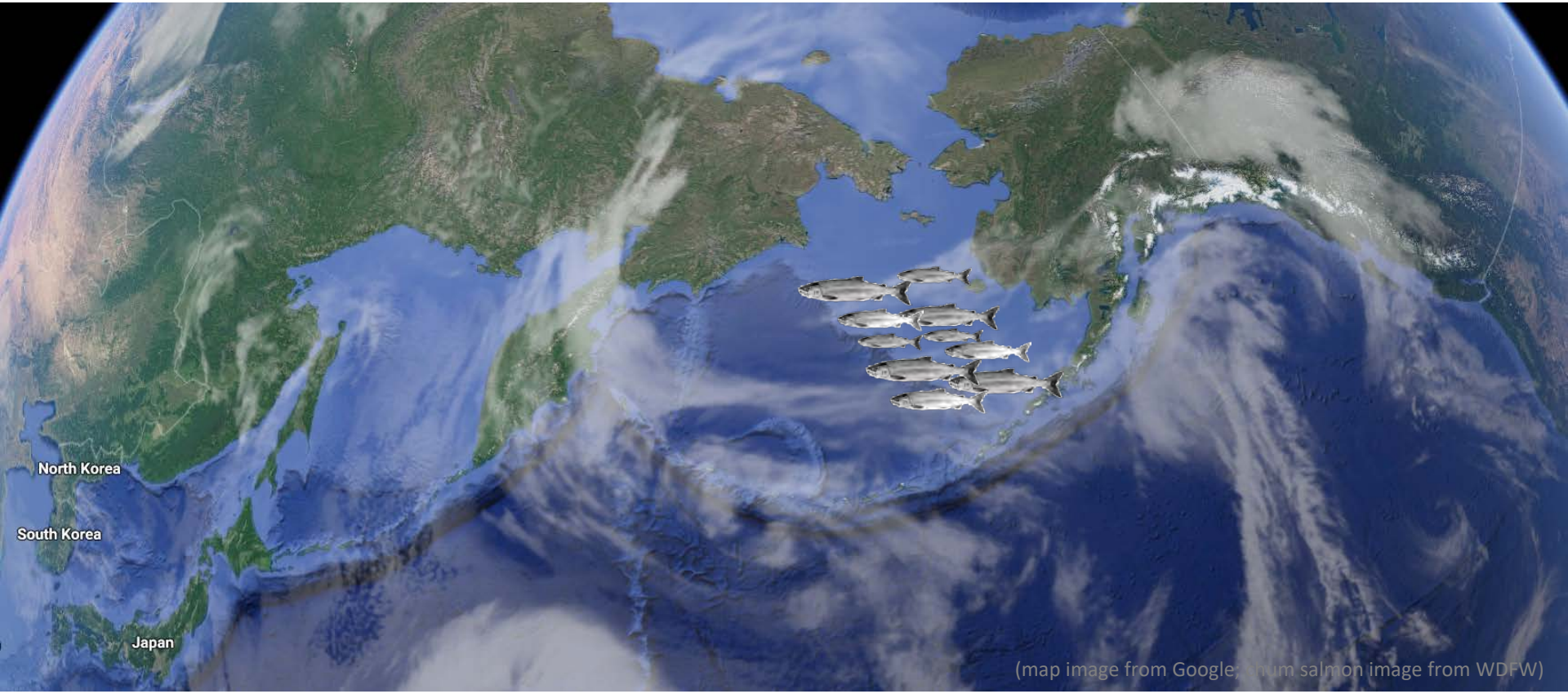
Improved Resolution of Chum Salmon Genetic Stock ID

Final Report to PCCRC

1 Feb 2019

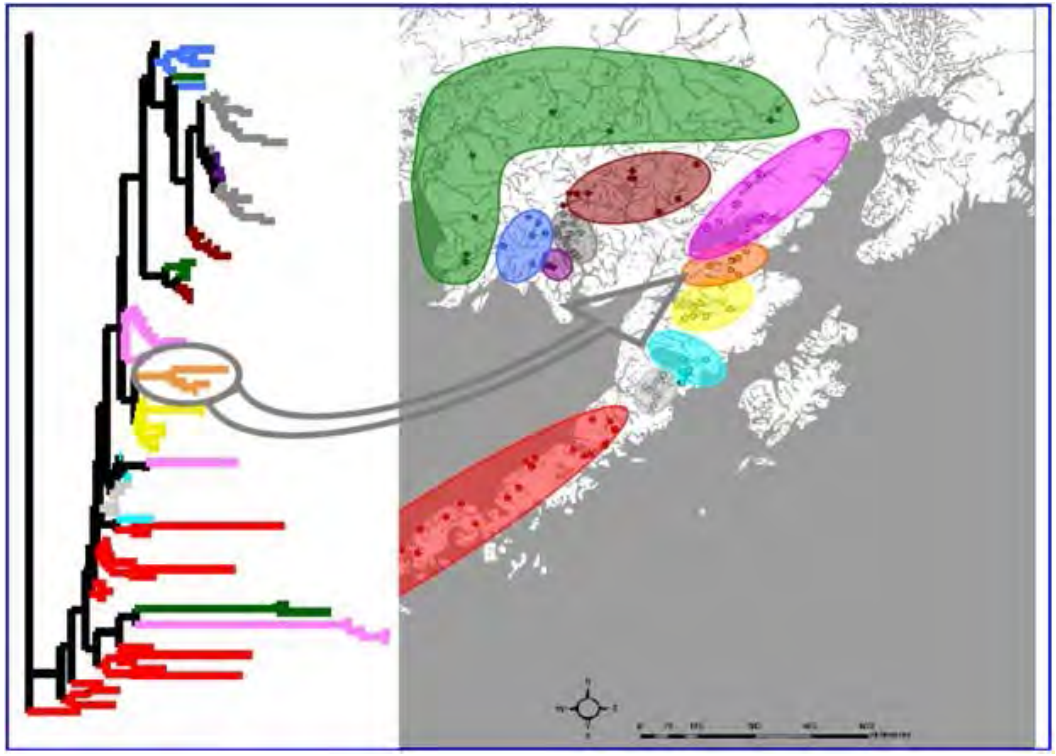
Megan McPhee & Garrett McKinney*
Juneau Fisheries Center
College of Fisheries & Ocean Sciences
University of Alaska Fairbanks
(*currently @ NW Fisheries Science Center, NMFS)

The problem?


