Seascape Genetic Analysis of Chinook Salmon in the California Current Reveals Distinct Marine Distributions Among Stocks

West Coast Salmon Genetic Stock Identification Project

California: **John Carlos Garza**, Eric Crandall, Sarah Bates
Oregon: Pete Lawson, Renee Bellinger, Michael Banks

NOAA Fisheries
National Marine Fisheries Service

Salmon spend most of their lives in the Ocean

-Chinook salmon (*O. tshawytscha*) are born in freshwater but, in California, most migrate to the sea several months after emergence (subyearling migrants)

-They then spend 2-5 years at sea feeding and growing before returning to freshwater to spawn

-Much more difficult to study them in the ocean, both because of spatial scale, but also because of lack of ability to discriminate fish from different populations

-Recovery requires a better understanding of ocean distribution and migration.

Sacramento River Fall Chinook "Collapse"

-Chinook salmon escapement in 2008 and 2009 was lowest on record

-Fisheries completely closed for the first time in history

-Consensus explanation is that this was due to oceanographic anomalies-temperature and upwelling-off California in critical period for entering juveniles

-While cause primarily ocean, hatchery homogenization of life history (e.g. outmigration timing) meant larger effect on CV fall run.
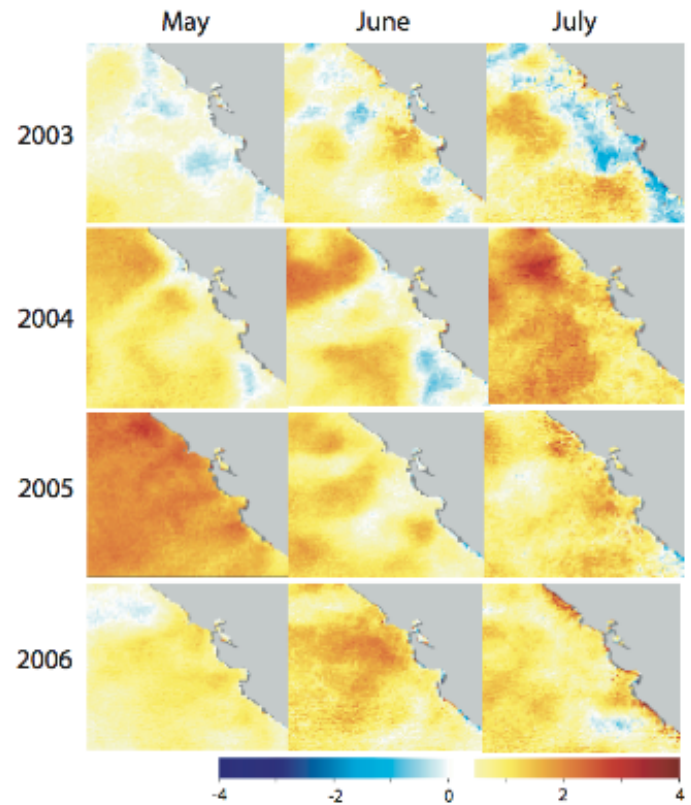
Sea surface temperature off Central California
Lindley et al. 2009

Genetic Stock Identification in Chinook Salmon

-Chinook salmon (*O. tshawytscha*) range from California, USA to Russia

-At least 200 genetically distinct populations

-Genetically distinct populations are NOT visually distinguishable

-Historically huge fishery for salmon in California ocean. Currently $100-200M/year. Over $1B/year in entire Pacific Ocean

-Mixed fishery: fish from many populations (stocks) comingle in the ocean

D. Pearse

Mixed Stock Salmon Fisheries

•Fish from multiple stocks co-distributed in the ocean, but largely isolated when reproducing in freshwater

•Some stocks rare and protected, others are abundant and can be harvested

•In California fisheries the main stocks are: Central Valley Fall, Klamath River, Coastal California & Central Valley Spring, usually in that order of abundance.

Inference in Mixed Fisheries
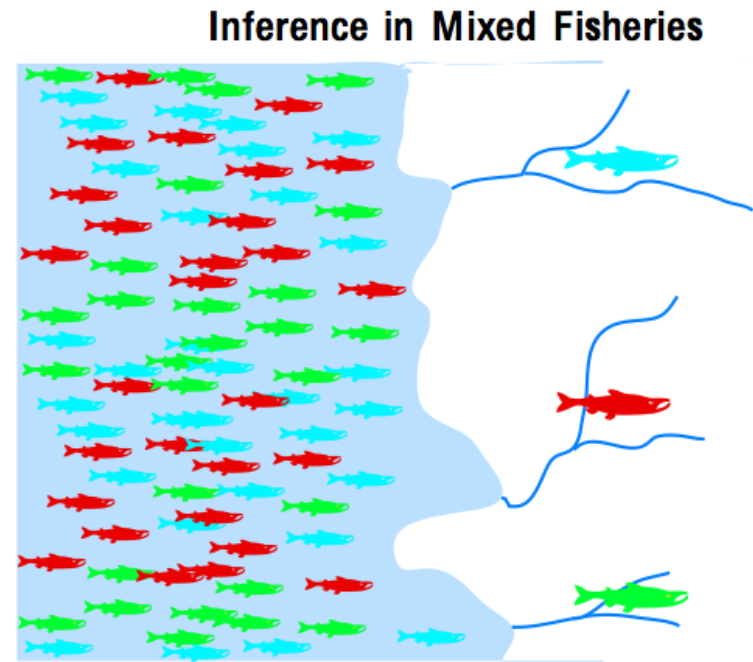
Genetic Stock Identification in Chinook Salmon

-Commercial fisheries in California heavily restricted in 2006-2007, closed 1st time ever in 2008-2009: 1st Klamath River & California Coastal stocks, then Central Valley constraint

-Genetic stock identification can help fishery managers by defining management units

-Genetic Stock Identification (GSI) uses a baseline reference database with genotypes of many fish of known origin to identify an unknown fish of population of origin

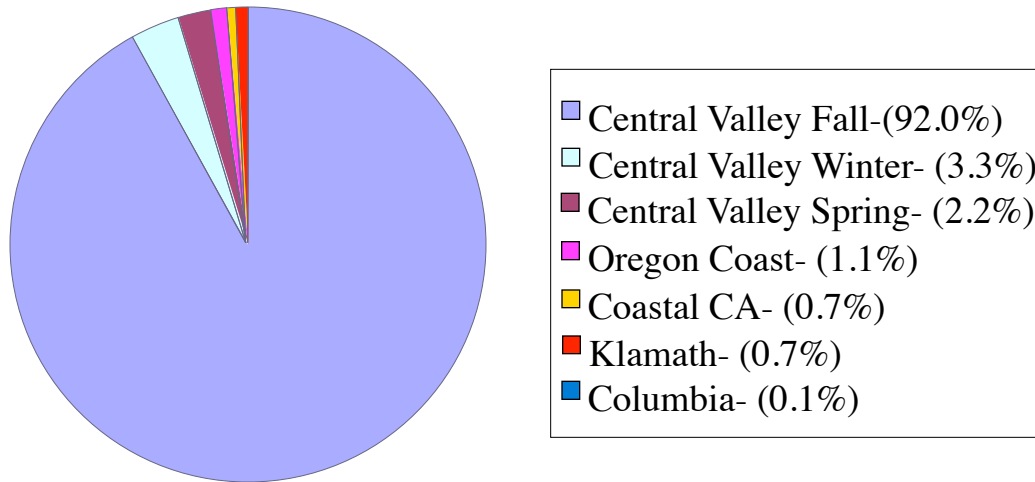-Baseline database of 96 SNPs for 8,000 fish covering all salmon stocks found at more than 0.1% in California fishery.

California Chinook Salmon Genetic Stock ID

Monterey Bay Sport Fishery-2006

- Central Valley Fall-(92.0%)
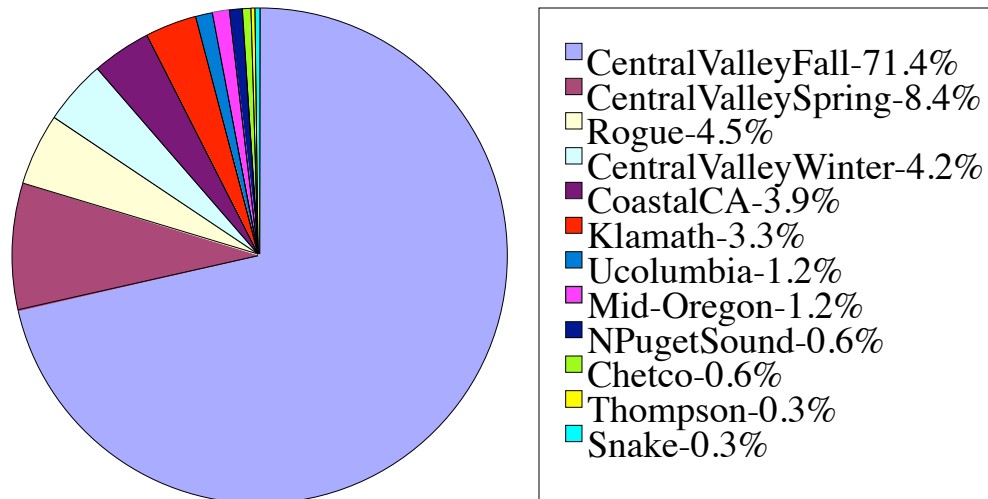- Central Valley Winter- (3.3%)
- Central Valley Spring- (2.2%)
- Oregon Coast- (1.1%)
- Coastal CA- (0.7%)
- Klamath- (0.7%)
- Columbia- (0.1%)

Monterey Bay Sport Fishery-2006

April 2006
Cumulative total:
N=735 fish

Monterey Bay Sport Fishery-2007

April-Aug 2007
Cumulative total:
N=340 fish

- CentralValleyFall-71.4%
- CentralValleySpring-8.4%
- Rogue-4.5%
- CentralValleyWinter-4.2%
- CoastalCA-3.9%
- Klamath-3.3%
- Ucolumbia-1.2%
- Mid-Oregon-1.2%
- NPugetSound-0.6%
- Chetco-0.6%
- Thompson-0.3%
- Snake-0.3%
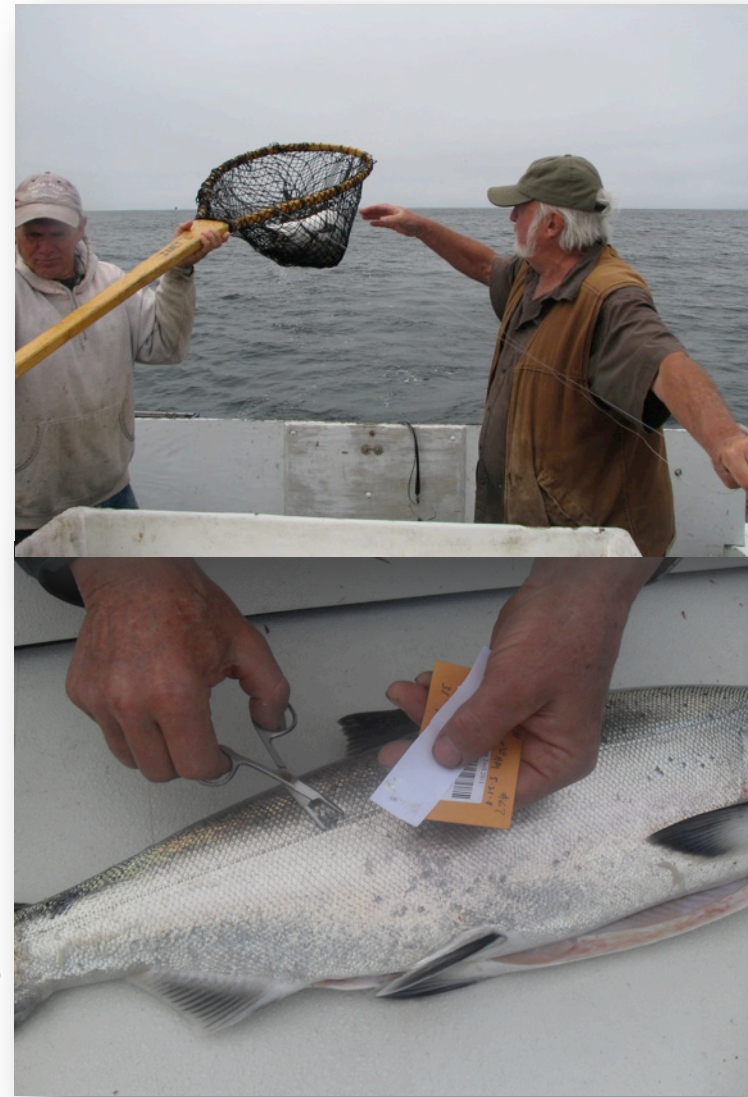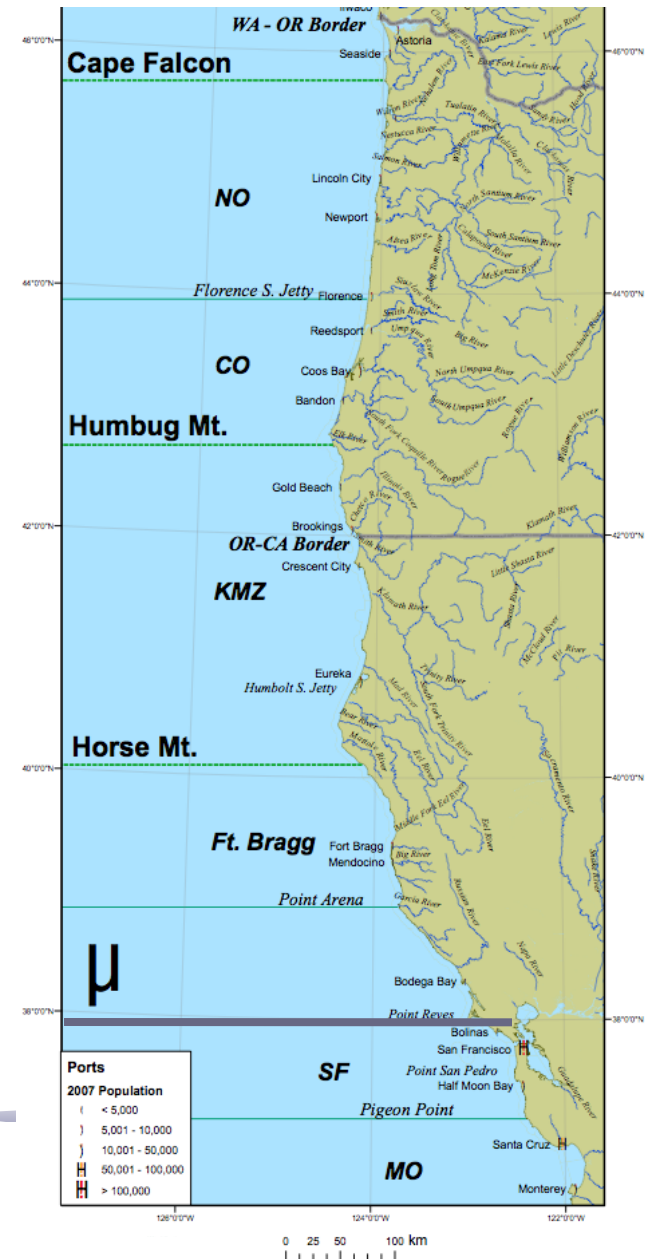
West Coast Genetic Stock Identification Collaboration

