g. commas in comment fields
- Non-‘ATCG’ genotypes e.g., “0 –”, “0 G”
- Duplicate sample IDs
- Individuals with duplicate genotypes



# Before FishGen Transform

- Alter metadata field names (62 fields)
- Combine multiple entries for 'Cross Data' and 'Spawn Date'
- Determine genetic sex
- Adjust 'Hatchery' field for compliance with FishGen
- Insert PBT 'Tag rate'
- Insert correct 'Lineage' and 'Run Description'
- Designate 'Collection Name' field for groups of samples
  - Fundamentally different than

# What is FishGen?

- Shared database developed by IDFG
- Repository for GSI and PBT data
  - GSI = Genetic Stock Identification
  - PBT = Parent based tagging
- Why?
  - Reporting requirements
  - Simplifies data sharing
  - Decreases genetic analysis cost

# Previous shared databases

- GAPS (Genetic Analysis of Pacific Salmonids)
- SPAN (Steven Phelps Allele Nomenclature)
- However, these were limited
  - Microsatellites only
  - Species specific
  - Static data



# Home Page FishGen.net

Logged in as user idfgadmin@fishgen.net , last login was 7/16/2015 at 9:36 AM

[Home](#) | [Logout](#) | [My Account](#) | [User Manual](#) | [Contact](#)



# FishGen

[Home](#)

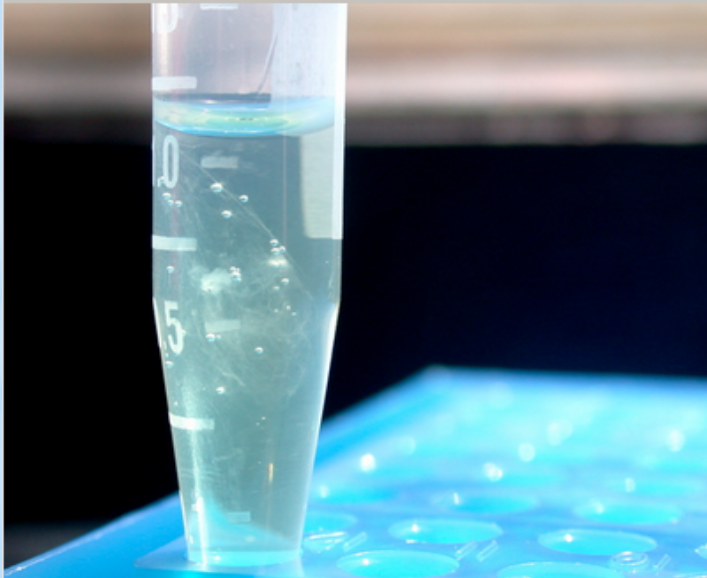
[Contributor](#)

[Search](#)

[Data Sets](#)

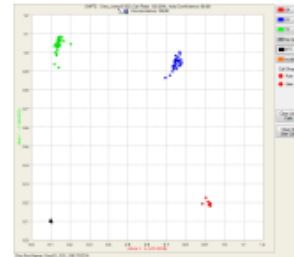
[Marker Sets](#)

[Admin](#)



## FishGen.Net

FishGen.net is a final repository for genetic information of fish species that are of conservation and management importance to federal, state and tribal agencies in the United States and Canada. The repository currently houses salmon and steelhead genetic data as part of Genetic Stock Identification and Parentage Based Tagging projects in the Columbia River basin and throughout the Pacific Coast of North America. FishGen.net was developed by Resource Data, Inc. (RDI) for the Idaho Department of Fish and Game with funding from the Pacific Coast Salmon Recovery Fund and the Bonneville Power Administration.



[Upload Data](#)



[Learn about collaborating labs](#)



**NOAA** PACIFIC COASTAL SALMON  
RECOVERY FUND PROJECT DATABASE  
NOAA FISHERIES

**IDAHO**

Idaho Governor's Office of  
Species Conservation



# Request an Account

Welcome, Guest.

Search

## Request Account

*\* denotes a required field*

*Email Address:	<input type="text"/>	*Organization:	<input type="text"/>
*Password:	<input type="password"/>	Other Organization:	<input type="text"/>
*Re-enter Password:	<input type="password"/>	Street Address 1:	<input type="text"/>
Password Hint:	<input type="text"/>	Street Address 2:	<input type="text"/>
*First Name:	<input type="text"/>	City:	<input type="text"/>
*Last Name:	<input type="text"/>	State:	<input type="text"/>
Phone:	<input type="text"/>	Zip Code:	<input type="text"/>

