Update Report

Naish, Kerry

Period: 2/1/2012 - 1/31/2013
Project: R/LME-1 - *Linking genetic variation, selection and adaptation in Chinook salmon: next generation sequencing and an oligonucleotide microarray with complete genome coverage*

:: STUDENTS SUPPORTED

Brieuc, Marine, mbrieuc@uw.edu, University of Washington, School of Aquatic and Fishery Sciences, status:cont, field of study:Genetics, advisor:Kerry Naish, degree type:PhD, degree date:2013-06-01, degree completed this period:No
Student Project Title:
An assessment of the frequency and extent of adaptive evolution in salmon using genomic approaches

Involvement with Sea Grant This Period:
Research Associate

Post-Graduation Plans:
Postdoctoral researcher
Waters, Charles, cwaters8@uw.edu, University of Washington, School of Aquatic and Fishery Sciences, status:new, field of study:Genetics, advisor:Kerry Naish, degree type:MS, *no degree date*, degree completed this period:No
Student Project Title:
Role of managed gene flow in reducing genetic divergence in hatchery fish

Involvement with Sea Grant This Period:
Coauthor on genome mapping research (not directly supported)

Post-Graduation Plans: *none*

:: CONFERENCES / PRESENTATIONS

Brieuc MSO, Waters CD, Seeb JE, Naish KA (2012) Mapping genomes in the age of genomics School of Aquatic and Fishery Sciences Graduate Student Symposium Seattle (poster), public/profession presentation, 70 attendees, 2012-11-17
Naish KA (2012) Evolution of fitness traits within salmon populations. Invited speaker Department of Biology, University of Southern California Los Angeles, California,, public/profession presentation, 75 attendees, 2012-03-20
Naish KA, Brieuc MSO, Kodama M, Hard JJ (2012) Fitness consequences of inbreeding in a pedigreed wild population of salmon 1st International Conference on Integrative Salmonid Biology Oslo, public/profession presentation, 150 attendees, 2102-06-17
Brieuc MSO, Naish KA (2012) Genomic analysis of parallel evolution of Chinook salmon (Oncorhynchus tschawytscha) in the Columbia River basin 1st International Conference on Integrative Salmonid Biology Oslo, public/profession presentation, 150 attendees, 2012-06-17

:: ADDITIONAL METRICS

K-12 Students Reached: 75

Husky Fest Open Day 21 April 2012. Booth on fisheries genetics

Acres of degraded ecosystems restored as a result of Sea Grant activities: 0

Curricula Developed: 1

400 level Conservation Genetics class, research findings integrated into data analysis laboratories

Resource Managers who use Ecosystem-Based Approaches to Management: 0

Volunteer Hours: 0

HACCP - Number of people with new certifications: 0

Cumulative Clean Marina Program certifications: 0

:: PATENTS AND ECONOMIC BENEFITS

No Benefits Reported This Period

:: TOOLS, TECH, AND INFORMATION SERVICES

Description		Developed	Used	Names of Managers	Number of Managers
Laboratory protocols for preparing samples for genotype by sequencing technology, R/LME-1	**Actual** (2/1/2012 - 1/31/2013) :	1	1	5 NOAA scientists (3 PIs, 2 technicians). Count as 3	3
	Anticipated (2/1/2013 - 1/31/2014) :	1	1		
Database of sequence tagged sites for use by scientist and managers in genomic studies in Chinook salmon. R/LME-1	**Actual** (2/1/2012 - 1/31/2013) :	1	1	Nichols, Krista, NOAA More than 50 anticipated	1
	Anticipated (2/1/2013 - 1/31/2014) :	1	1		
Chinook genome map that can be used to enhance genetic improvement of aquaculture populations and serve as a framework for other species genome maps, notably coho. R/B-51, R/LME-1	**Actual** (2/1/2012 - 1/31/2013) :	1	1	Nichols, Krista, NOAA More than 50 anticipated	1
	Anticipated (2/1/2013 - 1/31/2014) :	0	1		