-Collaboration between scientists and commercial salmon fishing fleet began as pilot in 2007

-Unique opportunity to characterize ocean distribution and migration patterns of Chinook salmon in the California Current Marine Ecosystem

-Large scale non-retention sampling of fish stratified in time (May-September) and space (Cape Falcon-Monterey)

-Exact GPS coordinates of catch locations and tracklines of all fishing effort.

-Genetic stock identification and scale age analysis for all fish.

West Coast Genetic Stock Identification Collaboration

-Collaboration between scientists and the commercial salmon fishing fleet begun in 2007

-Unique opportunity to characterize ocean distribution and migration patterns of Chinook salmon in the California Current Marine Ecosystem

-Large scale non-retention sampling of fish stratified in time (May-September) and space (Cape Falcon-Monterey)

-Exact GPS coordinates of catch locations and tracklines of all fishing effort.

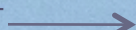-Genetic stock identification and scale age analysis for all fish.
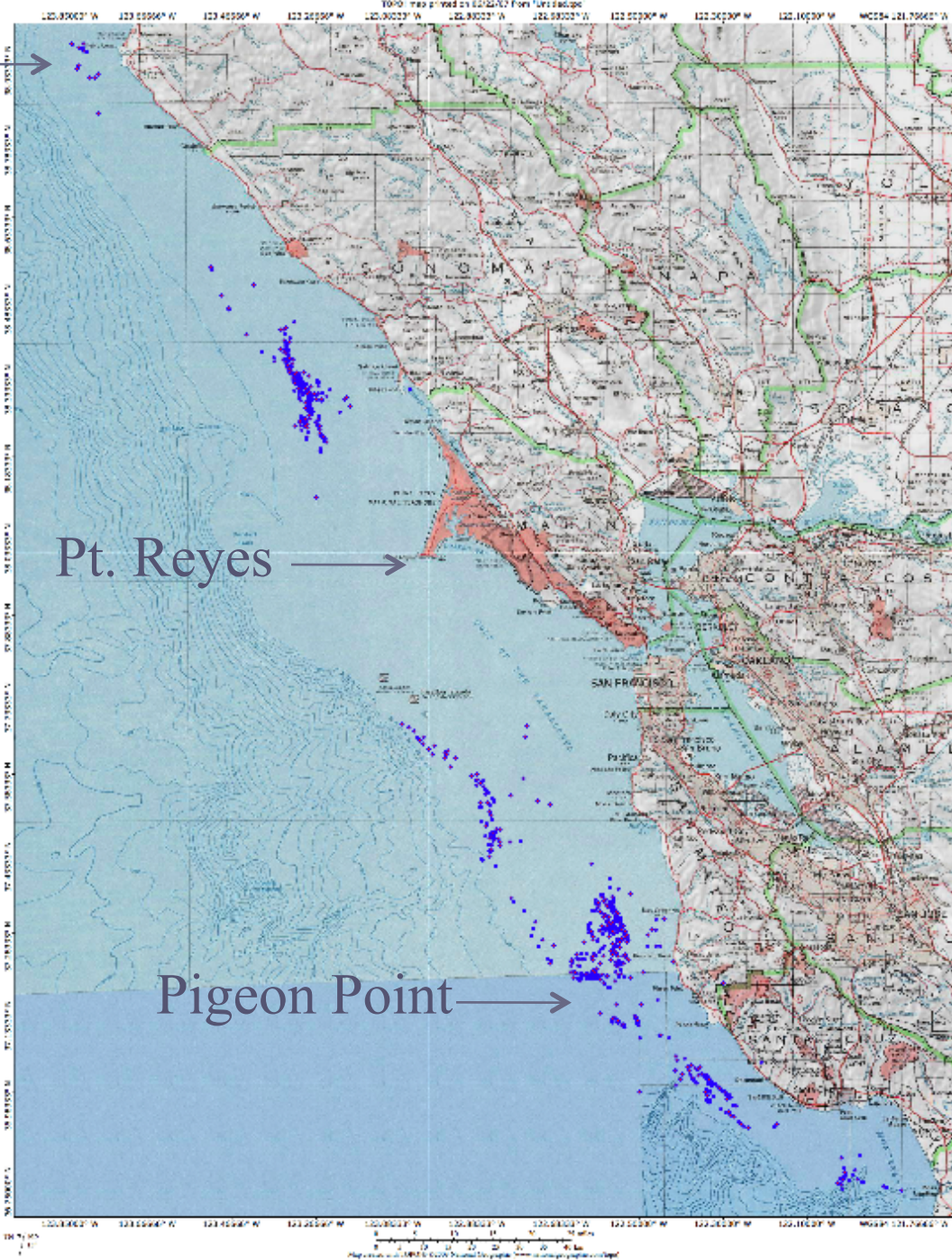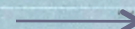
Catch
locations-CA
GSI Project
May 2007

N=1075

Pt. Arena

Pt. Reyes

Pigeon Point

Stock proportions by latitudinal region-May commercial fishery

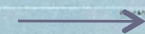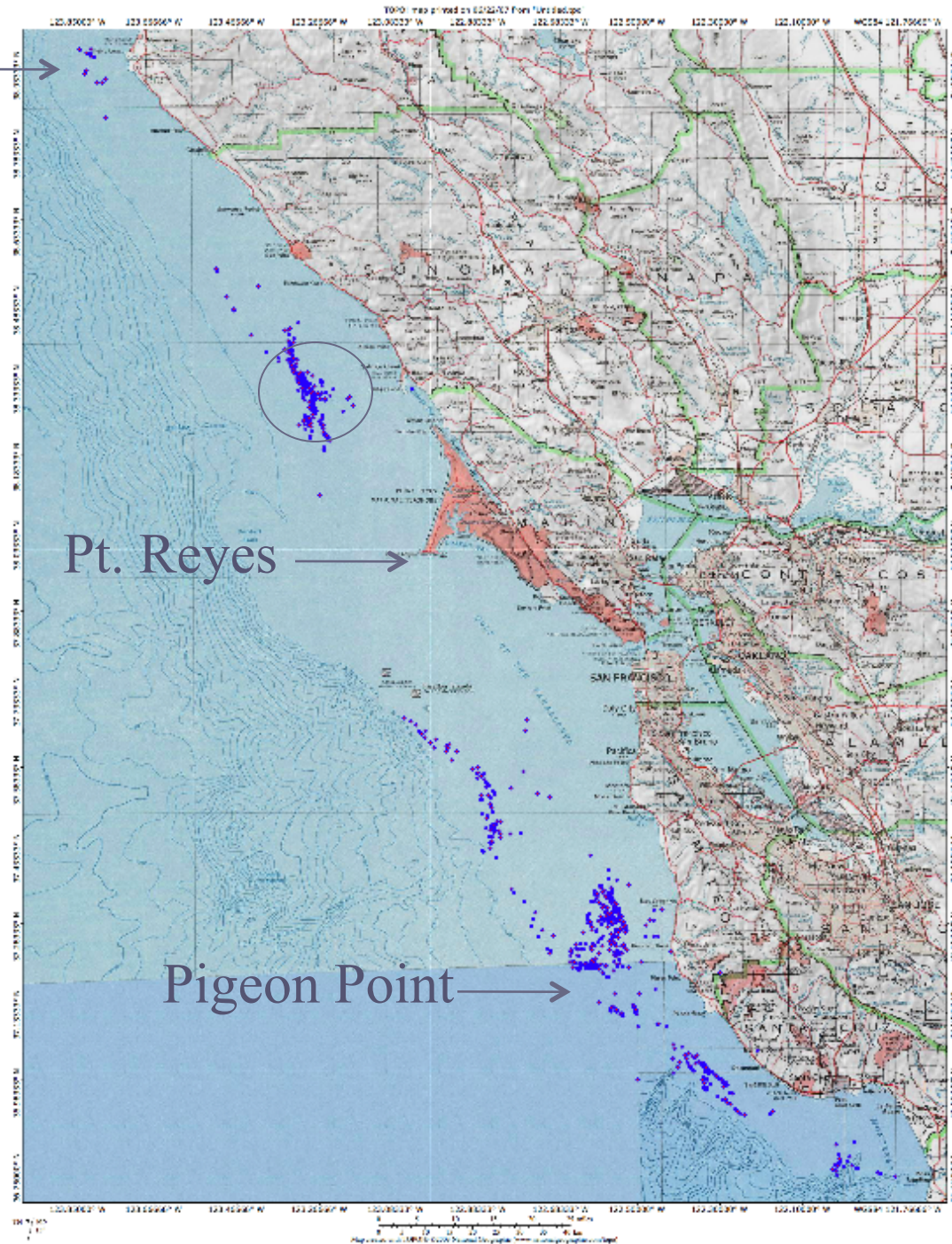|	Overall (N=1068)	Pt Arena-Pt Reyes (N=422)	Pigeon Pt-Pt Reyes (N=460)	South of Pigeon Pt (N=186)
Coastal CA	12.65	22.27	6.96	5.38
Klamath	9.21	12.80	7.83	4.30

Catch
locations-CA
GSI Project
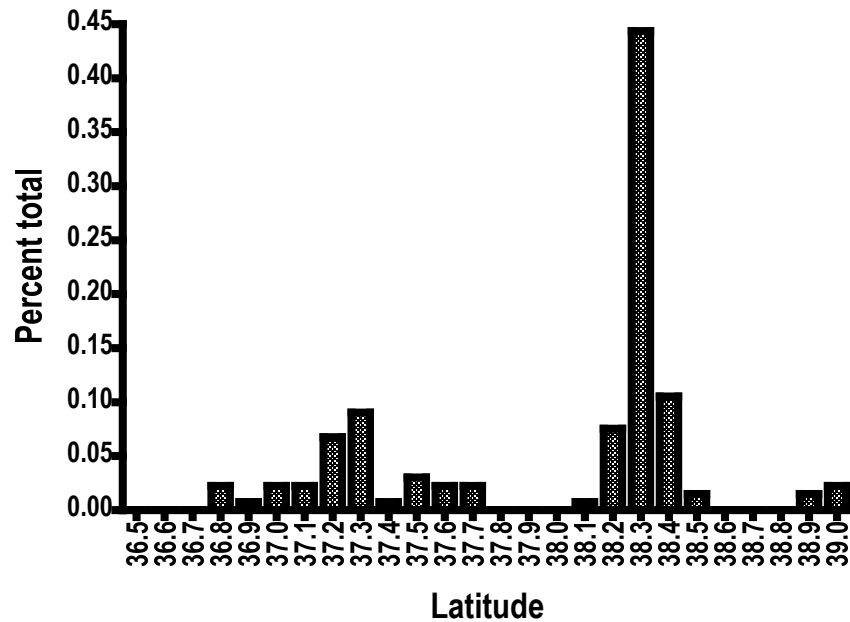May 2007

N=1075

Pt. Arena

Pt. Reyes

Pigeon Point

CoastalChinookbyLatitude

Comparison of latitudinal distribution of California Coastal conservation stock and overall catch distribution

AllChinookbyLatitude
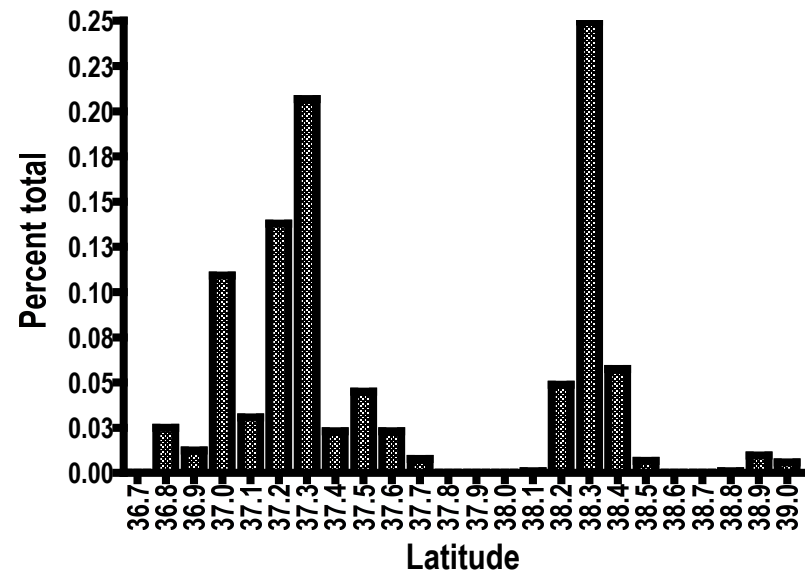
West Coast Salmon Genetic Stock Identification Project

-Coastwide sampling for fishery areas managed by Pacific Fishery Management Council
-Spatial and oceanographic patterns of stock abundance.

