

# PROGRAM & ABSTRACTS

## SNP Workshop III

Applications of SNP Genotyping  
in Non-Model Organisms

Semiahmoo Resort  
Blaine, Washington USA  
March 22 – 24, 2010

---

Welcome to SNP III! We thank all of the contributors who made the workshop possible, especially our major sponsors: **Roche 454 Life Sciences, Fluidigm Corporation, Gordon and Betty Moore Foundation, University of Washington School of Aquatic and Fishery Science, and the Washington Department of Fish and Wildlife**. The objective of the workshop is to assemble leading laboratories working on non-model organisms to exchange new developments in SNP discovery, next generation sequencing, data analyses, or high-throughput genotyping.

SNP Workshop III traces its origin to the first two workshops, held in 2005 and 2006 by the Department of Fish and Game in Alaska. SNP applications in non-model organisms were new at the time, and Alaska wanted to promote the advantages of SNPs for data sharing in international collaborations aimed at tracking migration and harvest of Pacific salmon in the treaty areas defined by the North Pacific Anadromous Fish Commission and the Pacific Salmon Commission. Alaska, Washington, and other Pacific salmon laboratories continue to play a lead role discovering and applying SNPs for these purposes.

The Gordon and Betty Moore Foundation identified activities of these workshops as a useful component of their Wild Salmon Ecosystems Initiative (<http://www.moore.org/salmon.aspx>) and founded the International Program for study of Salmon Ecological Genetics at the University of Washington in 2007. The program's goal is to promote SNP discovery, train scientists, and promote public access databases for Pacific salmon to be shared by any interested laboratories. Results of these activities are observable in many workshop abstracts, and we thank the Moore Foundation for their support of this workshop to promote open access data for Pacific salmon.

This workshop is made richer with two initiatives. First, Washington Sea Grant is funding (<http://www.wsg.washington.edu>) SNP-based studies that focus on using genomic approaches to detect ecologically and evolutionarily relevant genes in Pacific salmon. Second, FishPopTrace (<https://fishpoptrace.jrc.ec.europa.eu/>) assembles a consortium of European laboratories to develop a forensically validated framework to identify and monitor the distribution of stocks of marine fish. These projects will share exciting advances using both SNP discovery and array genotyping.

The workshop also benefits substantially from the contributions of our participants. The organizing committee (housed in aquatic research institutions) recognized that significant progress in SNP discovery and SNP genotyping has been made by the larger non-model community in the past two years. Substantial advances are evident in recent research conducted on microorganisms, insects, mammals, birds, and plants. The committee has thus made a special attempt to broaden the SNP workshop to those working on non-fish taxa, with the aim of exchanging information between communities. We are pleased to welcome participants from leading laboratories from five continents to explore opportunities now offered through next generation sequencing and high-throughput genotyping!

Jim Seeb, Lisa Seeb, Lorenz Hauser, Kerry Naish, Steven Roberts, Gary Carvalho

## CONTENTS

Meeting Agenda	
Sunday .....	2
Monday.....	2-3
Tuesday.....	4-6
Posters .....	5
Wednesday.....	7
Abstracts ( <i>organized by presenting author</i> ) .....	9-41
Workshop Participants.....	42-45

# SNP Workshop III: SNP Discovery and Applications in Non-model Organisms

**Sunday, March 21, 2010**

## **Welcome**

4:00–7:00 PM **Registration**

7:00–9:00 PM **Buffet Reception Pierside Room**

**Monday, March 22, 2010**

## **Plenary Session**

*Blakely/Cypress Rooms*

7:00–8:15 AM **Breakfast**

7:45 AM **Registration**

8:15 AM **Jim Seeb (moderator)**

Rapid advances in SNP discovery and genotyping in non-model organisms

8:45 AM **Willie Davidson**, Krzysztof Lubieniecki, Lisa Howard, Ben Koop, Matthew Kent, Sigbjorn Lien

Lessons gained from characterizing SNPs in Atlantic salmon

9:30 AM **Rob Ogden**

Unlocking the potential of genomic technologies for conservation management and wildlife forensics

10:15 AM **Break**

10:45 AM **Matthew Kent**, Paul Berg, Sigbjørn Lien

Development and application of an Atlantic salmon SNP chip

11:30 AM **Gordon Luikart**, Tiago Antao, Steve Amish Amish, Matthew Gruber

Power of SNPs for Estimation of Population Genetic Parameters

12:15 PM **Lunch Pierside Room**

1:15 PM **Einar Nielsen**, Jakob Hemmer-Hansen, Nina Therkildsen

Application of candidate gene SNP's for detection of adaptive population divergence in high gene flow marine species: Insights from targeted genome scans in Atlantic cod (*Gadus morhua*)

## SNP Workshop III: SNP Discovery and Applications in Non-model Organisms

Monday, March 22, 2010

### Databases & Analytical Approaches

- 2:00 PM **Lisa Seeb** (moderator) and the *PacSNP* collaborators of Japan and USA  
SNPs across a species' range: implications for conservation studies of Pacific salmon
- 2:20 PM **Gary Carvalho**, The *FishPopTrace* Consortium (<https://fishpoptrace.jrc.ec.europa.eu/>)  
FishPopTrace- An integrated European programme for tackling Illegal, Unregulated and Unreported Fishing
- 2:50 PM **Phillip Morin**, Eric Archer  
Incorporating uncertainty into inference of haplotypes and population analysis of linked SNP genotypes
- 3:10 PM **Break**
- 3:40 PM **Anders Gonçalves da Silva**, Phillip England  
How far can a SNP go? Testing the power of SNPs in fisheries management, a case for use in Orange Roughy (*Hoplostethus atlanticus*)
- 4:00 PM **Eric Anderson**  
Single nucleotide polymorphisms for intergenerational tagging
- 4:20 PM **Robin Waples**, Ryan Waples  
Inbreeding effective size and parentage analysis without parents
- 4:40 PM **John Carlos Garza**, Anthony Clemento, Alicia Abadia-Cardoso  
Development of SNP panels for intergenerational tagging and stock ID of salmon and trout
- 5:30-6:30 PM **Breakout: Chinook salmon panels**  
**Breakout: Atlantic Cod**  
**Breakout: Parentage and Kinship**
- 7:00 PM **Dinner** Pierside Room
- 8:30 PM **Bonfire on the beach**

## SNP Workshop III: SNP Discovery and Applications in Non-model Organisms

Tuesday, March 23, 2010

### Nextgen Sequencing & SNP Discovery

Blakely/Cypress Rooms

7:00–8:15 AM **Breakfast**

7:45 AM **Registration**

8:15 AM **Steven Roberts** (moderator), Lorenz Hauser, Lisa Seeb, Jim Seeb

SNP Discovery in Pacific herring using 454 transcriptome sequencing

8:35 AM **Armando Gerald**es, Ji Yang, Jan Jan Hannemann, Nina Thiessen, Johnson Pang, Michael Friedmann, Juergen Ehling, Carl J. Douglas, Quentin Cronk

SNP discovery from transcriptome resequencing in black cottonwood (*Populus trichocarpa*), a biofuels feedstock

8:55 AM **Sylvie Lapègue**, Serge Heurtebise, Estelle Harrang, Benjamin Morga, Emilie Flahauw, Christopher Sauvage, Pierre Boudry

SNPs detection and genotyping in oysters

9:15 AM **Meredith Everett**

Ultra short reads and non-model species: Exploring the complexities next generation sequence data in the absence of a reference genome.