North Korea
South Korea
Japan
(map image from Google; chum salmon image from WDFW)

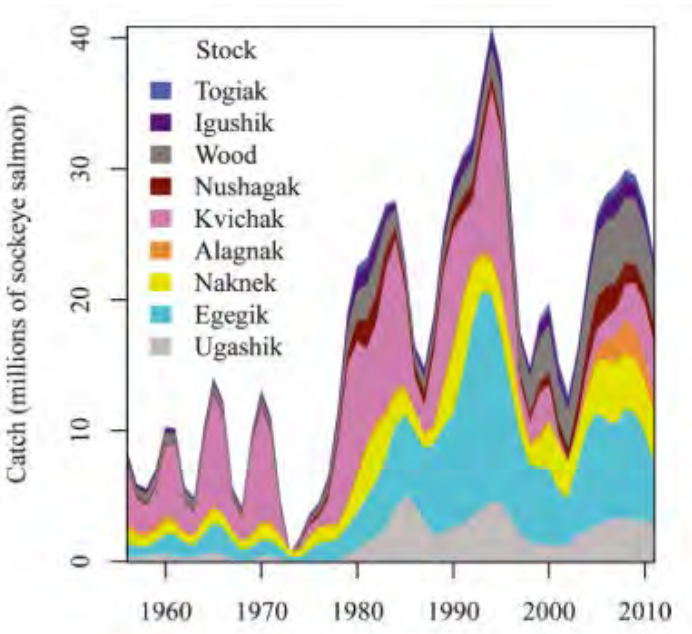
Genetic stock identification

… in a perfect world (Bristol Bay sockeye salmon)



(from ADF&G Gene Conservation Lab)

(NB: colors don't match between figures)

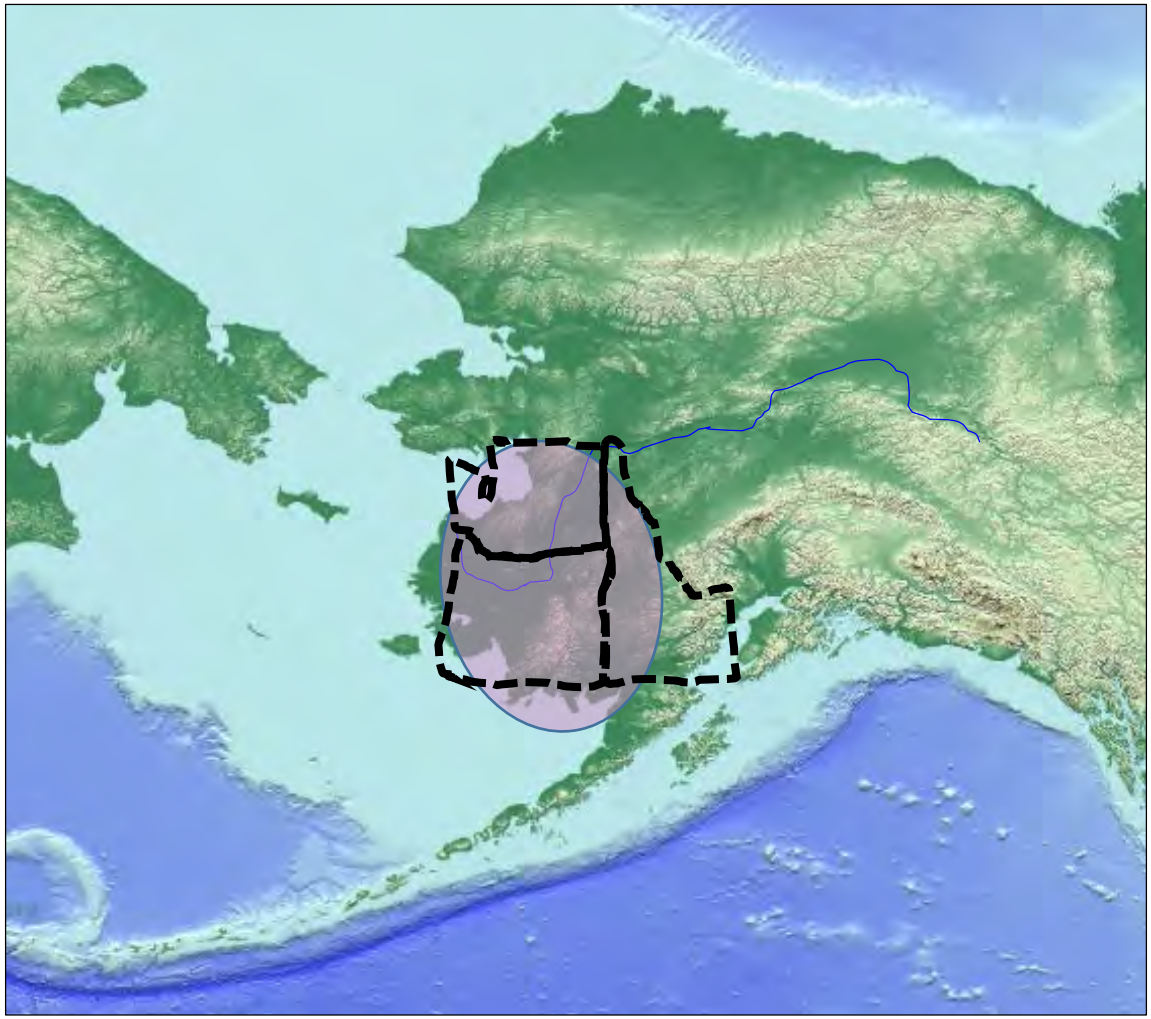


(Dann et al. 2009)