Year	N	Effort Hours
2010	6,514	14,507
2011	8,129	9,058
2012	11,607	10,521
Total	26,250	34,086

West Coast GSI - May - September 2010: Effort and Samples

Monthly differences in stock proportions-July & August 2011

Eureka n = 398
July

- Rogue n=155
- Central Valley Fall n=102
- Klamath n=59
- California Coast n=29
- Chetco n=28
- Upper Columbia Summer Fall n=11
- Snake River Fall n=7
- Central Valley Spring n=2
- Columbia River - Spring Creek n=2
- Lower Columbia Fall n=1
- Mid Oregon Coast n=1
- Washington Coast n=1

38.94%, 25.63%, 14.82%, 7.29%, 7.04%, 2.76%, 1.76%, 5 stocks < 1%

Eureka n = 48
August

- Klamath n=18
- Rogue n=17
- Central Valley Fall n=7
- Snake River Fall n=3
- California Coast n=1
- Mid Oregon Coast n=1
- Chetco n=1

37.5%, 35.42%, 14.58%, 6.25%, 2.08%, 2.08%, 2.08%

Fort Bragg n = 2060
July

- Central Valley Fall n=779
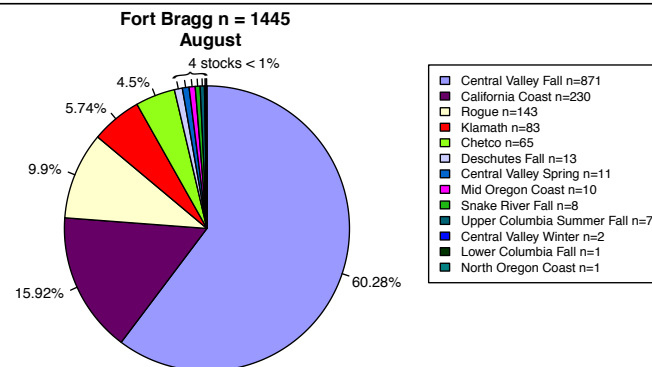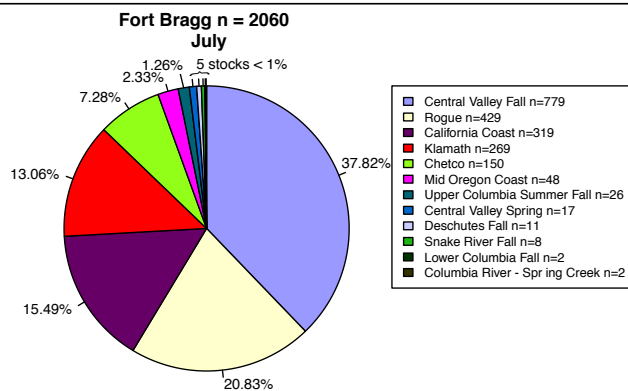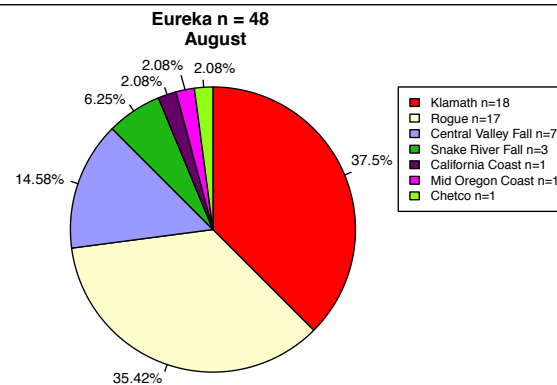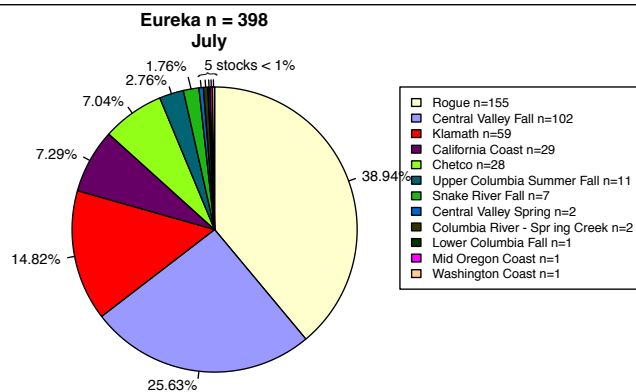- Rogue n=429
- California Coast n=319
- Klamath n=269
- Chetco n=150
- Mid Oregon Coast n=48
- Upper Columbia Summer Fall n=26
- Central Valley Spring n=17
- Deschutes Fall n=11
- Snake River Fall n=8
- Lower Columbia Fall n=2
- Columbia River - Spring Creek n=2

37.82%, 20.83%, 15.49%, 13.06%, 7.28%, 2.33%, 1.26%, 5 stocks < 1%

Fort Bragg n = 1445
August

- Central Valley Fall n=871
- California Coast n=230
- Rogue n=143
- Klamath n=83
- Chetco n=65
- Deschutes Fall n=13
- Central Valley Spring n=11
- Mid Oregon Coast n=10
- Snake River Fall n=8
- Upper Columbia Summer Fall n=7
- Central Valley Winter n=2
- Lower Columbia Fall n=1
- North Oregon Coast n=1

60.28%, 15.92%, 9.9%, 5.74%, 4.5%, 4 stocks < 1%

Bodega Bay n = 856
July

- Central Valley Fall n=679
- Rogue n=66
- California Coast n=38
- Chetco n=19
- Klamath n=18
- Mid Oregon Coast n=8
- Upper Columbia Summer Fall n=7
- Deschutes Fall n=6
- Central Valley Spring n=5
- Snake River Fall n=5
- Lower Columbia Fall n=3
- Central Valley Winter n=2

79.32%, 7.71%, 4.44%, 2.22%, 2.1%, 7 stocks < 1%

Bodega Bay n = 131
August

- Central Valley Fall n=130
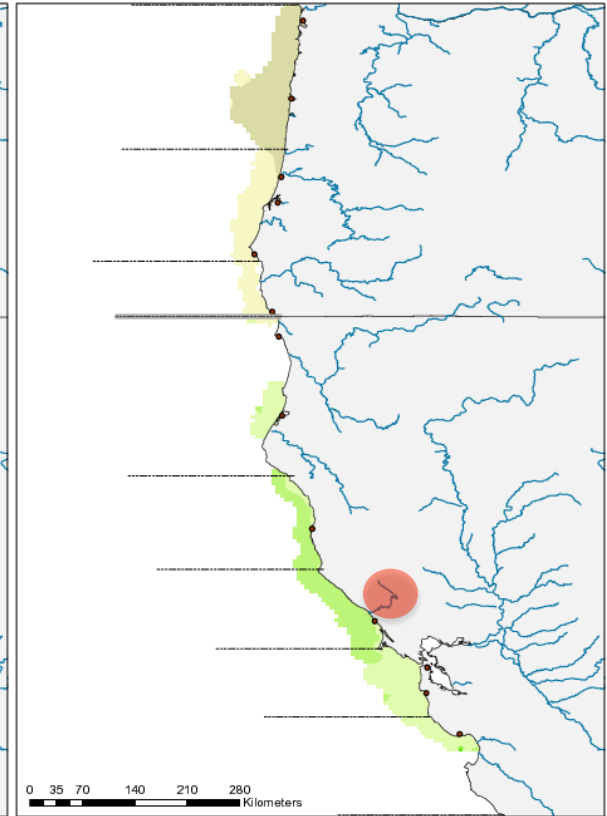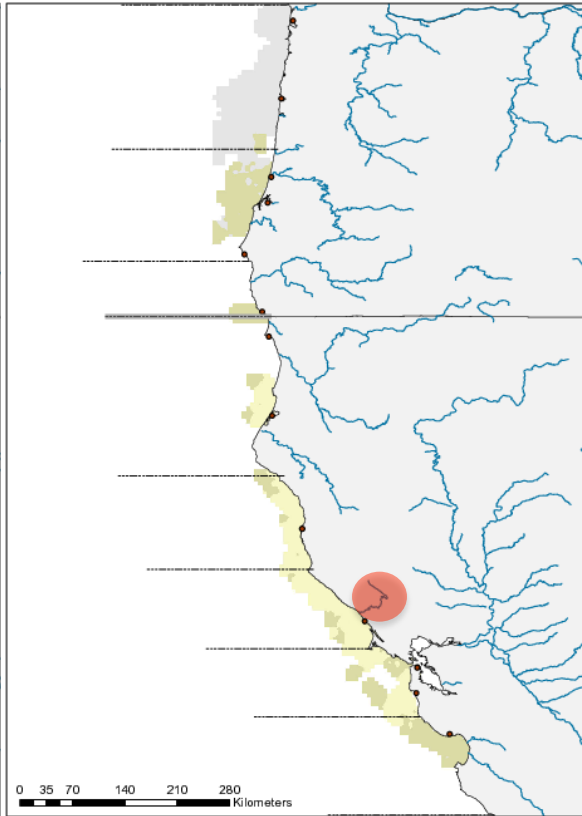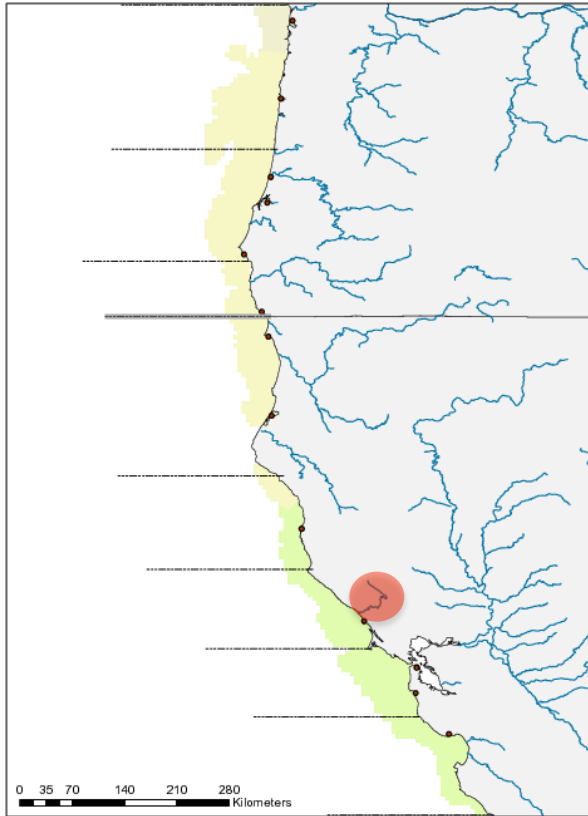- Rogue n=1