9:35 AM **Matthew Settles**, Barrie Robison, Terence Soule

Single feature polymorphisms, transcript presence/absence and differential expression between strains of *Danio rerio* (Zebrafish).

9:55 AM **Break**

10:25 AM **Morten T. Limborg**, FishPopTrace Consortium

SNP Discovery by 454 sequencing in non-model organisms

10:45 AM **Sarah Helyar**, The FishPopTrace Consortium

in silico validation: a viable approach to re-sequencing?

11:05 AM **Discussion**

12:00 noon **Lunch and Poster Session** Prefunction Room, Blakely/Cypress

## SNP Workshop III: SNP Discovery and Applications in Non-model Organisms

Tuesday, March 23, 2010, 12:00 noon – 1:30 PM

### Posters

- P1 **M. Renee Bellinger**, *Project CROOS*, Collaborative Research on Oregon Ocean Salmon SNPs for stocks: the application of genetic stock identification to studies of the ocean distribution and ecology of Chinook salmon
- P2 **Nathan Campbell**, Shawn Narum  
Development of 55 novel SNP assays for sockeye and coho salmon and assessment of their ability to differentiate stocks within the Columbia River
- P3 **Jennifer Eichelberger**, Matthew Krampe, Edward Heist  
Single nucleotide polymorphism (SNP) marker development in *Scaphirhynchus* sturgeons
- P4 **Matthew Leslie**, Amy Frey, Brittany Hancock, Kelly M. Robertson, Phillip A. Morin  
Using custom-designed capture arrays and next-generation sequencing to streamline SNP discovery: An example from the study of global population structure of fin whales (*Balaenoptera physalus*)
- P5 **Ken Livak**, Lisa Seeb, Jim Seeb, Jun Wang, Krishna Datta, Carita Pascal  
High-throughput melt curve analysis for verification and genotyping of SNPs in four populations of chum salmon
- P6 **Andrew Matala**, Shawn Narum  
The distribution of genetic variation among Chinook salmon lineages in the Columbia River Basin: a landscape genetics perspective
- P7 **Nikolai Mugue**  
SNP markers development in polyploid fish species
- P8 **Shunpei Sato**, William D. Templin, Lisa W. Seeb, Hiroyuki Nagoya, Shigehiko Urawa  
An application of SNP markers to stock identification of chum salmon in the Bering Sea during the summer of 2007
- P9 **Piper Schwenke**, Jon Baker, Linda Park  
Congeneric species identification using TaqMan® SNP genotyping assays
- P10 **Caroline Storer**, Carita Pascal, Eric Grau, Steven Roberts, Jim Seeb  
Searching for SNPs: Mining the sockeye salmon, *Oncorhynchus nerka*, transcriptome for high-resolution molecular markers
- P11 **Daria Zelenina**, Diana Stoklitskaya, Dmitry Shchepetov  
Russian Pacific salmon SNP database
- P12 **Mette H. H. Hansen**, Hanne Jørgensen, Carita Pascal, Mark Henryon, Jim Seeb  
SNP based association mapping of disease resistance to VHS virus in rainbow trout

## SNP Workshop III: SNP Discovery and Applications in Non-model Organisms

Tuesday, March 23, 2010, 12:00 noon – 1:30 PM

### Population Genetics

- 1:30 PM **Lorenz Hauser** (moderator) The A & C of parentage and kinship
- 1:50 PM **Christopher Habicht**, Michael Link, Tim Baker, Jim Seeb  
SNPs in fishery management: patterns of stock composition in the Port Moller test fishery in 2006 – 2008
- 2:10 PM **Matt Smith**, Lowell Fair, Chris Habicht, Tyler Dann, Lisa Seeb  
Multiplex preamplification and SNP genotyping of sockeye salmon (*Oncorhynchus nerka*) scales: a forty-five year retrospective of catch compositions in Bristol Bay
- 2:30 PM **Bjorn Erickson**, Molly Stephens, Bernie May  
The search for the Kern River rainbow trout
- 2:50 PM **Suzanne Roden**, Peter Dutton, Phillip Morin, George Balazs, Patricia Zarate, I.J. Cheng  
Detecting green turtle population structure in the Pacific using single nucleotide polymorphisms (SNPs)
- 3:10 PM **Break**
- 3:40 PM Jeong-Nam Yu, Moongeun Yoon, **Syuiti Abe**  
Comparative phylogeography of chum and masu salmon using mitochondrial and microsatellite DNA markers
- 4:00 PM **William Templin**, Lisa Seeb, James Murphy, Jim Seeb  
High-resolution stock identification for migratory studies of Chinook salmon
- 4:20 PM **Kenneth Warheit**, Todd Kassler, Jennifer Von Bargen, Sewall Young  
Population differentiation of Puget Sound Chinook salmon
- 4:40 PM **Jon Hess**, Andrew Matala, Shawn Narum  
Comparison of SNP and microsatellite markers for application of genetic stock identification for Chinook salmon in the Columbia River Basin
- 5:30–6:30 PM **Breakout: Chum Salmon Panels**  
**Breakout: Coho and Mykiss Panels**  
**Breakout: FishPopTrace**
- 7:00 PM **SNP Workshop Banquet** Pierside Room

## SNP Workshop III: SNP Discovery and Applications in Non-model Organisms

Wednesday, March 24, 2010

### **Adaptation**

*Blakely/Cypress Rooms*

7:00–8:15 AM **Breakfast**

8:15 AM **Kerry Naish** (moderator)

8:25 AM Heather Freamo, Patrick O'Reilly, Mark Culling, Paul Berg, Sigbjørn Lien, **Elizabeth G Boulding**  
Identification of non-neutral SNPs in Bay of Fundy Atlantic salmon (*Salmo salar*) and their use in population identification