Western Alaskan chum salmon

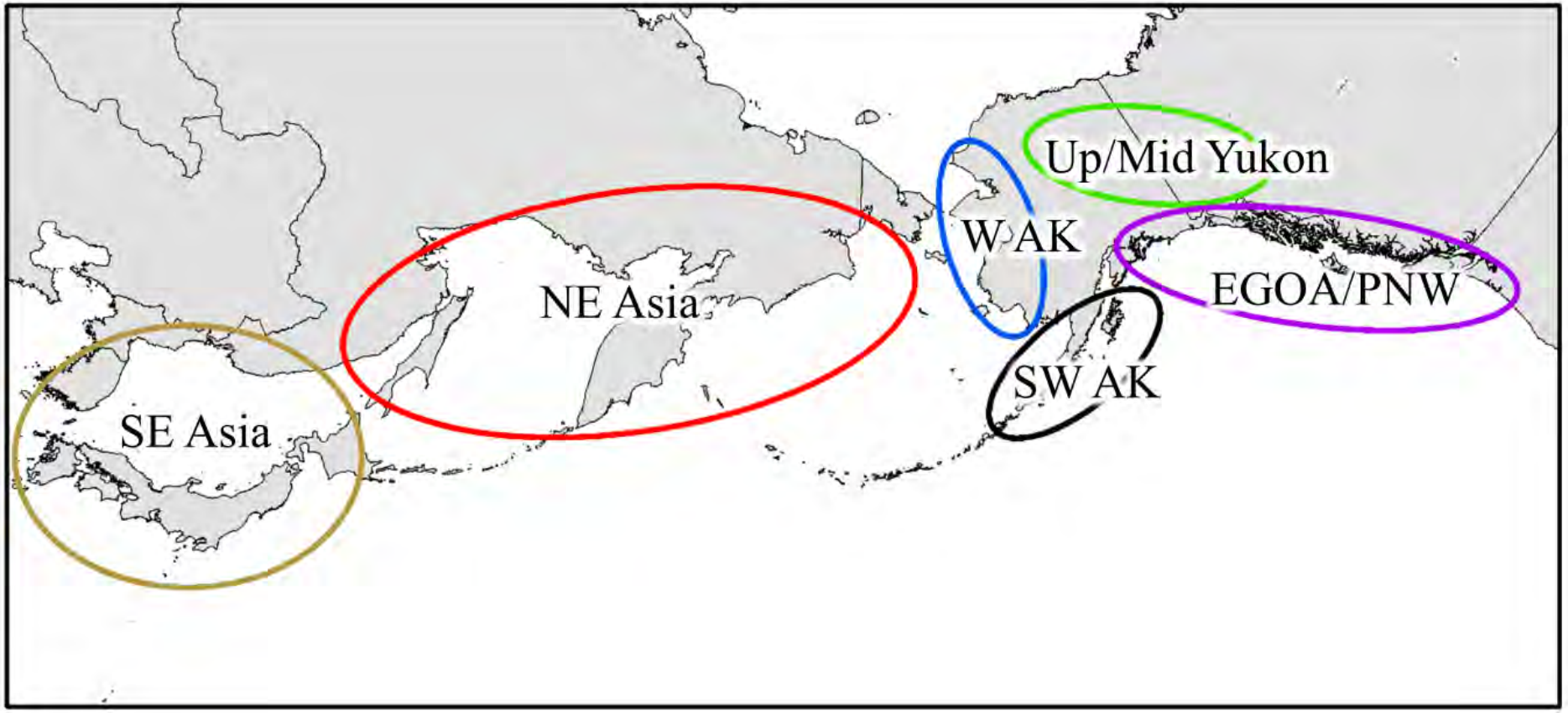
… not so perfect!

"Coastal Western Alaska" – summer-run chum are genetically indistinct across Norton Sound, Lower Yukon, Kuskokwim, and Bristol Bay



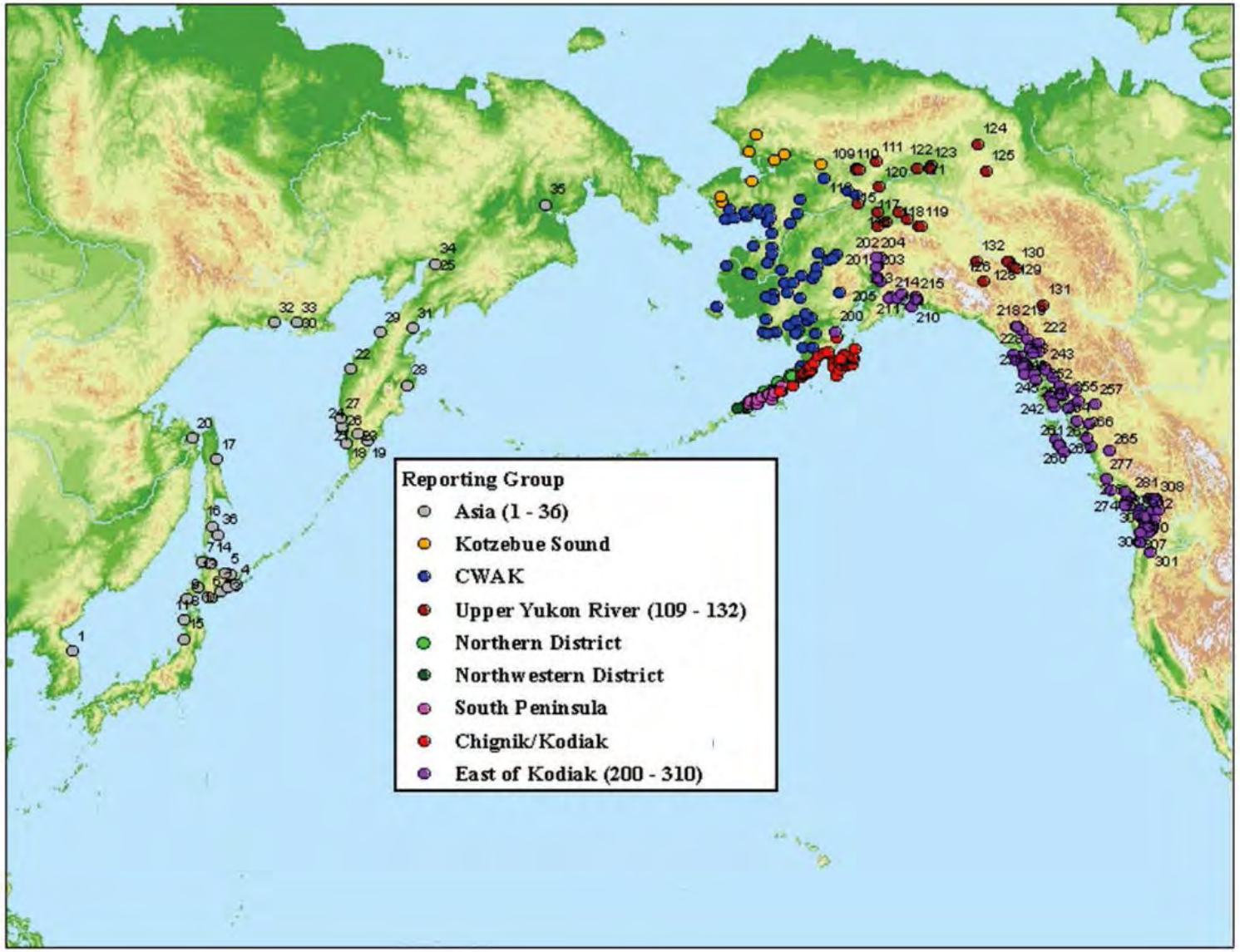
(from M. Garvin)