:: HAZARD RESILIENCE IN COASTAL COMMUNITIES

No Communities Reported This Period

:: ADDITIONAL MEASURES

Safe and sustainable seafood

Number of stakeholders modifying practices
Actual (2/1/2012 - 1/31/2013) :
Anticipated (2/1/2013 - 1/31/2014) :

Number of fishers using new techniques
Actual (2/1/2012 - 1/31/2013) :
Anticipated (2/1/2013 - 1/31/2014) :

Sustainable Coastal Development
Actual (2/1/2012 - 1/31/2013) :
Anticipated (2/1/2013 - 1/31/2014) :

Coastal Ecosystems
Actual (2/1/2012 - 1/31/2013) :
Anticipated (2/1/2013 - 1/31/2014) :

:: PARTNERS

Partner Name: Cle Elum Supplementation and Research Facility, Yakama Tribes, type: government, scale: tribal
Partner Name: NOAA Northwest Fishery Sciences Center, type: government, scale: federal
Partner Name: Northwest Events
Partner Name: University of Montana, type: academic, scale: regional
Partner Name: Washington State Department of Fish and Wildlife

:: IMPACTS AND ACCOMPLISHMENTS

Title: **Washington Sea Grant research maps the Chinook salmon genome and helps hatchery managers determine best conservation practices**

Type: impact
Relevance, Response, Results:
Relevance: Chinook, the largest Pacific salmon species, are threatened by development, habitat loss, over-harvesting, and short-sighted hatchery management. Although the scientific basis for hatchery operations has advanced significantly, their genetic effects on salmon runs, including the loss of diversity and adaptive fitness, may jeopardize the recovery and rebuilding of Chinook populations. Genetic tools that survey the Chinook genome are key to addressing such concerns. They can be used to monitor the population's genetic capacity to adapt to environmental and anthropogenic changes, and to inform hatchery management decisions.
Response: Washington Sea Grant-supported researchers have sequenced a large portion of the Chinook salmon genome, identified markers spanning the entire genome (including specific markers associated with adaptive evolution and population differentiation), and developed a comprehensive DNA sequence database.
Results: The identified markers can be used in a wide range of applications with significant benefits for resource management. They are already providing useful tools. Tribal, state and academic researchers are using them to test different hatchery approaches. They have found evidence that certain practices, such as mixing a few wild individuals with the hatchery fish every year, can keep hatchery stocks much closer genetically to wild populations. This could reduce the hatchery stocks' genetic drift and forestall the development of completely isolated, human-dependent stocks.

Recap:
Washington Sea Grant research has mapped the Chinook salmon genome and developed a DNA dataset that can work to guide hatchery management and protect biodiversity.

Comments:
Primary Focus Area – LME (SSSS)
Secondary Focus Area – LME (HCE)

Associated Goals: Support conservation and sustainable use of living marine resources through effective and responsible approaches, tools, models and information for harvesting wild and cultured stocks and preserving protected species (SSSS Supply).
Support conservation and sustainable use of living marine resources through effective and responsible approaches, tools, models and information for harvesting wild and cultured stocks and preserving protected species (HCE Science).

Related Partners:

Cle Elum Supplementation and Research Facility, Yakama Tribes
University of Washington, School of Aquatic and Fishery Sciences, College of the Environment (UW)
Washington State Department of Fish and Wildlife
Washington State University, Department of Biology, School of Biological Sciences, College of Sciences (WSU)

:: PUBLICATIONS

No Publications Reported This Period

:: OTHER DOCUMENTS

No Documents Reported This Period

:: LEVERAGED FUNDS

No Leveraged Funds Reported This Period

R/LME-1 Progress Report 2/1/2012-1/31/2012

Linking genetic variation, selection and adaptation in Chinook salmon: next generation sequencing and an oligonucleotide microarray with complete genome coverage,

Kerry Naish & Jim Seeb, University of Washington, School of Aquatic and Fishery Sciences

1. **Activities carried out**

The aim of our study is to develop molecular markers that span the genome of Chinook salmon so that these markers can be used in standardized genome-wide population surveys. This aim will be attained by sequencing significant fractions of the genome in individuals that represent the range of Chinook salmon (Task 1), identifying polymorphic sites (Task 2), creating assays for these markers (Task 3) and mapping these markers to the Chinook map (Task 4). Our activities to date for each of these tasks are as follows:

Task 1: We used RAD sequencing, a targeted approach that identifies thousands of DNA loci simultaneously, to sequence 11 populations from the Columbia Basin and 3 from Puget Sound.