Send Request

# Search Screen

Home Contributor Search Data Sets Marker Sets Admin

**Search Filter**

Search

► Species Options

Species

Oncorhynchus mykiss ✕

▼ Hatchery Options

▼ Individual Options

► Collection Options

From Sample Year

2010 ▼

To Sample Year

2012 ▼

HUC4

Select Some Options

+

-

Switch Basemap

Esri, DeLorme, FAO, USGS, NOAA, EPA, NPS

▼ Switch Basemap

Imagery With Labels

Streets

Topographic

National Geographic




HUC



















Open Street Map



# Saved Datasets

## Data Set Find and Export

Name	Organization
<input type="text"/>	Please select 
Species	BPA Project
Please select 	Please select 
<input type="button" value="Search"/>	

Search Results									
Dataset Name	Private	Dataset Owner	Organization	Created (MST)	# of Collections	# of Individuals	Notes		
<a href="#">Maurens first PBT upload gender added 23May14</a>	<input type="checkbox"/>	Jeff Stephenson	Columbia River Inter-Tribal Fish Commission, Hagerman, ID	5/23/2014 10:24 AM	2	1758	<input type="text"/>	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">Jeffs Test Data Set</a>	<input type="checkbox"/>	Jeff Stephenson	Columbia River Inter-Tribal Fish Commission, Hagerman, ID	5/7/2014 1:53 PM	4	350	Don't use this,	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">Ots Baseline 7May1207</a>	<input type="checkbox"/>	Jeff Stephenson	Columbia River Inter-Tribal Fish Commission, Hagerman, ID	5/7/2014 12:30 PM	52	4073	<input type="text"/>	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">Reference Omykiss for Mo</a>	<input type="checkbox"/>	Matthew Campbell	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	4/22/2014 11:42 AM	37	1741	This saved dat	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">EFGl Snake River basin steelhead genetic baseline v3.0</a>	<input type="checkbox"/>	Ninh Vu	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	3/28/2014 10:13 AM	139	8068	Snake River ba	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">EFGl Snake River basin Chinook salmon genetic baseline v3.0</a>	<input type="checkbox"/>	Ninh Vu	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	3/28/2014 10:11 AM	151	6165	Snake River ba	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">Blankenship et al 2011</a>	<input type="checkbox"/>	Matthew Campbell	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	1/30/2014 10:55 AM	339	15658	Blankenship, S.	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">SNP Test-Warm Springs Chinook 16jan14</a>	<input type="checkbox"/>	Matthew Campbell	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	1/16/2014 2:39 PM	22	1823	Test Dataset fr	 	<a href="#">Export</a> <a href="#">Delete</a>
<a href="#">Hess et al 2014 GAPS subset</a>	<input type="checkbox"/>	Matthew Campbell	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	12/30/2013 5:00 PM	29	2025	This saved dat	 	<a href="#">Export</a> <a href="#">Delete</a>

# Saved Dataset: Blankenship 2011

## Data Set Find and Export

Name

Species

Please select

Search

## Dataset Name

[Maureens first PBT upload gender a](#)

[Jeffer's Test Data Set](#)

[Ots Baseline 7May1207](#)

[Reference Omykiss for Mo](#)

[EFGL Snake River basin steelhead g](#)

[EFGL Snake River basin Chinook salmon genetic baseline v3.0](#)

[Blankenship et al 2011](#)

[SNP Test-Warm Springs Chinook 16jan14](#)

[Hess et al 2014 GAPS subset](#)



## Transactions of the American Fisheries Society

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/utaf20>

## Major Lineages and Metapopulations in Columbia River *Oncorhynchus mykiss* Are Structured by Dynamic Landscape Features and Environments

Scott M. Blankenship <sup>a</sup>, Matt R. Campbell <sup>b</sup>, Jon E. Hess <sup>c</sup>, Maureen A. Hess <sup>c</sup>, Todd W. Kassler <sup>a</sup>, Christine C. Kozfkay <sup>b</sup>, Andrew P. Matala <sup>c</sup>, Shawn R. Narum <sup>c</sup>, Melanie M. Paquin <sup>d</sup>, Maureen P. Small <sup>a</sup>, Jeff J. Stephenson <sup>c</sup>, Kenneth I. Warheit <sup>a</sup> & Paul Moran <sup>d</sup>

<sup>a</sup> Washington Department Fish and Wildlife, Genetics Section, 600 Capitol Way North, Olympia, Washington, 98501-1091, USA

<sup>b</sup> Idaho Department of Fish and Game, 1800 Trout Road, Eagle, Idaho, 83616, USA

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

<sup>d</sup> NOAA Fisheries Service, Northwest Fisheries Science Center, 2725 Montlake Boulevard East

