Completion Report

Period: 2/1/2014 - 1/31/2015

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

No Students Reported This Period

CONFERENCES / PRESENTATIONS

Brieuc MSO, Naish KA* (2014) "An evaluation of the role of adaptation in Chinook salmon (O.tshawytscha) evolution using genome based approaches", 2nd International Conference on Integrative Salmonid Biology, Vancouver, Canada, June 10-12., public/profession presentation, 400 attendees, 2014-06-10

Naish KA (2014) "An overview of our understanding on the location and frequency of recombination between homeologs in Pacific salmon ", "Homeolog Hell" workshop, American Genetics Association, Seattle, WA, June 30 - July 1., public/profession presentation, 40 attendees, 2014-07-01

McKinney, G (2014) Mapping duplicated genes in Chinook salmon. Oral presentation at the Homeolog Hell workshop, Seattle, WA, June 30-July 1, public/profession presentation, 40 attendees, 2014-07-01

Seeb, J (2014) Alaska-Seattle, ties that bind. Seattle Chamber of Commerce, public/profession presentation, 75 attendees, 2014-11-01

Seeb, J (2014) Haploid - assisted mapping disentangles complexities of duplicated salmonid genomes. International Conference on Integrative Salmon Biology (ICISB), public/profession presentation, 150 attendees, 2014-06-15

Seeb, J (2014) Keynote: Homeolog Hell: transmission genetics in salmondis. Oral presentation at the Homeolog Hell workshop, Seattle, WA, June 30-July 1, public/profession presentation, 40 attendees, 2014-07-01

Brieuc MSO (2014) Using gynogenetic haploids _and diploids for genome mapping. Oral presentation at the Homeolog Hell workshop, Seattle, WA, June 30-July 1, public/profession presentation, 40 attendees, 2014-07-01

ADDITIONAL METRICS

P-12 Students Reached:	400	**P-12 Educators Trained:**	0
Marine Brieuc and Kerry Naish participated at the Discover Science Weekend at the Seattle Aquarium, November 8 2014.			
Participants in Informal Education Programs:	0	**Volunteer Hours:**	0

Acres of coastal habitat protected, enhanced or restored:	0	**Resource Managers who use Ecosystem-Based Approaches to Management:**	0
Annual Clean Marina Program - certifications:	0	**HACCP - Number of people with new certifications:**	0

ECONOMIC IMPACTS

No Economic Impacts Reported This Period

SEA GRANT PRODUCTS

No Sea Grant Products Reported This Period

HAZARD RESILIENCE IN COASTAL COMMUNITIES

No Communities Reported This Period

ADDITIONAL MEASURES

Number of stakeholders modifying practices: 0

Sustainable Coastal Development

# of coastal communities: 0

PARTNERS

Partner Name: Alaska Department of Fish and Game

IMPACTS AND ACCOMPLISHMENTS

No Impacts or Accomplishments Reported

PUBLICATIONS

No Publications Reported This Period

OTHER DOCUMENTS

No Documents Reported This Period

LEVERAGED FUNDS

No Leveraged Funds Reported This Period

COMPLETION NARRATIVE

Uploaded File: Naish_5225_completion_....3.pdf, 606 kb

Final Report April 29, 2015

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

R/LME-1

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

Project Objectives (from the original proposal)

Our aim is to develop a SNP microarray, comprising markers placed in hundreds of genes that span the Chinook genome, so that comprehensive genome-wide studies can be used to develop species-wide datasets for monitoring and forecasting. Our specific objectives are:

1. To survey Chinook salmon genes for thousands of genetic variants that can be used as SNP markers
2. To identify putative SNPs in genes that show evidence of adaptive and neutral divergence
3. To create a custom SNP microarray using a subset of these markers, and to verify SNP assays by surveying individuals across a significant part of the species range
4. To identify the chromosomal location of the SNP markers by mapping the loci in segregating crosses of Chinook salmon

Rationale and background

One of the key challenges remaining in the conservation and management of fish populations is the ability to anticipate their adaptive response to human activities such as climate change, fishing, habitat changes, domestication, and hatchery releases (Naish & Hard 2008; Waples & Naish 2009). Understanding this response will allow us to track the impact of these activities and assess the success of any remedial actions taken. Geneticists have been successful at discriminating individuals and populations at a fine scale (Waples & Gaggiotti 2006), assigning mixtures of individuals to source populations (Seeb *et al.* 2004; Habicht *et al.* 2007) and exploring the relationship between population size and extinction risk (Hauser & Carvalho 2008). However, we have yet to gain a comprehensive understanding of how variation at the genetic level is associated with variation in fitness traits relevant to a species' response to change, because these traits are encoded by many genes that interact with each other and with the environment. In most cases, the identity of these genes is unknown. *Here, we used "next-generation sequencing" to develop a hundreds of single genetic variants (single nucleotide polymorphisms, SNPs; Morin et al. 2004) that span the genome of an important salmon species.* These SNPS will permit the rapid survey of genetic variation at the individual level in a very wide number of applications. *Specifically, we have developed powerful tools that will provide insight into a fundamental component of marine biodiversity and will contribute to our ability to sustainably manage and conserve a key marine resource.* We have elected to develop this technology in Chinook salmon in the first instance, because of the species' importance to State and Coastwide issues.