8:45 AM **Shawn Narum**, Jon Hess  
Testing for selection and local adaptation with SNP markers

9:05 AM **Marine Brieuc**, Kerry Naish  
Running before we can walk: identifying SNP loci under selection in partially sequenced genomes

9:25 AM **Sebastien Rioux Paquette**, Fred Allendorf, Peter Ritchie  
Searching for evidence of selection in the genome of a marine fish: hoki (*Macruronus novaezelandiae*)

9:45 AM **Break**

10:15 AM **Svein-Erik Fevolden**, Kim Præbel, Helge Meissner, Jørgen S. Christiansen  
Alaska pollock (*Gadus chalcogrammus*) off the coast of northern Norway - evidence for an Atlantic-Pacific link revealed by SNPs and microsatellites

10:35 AM **Daniel Gomez-Uchida**, Matthew Smith, Chris Habicht, James E. Seeb, Lisa W. Seeb  
Single nucleotide polymorphisms unravel hierarchical divergence among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations

10:55 AM **Ian Bradbury**, Paul Bentzen  
Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature

11:15 AM **Discussion**

12:00 NOON **Lunch and Adjourn** Pierside Room



# ABSTRACTS

*(organized by presenting author)*

## SNP Workshop III

Applications of SNP Genotyping in Non-Model Organisms

---

## Comparative phylogeography of chum and masu salmon using mitochondrial and microsatellite DNA markers

Jeong-Nam Yu<sup>1</sup>, Moongeun Yoon<sup>2</sup>, **Syuiti Abe**<sup>1</sup>  
Hokkaido University<sup>1</sup>, Pukyong National University<sup>2</sup>, Japan

Chum salmon *Oncorhynchus keta*, demonstrate anadromous life history only and extend widely from the Far East to North America around the Pacific Rim. In comparison, masu salmon *O. masou* have two life history forms of anadromous and river-resident, and are endemic to the Far East or western side of the North Pacific. The genetic population structure and phylogeography of the two Pacific salmon species were investigated in order to understand their evolutionary process, using nucleotide sequence variation of the mitochondrial (mt) DNA control region and NADH dehydrogenase subunit 5 gene and allelic variation at several polymorphic loci of microsatellite (ms) DNA. Both mtDNA and msDNA markers were employed for more than 4,200 chum salmon representing a total of 96 populations from the Pacific Rim and more than 1,100 masu salmon representing 24 populations from the Far East.

Both DNA markers disclosed substantial genetic variation and clear structure of populations among and within regions after AMOVA and other population genetic analyses. The analysis described known geographical groupings of chum and masu salmon into Japan, Korea, Russia or North America, but also revealed novel large population groups in the two species, including the Rim or coasts of the Sea of Japan, the Sea of Okhotsk, West Bering Sea, Northwest Alaska, or Gulf of Alaska. The observed population structure appears to reflect isolation by distance with limited gene flow between regions and larger amounts of gene flow between populations within these regions, as suggested by the Mantel test with mtDNA in chum salmon and with msDNA in masu salmon. The differences in results between mtDNA and msDNA in estimate the influence of isolation by distance could be associated with different life history of the two species; the possible male-biased restriction of gene flow was suggested by a mild but significant genetic differentiation between male-biased river-resident and female-biased anadromous form of masu salmon in the same river.

The mismatch distribution analysis, neutrality test, and nested clade phylogeographical analysis suggested population expansion in the middle to late Pleistocene with primarily contiguous range expansion over the Pacific Rim in chum salmon. Similar demographic inference also was obtained for masu salmon, in which the late Pleistocene population expansion in the Sea of Okhotsk was followed by colonization to the Sea of Japan explained primarily isolation by distance. Thus, the observed genetic population structure of chum and masu salmon is likely reflecting the demographic histories influenced by the past glacial movement in the North Pacific.

The current phylogeographical data will be useful for the identification of effective management units and hence for sustainable use of chum and masu salmon.

---

## Single nucleotide polymorphisms for intergenerational tagging

**Eric Anderson**, Southwest Fisheries Science Center, USA

SNP markers are ideal for large scale parentage inference because they are easily standardized between different laboratories and they can be genotyped accurately and inexpensively. This allows for a system of intergenerational genetic tagging of wide-ranging cultured, or partially cultured, species. This talk describes the use of SNPs to track hatchery-born Pacific salmon using parentage-based tagging (PBT). The underlying concept is simple: genotypes of hatchery broodstock are first placed in a centralized parent data base. Then, the genotypes of ocean-caught fish can be compared to this data base to determine their parents, providing information needed to parameterize fishery forecasting models, plus a great deal more. The scale of the associated parentage inference problem is quite large and overwhelms currently available software. I describe several advances implemented in my software SNPIT (SNPs for Intergenerational Tagging) that allow such large problems to be handled with ease. A simulation study parameterized with allele frequencies from a standardized panel of 96 SNPs demonstrates there is ample power with 96 SNPs to conduct PBT on a large scale.

---

## SNPs for stocks: the application of genetic stock identification to studies of the ocean distribution and ecology of Chinook salmon

**M. Renee Bellinger**<sup>1</sup>, Jeff Feldner<sup>2</sup>, Nancy Fitzpatrick<sup>3</sup>, Gil Sylvia<sup>1</sup>, Robert Ireland<sup>1</sup>, Peter Lawson<sup>4</sup>, Michael A. Banks<sup>1</sup>  
Oregon State University<sup>1</sup>, Oregon Sea Grant<sup>2</sup>, Oregon Salmon Commission<sup>3</sup>, NOAA Fisheries<sup>4</sup>, USA