99.24%, 0.76%

Maximum Entropy Model Results – Per Hour Catch Probability

Russian River 2010

Russian River 2011
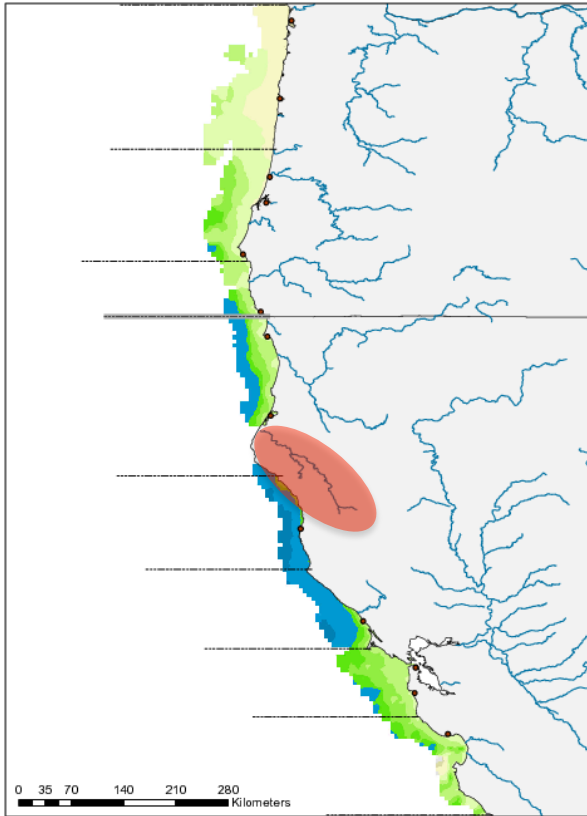
Russian River 2012

0	0.0001 - 0.001	0.01 - 0.025
0 - 0.0001	0.001 - 0.01	0.025 - 0.05

0.05 - 0.075   0.1 - 0.25   0.5 - 0.75
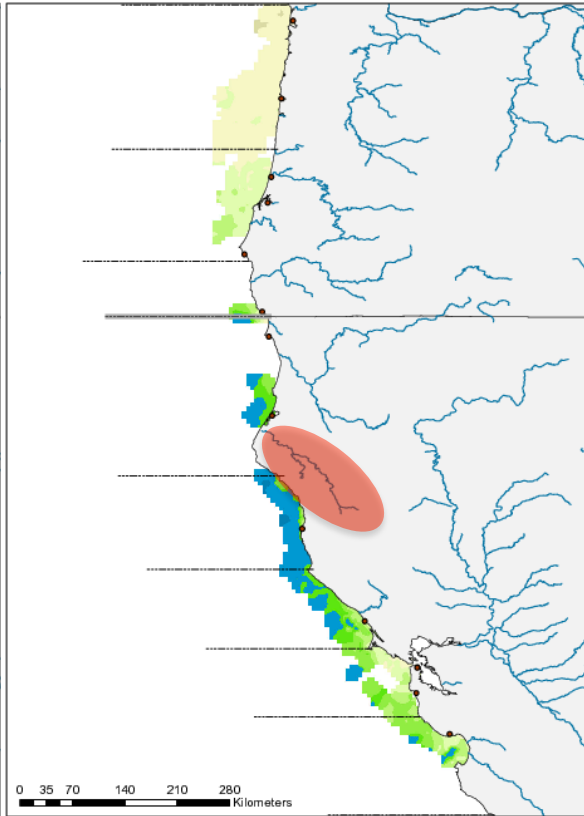0.075 - 0.1   0.25 - 0.5   0.75 - 1

Maximum Entropy Model Results – Per Hour Catch Probability

Eel River 2010   Eel River 2011   Eel River 2012

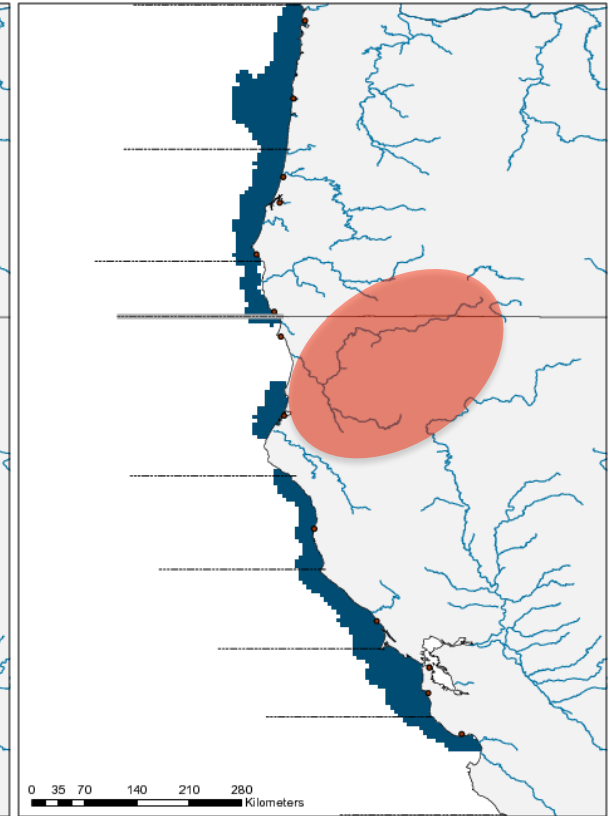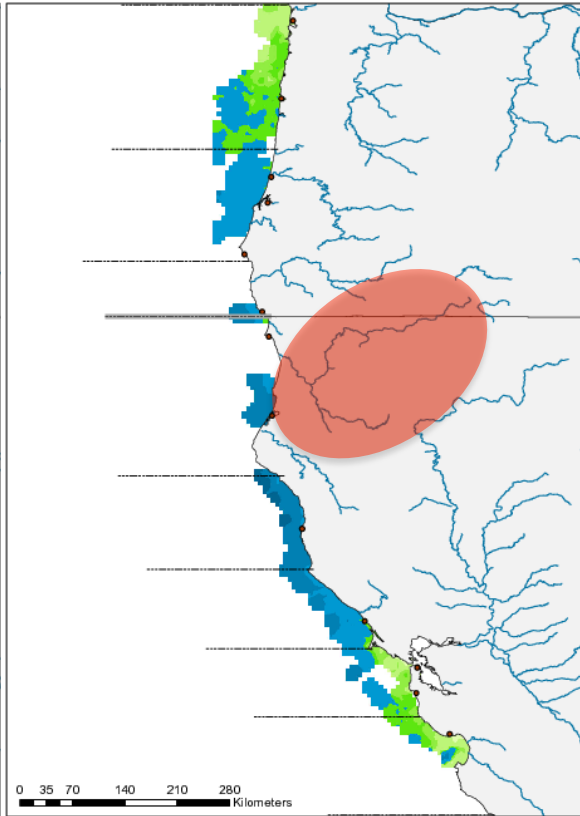| Color	Value
	0
	0 - 0.0001
	0.0001 - 0.001
	0.001 - 0.01
	0.01 - 0.025
	0.025 - 0.05
	0.05 - 0.075
	0.075 - 0.1
	0.1 - 0.25
	0.25 - 0.5
	0.5 - 0.75
	0.75 - 1
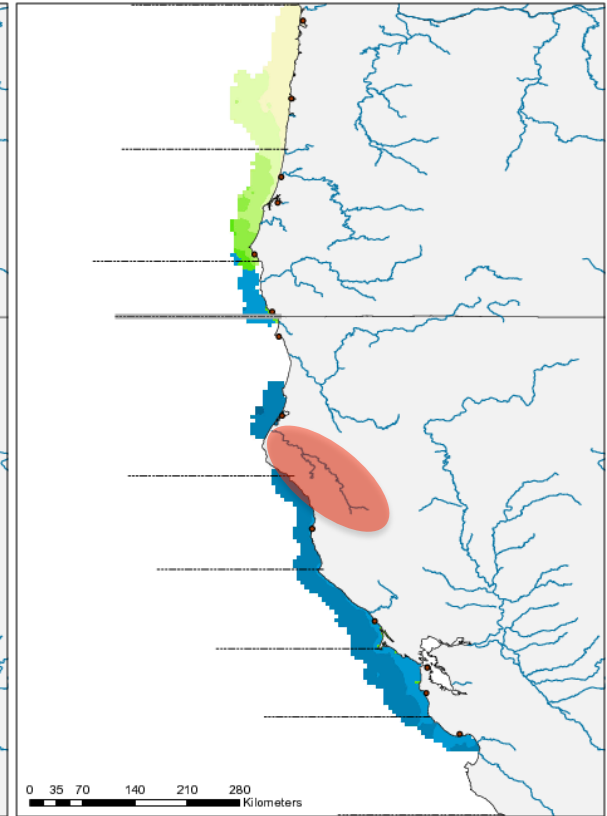
Maximum Entropy Model Results – Per Hour Catch Probability

Klamath River 2010   Klamath River 2011   Klamath River 2012

0	0.0001 - 0.001
0 - 0.0001	0.001 - 0.01
0.01 - 0.025	0.05 - 0.075
0.025 - 0.05	0.075 - 0.1
0.1 - 0.25	0.5 - 0.75
0.25 - 0.5	0.75 - 1
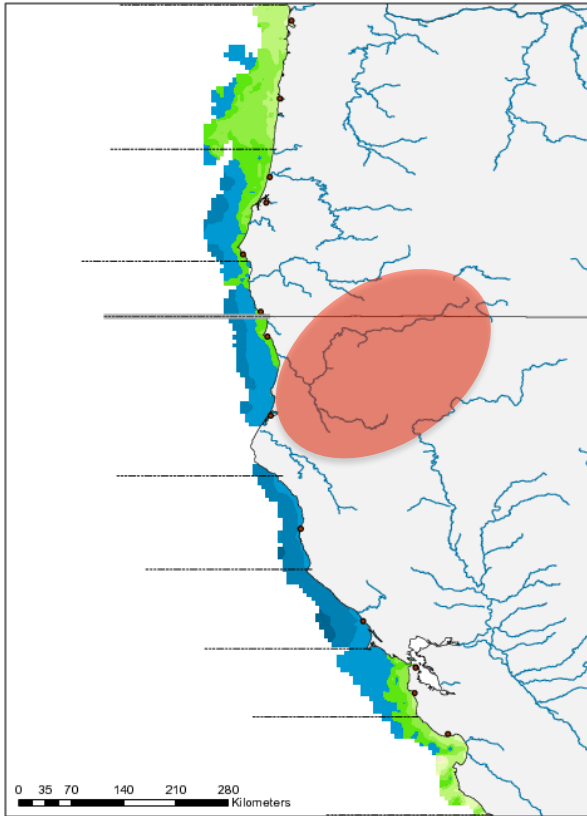
West Coast Salmon Genetic Stock Identification Project

2010 results

-Fish color coded by stock

-Spatial and oceanographic patterns examined
- •Simple aggregation
- •Ocean depth
- •Catch depth
- •Sea Surface Temperature
- •Productivity