The recent and rapid improvement in sequencing technologies (Hudson 2008; Shendure & Ji 2008; Metzker 2010) has facilitated the characterization of thousands of variable markers, or species with little or no available genetic information (Davey *et al.* 2011) – such as Chinook salmon. Several of these approaches take advantage of the large amount of information afforded by sequencing a reduced portion of the genome. Restriction site associated DNA (RAD) sequencing (c.f. Miller *et al.* 2007; Baird *et al.* 2008) targets a consistent portion of the genome across individuals.

Application of RAD sequencing to genome mapping in Chinook salmon also provides an opportunity to develop analytical approaches relevant to mapping a species with polyploidy ancestry. Salmonids are descended from an autopolyploid ancestor that underwent a duplication even 58-63 million years ago (Crete-Lafreniere *et al.* 2012). Partial diploidy has been restored in this lineage through chromosomal rearrangements and divergence of duplicated chromosomes. Restoration of diploidy is incomplete in Salminidae (Wright *et al.* 1983; Allendorf & Thorgaard 1984; Allendorf & Danzmann 1997). However, the distribution of duplicated loci across the Chinook salmon genome is unknown. Mapping duplicated markers across chromosomes will reveal the rates of divergence between duplicated

chromosomes. A comparative analysis will also identify whether chromosomal divergence is conserved across other salmon species. Finally, identification of the genomic location of duplicated loci will assist researchers interested in differentiating duplicated and non-duplicated markers in population studies.

Here, we created a publically available reference database of RAD loci for Chinook salmon, where duplicated loci have been identified. This database rapidly facilitates alignment of newly sequenced individuals in related studies, and promotes data sharing across research groups. We then used these markers to map the genome of Chinook salmon, thereby creating a reference framework for genome-wide surveys of genetic variation in the species.

Methodology

Objective 1. Identification of loci and creation of a database of RAD loci for Chinook salmon

Creation of the reference database of RAD loci was carried out using three steps: the generation of a preliminary database of loci for Chinook salmon, the screening of the preliminary database for loci in repetitive regions and loci with repeat sequences, and the identification of duplicated loci.

All individuals were sequenced using RAD sequencing (Baird *et al.* 2008). Briefly, SbfI was used as a restriction enzyme, DNA was sheared to reduce fragment size and six nucleotide-specific barcodes were added to every DNA fragment. The DNA was then sequenced on the Illumina platform using 100-nucleotide single-read sequencing. The last 20 nucleotides were subsequently trimmed because the last 20 base pairs of the sequence had a consistently lower quality. For the purpose of this study, we defined a locus as a 74-nucleotide RAD sequence.

We first used 159 individuals from 10 populations across the Columbia River Basin to create the preliminary database. Reads were sorted into polymorphic and monomorphic loci de novo using STACKS 0.9995 (Catchen *et al.* 2011) with a minimum of three nucleotide mismatches between loci within an individual. We retained a consensus sequence for every locus that had been sequenced with a depth greater than 5X in more than 135 individuals (85%) as a temporary database of loci: these loci were used for further screening. The screening steps were aimed at identifying repetitive loci and loci with tandem repeat units, such as microsatellites and minisatellites. We used two alignment-based strategies. First, loci within the temporary database were aligned against themselves using BOWTIE 0.12.9 (Langmead *et al.* 2009), allowing up to three nucleotide mismatches. We expect that most homeologous loci with three or fewer mismatches between the paralogs would have been identified as a single locus during the creation of the preliminary database. Paralogs with more than three mismatches would not be detected using the BOWTIE alignment criteria we used here. Therefore a locus that aligned to several loci, including itself, was likely a repeat sequence and was excluded from the database. Second, a BLAST search (Basic Local Alignment Search Tool; Altschul *et al.* 1990) of the temporary database was conducted against itself using the low-complexity filter implemented in the search algorithm. This filter masks regions of low complexity, such as repeat nucleotides or motifs, within the query sequence. When this filter is used, a BLAST search that compares sequences with low complexity with themselves will rarely return a match or might return a match with another sequence, because the flanking sequence will be short. Therefore, loci within the database that did not return a match or where the best match (E-value less than 10^{-15}) for a given locus was not itself were discarded from the temporary database. Finally, polymorphic duplicated loci were identified using three haploid families (46-72 individuals per family). In haploid offspring, all unique loci will be homozygous; polymorphic duplicated loci will be heterozygous. Reads for all the haploid individuals were sorted into loci by alignment to the temporary database using BOWTIE. Individual reads from the haploids that aligned to more than one locus in the database could not be confidently relied upon in further analyses, and so were removed from the database. Loci with a depth of less than 10 reads for an individual were discarded for that individual. Genotypes for each individual were obtained using STACKS, which uses a maximum likelihood approach to identify polymorphisms (Catchen *et al.* 2011). The presence of a single individual with a heterozygous genotype at a locus within each haploid family was considered as insufficient evidence for duplication, as this genotype could be the result of a potential sequencing error, and the locus was retained in the database.