Genetic stock identification can be used to elucidate patterns of ocean distribution and migratory timing of Chinook salmon populations on the West Coast of North America. Low escapement levels of Klamath and Sacramento River stocks and harvest limits on ESA listed salmon have substantially constrained harvests of Oregon and West Coast salmon troll fisheries. This has resulted in the loss of hundreds of jobs and millions of dollars in annual coastal income. Presently, salmon ocean managers have no access to “real time” data on stocks and are unable to differentiate stocks at local spatial and temporal scales. Consequently, managers have implemented large area closures for entire seasons. New scientific and management tools are needed that can differentiate stocks in “real time” at refined spatial areas to protect weak stocks, and to increase access to healthy stocks. These data can also be used to understand the influence of oceanographic conditions on movements of salmon during their ocean residency. Project CROOS, Collaborative Research on Oregon Ocean Salmon, is a State, Federal and University collaboration with commercial fishermen that aims to understand the migration and at-sea distribution of Chinook salmon stocks encountered off the coast of Oregon. Fine-scale harvest locations and fishing effort data are coupled with oceanographic information to investigate factors that contribute to at-sea distribution patterns. Using the Genetic Analysis of Pacific Salmonids (GAPS) standardized microsatellite baseline for Chinook salmon, estimates of stock mixture compositions (Mixed Stock Analysis, MSA) are used to assess impacts of fisheries in a given region during a limited time-frame. These mixture composition estimates can be incorporated into fisheries management decisions aimed at reducing harvest of stocks of concern. Although bi-allelic SNPs present less information per locus than more polymorphic microsatellites, their lower cost and ease for scoring and standardization across labs means that many more SNPs can be assayed for the same cost, which through their greater number provides sampling of more of the genome (through linkage) and thus in the end likely more stock discriminatory power. The number of SNPs available for Chinook salmon is rapidly increasing. Therefore, once an adequate baseline has been developed, SNPs may become an integral part of studies of ocean ecology and at-sea distribution of Chinook salmon along the Oregon coast.

---

## Identification of non-neutral single nucleotide polymorphisms in Bay of Fundy Atlantic salmon (*Salmo salar*) and their use in population identification

Heather Freamo<sup>1</sup>, Patrick O'Reilly<sup>2</sup>, Mark Culling<sup>3</sup>, Paul Berg<sup>3</sup>, Sigbjørn Lien<sup>3</sup>, **Elizabeth G Boulding**<sup>1</sup>  
University of Guelph<sup>1</sup>, Dept. Fish. Oceans Canada; Bedford Institute of Oceanography<sup>2</sup>, Canada  
CIGENE, Norwegian University of Life Sciences<sup>3</sup>, Norway

Atlantic salmon from the Inner Bay of Fundy (iBoF) on the East coast of Canada have declined during the last two decades. In 2002, they were designated by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) as endangered. Distinguishing wild iBoF salmon from wild outer Bay of Fundy (oBoF) salmon and from aquaculture escapees would benefit current conservation and captive-breeding programs for iBoF salmon but has been difficult with existing neutral markers. We used CIGENE's Sequenom Mass Array™ system to genotype 11 hatchery-reared Bay of Fundy (BoF) populations for 320 SNPs within ESTs. The edited SNP data was analyzed with five different Fst-outlier detection programs. Nine non-neutral SNP markers were chosen that 1) had high-Fst values, 2) were highly-significant outliers, and 3) had maximally different allele frequencies between the iBoF and oBoF populations. We then used these nine SNP markers to genotype archived DNA samples from historical wild BoF populations using Invader™ chemistry at the Department of Fisheries and Oceans laboratory and were able to detect significant genetic differentiation between wild iBoF and wild oBoF salmon populations. (Funded by Natural Sciences and Engineering Research Council of Canada).

---

## Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature

**Ian Bradbury**, Paul Bentzen  
Marine Gene Probe Lab, Dalhousie University, Canada

Despite the enormous economic and ecological importance of marine organisms, the spatial scales of population structure and adaptation remain largely unknown. The identification and preservation of local stocks that display adaptive diversity is critical to the long-term maintenance of productive stable fisheries and ecosystems as they allow species to persist in the face of environmental change. Here we document genomic evidence of widespread adaptive spatial differentiation in a broadcast spawning marine fish, Atlantic cod (*Gadus morhua*), using a transcriptome wide survey of 1641 single nucleotide polymorphisms. We identify a subset of gene-associated polymorphisms (~40 loci) for which allele frequencies show parallel temperature-associated clines ( $p < 0.001$ ,  $r^2 = 0.88$ ) on either side of the Atlantic. Temperature associations were robust to the statistical removal of geographic distance or latitude effects, and contrasted "neutral" loci (~1636 loci) which displayed no temperature association. These clinal loci tested positive for signatures of selection using a Bayesian approach, and linkage mapping revealed that they tightly cluster into several linkage groups. Results are consistent with the parallel evolution and/or selective sweep of multiple genes in response to ocean temperature, and support the possibility of a new conservation paradigm based on functional and adaptive genetic diversity.

---

## Running before we can walk: identifying SNP loci under selection in partially sequenced genomes

**Marine Briauc**, Kerry Naish  
University of Washington, USA

Next-generation sequencing has allowed a rapid expansion in the discovery of single-nucleotide polymorphisms (SNPs). In non-model species, the production of large numbers of SNP loci can be of great importance in tackling a diversity of questions. Some approaches, such as population assignment, association mapping and evolutionary studies, place an emphasis on differentiating between neutral and adaptive polymorphisms. One way to achieve this goal is to identify selected loci from incomplete sequences prior to the development of SNP assays. Salmonid species have been the subject of a number of genome wide studies, and the abundance of partial DNA sequences across species can now be used to identify key genes involved in their adaptive divergence. We initiated a study on seven salmonid species (Atlantic salmon, rainbow trout, chum salmon, sockeye salmon, Chinook salmon, brook trout and lake whitefish), with the aim of developing approaches to identifying SNP loci of evolutionary importance in partially sequenced genomes. An estimation of the rate of evolution for neutral (synonymous) and selected (non synonymous) substitutions permits the detection of positively selected genes and specific sites under selection, as well as the species lineages in which selection has occurred. Classification of these genes by function will identify key biochemical pathways related to adaptive evolution between species. Here we will describe the procedure, the results and the challenges of a study based on partial sequences, with the aim of informing similar studies within a species. The list of genes we have examined is not exhaustive, but the steps we have taken provide fundamental insight on identifying SNP loci involved in adaptive evolution, and insight on the importance of classes of genes in species evolution.

---

## Development of 55 novel SNP assays for sockeye and coho salmon and assessment of their ability to differentiate stocks within the Columbia River

**Nathan Campbell**, Shawn Narum, Columbia River Inter-Tribal Fish Commission, USA

Genomic and EST sequences from closely related salmonids (Chinook salmon, and rainbow trout) were used to design primers for amplification and sequencing of sockeye and coho DNA. One-hundred and six primer sets were designed and tested for amplification in each species. A panel of 32 diverse individuals from each species was used as template for PCR amplification of working primer sets. The products were sequenced using an Applied Biosystems 3730 capillary electrophoresis instrument and the sequences were analyzed using Sequencher v4.7 (GeneCodes). In total 20,784 bases of consensus sequence were screened in coho and 21,647 bases in sockeye with 149 and 93 SNP sites identified respectively. Sixty four SNP sites were chosen for assay development and 55 of the assays were validated by comparison of the genotyping data to the sequencing data. The validated SNP assays (coho = 32; sockeye = 23) were used to genotype collections of sockeye and coho from various sites in the Columbia River, located in the Pacific Northwest of the USA. Descriptive statistics such as  $F_{st}$  and  $G_{st}$  were used to rank informativeness of each assay.