West Coast Salmon Genetic Stock Identification Project
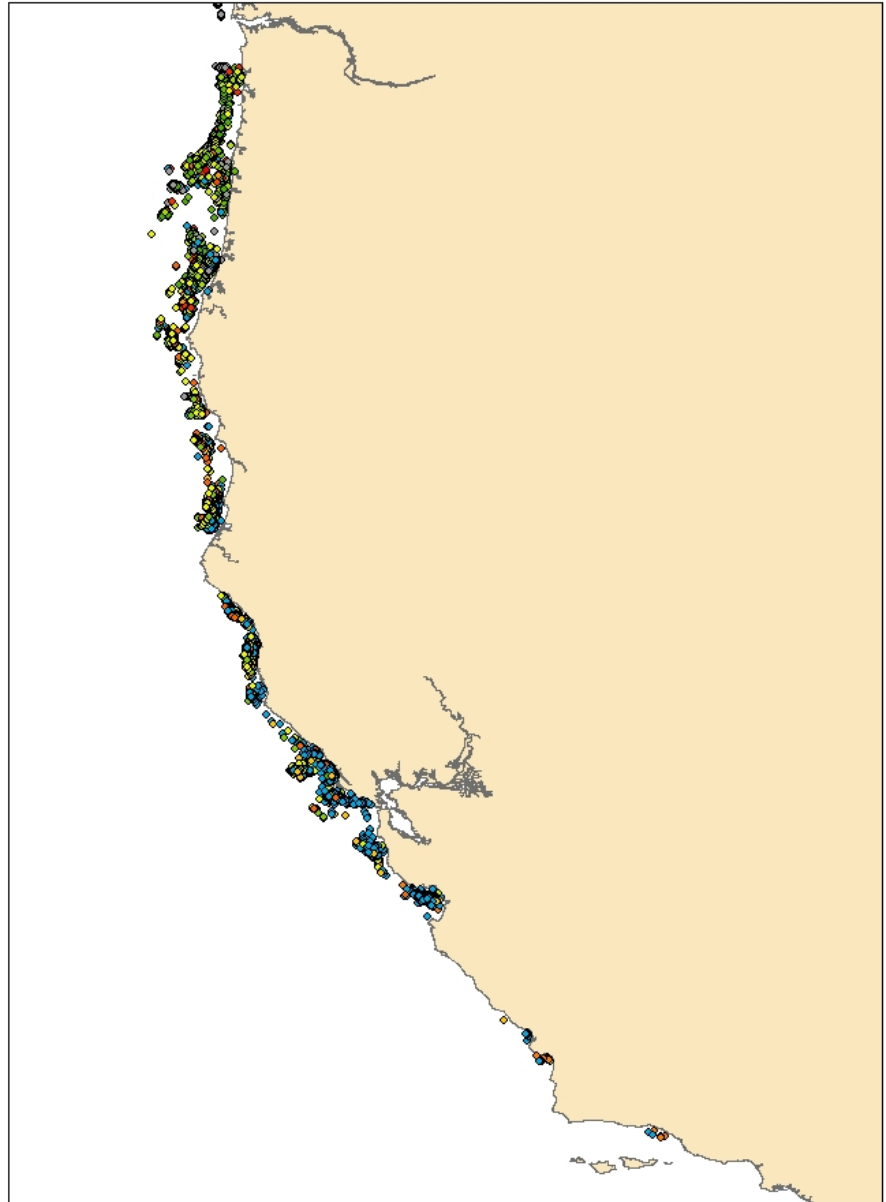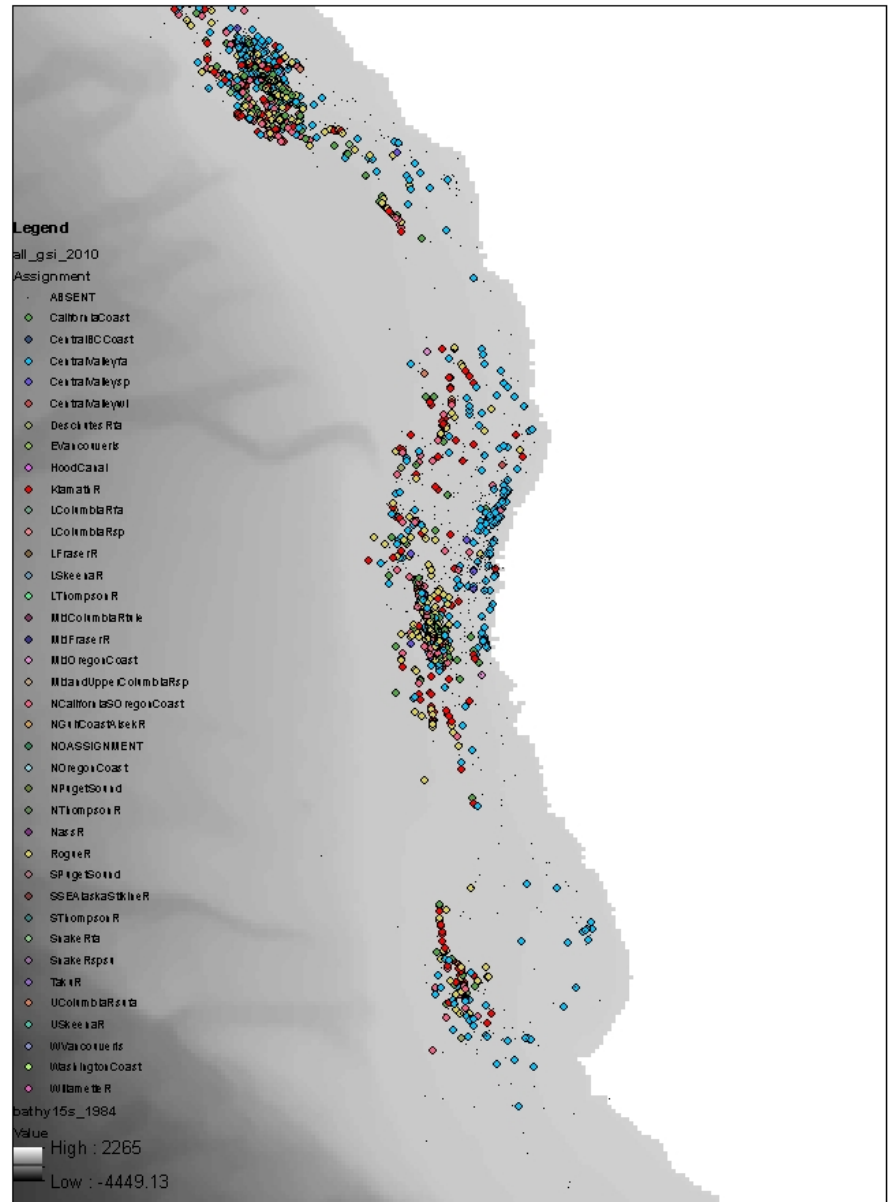
2010 results

-Ft. Bragg Management Area

-Significant spatial pattern of aggregation by stock found

-Strong correlated with ocean depth-Central Valley Fall and Spring shallower.

Legend
all_gsi_2010
Assignment
- · ABSENT
- ◇ CaliforniaCoast
- ◆ CentralBCCoast
- ◇ CentralValleyfa
- ◆ CentralValleysp
- ◆ CentralValleywi
- ◇ Deschutes Rfa
- ◇ EVancouverIs
- ◇ HoodCanal
- ◆ Klamath R
- ◇ LColumbiaRfa
- ◇ LColumbiaRsp
- ◆ LFraser rR
- ◇ LSkeena R
- ◆ LThompson R
- ◆ MidColumbiaRbrite
- ◆ MidFraser R
- ◇ MidOregonCoast
- ◇ MidandUpperColumbiaRsp
- ◇ NCaliforniaSOregonCoast
- ◇ NGulfCoastAlsekR
- ◇ NOASSIGNMENT
- ◇ NOregonCoast
- ◇ NPugetSound
- ◇ NThompson R
- ◆ Nass R
- ◇ Rogue R
- ◇ SPugetSound
- ◆ SSEAlaskaStikineR
- ◇ SThompson R
- ◇ Snake Rfa
- ◇ Snake Rspsu
- ◇ Takit R
- ◇ UColumbiaRsusfa
- ◇ USkeena R
- ◆ WVancouverIs
- ◇ WashingtonCoast
- ◇ Willamette R

bathy15s_1984
Value
High : 2265

Low : -4449.13

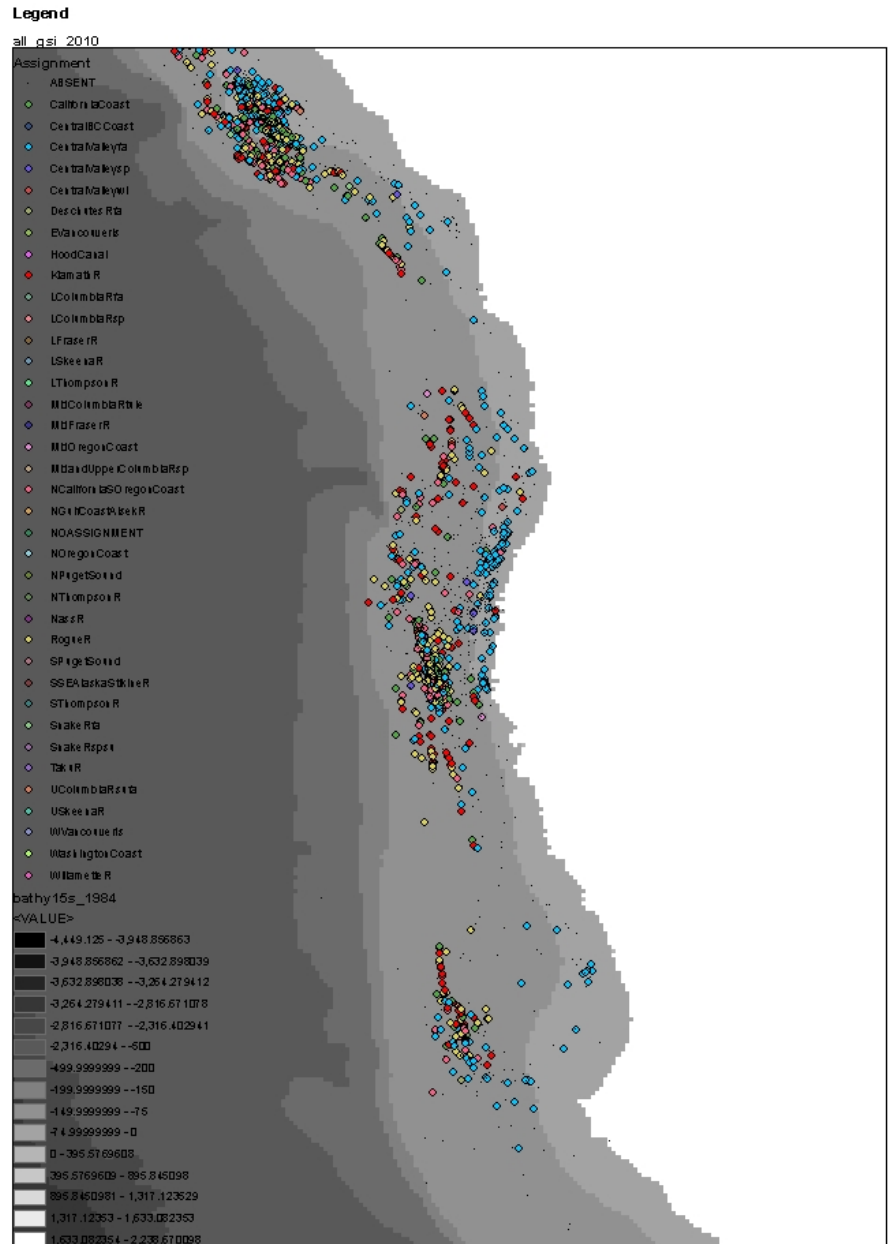# West Coast Salmon Genetic Stock Identification Project

2010 results

-Ft. Bragg Management Area

-Significant spatial pattern of aggregation by stock found

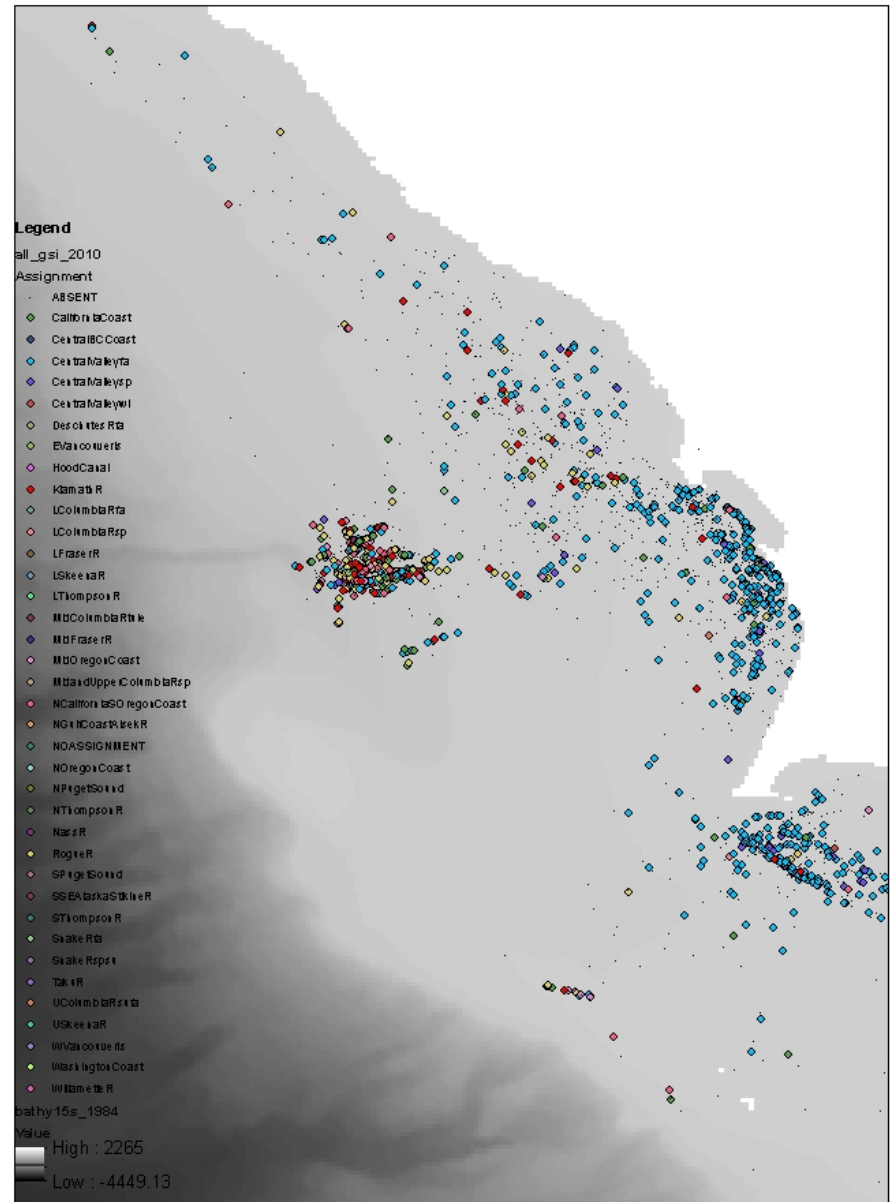-Strong correlated with ocean depth-Central Valley Fall and Spring shallower.

West Coast Salmon Genetic Stock Identification Project

2010 results

-San Francisco Management Area

-Significant spatial pattern of aggregation by stock found

-Strongly correlated with ocean depth-Central Valley Fall and Spring shallower

-May be associated with different species of krill.