However, if more than one haploid individual was heterozygous at a locus, this locus was identified as being duplicated, since the same error occurring in two individuals was viewed as unlikely. We did not weigh these choices by family sizes, because the recurrence of a heterozygote genotype caused by sequencing error was deemed unlikely, regardless of number of offspring. This final step provided the final database of RAD loci for Chinook salmon.

Objective 2. To identify putative SNPs in genes that show evidence of adaptive and neutral divergence.

Putative SNPs in genes that show evidence of adaptive and neutral divergence were described in two ways. All RAD genotyping was done according to protocols above.

First, during the first portion of this study we generated paired haploid and diploid families from the University of Washington strain of Chinook salmon for the mapping of quantitative trait loci (QTL). Detailed experimental design, following IACUC protocol 4229-01, can be found in (Everett & Seeb 2014a). Briefly, the diploid families were reared for ten weeks post hatch and sampled for body weight and length. The families were then starved for 24h and subjected to increasing temperatures; sensitive individuals that lost equilibrium were removed until approximately 50% were sampled. The remaining individuals were considered thermotolerant. QTL were identified by associating RAD genotypes of the diploid progeny, along with size and thermotolerance, to the map location of same RAD loci determined by the family's haploid counterpart. We successfully detected three QTL for temperature tolerance and one QTL for body size at the experiment-wide level, as well as additional QTL significant at the chromosome-wide level (Figure 1).

These same mapping data were used by another, non-Sea Grant supported student, to test for pairwise signals of adaptive significance in five wild populations of Chinook salmon from Alaska (Larson *et al*. 2014b). Larson et al. (2014b) isolated three hot spots of divergence onto three linkage groups, providing focus for the second portion *Objective 2*.

Second, we used improvements in methodology to re-analyze and combine data across multiple sources including those data available from the University of Washington QTL map (Objective 2) and the much larger Columbia Basin map (Objective 4). An integrated linkage map was produced following the methods of Waples et al. (2015); the integrated map doubled the map density and consisted of 14,620 SNP loci, including 2,336 loci that are duplicated in subtelomeric regions. This dense integrated map enabled anchoring of scaffold sequence from the Atlantic salmon genome to provide a framework for locating genes of adaptive interest. We also had available for a resource more than 40,000 DNA sequences from protein coding loci (expressed sequence tags or ESTs; Gomez-Uchida *et al.* 2014).

Objective 3. To create a custom SNP microarray using a subset of these markers, and to verify SNP assays by surveying individuals across a significant part of the species range.

We identified and reported potential changes in project directions during the first portion of this project. The array technology that we originally envisioned evolved drastically as did the interest of potential fish and wildlife partners who would potentially capitalize on the array. Ultimately array tests that were funded by sister projects in our laboratory demonstrated imperfect scoring accuracy. At the same time agencies identified that near-term goals could better be met using either multiple low-density arrays, that we had developed prior to this project, or else RAD sequencing if the project scope warranted denser genomic data. As a result we abandoned this objective before Sea Grant funds were spent and reprogrammed efforts that resulted in the expanded Objective 2.

Objective 4. Characterization of the Chinook salmon genome

The three haploid families described in objective 1 were used to construct the initial linkage map. We also used three gynogenetic diploid families, created at the University of Washington hatchery facility, to map the centromere on each linkage group. Finally we used the diploid cross of Naish *et al.* (2013) to verify linkage group and chromosomal representation based on the microsatellite markers mapped previously, and to align the maps with the 34 chromosomes identified in Phillips *et al.* (2013).

All individuals were sequenced with RAD sequencing as previously described. Reads were aligned to the reference database using BOWTIE, and the polymorphic loci were identified with STACKS. This program uses a maximum likelihood approach to determine whether a polymorphism in an individual is true, or whether it is due to a sequencing error (Hohenlohe *et al.* 2010). This approach can be biased against the designation of heterozygous genotypes for individuals that differ in sequence depth between the two alleles. To correct this bias, we developed a Python script that called a heterozygote if both verified alleles had a depth of more than two and the total read depth at the locus was 10X or greater; this was the minimum depth we designated previously. Parental haplotypes for loci following Mendelian inheritance in the diploid cross were determined using linkage relationships with the previously mapped microsatellite markers.