NMFS Baseline – 11 microsatellites



(Whittle et al. 2018)

ADF&G Baseline – 96 SNPs

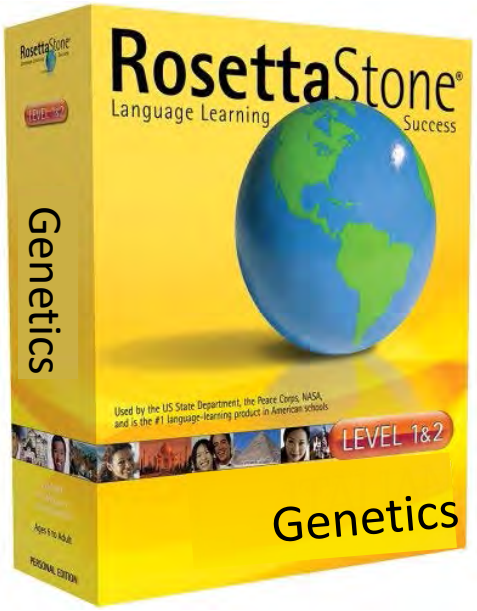


(DeCovich et al. 2012)

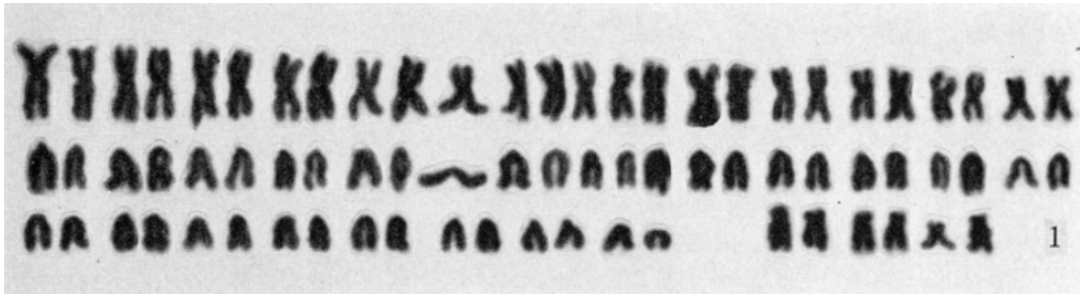
Our objectives

- To develop, test, and optimize a ‘GT-seq’ panel of 500 SNPs for Coastal Western Alaska chum salmon;

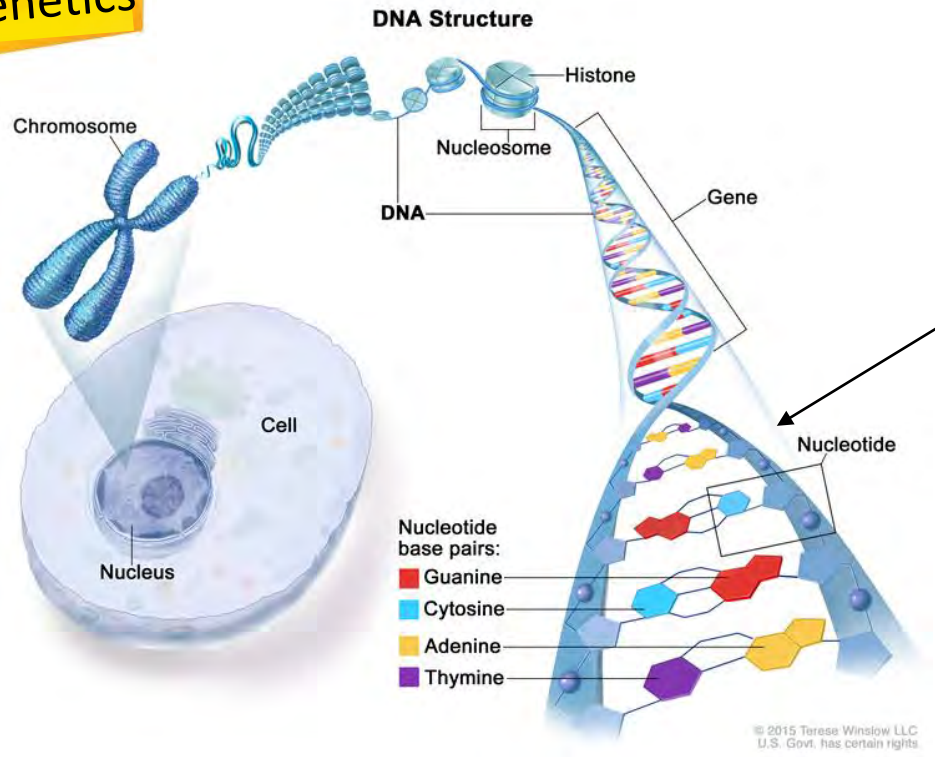
- Provide the optimized panel and protocols to resource managers throughout Alaska and adjacent areas



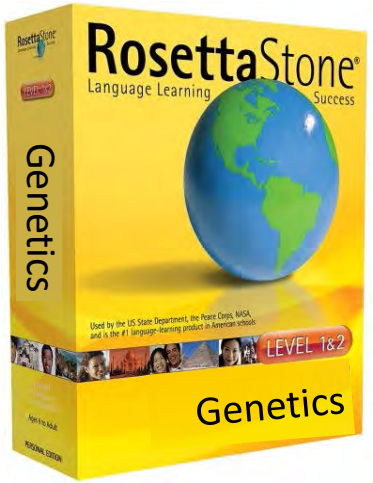
Genome: creature's DNA packaged in many chromosomes



(chum salmon 2N = 70, Sasaki et al. 1968)



DNA sequence: order of base pairs
e.g., TACGA



Marker: a place in the genome where we can assess genetic differences (sometimes also called 'locus/loci')

Microsatellite: differences are based on number of repeated motifs

5'-CGTAGGTATTTATTTATTTATTGCTAGGTT
5'-CGTAGGTATTTATTTATTGCTAGGTT

SNP ('snip'): differences are based on a single base pair in the DNA sequence (single nucleotide polymorphism)

5'-CGTAGGGCTAGGTT vs. 5'-CGTAGAGCTAGGTT

Allele: a variant at a given marker (e.g., G & A above)

Genotype: what 2 alleles does an individual possess? (1 from ma, 1 from pa)

GG 

GA 

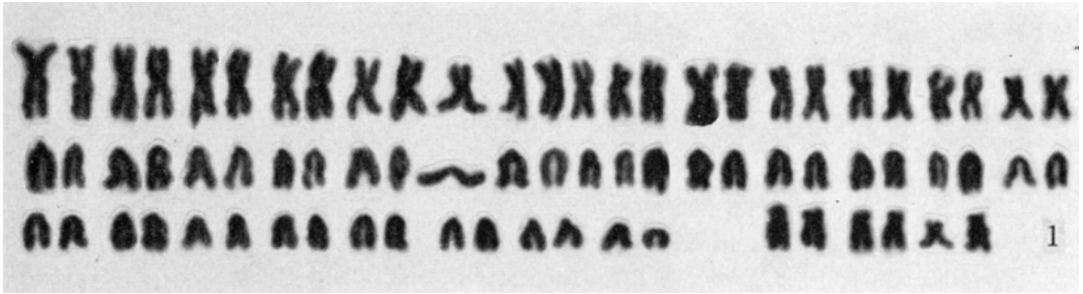
AA 

What is a GT-seq panel??