Task 2: Bioinformatic approaches for identifying polymorphic DNA sites across the Chinook genome were developed and implemented. Reference databases comprising all sequences from Task 1 were created and checked for quality and consistency. This reference database is a key resource for aligning all sequences generated by future projects, and will be made freely available on publication.

Task 3: We provided transcriptome sequence to the University of Montana who provided a 10,000 feature array for our testing.

Task 4: We sequenced three haploid families, three gynogenetic diploid families and one diploid family previously mapped with microsatellite markers. Variable DNA sequences were mapped to all Chinook chromosomes using the haploid and gynogenetic diploid families. Microsatellites in the diploid family were used to identify chromosome arms and align the map with the rainbow trout chromosomes. We also mapped the SNP markers that have previously developed for the Chinook Coastwide SNP database.

Participants

We collaborated with the Cle Elum Supplementation and Research Facilities, managed by the Yakama Nations, and the Washington Department of Fish and Wildlife to create the haploid families for genome mapping. Use of these populations facilitated collaborations with these agencies and with NOAA on related work, researching the effects of domestication selection in Chinook hatchery lines. We also collaborated with Washington Department of Fish and Wildlife, Alaska Department of Fisheries, and the Northwest Fisheries Science Center who are using SNPs developed by this project for studies of stock identification and reproductive success.

Finally, scientists from the University of Montana developed, from our data and for our joint use, a 10,000 feature array that we are using to test Chinook salmon from this project as well as other salmonids of interest in the Pacific Northwest.

Results

Task 1: We successfully sequenced 31 to 48 individuals per population. Our libraries comprised 36 individuals per sequencing lane and yielded an average of 80 million high quality reads per library. We identified 62,449 loci with high representation (80% or greater) across all populations and with a depth greater than 10 reads per individual. All of these loci were screened for quality and aligned, where possible, with the existing rainbow trout data (Miller *et al.* 2012). Loci passing this initial screen were added to a reference sequence database.

Task 2: We genotyped individuals using a bioinformatic pipeline based on PYTHON scripts and STACKS (Catchen *et al.* 2011). We identified 12,464 polymorphic loci across all populations with a minor allele frequency greater than 5% in at least one of the populations. These loci were screened using BLAST searches and haploid mapping was used (Task 4) to determine whether they were duplicated across chromosomes.

Task 3: With the help of a sister lab at the University of Montana we created and are initiating testing of a 10 000 feature array in haploid Chinook salmon (and other salmonid species).

Task 4: A substantial fraction of the polymorphic loci identified in task 2 were also variable across all of our mapping families. The three haploid families were used to map and identify duplicated loci. The salmon genome underwent a recent duplication – twice the number of chromosomes compared to the ancestral species. These loci can be used to identify the two matching chromosomes, but can complicate population genetic analyses. These haploid maps were then compared to the diploid map with microsatellite markers to identify the chromosome arms. Finally, we used the gynogenetic diploid crosses to identify the centromere location for each chromosome. Recombination probability increases as the distance from the centromere increases. Therefore the number of heterozygote offspring from a gynogenetic diploid cross is about 0 on or close to the centromere and close to 1 at the telomeres. Identifying the centromere and the chromosome arms is essential for ensuring complete coverage, but also gives us a means of comparing the salmonid genomes as each species is sequenced. Out of the 62,449 loci identified in Task 1, 9,100 appeared to be duplicated. We mapped a total of 7107 non duplicated RAD loci and 149 of the previously identified SNP loci. We were able to identify all 34 Chinook chromosomes using the microsatellite markers and successfully located the centromere for each chromosome using the gynogenetic diploid families (Figure 1). The 149 SNPs mapped in this study were distributed fairly evenly across the genome (Figure 2), which confirms that these markers are independent in population studies. This work is currently being prepared for publication. We intend expanding our mapped loci by sequencing and mapping additional Chinook diploid families from other populations, so that we can expand the marker set, examine sex-specific differences in map size, and screen for associations between markers and fitness-related phenotypes.