4a. Construction of the linkage map

Duplicated markers identified in the haploids during the database development in objective 1 were used for mapping when one of the paralogs was polymorphic (one paralog polymorphic, parental genotype aa and ab, or aa and bc) or when both paralogs were polymorphic for different alleles (both paralogs polymorphic, parental genotype ab and ac and ab and cd). We also observed loci with ab and ab parental genotypes, but did not map these loci because heterozygous offspring were uninformative. Additionally we used 59-nuclease genotyping as in Seeb *et al.* (2011) to screen and map 384 SNPs that originated from other labs (Smith *et al.* 2005a; Smith *et al.* 2005b; Campbell & Narum 2008; Abadia-Cardoso *et al.* 2011; Larson *et al.* 2014a) in two haploid families. Many of these loci are polymorphic expressed sequence tags (ESTs) that are used in conservation and management applications for Chinook salmon across Pacific North America.

We used Onemap 2.0-3 (Margarido *et al.* 2007) for genome mapping in the haploid crosses and the F2 diploid cross. The Chinook salmon karyotype comprises 34 pairs of chromosomes (Phillips & Rab 2001). We therefore predicted 34 linkage groups per mapping cross. Linkage groups were identified independently for each haploid and diploid family using Onemap with a maximum recombination fraction of 0.25 and a starting LOD of 3.0. This LOD was subsequently increased by increments of 1.0 until the number of linkage groups identified was 34 or greater. We then used the microsatellite markers previously mapped and the RAD loci polymorphic in the diploid cross and the haploid crosses to identify each chromosome. Markers on each linkage group were subsequently ordered using ONEMAP for each haploid family. Individual haploid maps were merged using MERGEMAP (Wu *et al.* 2011) to create a consensus map.

4b. Centromere mapping and orientation of the chromosomes

We used three gynogenetic diploid families (84, 90 and 93 progeny per family) to identify the centromere. The second polar body is retained during the creation of gynogenetic diploid progeny. Therefore, a progeny will be heterozygous at a locus if a crossover event occurred in the female parent between a given marker and the centromere during meiosis I. The percentage of heterozygous offspring at a locus is expected to be 0% at the centromere, increasing to 100% in the telomeric region, because salmonids exhibit complete to near complete interference and typically have one crossover event per chromosome arm (Thorgaard *et al.* 1983).

We estimated the proportion of heterozygous progeny in each gynogenetic diploid family at every non-duplicated marker mapped on the haploid map and polymorphic in the gynogenetic diploid crosses. This information was used to identify the centromere and the chromosome type (acrocentric or metacentric) for each haploid family. Comparison with the diploid map was used to characterize the short (p) arm and long (q) arm for each chromosome as defined in Phillips *et al.* (2013).

4c. Distribution of duplicated markers across the genome

Two types of duplicated markers were used in this study. Duplicated loci with both paralogs polymorphic (BPP) were used to infer homeologies, because both paralogs could be mapped (Table 1). Occasional homeologous chromosome pairing in salmon may result in reduced divergence between the arms involved. We examined the position of the duplicated loci on the consensus haploid map to determine whether there was a bias in distribution of these loci. We reasoned that this analysis would identify chromosomal regions of reduced divergence between homeologs, indicating possible map

positions where homeologs have a tendency to pair. Here, we estimated the relative proportion of duplicated loci along the linkage groups. Because map positions are not uniformly distributed along the chromosomes, we used a kernel smoothing sliding window approach with a bandwidth of 2cM to determine the relative proportion of duplicated loci along the linkage groups.

Major Findings

Objective 1. Identification of loci and creation of a database of RAD loci for Chinook salmon

A total of 62,249 putative loci were sequenced in at least 135 individuals from the Columbia River basin with a minimum depth of five reads per locus per individual: these sequences formed the temporary database of RAD loci. Of these, 7312 loci were removed and 6409 were identified as potential duplicated loci using the haploid crosses. The final reference database comprised 48,528 putative non-duplicated loci and 6409 duplicated loci. This database was made publically available.

Objective 2. To identify putative SNPs in genes that show evidence of adaptive and neutral divergence.

In the first portion of the study we used the University of Washington data to successfully detect and map three QTL for temperature tolerance and one QTL for body size at the experiment-wide level, as well as additional QTL significant at the chromosome-wide level (Figure 1). We were unable to identify specific genes responsible for thermotolerance or body size until the integrated map became available the following year.