---

## FishPopTrace- An integrated European programme for tackling illegal, unregulated and unreported fishing

### Gary Carvalho

The FishPopTrace Consortium (<https://fishpoptrace.jrc.ec.europa.eu/>), Bangor University, UK

Although exploited fishes have traditionally been managed on a geographic basis, for conservation purposes they should be managed at the population level: the extent and dynamics of population structuring underlies resilience and sustainability. More effective enforcement and conservation demands a focus on identification and monitoring of wild fish populations and traceability of products. An introductory overview of the European consortium project, FishPopTrace is presented. The programme brings together expertise in fish biology, genetics and forensics to improve the traceability of fish and fish products and protection of consumer interests through enhanced understanding of the dynamics, temporal stability and distribution of major populations of four key exploited fish species: Atlantic cod (*Gadus morhua* L.), European hake (*Merluccius merluccius* L.), Atlantic herring (*Clupea harengus* L.), and common sole (*Solea solea* L.). Three primary traceability tools will be developed to incorporate both on-board sampling (otolith microchemistry and morphometrics) and sampling throughout the food supply chain (single copy nucleotide polymorphisms, SNPs). The latter are the only current widely-used molecular tool that will allow detection of population variability across a range of spatial scales, and yet has sufficiently high reproducibility and robustness for forensic validation. The framework provided by FishPopTrace will thereby enhance the European Common Fisheries Policy aim to promote sustainability through conservation of genetic resources. Incorporation of population biodiversity (“biocomplexity”) into management instruments and policies will further underpin an ecosystem-based approach to fisheries through increased potential for recovery of declining stocks and associated resilience in trophic interactions.

---

## Lessons gained from characterizing SNPs in Atlantic salmon

**Willie Davidson**<sup>1</sup>, Krzysztof Lubieniecki<sup>1</sup>, Lisa Howard<sup>1</sup>, Ben Koop<sup>2</sup>, Matthew Kent<sup>3</sup>, Sigbjorn Lien<sup>3</sup>  
Simon Fraser University<sup>1</sup>, University of Victoria<sup>2</sup>, Canada, CIGENE<sup>3</sup>, Norway

One of the goals of the Consortium for Genomic Research on All Salmonids Project (cGRASP) was to construct a very dense genetic map for Atlantic salmon. This called for identifying and validating thousands of SNPs that could be placed on the microsatellite based linkage map. These SNPs will be used in breeding programs to identify quantitative trait loci (QTL) and selectable haplotypes as well as for population studies and the assignment of a salmon to its natal river. Several approaches were taken to identify putative SNPs. These include: searching the extensive Atlantic salmon EST database (~500,000 ESTs from different tissues and individuals); re-sequencing BAC end sequences from a selection of individuals from different populations; 454 sequencing of overlapping BACs from minimum tiling paths; and sequencing reduced genome complexity libraries prepared from a group of individuals representing populations of interest. All of these approaches produced many putative SNPs, but the key is in their validation. Analysis of the Atlantic salmon genome is complicated by its fairly recent whole genome duplication as well as the abundance of repetitive element families. A wide variety of techniques were used to test and validate putative SNPs. These ranged from: a simple low throughput method involving PCR followed by restriction enzyme challenges and agarose gel electrophoresis; medium throughput methods such as single TaqMan assays; the higher throughput BioTrove array; and finally the use of an Illumina 16,000 SNP chip. Among the many take-home messages we have learned is the need to assess putative SNPs using families to follow their transmission. This identifies single locus SNPs, SNPs from duplicated loci and putative SNPs from paralogous sequence variants (PSVs) and multi-sequence variants (MSVs). The PSVs and MSVs are the products of gene families and repetitive elements, and should be discarded from a final working list of SNPs. The different technologies allow different questions to be answered concerning population dynamics, fisheries management and the selection of broodstock. However, all depend on a suite of robust, well-characterized SNPs. The current tally for Atlantic salmon is approximately 6,500. Lessons gained from characterizing SNPs in the Atlantic salmon genome will be shared.

---

## Single nucleotide polymorphism (SNP) marker development in *Scaphirhynchus* sturgeons

**Jennifer Eichelberger**, Matthew Krampe, Edward Heist  
Southern Illinois University USA

The pallid sturgeon (*Scaphirhynchus albus*) is a federally endangered species endemic to the Missouri and Mississippi river drainages in the United States. It is found in sympatry throughout its range with the more common shovelnose sturgeon (*S. platorhynchus*). Discrimination of adults is increasingly difficult in the southern parts of their shared range due to convergent morphology, perhaps due in part to hybridization. Larvae cannot be reliably distinguished based on morphology. A panel of 19 DNA microsatellites markers is currently used to assign species, with some individuals displaying intermediate genotypes indicative of hybridization and likely backcrossing. We are developing a panel of Single Nucleotide Polymorphism (SNP) markers as a tool for more efficient genotyping of *Scaphirhynchus* for species identification and further analyses of population structure. Several aspects of sturgeon biology have made SNP marker development difficult. Sturgeons are very slow evolving and exhibit very little sequence divergence among species. Karyological evidence suggests that all extant sturgeons are derived from a tetraploid ancestor. All nuclear gene loci examined to date occur as pairs of isoloci with much greater divergence among pairs of loci than between alleles within loci. Nevertheless, the paired isoloci are similar enough to make design of locus-specific assays challenging. Expressed sequences have shown inadequate variability for SNP assay design. SNP discovery has been achieved by obtaining sequences from a *Scaphirhynchus* cDNA library and designing Exon Priming Intron Crossing (EPIC) primers to amplify introns from genomic DNA. Loci are isolated by additional cloning and sequencing before genotyping assays can be developed. If both isoloci of a gene are expressed, high throughput resequencing of cDNA from multiple individuals would reveal SNPs that distinguish isoloci but would not identify the small fraction of SNPs that distinguish only alleles within a single locus. If only one isolocus is expressed and thus represented in a cDNA library, successful locus-specific primer design is unlikely and cloning would still be required to separate isoloci. To date, we have not invested in a large-scale resequencing effort. Despite these difficulties, we have developed several single-locus polymorphic SNP loci that exhibit allele frequency differences between species.