- GT-seq: “genotyping in thousands” by DNA sequencing – Nate Campbell & colleagues at Columbia River Inter-Tribal Fish Commission
- pool 1000s of individuals/100s of SNPs into single DNA sequencing run
- use DNA “bar codes” to identify individual and marker
- rapidly generate genotype data for 800-1000 individuals at 300-500 SNPs
- now running at ADF&G Gene Conservation Lab

Methods

1. identify SNPs across the chum salmon genome using “RAD-seq”



= 2.4 billion base pairs!

RAD-seq splices up the genome into manageable fragments, which are then sequenced

“Restriction-site associated DNA sequencing” – restriction enzymes cut at specific DNA sequences interspersed throughout genome; pieces are then sheared to smaller size for sequencing

Samples for SNP discovery

48 individuals/site

- Norton Sound (2)
- Yukon (2)
- Kuskokwim (1)
- Bristol Bay (1)

Methods

1. identify SNPs across the chum salmon genome using “RAD-seq”
2. Filter SNPs down to manageable number of useful markers (well-behaved and informative)

Filtering

initial: **94,002 RAD tags | 222,668 SNPs**

removed loci that did not genotype well *(including paralogs)*
removed loci that had very uneven allele frequencies *(MAF < 0.05)*

step 2: **31,919 RAD tags | 45,639 SNPs**

applied more stringent criteria for genotyping rate *(>90% of samples)*

step 3: **22,693 RAD tags | 30,006 SNPs**

Next: evaluated all 30,006 SNPs plus 500-SNP subsets using simulations

Results - all 30,006 SNPs

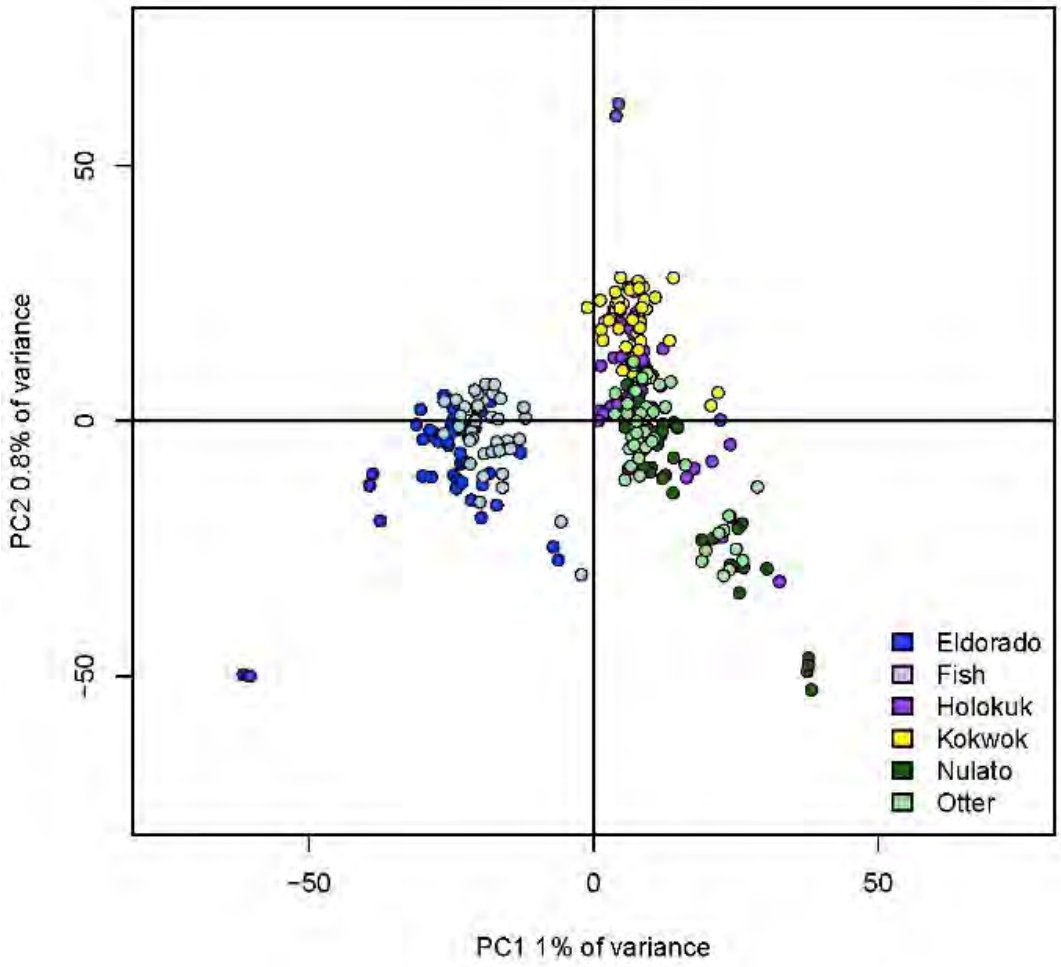
Norton Sound emerges as distinct

Norton Sound

Lower Yukon

Kuskokwim

Bristol Bay



500-SNP panel evaluations

Three different modes of marker selection:

1) individual SNP F_{ST}

2) Linked-SNP F_{ST}

3) “Random forest” marker selection *(Sylvester et al. 2016)*

Simulated individuals based on observed allele frequencies & evaluated accuracy of self-assignment in mixture analyses *(100% simulations, GSIsim)*

Results – panel evaluations

Accuracy (and 95% confidence intervals) of self-assignment by marker selection method

Selection Method	Norton Sound	Lower Yukon	Kuskokwim & Bristol Bay
Individual SNP F_{ST}	0.90 (0.87-0.94)	0.92 (0.89-0.95)	0.83 (0.78-0.88)
Linked-SNP F_{ST}	0.85 (0.82-0.89)	0.92 (0.89-0.95)	0.83 (0.78-0.88)
Random forest marker selection	0.92 (0.89-0.95)	0.82 (0.78-0.87)	0.68 (0.63-0.74)