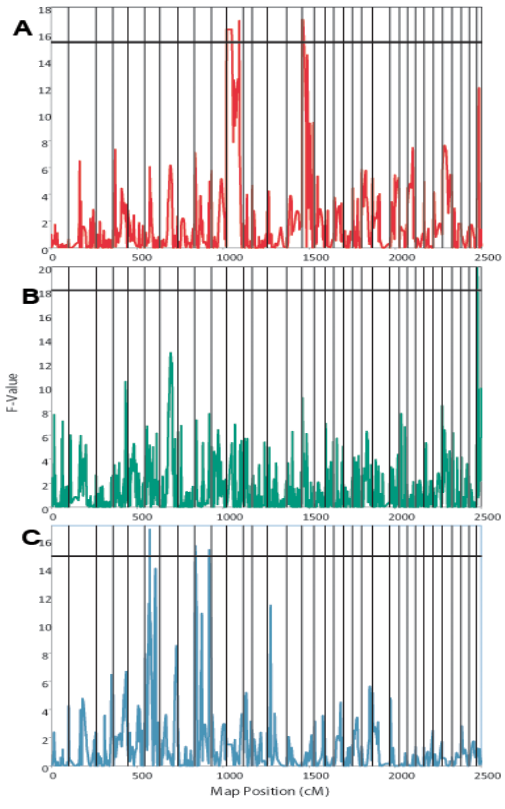

Figure 1 – Identification of genomic regions responsible for thermal tolerance (A and B) and growth (C). The distribution of F-values containing quantitative trait loci (QTL) significant at the experiment-wide level in the male (M7) and female (F14) parents. (A) Fvalues for thermotolerance in the male. (B) F-values for thermotolerance in the female F14. (C) F-values for body weight in the male. The dark horizontal line in all three figures is the experiment-wide ($P < 0.05$) significance threshold determined by a 10 000 permutation test (Churchill and Doerge 1994). Vertical lines designate individual linkage groups. Linkage group 11 on plot A contains two peaks, our model specified one QTL per linkage group, so only the higher of the two peaks was determined to be a QTL.

Objective 4. Characterization of the Chinook salmon genome

We identified 7151 polymorphic RAD loci in three haploid families that were used, along with 153 SNP loci currently used in conservation and management studies, to create a consensus map with a length of 4163 cM (Figure 2). The map comprised 34 linkage groups, which were anchored to all Chinook salmon chromosome arms using microsatellite loci that have been physically mapped in previous studies. The centromere was identified for all 34 chromosomes.

Figure 2– Consensus Chinook salmon female linkage map – 34 linkage groups corresponding to the 34 Chinook salmon chromosomes. Ots01 to Ots16 are metacentric; Ots17 to Ots34 are acrocentric. The size of each linkage group varies from 58 to 173.2 cM. Each line corresponds to the location of one or more markers. The centromere is represented in pink. All the chromosomes are oriented with the shorter arm (p arm) before the centromere, longer arm (q arm) after the centromere.

The placement of 799 duplicated loci on the linkage map revealed an uneven distribution of these loci across chromosomes (Figure 3), suggesting that homeologs diverged at different rates following whole genome duplication. Crossover frequency measured in one haploid family confirmed near complete interference across chromosome arms. Finally, the genome map supports previously published homologies among rainbow trout and Chinook salmon chromosome arms, but these homologies are supported using more extensive data and centromere placement.

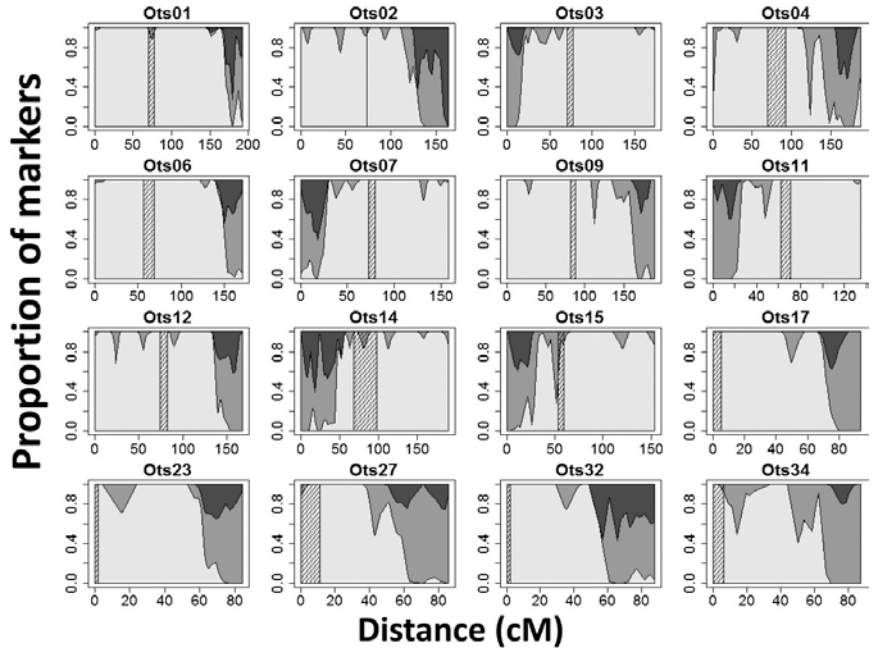

Figure 3– Proportion of duplicated and non-duplicated loci along the 16 linkage groups (denoted by chromosome number) with a high number of duplicated loci. Non-duplicated loci are represented in white; Duplicated loci are represented in gray (loci with one paralog polymorphic) or in dark gray (loci with both paralogs polymorphic). The centromere is represented by the cross-hatched area. All chromosomes are oriented with the short arm (p) on the left where relevant.