---

## The Search for the Kern River Rainbow Trout

**Bjorn Erickson**, Molly Stephens, Bernie May  
UC Davis USA

The Kern River rainbow trout (“KRRT” - *Oncorhynchus mykiss gilberti*) is a subspecies of rainbow trout endemic to the mainstem of the Kern River in California, USA. Over the past century, angling interest has prompted the introduction of many different strains of rainbow trout into KRRT’s habitat. These introductions have included subspecies native to other sections of the Kern River basin (*O. m. aguabonita* and *O. m. whitei*), as well as coastal and hatchery-raised rainbow trout. With all strains readily interbreeding, introgression has become rampant, making identification of “pure” KRRT difficult. Due to these concerns about hybridization, KRRT has been listed as a California Species of Special Concern. As interest in conserving native fish taxa grows, accurate detection of introgression throughout KRRT’s range has become vital, with genetic tools as the best chance for success. Our research uses diagnostic SNP markers to distinguish KRRT from all of the other strains known to have been stocked into the Kern River. SNPs were designed by comparing sequence data from all strains of interest at both nuclear and mitochondrial loci. Analysis has shown variable levels of influence from non-KRRT strains throughout the habitat, with some populations appearing to be good candidates as “pure” KRRT.

---

## Ultra short reads and non-model species: Exploring the complexities next generation sequence data in the absence of a reference genome

**Meredith V Everett**  
University of Washington, USA

How practical is gene annotation and SNP discovery in a non-model species using ultra short read sequences? Next generation sequencing (NGS) technologies are being applied to an increasing number of species with no reference genome. Of the three primary NGS platforms available the Roche 454 or GFLX sequencing platform produces the longest reads. However, while the ultra short reads produced by both the ABI SOLiD system and the Illumina Genome Analyzer present challenges for assembly, one attraction of these technologies is the sheer volume of data produced at relatively low cost. For labs working with non-model species, the costs, availability of genetic resources including sequences from closely related species, sequence complexity, and the difficulties of performing de novo assembly must all be considered when selecting a sequencing platform. The goal of present study is to examine the feasibility and optimal methodology for SNP and gene discovery in the tetraploid sockeye salmon (*Oncorhynchus nerka*), using ultra short read sequences. SOLiD short reads were generated from single and pooled tissue transcriptome libraries from 10 *O. nerka* individuals. The selected individuals were from 5 distinct populations of interest, spanning a large geographic range. As no reference genome is available for *O. nerka*, resulting sequences were aligned to publically available EST references sequences from *O. nerka*, and two closely related species, rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). Additionally, de novo assembly of the SOLiD reads was carried out utilizing both the CLC Genomics workbench and Oases, a de novo transcriptome assembler for ultra short reads. While assembling sequences to the public databases may limit the discovery of novel transcripts, it allows efficient identification of genes shared between species, as well as identification of informative genetic markers, including SNPs. The results from each reference assembly were compared across species, as well as to the results of de novo assembly to determine the feasibility and optimal methodology for gene and SNP discovery using ultra short reads in a non-model species.

---

## Alaska pollock (*Gadus chalcogrammus*) off the coast of northern Norway - evidence for an Atlantic-Pacific link revealed by SNPs and microsatellites

**Svein-Erik Fevolden**, Kim Præbel, Helge Meissner, Jørgen S. Christiansen  
University of Tromsø Norway

Recently, the gadoid fish Berlevågfisk, *Theragra finnmarkica* (Kofoed, 1956), off northern Norway was acknowledged taxonomically indistinguishable from the Pacific Alaska pollock, *Theragra chalcogramma* (Pallas, 1814). Fishes of Pacific and Atlantic origin were compared both by sequencing the mitochondrial cytochrome c oxidase I (COI) gene and by extensive morphological analysis. Although a few morphological traits differ between the Pacific and Atlantic populations, this can be ascribed ontogenetic plasticity and environmental adaptations rather than genetic differences. Furthermore, given the genetic similarity between the genera *Theragra* and *Gadus*, *Gadus chalcogrammus* has been suggested as the common scientific name for the two populations. The first encounter of *G. chalcogrammus* off northern Norway dates back to 1932, and only about 65 adult specimens have been recorded since then. So, the mere existence of a very small but apparently sustainable population of *G. chalcogrammus* in the NE Atlantic remains enigmatic.

The question how and when was the NE Atlantic population established is in itself a mystery. Having dismissed deliberate introductions from the Pacific to the Atlantic, the outstanding question concerns the possible natural migration routes. One way to approach this question is to examine whether some of the Pacific populations are genetically more similar to the NE Atlantic population than others. Thus, the available specimens of *G. chalcogrammus* from Norwegian waters are presently being compared with specimens of Alaska Pollock from each of the western and eastern North Pacific populations for variation at a number of SNP and microsatellite loci. A closer genetic resemblance to either of these would at least indicate possible migration routes between the N Pacific and the NE Atlantic.

---

## Development of SNP panels for intergenerational tagging and stock ID of salmon and trout

**John Carlos Garza**, Anthony Clemento, Alicia Abadia-Cardoso  
NOAA Southwest Fisheries Science Center/University of California at Santa Cruz USA

The use of genetic data for fishery management and ecological investigation has been hampered by the lack of sufficiently powerful genetic techniques and markers to identify individuals consistently to the population and cohort of origin. The use of 96 locus SNP panels for genotyping of salmon and trout (and other species) allows unprecedented opportunity to reorganize fish population genetics to consider each genotype first as a biological tag that can identify each individual's parents or offspring if they have been sampled and genotyped with the same set of SNP markers, a technique termed parentage-based tagging (PBT). Those genotypes without pedigree relationships can then be used in traditional mixed stock and individual assignment analyses.

We describe the development of novel SNP markers in salmon and trout through targeted resequencing of ESTs in a balanced ascertainment panel of 24 individuals. We describe the discovery of over 1000 variable sites in Chinook salmon *Oncorhynchus tshawytscha* and *O. mykiss* and the development of over 300 TaqMan assays.

From these novel SNP markers, and others discovered by Genetic Analysis of Pacific Salmonids collaborators, we have developed panels of 96 markers in these two species that have considerable power for both PBT and traditional stock identification analyses in a wide variety of salmon and trout populations from the Central Valley of California through the Colombia River basin in the Pacific Northwest and beyond. We also discuss the benefits of adopting standard sets of SNP markers as genetic tags so as to develop large multi-user tagging databases.