Methods

1. identify SNPs across the chum salmon genome using “RAD-seq”
2. Filter SNPs down to manageable number of useful SNPs (well-behaved and informative)
3. Optimize and evaluate GT-seq panel

Panel selection

initial: **700 SNPs (highest individual F_{ST})**

removed loci unlikely to genotype well in GT-seq
(e.g., transposable elements, frequent primer sequences)

step 2: **533 SNPs (including 31 ADFG)**

Ran through 2 rounds of sequencing and optimization
Removed loci that did not work well when pooled

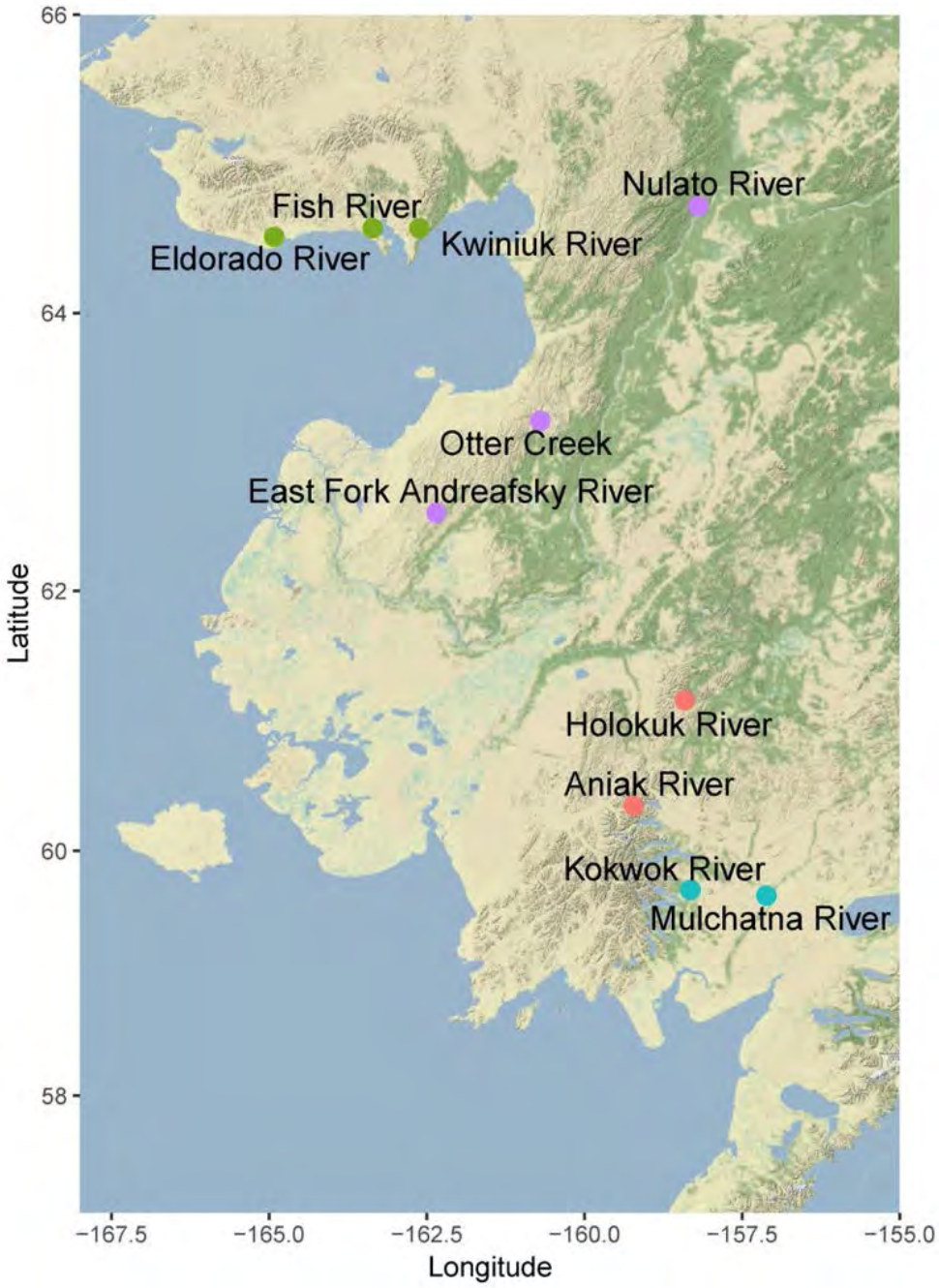
step 3: **448 SNPs (including 29 ADFG)**