For the second portion of the project we constructed the integrated linkage map that included 14,620 SNP loci. Scaffold sequence from the Atlantic salmon genome that was anchored to the linkage map provided a framework for locating 11,728 Chinook salmon ESTs (Figure 4).

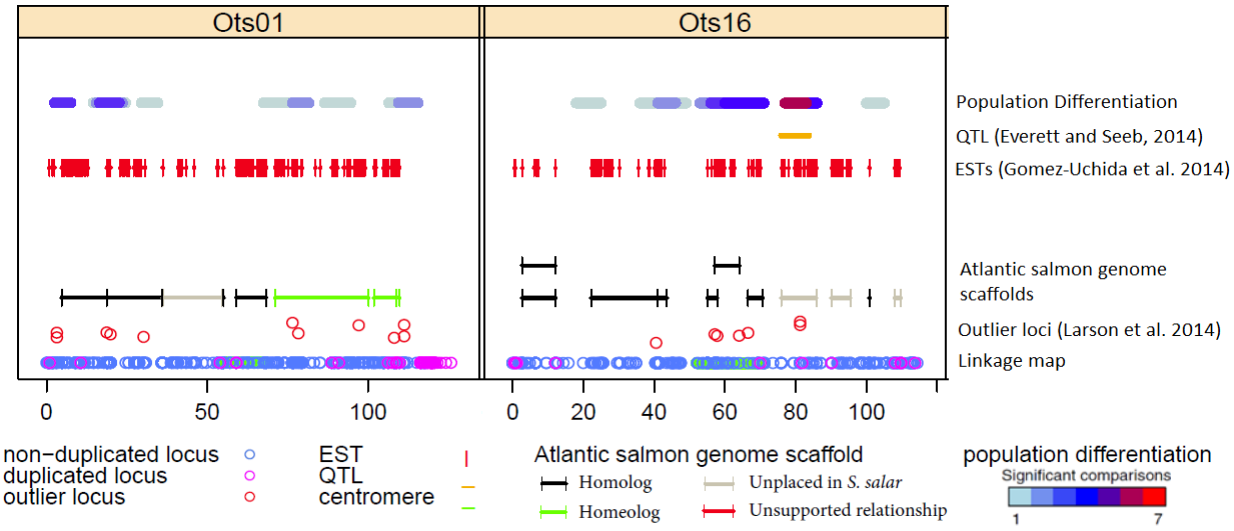

Figure 4. Example Plot of genetic resources overlaid onto the genetic map. Plot of genetic resources overlaid onto the genetic map. Linkage groups are plotted in separate panels with distance oriented along the x-axis and the different resources placed onto the map separated into different levels on the y-axis. Symbols and color-coding for map resources are denoted in the legend. The full plot of all linkage groups is in Fig. S1.

We further examined co-location of the thermotolerance and body weight QTL with annotations of the ESTs: QTL were found to co-localize with several candidate gene annotations including a gene known to be involved in thermotolerance as well as its inhibitor (McKinney et al. in review in house). Thermotolerance QTL were also found to contain annotations for functions such as stress response (MYLK), and DNA repair (APEX1). Annotations co-locating with QTL for length included functions such as muscle (TITIN, MYPC3) and bone development (SFRP3, PPBT) as well as metabolism (DHDH, NSDHL). The QTL for weight contained ESTs with annotations for functions such as hormone regulation and cell signaling (FKBP5, PLA2R), and cell development (CD109).

Multiple regions of the genome with increased divergence between populations described in Larson et al. (2014b) were also identified. Annotation of ESTs in these regions identified several genes with likely roles in adaptive processes. Two of the three genomic regions of divergence identified by Larson (2014c) were identified and placed on this map (Ots09p, Ots28), as well as an additional ten regions of divergence that were not previously identified (Ots02p, Ots05p, Ots09p (2 additional), Ots16q, Ots19, Ots25, Ots28, Ots30, Ots31). Several annotations of interest were associated with regions of divergence. These included genes for heat shock proteins: HSP47 (Ots09), HSP74 (Ots09), stress response: XRCC (Ots05), TNIK (Ots09), growth: NBL1 (Ots 07), CRYM (Ots25), PMP22 (Ots25), immune function: SHPK (Ots09), BAG6 (Ots31), behavior: GRM03 (Ots 07), RGRF1 (Ots09), STXB1 (Ots33), and homing ability: (OLF3A). Larson *et al*. (2014c) also identified 733 loci as F_{ST} outliers, 61 outliers were previously placed on the linkage map from Everett and Seeb (2014b) while 247 outliers were placed on our new linkage map. ESTs that co-locate with outlier loci can provide annotations for potential candidate genes linked with outlier loci. One-hundred thirty one outliers had at least one EST located within 0.1 cM, 79 outliers had at least one annotated EST located within 0.1 cM. Of the total 1,071 ESTs that co-located with outliers, 203 were annotated.