---

## SNP discovery from transcriptome resequencing in Black Cottonwood (*Populus trichocarpa*), a biofuels feedstock

**Armando Gerald**<sup>1</sup>, Ji Yang<sup>1</sup>, Jan Jan Hannemann<sup>2</sup>, Nina Thiessen<sup>3</sup>, Johnson Pang<sup>3</sup>, Michael Friedmann<sup>1</sup>, Juergen Ehling<sup>2</sup>, Carl J. Douglas<sup>1</sup>, Quentin Cronk<sup>1</sup>  
University of British Columbia<sup>1</sup>, University of Victoria<sup>2</sup>, BC Genome Sciences Centre<sup>3</sup>, Canada

Recent advances in sequencing technology enable unprecedented amounts of data from natural populations to be collected. In this study we take advantage of these new advances to undergo the first genome wide survey of nucleotide polymorphism in the black cottonwood (*Populus trichocarpa*). The black cottonwood is a tree native to the Northwest of North America, ranging from California to Alaska and displaying a wide range of adaptations to the different habitats where it thrives. It has a very fast growth rate even in marginal lands, a fully sequenced genome and is therefore ideal making it ideal for “accelerated domestication” scientific endeavors. Whole mRNA was extracted from developing xylem and sequenced at high depth. In order to maximize genetic diversity we sampled 20 accessions, 12 Southern genotypes (44-54° latitude) and 8 Northern genotypes (58-59° latitude). Data were assembled onto the reference Nisqually 1 genome sequence and polymorphisms in each individual were identified. Close to 10,000 transcripts were sequenced at a depth of at least 10X across all genotypes. We detected almost 400,000 SNP segregating in our sample. The majority of these were located in exons but ~30% were located in sequences annotated as introns suggesting that many novel transcripts in this species are still to be described. To estimate our error rate in SNP and genotype calling we performed Sanger resequencing in a number of genomic locations and found that these differed considerably between exons and introns. Whole genome SNP data was used to estimate the level of population differentiation between the North and Southern groups.  $F_{st}$  overall was very low likely due to high levels of gene flow, but the distribution had a long tail of moderately high  $F_{st}$  values suggesting that natural selection has shaped allele frequencies at some loci. We also looked at insertion and deletion (indel) polymorphisms in our data set and found that their frequency decreased with size. Small indels (1bp) are common while large indels are rare. Once more the pattern was markedly different between exons and introns. Indels with size multiple of 3bp were overrepresented in exons (as expected) and these might be functionally relevant. We genotyped six such polymorphisms in a large sample (~100 genotypes) and found that they were indeed polymorphic but few alleles were found.

---

## Single nucleotide polymorphisms unravel hierarchical divergence among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations

**Daniel Gomez-Uchida**<sup>1</sup>, Matthew Smith<sup>1</sup>, Chris Habicht<sup>2</sup>, James E. Seeb<sup>1</sup>, Lisa W. Seeb<sup>1</sup>  
School of Aquatic and Fishery Sciences, University of Washington<sup>1</sup>  
Alaska Department of Fish and Game<sup>2</sup>, USA

In spite of the broad utilization of single nucleotide polymorphisms (SNPs) among model organisms, applications of this increasingly popular marker to resolve ecological and evolutionary questions in nonmodel species have been comparatively scant. Here we typified 31 *Oncorhynchus nerka* spawning populations ( $n = 3945$ ) from the Kvichak River drainage in southwest Alaska, USA, using a panel of 45 SNPs. Our main goal was to jointly evaluate the efficacy of SNPs in describing geographic genetic patterns in this spatially structured population system, and the extent to which putatively neutral or selected (outlier) SNPs were responsible for such patterns. Although exploratory genome scans initially suggested eight outliers, only two SNPs found within the Major Histocompatibility Complex (MHC) class II locus deviated consistently at varying spatial scales, suggesting higher estimates of divergence than under neutral expectations. Population differentiation between SNP datasets with and without outliers was nonetheless significantly correlated ( $R = 0.97 - 0.98$ ) and implied interlake divergence greatly exceeded intralake differences among locations, a result validated through hierarchical AMOVA and individual-based clustering. Yet, we uncovered substantial intralake genetic structure, in particular for the large Iliamna Lake, wherein genetic isolation-by-distance plots indicated dispersal was more likely to occur among populations (i) spawning in similar than different habitat types (tributaries or beaches) and (ii) reproducing at similar than different times during the year, as evident in one highly differentiated population that reproduces 'late' in the season. We argue about the relative merits of neutral vs. adaptive divergence explaining these findings, the importance of local adaptation at varying spatial scales, and the implications of hierarchical structure in the interpretation of tests for selective neutrality.

---

## How far can a SNP go? Testing the power of SNPs in fisheries management, a case for use in Orange Roughy (*Hoplostethus atlanticus*)

**Anders Gonçalves da Silva**, Phillip England  
CSIRO Marine and Atmospheric Research, Australia

Two important parameters in fisheries management are (1) number of stocks and (2) degree of connectivity among them. Due to the difficulty of observing exploited marine fish species such as Orange Roughy (*Hoplostethus atlanticus*), genetics is critical to obtaining reliable estimates of these parameters. Efforts to identify stock structure and connectivity in Roughy, found from the North Atlantic to New Zealand, have yielded conflicting results – from global panmixia to localized structure. This situation is not unique to Roughy, a consequence of large effective population sizes ( $N_e$ ) in marine fishes, which limits the power to measure such parameters. Single nucleotide polymorphisms (SNPs) promise to bring vastly greater power to such studies. Here, we use simulations to examine the power of SNPs to identify structure in fisheries-based scenarios. A number of parameters can affect the sensitivity of SNPs, including: effective population size, migration rate, local adaptation, genetic drift-gene flow equilibrium, sample size, minor allele frequency (MAF) threshold, number of SNPs, and how SNP are collected (in linkage or independent). Of these, only the last four are controllable by the investigator. In this first step, we employ a factorial design to examine the ability of 3K neutral SNPs to detect different effective population sizes and migration rates using varying sample sizes.

---

## SNPs in fishery management: Patterns of stock composition in the Port Moller Test Fishery in 2006 - 2008

James E. Seeb<sup>1</sup>, Michael Link<sup>2</sup>, Tim Baker<sup>3</sup>, **Christopher Habicht**<sup>3</sup>

University of Washington<sup>1</sup>, Bristol Bay Science and Research Institute<sup>2</sup>, Alaska Department of Fish and Game<sup>3</sup>, USA

Sockeye salmon returning to spawn in Bristol Bay, Alaska drainages support the most valuable Pacific salmon fishery in the world. Almost all the fish return within a few weeks to the nine major drainages, but pre-season forecasting the return to each drainage is imprecise. As a result fishery managers must be conservative in allowing the fleet to fish. The Port Moller Test Fishery samples fish returning to Bristol Bay when they are about a week away from entering the fishing districts which are located at the mouths of the major drainages. The test fishery provide some insight into the number of fish returning to Bristol Bay, but little information on their drainage destination. We used genetic stock identification (GSI) methods using 38 SNPs to determine the stock composition (to drainage level) of mixtures for sockeye salmon returning to Bristol Bay. We used these GSI methods to analyze mixtures from the Port Moller Test Fishery in-season during 2006-2008 to assess relative run strength to help shape fisheries openings. Here we compare stock compositions, and associated CPUE, to inshore returns to better understand spatial (distance from shore) and temporal (early-mid-late season) influences on stock composition.