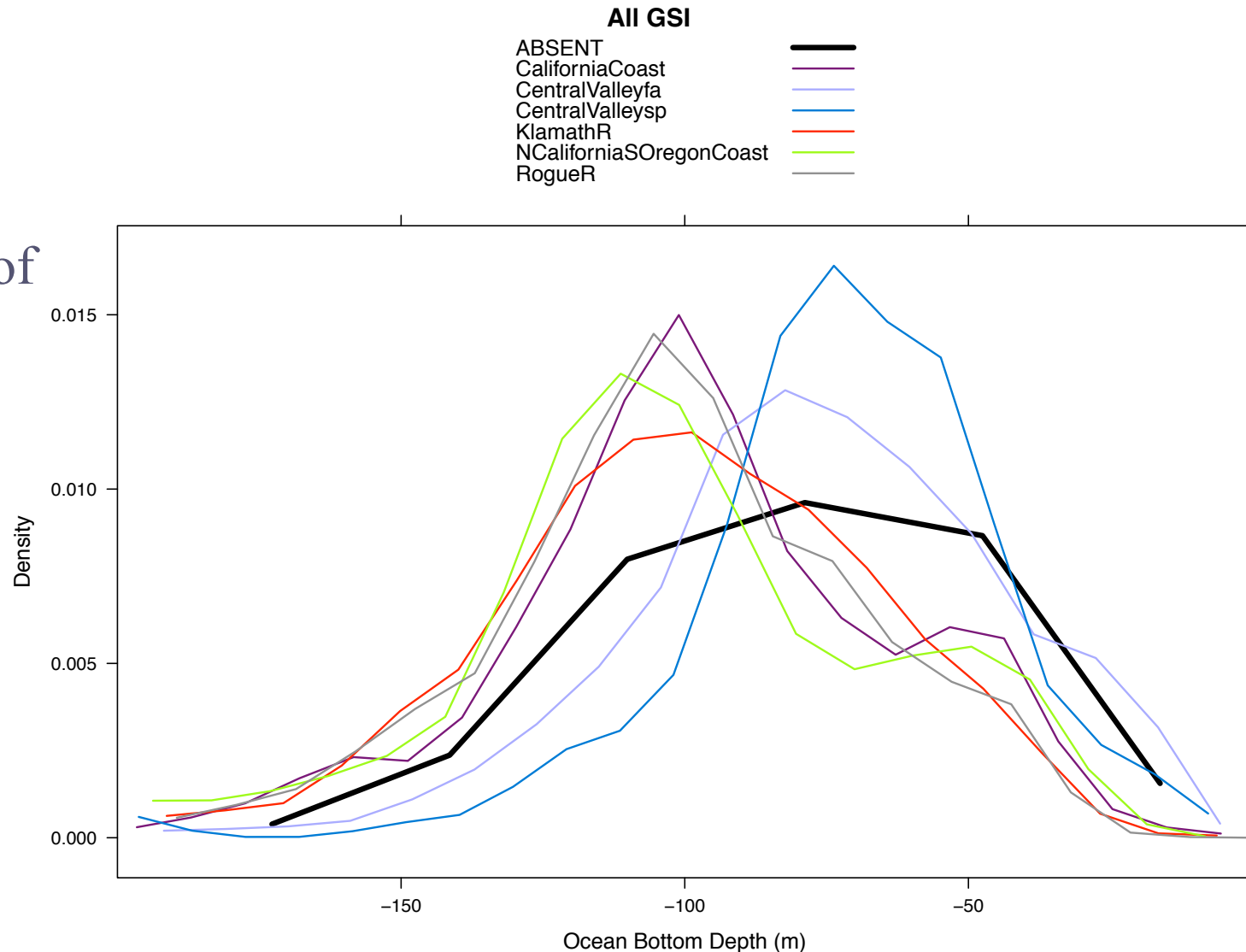# West Coast Salmon Genetic Stock Identification Project
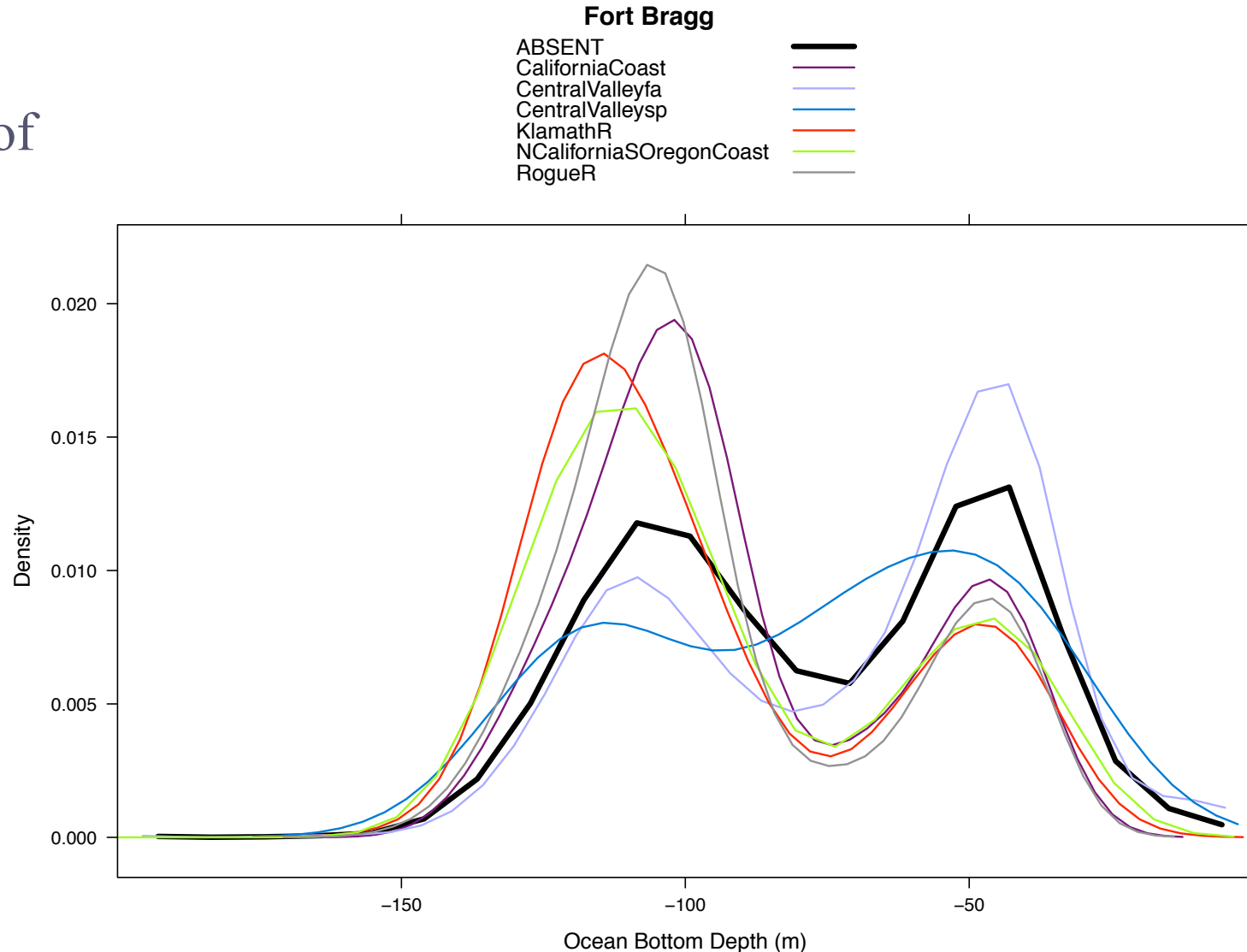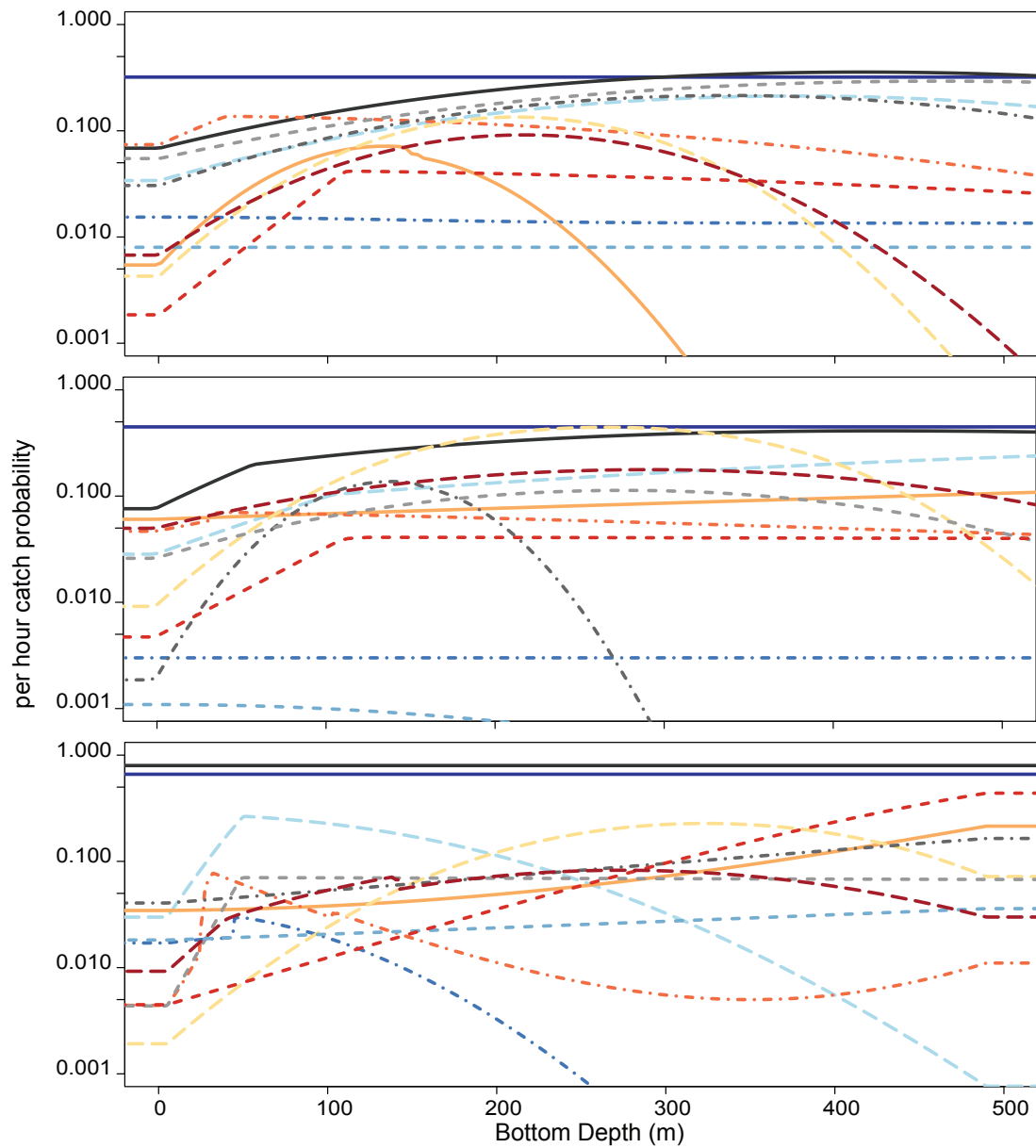# 2010 results

-Coastwide (California and Oregon)

-Significant spatial pattern of aggregation by stock found

-Strongly correlated with ocean depth

-Not correlated with age (preliminary).

West Coast Salmon Genetic Stock Identification Project
2010 results

-Fort Bragg

-Significant spatial pattern of aggregation by stock found

-Strongly correlated with ocean depth

-Not correlated with age.