Next: evaluated panel using increased number of populations and individuals

Panel evaluation

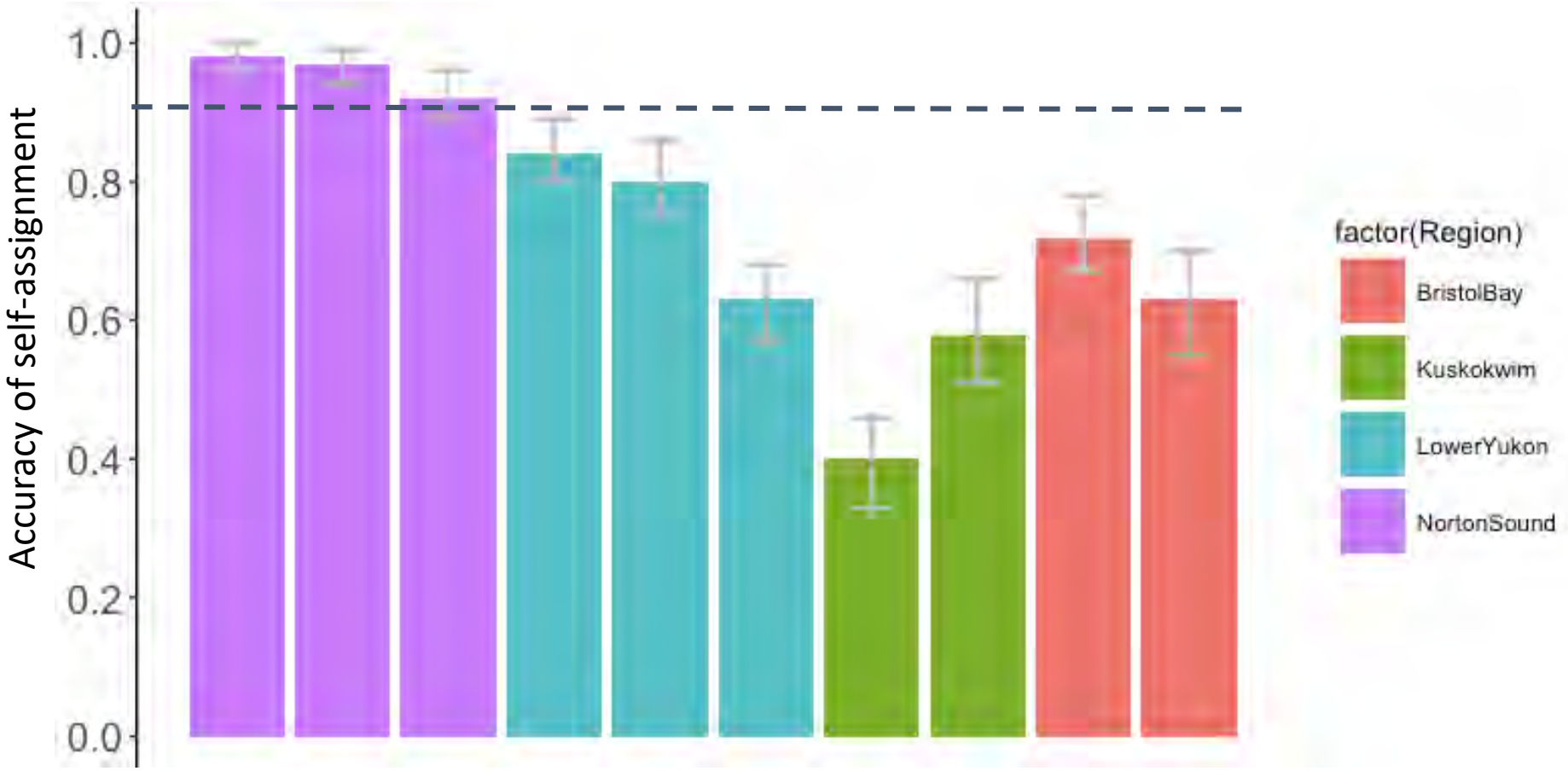
~80 individuals/site

- Norton Sound (3)
- Yukon (3)
- Kuskokwim (2)
- Bristol Bay (2)