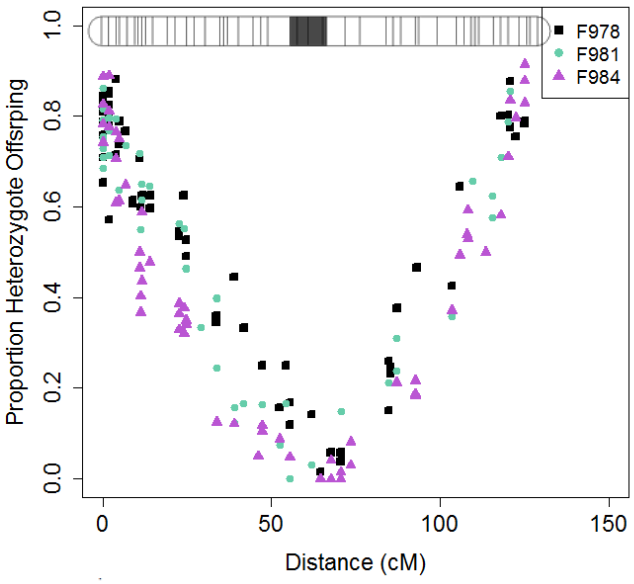

Figure 1 – Identification of centromere location on linkage group Ck04 by plotting offspring heterozygosity in three gynogenetic diploid families for each of the mapped markers
The centromere is located where proportion of heterozygotes are equal to zero. The proportion increases as the distance from the centromere increases
The bar on top corresponds to the mapped metacentric chromosome , displaying the two arms (white) and the centromere (grey). Each line on the chromosome represents the position of one or more markers

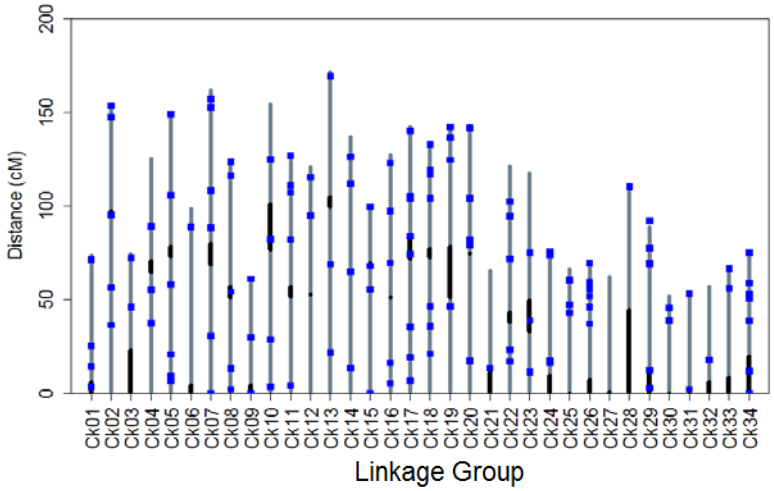

Figure 2 – Location of previously characterized SNPs on the mapped linkage groups. Each linkage group (grey bar, Ck01-Ck34) represents one chromosome.
The centromere locations are represented in black. Each SNP is represented in blue The y axis gives the length of the linkage group.

Challenges encountered: One challenge was developing the relevant bioinformatics tools for analyzing the data. However, we have established a full and reliable approach to attaining our goals. We have also assisted two laboratories in data analysis, and anticipate hosting International researchers in the coming year.

Changes in project directions: The objectives and deliverables remain intact, although the custom microarray originally envisioned was to be built for only 1 536 SNPs. This number was obsoleted by rapid advances in SNP technology during this reporting period. Additionally, the transcriptome sequencing approach to SNP discovery originally proposed was upgraded to the RAD sequencing approach described above. Consequently, we are describing many thousands of SNPs using RAD sequencing, and we are now testing the 10,000 feature array developed by the University of Montana from our data (instead of the 1 536 array originally envisioned). These project upgrades came at a steep price that was paid by collateral research (KN funded by NOAA and JS funded by Moore Foundation).