Legend:
- Central Valley Fall
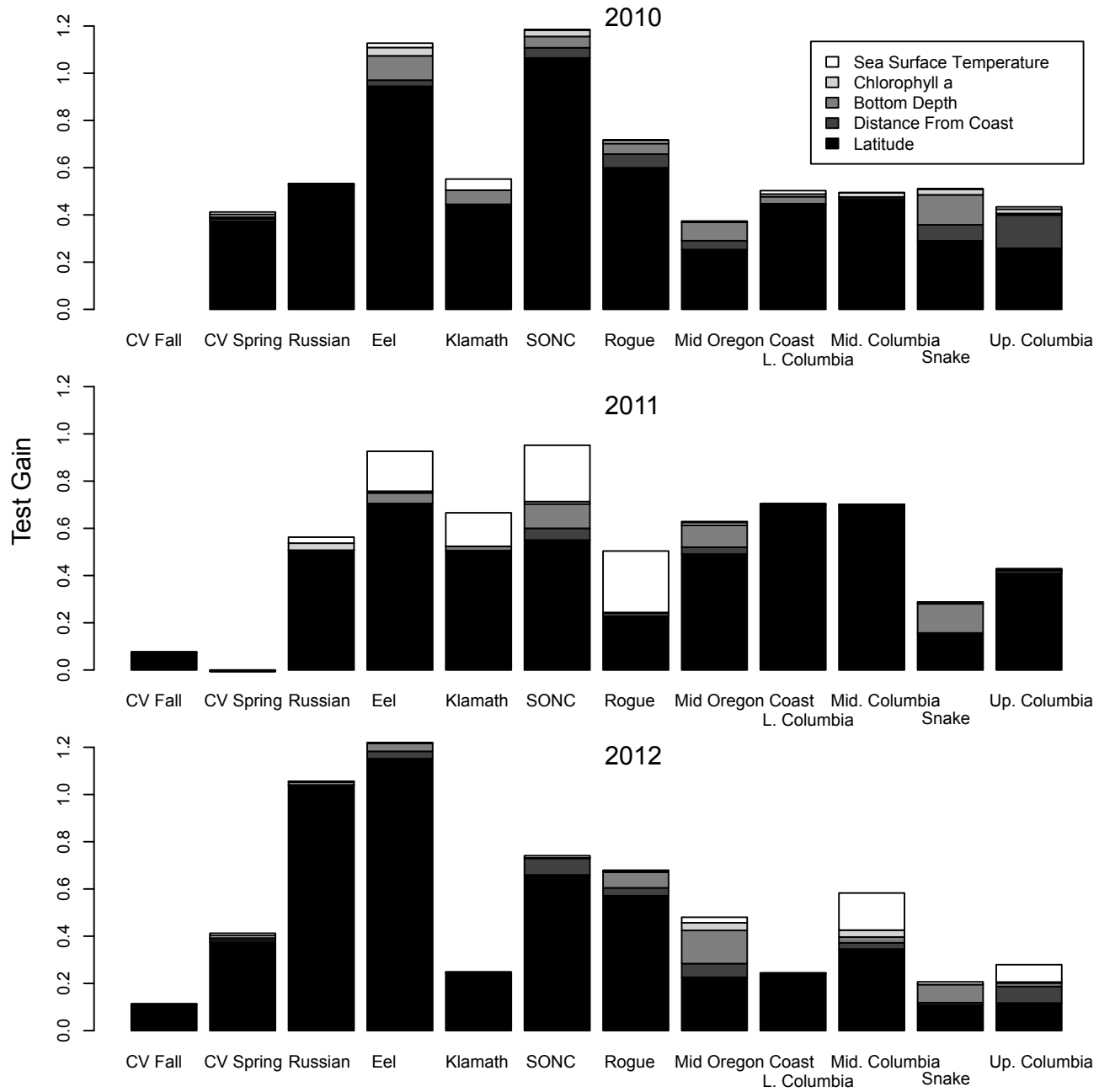- Central Valley Spring
- Russian
- Eel
- Klamath
- SONC
- Rogue
- Mid Oregon Coast
- Lower Columbia Fall
- Columbia River - Tule
- Snake
- Upper Columbia Summer Fall

Maximum Entropy Model Covariate Contributions

Chinook Salmon Conclusions

-North/south gradient in abundance for Coastal and Klamath River Chinook (conservation) stocks apparent

-Most conservation status fish caught in 15km latitudinal zone in San Francisco Management Area in 2007

-Clustering of Coastal Chinook and Klamath stocks could lead to finer scale management areas: San Francisco Management Area could be cut in half

-Significant association of non-Central Valley stocks with ocean bottom depth- found in deeper water.

-Opportunity to increase both harvest and progress on conservation objectives.

Acknowledgements

SWFSC/UC Santa Cruz

- Eric Anderson
- Vanessa Apkenas
- Anthony Clemento
- Libby Gilbert-Horvath
- Devon Pearse

NWFSC/Oregon State Univ.

- Gil Sylvia
- Jonathan Minch
- Nancy Fitzpatrick

California Salmon Council

- David Goldenberg
- Jim Anderson
- Many, many fishermen.

Most salmonid tagging currently employs coded wire tags

- Mechanically inserted and manually extracted metal tags that are manually read under a microscope

- Since 1968, at least 71 agencies in 5 states and B.C. have inserted ~600 miles of wire and tagged ~ 1 billion salmon and steelhead

- Until 1996, only fish with CWTs generally received adipose fin clips

- Nearly 1 million heads analysed at Juneau head lab alone.

Images from www.nmt.us

Current CWT tagging system

- Very useful tagging system over its 30+ year life

- Crucial to management objective of estimating multiple stock fishery mortality

- Provides stock of origin AND cohort of origin

- Large historical databases of tag recoveries provide comparative baseline

Challenges to CWT system

- Very low tag recovery rates (1.6 per 1,000 in chinook)

- Tag loss rates are poorly known

- CWT harvest may be underreported

- Mass-marking - Not all Ad-clipped fish have CWTs

- Assumption of equivalency of hatchery indicator stocks and genetically similar naturally spawning stocks (the gorilla assumption) - can be large areas.

Emerging framework for intergenerational genetic tagging of salmonids

- Tagging crucial to fishery management, population monitoring, ecological investigation, hatchery evaluation, etc.

- Most tagging involves physical manipulation of juveniles

- Tag recovery rates are dismally low (<<1%), because of high cumulative mortality

- Intergenerational tagging would be a huge efficiency, since adults could be tagged and transmit tags to their offspring, removing need for inefficient manipulation of juveniles and consequent tag loss.

Parentage-based Tagging

– Highly efficient, intergenerational, genetic tagging method

- Genotype (preferably) all parents from a hatchery or natural population
- Create reference (parent) database of all possible parent pairs
- Fishery/ecological sampling and genotyping in offspring generation
- Query of reference (parent) database to determine if parents are present
- Determine parental pair and, therefore, hatchery stock of origin and exact age

– Information obtained for each tag recovery is the same as for a CWT (+more)

– By genotyping two parents, you effectively tag all of their 1,000s of offspring

– Requires no juvenile tagging, but MUCH higher tagging rates feasible.

Parentage-based Tagging

- Fundamentally different than genetic stock identification: matches fishery sample to pairs of parents in reference (parent) database that have Mendelian compatiblity. GSI uses frequency-based probability assessments

- Can be done using either traditional exclusion or maximum likelihood

- Power comes from number of loci, since each locus is an opportunity for incompatibility

- Marking and sampling issues with other tagging systems don't entirely go away

- Do not need to record matings or store tissue samples separately: day buckets.

Power analysis for large scale parentage-inference

- Anderson and Garza (2006; Genetics) evaluated the plausibility of large scale PBT through evaluation of power of SNP markers to infer parentage

  - Determine false positive rates in large-scale parentage inference studies
  - Evaluate number of SNP loci necessary to correctly ID parent pairs
  - Describe new analytical method for fast ML parentage analysis
  - Evaluate effects of allele frequency, genotyping error and presence of kin

- 100 SNP genotype can identify parental pairs with false positive rate less than 1 fish per 300,000 fishery samples.

- False positive rates decrease exponentially with number of loci!

Advantages of SNPs for large scale parentage-inference

- Low genotyping error rate

- Allele calls (nomenclature) are easily standardized between labs

- Minimal human interaction with the raw genotyping data

- Amenable to high-throughput / low-cost genotyping- new technology: material costs for 100 SNPs similar to 12-18 microsatellites, and labor costs are MUCH lower.

Exponential decline in false positive rate with number of SNP loci

P = 0.2   GtypErrorRate = 0.01   POWER = 0.9
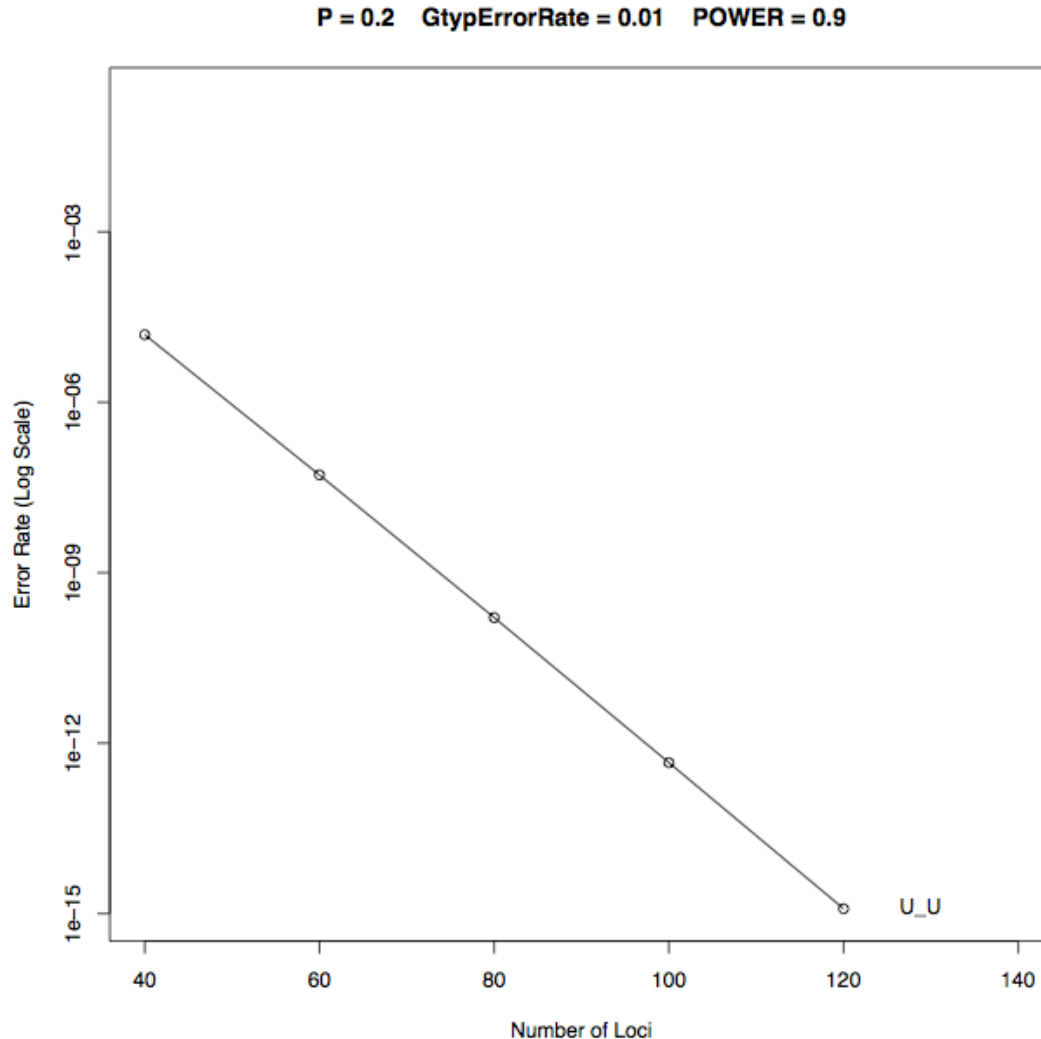
Per-trio false positive rates
80 loci = 4.6 x 10^{-10}
100 loci = 4.5 x 10^{-13}

Genotypes as DNA Fingerprints

96 single nucleotide polymorphisms (SNPs) in separate genes

Paternal C,A,G,C,T,T,C,T,A......T,A,G,A,T,C,A,C,T
Maternal C,A,A,T,T,G,C,A,A......T,C,C,A,G,C,A,T,T

– Transform the way that we use population genetic data from allele frequency-based framework to an individual DNA fingerprint framework.

– First line of analysis now considers all genotypes separately, as potential direct matches (recaptures) or as nodes in pedigrees involving other genotypes in the database(s).

– Various options for second line of analysis: roll unidentified genotypes into frequency-based analyses, such as standard genetic stock idenfication.
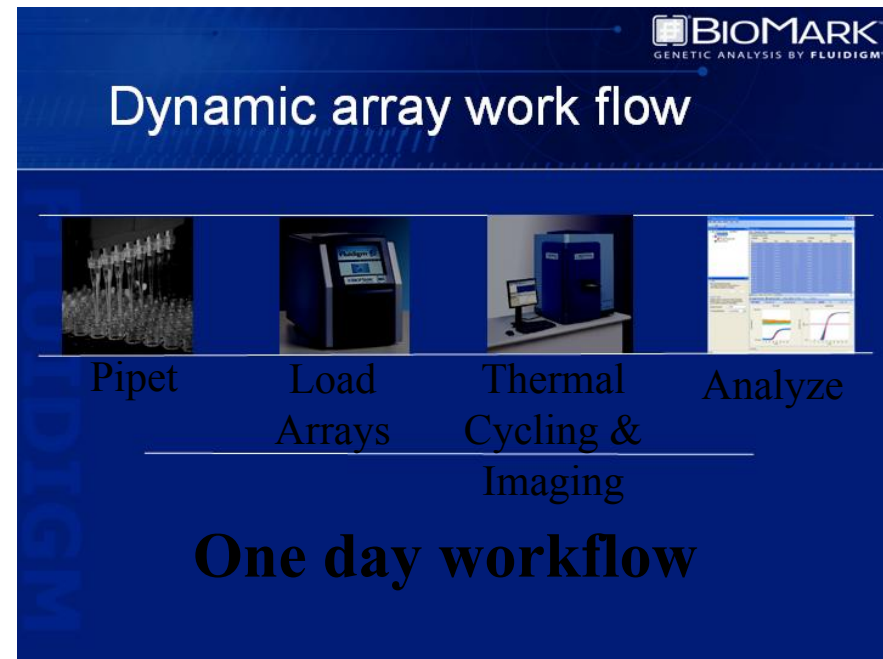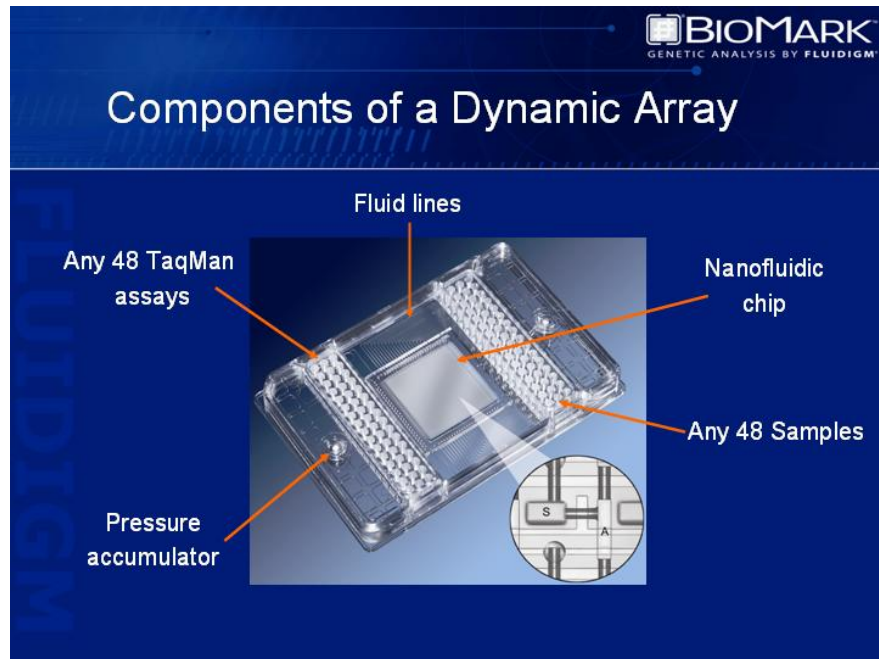
Genotypes as DNA Fingerprints

96 single nucleotide polymorphisms (SNPs) in separate genes

Paternal C,A,G,C,T,T,C,T,A......T,A,G,A,T,C,A,C,T
Maternal C,A,A,T,T,G,C,A,A......T,C,C,A,G,C,A,T,T

– Transform the way that we use population genetic data from allele frequency-based framework to an individual DNA fingerprint framework.