Significance of Results

The benefit of the genomic resources developed is the ability to generate large and readily standardized datasets that address a broad range of questions in Chinook salmon. These questions include identifying source populations in mixed stock harvest fisheries, using pedigrees to determine reintroduction success in conservation hatcheries, tracking the impacts of hybridization between different populations and detecting genetic variation that plays an important functional role in population fitness and adaptation. The reference database of RAD loci in Chinook salmon is extensive and provides a resource against which future RAD sequences can be aligned. We also developed several analytical approaches and resources relevant to analyzing similar datasets.

A significant academic outcome of the study is the description of factors that influence the retention of duplicated regions across salmon species. This work is significant, because it identified key events in the evolution of the salmonid genome, and laid the groundwork for rapidly identifying these regions in other species. Duplicated markers are difficult to identify in population studies, but we found that a significant fraction of the genome might be missing in genome wide surveys. We thus recommend using mapped ESTs or microsatellites to target these regions.

Student supported

Marine Brieuc, PhD student, School of Aquatic and Fishery Sciences. Graduated in Summer 2013.
Garrett Mckinney, postdoctoral researcher, School of Aquatic and Fishery Sciences.

Partnerships

We collaborated with the following:

- Scientists at Northwest Fisheries Science Center, NOAA Fisheries Columbia River Intertribal Fisheries Commission provided DNA samples from Columbia River Populations
- The Cle Elum Supplementation and Research Facility , Yakama Tribes and Washington Department of Fish and Wildlife provided crosses for genome mapping
- Scientists at Washington Department of Fish and Wildlife
- Scientists at Alaska Department of Fish and Game
- Faculty at University of Montana

Outreach

- We have presented our findings in several conferences of national and international scientists.
- We have participated in public outreach activities at the University of Washington (Orca Bowl, 2011 and Husky Fest Open day, 2012) and at the Seattle Aquarium (Discover Science Weekend, 2013 and 2014).
- Along with support from other funding sources, we organized a workshop on the application of SNP technologies in non-model organisms in 2011.
- We hosted an International Workshop *Escape From Homeologue Hell* in 2014 based in large part upon results from Chinook salmon map

Information/technology transfer activities

- A database of genomic resources for Chinook salmon, including markers mapped to Chinook salmon chromosomes and putative duplicated loci has been made publicly available. These resources have already been used by partners and agency scientists.
- Bioinformatic procedures developed in this study have been made publically available
- The genomic map enabled a later study that detected markers linked to an important fitness trait – run timing – in Columbia River Chinook salmon. Therefore, we have gained a greater understanding of the genetic diversity underlying a trait of key interest to conservation planning.

- The resources developed for Chinook salmon have been transferred to two other species of salmon – coho salmon (Kodama *et al.* 2014) and steelhead, thereby enabling similar molecular tools in these species.
- The work on Chinook salmon was also the first that permitted identification of conserved duplicated chromosomal regions across Pacific salmon, thus forming the basis for a clearer understanding of evolutionary events in these species.

References

Abadia-Cardoso A, Clemento AJ, Garza JC (2011) Discovery and characterization of single-nucleotide polymorphisms in steelhead/rainbow trout, Oncorhynchus mykiss. *Molecular Ecology Resources* **11**, 31-49.

Allendorf FW, Danzmann RG (1997) Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. *Genetics* **145**, 1083-1092.

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (ed. B.J. T). Plenum Press, New York.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410.

Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *Plos One* **3**.

Campbell NR, Narum SR (2008) Identification of novel single-nucleotide polymorphisms in Chinook salmon and variation among life history types. *Transactions of the American Fisheries Society* **137**, 96-106.

Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) STACKS: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171-182.

Crete-Lafreniere A, Weir LK, Bernatchez L (2012) Framing the Salmonidae Family Phylogenetic Portrait: A More Complete Picture from Increased Taxon Sampling. *Plos One* **7**, 19.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Everett MV, Seeb JE (2014a) Detection and mapping of QTL for temperature tolerance and body size in Chinook salmon (*Oncorhynchus tshawytscha*) using genotyping by sequencing. *Evolutionary Applications* **7**, 480-492.

Everett MV, Seeb JE (2014b) Detection and mapping of QTL for temperature tolerance and body size in Chinook salmon (Oncorhynchus tshawytscha) using genotyping by sequencing. *Evol Appl* **7**, 480-492.