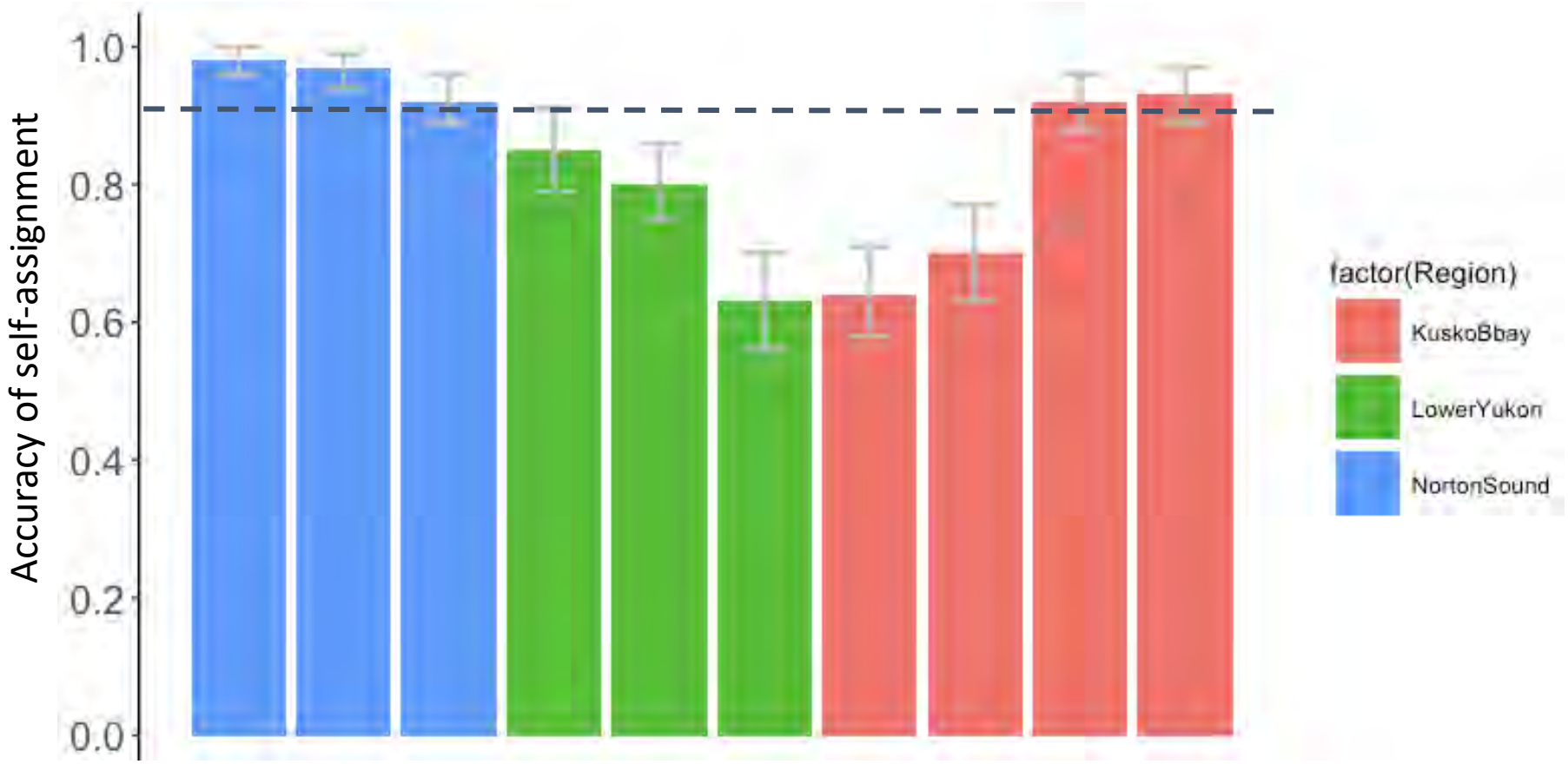


Region:
- Norton Sound
- Lower Yukon
- Kuskokwim
- Bristol Bay

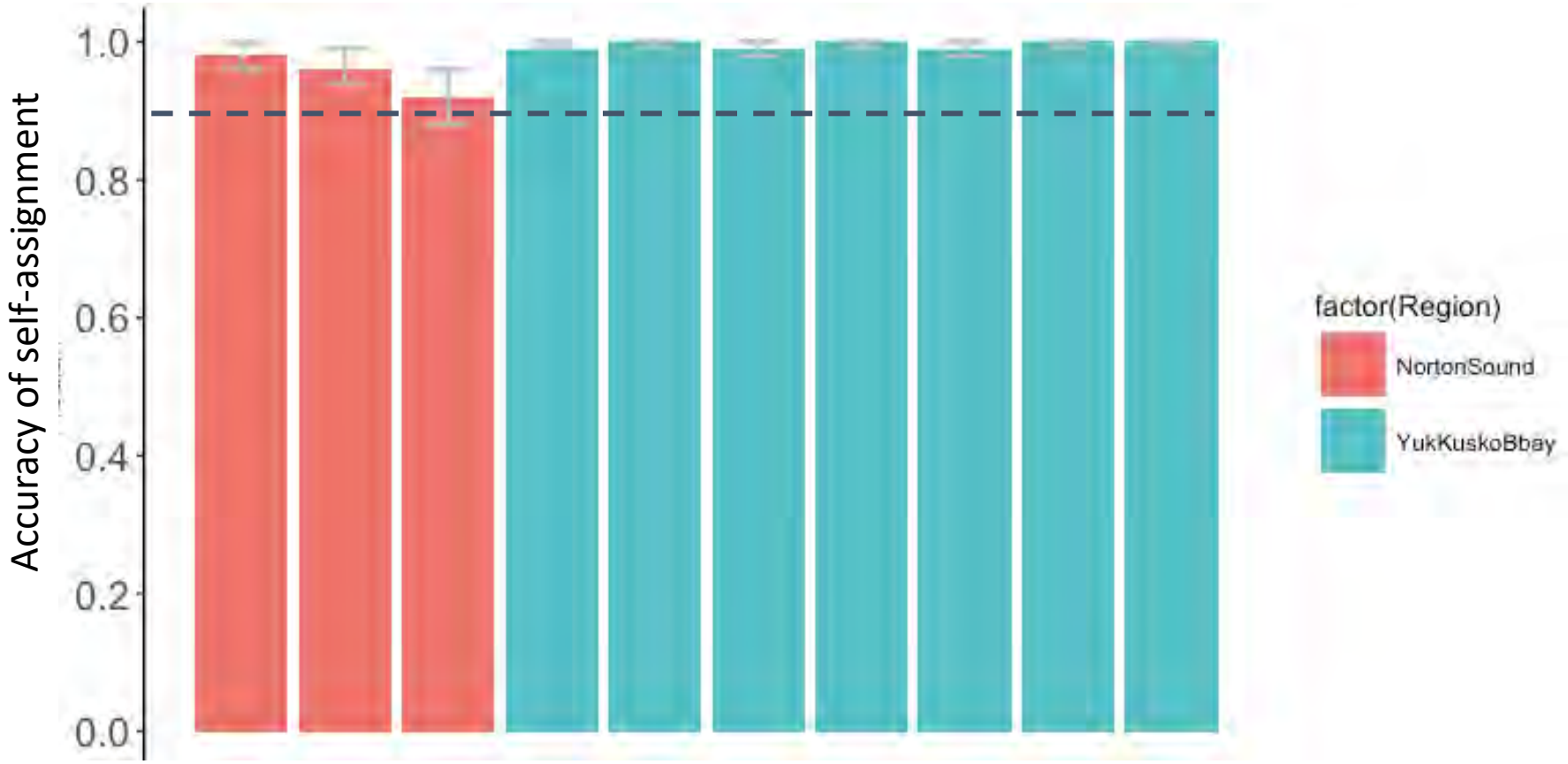
Results – 4 reporting groups



Results – 3 reporting groups



Results – 2 reporting groups



Why is coastal Western Alaska so difficult?

(Llamas et al. 2016)

Why is coastal Western Alaska so difficult?

(Llamas et al. 2016)

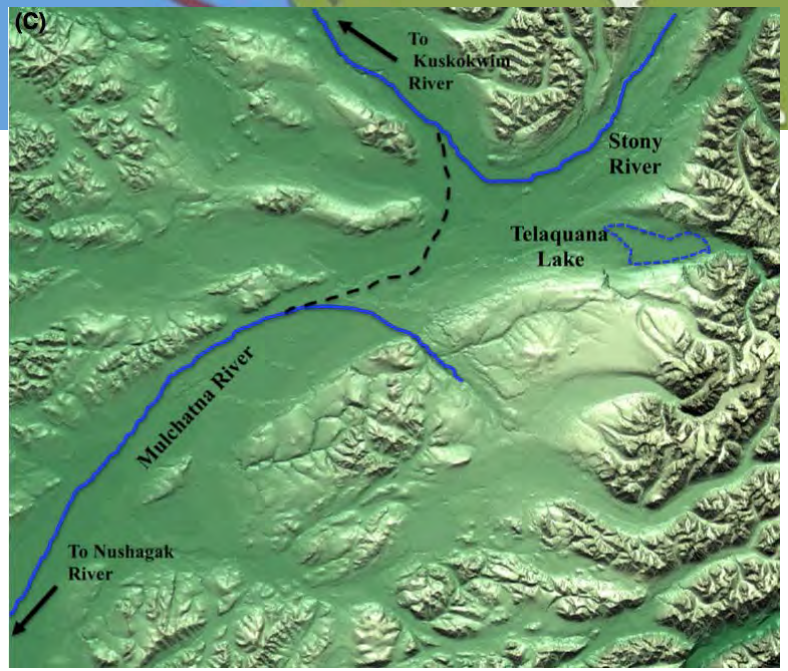
(Garvin et al. 2012)

Connection between Yukon & Kuskokwim

Why is coastal Western Alaska so difficult?

(Llamas et al. 2016)

(Garvin et al. 2012)



Connection between Kuskokwim & Bristol Bay (Nushagak R.)

Conclusions

- New panel allows for splitting of Coastal Western Alaska summer chum into 2 reporting groups: Norton Sound + Yukon/Kusko/Bristol Bay
- RAD-seq panel of 448 SNPs is available now for future genetic stock identification work
- Additional applications of panel – parentage-based tagging (e.g.)

Acknowledgments

- PCCRC (+ BOEM for previous funding)
- Jim & Lisa Seeb, Carita Pascal (UW)
- ADF&G Gene Conservation Lab
- NMFS Genetics Lab at Auke Bay
- Mike Garvin (UAF)
- Rose Fosdik, Tim Andrew, Verner Wilson - Western Alaska Salmon Coalition