– First line of analysis now considers all genotypes separately, as potential direct matches (recaptures) or as nodes in pedigrees involving other genotypes in the database(s).

– Various options for second line of analysis: roll unidentified genotypes into frequency-based analyses, such as standard genetic stock idenfication.

High throughput SNP Genotyping

With 96.96 Arrays and only one controller and thermal cycler, can genotype almost 300 fish per day w/96 SNPs!
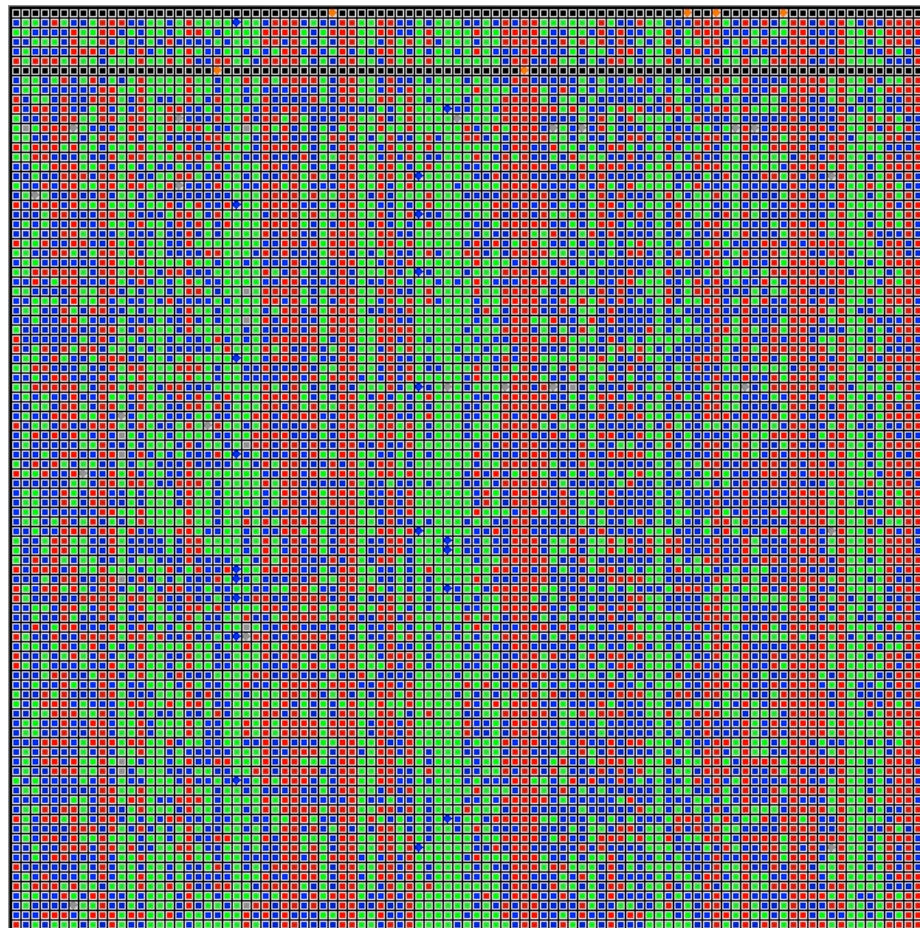
Materials cost: <$10/fish

Digital DNA Fingerprints
Fluidigm EP-1 w/ 96.96 dynamic arrays

Loci

Individual

Steelhead
from Russian
River

Chinook genetic tags are here!!!
Power analysis of 96 SNP loci

False Negative Rate	False Positive Rate		
	0.1	0.05	0.01
ButteSp	3.80E-13	2.46E-12	4.95E-11
MillSp	1.37E-13	7.65E-13	1.78E-11
DeerSp	1.45E-13	8.59E-13	1.77E-11
FRHsp	1.04E-13	5.95E-13	1.45E-11
CVFall	1.74E-13	1.03E-12	2.32E-11
UpSacLFall	1.82E-13	1.15E-12	2.38E-11
SacWinter	2.99E-12	1.50E-11	3.02E-10
EelVA	2.99E-12	1.57E-11	3.62E-10
KlamIGH	1.50E-12	8.29E-12	1.83E-10
TrinityH	2.30E-12	1.36E-11	2.80E-10
RogueSp	3.60E-13	2.28E-12	4.85E-11
Chetco	3.02E-13	1.59E-12	4.25E-11
Umpqua	1.20E-13	6.97E-13	1.72E-11
Kalama	2.36E-13	1.40E-12	2.95E-11
Cowlitz	2.79E-13	1.61E-12	4.24E-11
Snake	2.57E-13	1.47E-12	3.09E-11

New standardized genetic tags provide more than sufficient power for PBT in the Central Valley, Klamath and Columbia.

Chinook genetic tags are here!!!
Power analysis of 96 SNP loci

Interbasin Assignment Accuracy

FeatherSP	1
ButteSP	0.99
Mill/DeerSP	1
CV Fall/late Fall	0.97
Winter	1
CoastalCA	0.98
Klamath/Trinity	0.99
Oregon Coast	0.96
Lower Columbia	0.95
Upper Columbia	1
Fraser	1
Situk	1

The same genetic tags also provide accuracy of assignment to basin of more than 99% overall. Many fish misassigned to out of basin stocks were missing data.

Parentage-based tagging in salmonids
Where are we now?

Steelhead/rainbow trout tags too!!!
Power analysis of 96 SNP loci

False Negative Rate	False Positive Rate		
	0.1	0.05	0.01
BattleCreek	2.01E-13	1.20E-12	2.74E-11
Trinity	4.07E-12	2.32E-11	4.19E-10
Russian	2.31E-13	1.37E-12	3.38E-11
ScottCr-SC	2.06E-13	1.22E-12	3.22E-11
Eel	1.09E-12	6.41E-12	1.29E-10
Kamloops	7.33E-10	3.74E-09	5.53E-08
MtWhitney	8.28E-12	4.40E-11	8.73E-10
Eagle	6.96E-11	3.57E-10	8.25E-09
N.Zealand-Taupo	2.34E-12	1.09E-11	2.03E-10
Willamette	4.93E-09	2.34E-08	3.48E-07
Patagonia-SC	1.04E-11	5.52E-11	1.15E-09

New standardized genetic tags provide more than sufficient power for PBT with common CA steelhead and RBT stocks.

Applications of parentage-based tagging with salmon and steelhead

In addition to tagging data, PBT gives you large pedigrees

- Near parametric estimates of variance in family size

- Conduct large quantitative genetic studies of phenotype: run timing, age at maturity, disease resistance

- Map genes for phenotypic traits to locations in the genome

- Evaluate different hatchery/release practices and consequences for fecundity, marine survival and straying

- Estimate straying and reproductive success of strays

- Study relative productivity of hatchery and natural fish by sampling at weirs, fish ladders and carcasses (carefully).

PBT in Chinook salmon

Validation of PBT for Chinook salmon started in 2006 at Feather River Hatchery

- Spring Chinook program chosen: 1/5 size of Fall Chinook program and 100% CWTs
- Assign 2010 broodstock (N=1,741) to parents: 2006-2008
- Assess False Discovery Rates.
- Assess accuracy with ocean fishery caught Chinook salmon

Estimation finds that 1234 of 1741 fish assigned with FDR of 1/200. Given proportion of broodstock sampled, this is consistent with expectation.
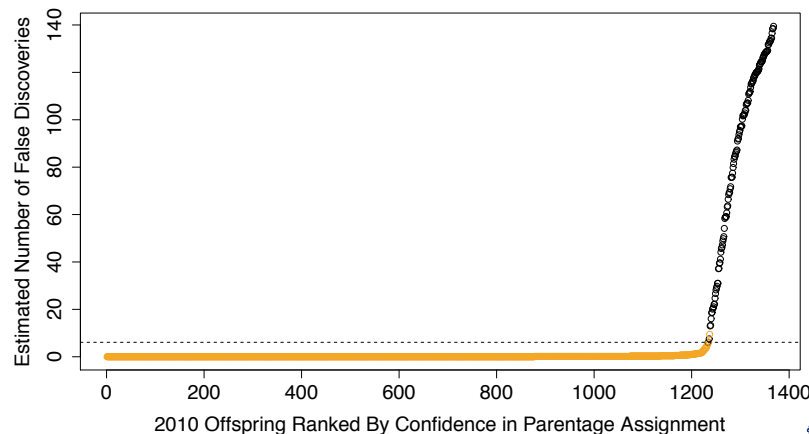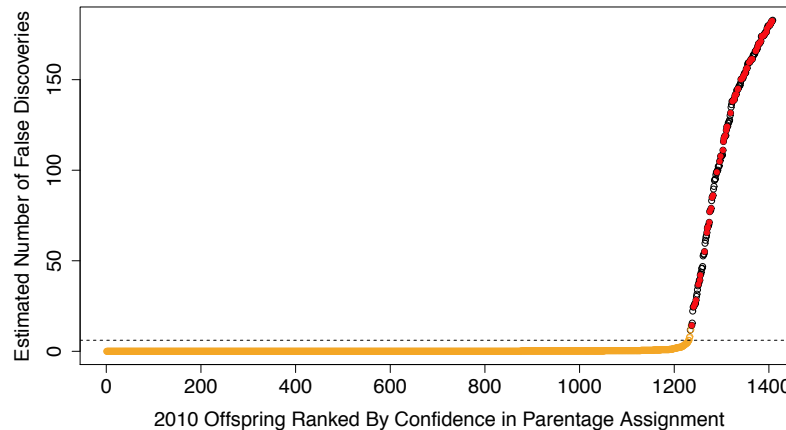
PBT in Chinook salmon

Validation of PBT for Chinook salmon started in 2006 at Feather River Hatchery

• Spring Chinook program chosen: 1/5 size of Fall Chinook program and 100% CWTs
• Assign 2010 broodstock (N=1,741) to parents: 2006-2008
• Assess False Discovery Rates.
• Assess accuracy with ocean fishery caught Chinook salmon

"Decoy" samples, 2009 broodstock, indicate that FDR of 1/200 is highly conservative.

Assignments in red involve 1 or 2 "decoy" samples as parents.

PBT on mixed-stock, ocean-fishery Chinook salmon samples

Chinook salmon port-sampled from California commerical and sport fisheries

- 1,806 of these had tissue samples suitable for genotyping.
- 968 of these carried CWTs to give them stock of origin
- 56 of these CWTs are FRH spring run, which is tagged at a rate of 100%.
- 41 of these were identified by PBT using our FRH samples.

Note: given age distribution and sampling fractions, we expect to assign parentage to 44.3 out of 56 fish from FRH Spring. 41 is not significantly different than 44.3 (binomial distribution), which implies false negative rate is near 0.

PBT on mixed-stock, ocean-fishery Chinook salmon samples

We can use PBT to assess how accurate the CWTs are for FRH Spring:

- 1806 fish genotyped and compared to FRH Spring broodstock
- 51 of these have parents in FRH Spring
- 41 of these 51 have CWTs indicating FRH Spring
- 8 of these 51 were sampled by state, but no CWTs from the fish.
- 2 of these 51 had CWTs that were read to be from
    - i) Coleman National Fish Hatchery Late Fall
    - ii) Feather River Hatchery Fall.

8/51 ≈ 16% CWT loss / failure-to-recover rate!
2/43 ≈ 5% CWT misreading rate!