Dataset Name	Blankenship et al 2011	Blankenship et al 2011	Blankenship et al 2011	Blankenship et al 2011	Blankenship et al 2011	Blankenship et al 2011	Blankenship et al 2011	Blankenship et al 2011	Blankenship et al 2011
EFGL Snake River basin Chinook salmon genetic baseline v3.0	<input type="checkbox"/>	Ninh Vu	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	3/28/2014 10:11 AM	151	6165	Snake River ba	<input type="checkbox"/>	<a href="#">Export</a> <a href="#">Delete</a>
Blankenship et al 2011	<input type="checkbox"/>	Matthew Campbell	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	1/30/2014 10:55 AM	339	15658	Blankenship, S.	<input type="checkbox"/>	<a href="#">Export</a> <a href="#">Delete</a>
SNP Test-Warm Springs Chinook 16jan14	<input type="checkbox"/>	Matthew Campbell	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	1/16/2014 2:39 PM	22	1823	Test Dataset fr	<input type="checkbox"/>	<a href="#">Export</a> <a href="#">Delete</a>
Hess et al 2014 GAPS subset	<input type="checkbox"/>	Matthew Campbell	Idaho Department of Fish and Game, Eagle Fish Genetics Lab, Eagle, ID	12/30/2013 5:00 PM	29	2025	This saved dat	<input type="checkbox"/>	<a href="#">Export</a> <a href="#">Delete</a>

# Saved Dataset: PBT baseline

- Will include
  - Samples from 2008 through 2016
  - 263 Collections
  - 155,000 Individuals
- Static text file that can be downloaded by anyone
- Ensures that everyone uses the same baseline data
  - Straying of hatchery stocks into natural spawning areas
  - Stock composition returning to Snake River basin
  - B-run hatchery stock composition in Zone 6
  - Hatchery origin of kelt steelhead at Lower Granite dam

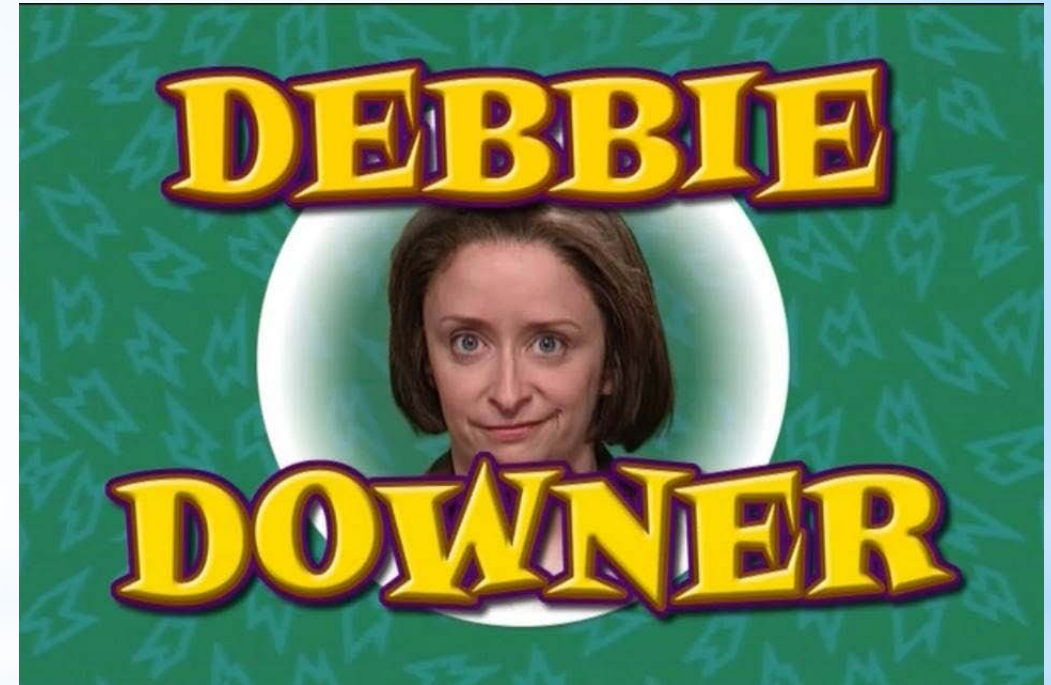
# Presentation Subjects

- Hagerman genetics program overview
- Genetic marker types
- Genotyping methods
- Hagerman database needs: Progeny Clinical
- Shared databases / FishGen
- **Data use in other databases**



# Data Sharing Problems

- Data transfers are difficult and I'm lazy
- Need a common unique field. Our Progeny ID doesn't always match the field ID
- Data may be meaningless without context
- Data may change with different analysis effort



# What type of data would be available

- **Genotypic sex**
- **Genotypic age (parentage results)**
- **Hatchery of origin (PBT or GSI)**
- **Non-Neutral markers developed from PoolSeq**
  - Run timing
  - Growth rate/size
  - Thermal tolerance
  - Age at maturity
  - Any heritable trait (with enough effort).

# Data Sharing Suggestions

- **Talk to a geneticist about specific needs**
- **Choose a common unique field ahead of time.**
  - **Pre-printed whatman sheets with ID's**
- **Treat everything on a case by case basis**