Gomez-Uchida D, Seeb LW, Warheit KI, McKinney GJ, Seeb JE (2014) Deep sequencing of the transcriptome and mining of single nucleotide polymorphisms (SNPs) provide genomic resources for applied studies in Chinook salmon (*Oncorhynchus tshawytscha*). *Conservation Genetics Resources* **6**, 807-811.

Habicht C, Seeb LW, Seeb JE (2007) Genetic and ecological divergence defines population structure of sockeye salmon populations returning to Bristol Bay, Alaska, and provides a tool for admixture analysis. *Transactions of the American Fisheries Society* **136**, 82-94.

Hauser L, Carvalho GR (2008) Paradigm shifts in fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**, 333-362.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics* **6**.

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8**, 3-17.

Kodama M, Brieuc MSO, Devlin RH, Hard JJ, Naish KA (2014) Comparative mapping between coho salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a Whole Genome Duplication event. *G3: Genes|Genomes|Genetics* **4**, 1717-1730.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**.

Larson WA, Seeb LW, Everett MV, Waples RK, Templin WD, Seeb JE (2014a) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (Oncorhynchus tshawytscha). *Evolutionary Applications* **7**, 355-369.

Larson WA, Seeb LW, Everett MV, Waples RK, Templin WD, Seeb JE (2014b) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (Oncorhynchus tshawytscha). *Evol Appl* **7**, 355-369.
Larson WA, Seeb LW, Everett MV, Waples RK, Templin WD, Seeb JE (2014c) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (Oncorhynchus tshawytscha). *Evol Appl*, n/a-n/a.
Margarido GRA, Souza AP, Garcia AAF (2007) OneMap: software for genetic mapping in outcrossing species. *Hereditas* **144**, 78-79.
Metzker ML (2010) APPLICATIONS OF NEXT-GENERATION SEQUENCING Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46.
Miller CT, Beleza S, Pollen AA, *et al.* (2007) cis-regulatory changes in kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**, 1179-1189.
Morin PA, Luikart G, Wayne RK, Grp SNPW (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19**, 208-216.
Naish KA, Hard JJ (2008) Bridging the gap between the phenotype and the genotype: linking genetic variation, selection, and adaptation in fishes. *Fish and Fisheries* **9**, 396-422.
Naish KA, Phillips RB, Brieuc MSO, Newton LR, Elz AE, Park LK (2013) Comparative Genome Mapping Between Chinook Salmon (Oncorhynchus tshawytscha) and Rainbow Trout (O. mykiss) Based on Homologous Microsatellite Loci. *G3-Genes Genomes Genetics* **3**, 2281-2288.
Phillips R, Rab P (2001) Chromosome evolution in the Salmonidae (Pisces): an update. *Biological Reviews* **76**, 1-25.
Phillips RB, Ventura AB, DeKoning JJ, Nichols KM (2013) Mapping rainbow trout immune genes involved in inflammation reveals conserved blocks of immune genes in teleosts. *Animal Genetics* **44**, 107-113.
Seeb JE, Pascal CE, Grau ED, *et al.* (2011) Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Molecular Ecology Resources* **11**, 335-348.
Seeb LW, Crane PA, Kondzela CM, *et al.* (2004) Migration of Pacific Rim Chum Salmon on the High Seas: Insights from Genetic Data. *Environmental Biology of Fishes* **69**, 21-36.
Shendure J, Ji HL (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-1145.
Smith CT, Seeb JE, Schwenke P, Seeb LW (2005a) Use of the 5 '-nuclease reaction for single nucleotide polymorphism genotyping in Chinook salmon. *Transactions of the American Fisheries Society* **134**, 207-217.
Smith CT, Templin WD, Seeb JE, Seeb UW (2005b) Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of US and Canadian Chinook salmon caught in Yukon River fisheries. *North American Journal of Fisheries Management* **25**, 944-953.
Thorgaard GH, Allendorf FW, Knudsen KL (1983) Gene-centromere mapping in Rainbow-trout - High interference over long map distances. *Genetics* **103**, 771-783.
Waples RK, Seeb LW, Seeb JE (2015) Linkage mapping with paralogs exposes regions of residual tetrasomic inheritance in chum salmon (Oncorhynchus keta). *Molecular Ecology Resources*, n/a-n/a.
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.
Waples RS, Naish KA (2009) Genetic and evolutionary considerations in fishery management: Research needs for the future. In: *The Future of Fisheries Science in North America* (eds. Beamish R, Rothschild B), pp. 427-451. Springer, Fish and Fisheries Series.
Wright JE, Johnson K, Hollister A, May B (1983) Meiotic models to explain classical linkage, pseudolinkage, and chromosome-pairing in tetraploid derivative Salmonid genomes. *Isozymes-Current Topics in Biological and Medical Research* **10**, 239-260.

Wu YH, Close TJ, Lonardi S (2011) Accurate Construction of Consensus Genetic Maps via Integer Linear Programming. *Ieee-Acm Transactions on Computational Biology and Bioinformatics* **8**, 381-394.