---

## The A and C of SNPs in parentage and kinship

**Lorenz Hauser**, Ray Hilborn, Lisa Seeb, Jim Seeb

School of Aquatic and Fishery Sciences, University of Washington

Because of easy standardization and low genotyping error, SNPs have great potential for long-term, multigenerational parentage studies reconstructing pedigrees in wild populations for evolutionary studies. However, lower variability compared to microsatellites may limit the power of parentage assignment and kinship detection. Here, we applied a newly developed set of 83 nuclear SNPs in a small (N=300-500) sockeye salmon population (A creek) in the Wood River system, Alaska, and compared results to data from 12 microsatellites. We used two alternative approaches, a likelihood method for kinship inference, both with and without parental data (Colony) and pairwise likelihood for parentage assignment (Cervus). Data from both markers were used to estimate  $N_e$  and to identify immigrants. Kinship reconstruction in the offspring was compared with pedigrees reconstructed from the parentage analysis. Results from genetic assignments were verified by tagging data showing arrival date, life span, position in the creek and three mtDNA SNPs. Data from both marker classes showed high assignment success and biologically feasible parentage identification.

---

## *in silico* validation: a viable approach to re-sequencing?

**Sarah Helyar**

The FishPopTrace Consortium, Bangor University, UK

FishPopTrace is a European Union funded project which aims to develop a suite of SNP and otolith-based traceability markers for four commercially important species of marine fish; cod (*Gadus morhua*), herring (*Clupea harengus*), hake (*Merluccius merluccius*) and sole (*Solea solea*). SNP discovery for three of these species has been carried out using next generation (454) sequencing.

A novel, entirely *in silico*, SNP validation strategy was employed. The current approach for validating large SNP panels in non-model species generally entails screening and re-sequencing putative markers, which is both time-consuming and costly. Our approach eliminated the process of individual SNP validation in the laboratory. Instead, the panel of putative SNPs were converted directly into a high-throughput genotyping array (using the Illumina Golden Gate© platform).

A critique of the *in silico* approach is presented, together with a consideration of potential improvements in performance. An initial analysis of discriminatory power at the population level of the first generation SNP panel is discussed, highlighting those markers exhibiting signatures of selection.

Ultimately, a sub-set of the most informative markers will be chosen to provide a rapid, rigorous and efficient platform for the prevention of Illegal, Unregulated and Unreported (IUU) fishing, the enforcement of stock conservation measures and consumer protection.

---

## Characterization of new single nucleotide polymorphisms in candidate genes for growth and reproduction in Atlantic cod, *Gadus morhua*

**Jakob Hemmer-Hansen**, Einar Eg Nielsen, Dorte Meldrup

Technical University of Denmark,

The search for markers under selection in non-model species is currently receiving increased attention in evolutionary research. Here, we report the characterization of 85 single nucleotide polymorphisms (SNPs) and develop a genotyping assay for 35 SNPs in 18 candidate genes for growth and reproduction in Atlantic cod (*Gadus morhua*). These markers should be useful for scanning natural populations for signals of selection in both contemporary and historical samples, for example in retrospective studies assessing the effects of environmental changes, such as fishing pressure and temperature, through time.

---

## Comparison of SNP and microsatellite markers for application of genetic stock identification for Chinook salmon in the Columbia River Basin

**Jon Hess**, Andrew Matala, Shawn Narum  
Columbia River Inter-Tribal Fish Commission, USA

Genetic stock identification (GSI) is increasingly being applied to fisheries management issues, due to a combination of its ability to estimate stock proportions of hatchery and natural origin fishery mixtures and its overall cost-efficiency. The number and type of genetic markers available for GSI applications have also increased, and determining the ideal suite of markers to use for a particular question is an ever present challenge. In this study, we tested the performance of two genetic marker sets, 13 microsatellite and 94 SNP loci, for an application of GSI. We employed a newly developed genetic baseline for Chinook salmon consisting of 51 collections distributed across the Columbia River Basin in the Pacific Northwest, USA. This baseline was genotyped for both marker sets to analyze a real fishery mixture (n=3240) representing the total spring, summer, and fall-run of Chinook salmon passing Bonneville dam. We evaluated each marker set based on three criteria: 1) The level of spatial scale at which population structure could be detected among baseline collections based on clustering with neighbor-joining dendrograms and pairwise differentiation 2) level of correct assignment achieved in mixture simulation tests, and 3) average levels of individual assignment and span of confidence intervals around stock assignment proportions achieved with the real fishery mixture. Finally, we ranked all genetic markers according to their information content, and conducted these evaluations with only top-ranked markers. To determine the minimum loci required for best power, we compared the top-ranked markers with the microsatellite and SNP marker sets alone. Previous studies have tested fewer SNP markers on a much broader spatial scale. Here we evaluate a greater number of markers to identify which ones are most powerful at not only distinguishing among stocks of Chinook in the Columbia River Basin, but also discriminating these stocks according to their peak run-timing.

---

## Development and application of an Atlantic salmon SNP chip

**Matthew Kent**, Paul Berg, Sigbjørn Lien  
CIGENE, Norway

Data generated with high-throughput genotyping technologies can be used to measure variation within and between populations, characterize genetic variation, fine map Quantitative Trait Loci (QTL) and, with the implementation of Whole Genome Association (WGA) studies, may be used to assist breeding programs through genomic selection. A successful genotyping product is characterized by having a large number of evenly distributed and informative SNPs that can be robustly genotyped. At Cigene we have constructed an Illumina iSelect bead-array containing probes for 15,225 Atlantic Salmon putative SNPs. These were gathered from three sources, (i) the alignment of EST sequences (n=8559), (ii) the resequencing and alignment of PCR amplifiable sequences derived from mitochondrial DNA and BAC end sequences (n=277), and finally (iii) Genome Complexity Reduction (GCR) (n=6389). In this presentation we will describe the development of the 15K Salmon SNP chip and discuss the subsequent challenge of establishing SNP validity.

One of the first uses of this array has been the genotyping of 2760 individuals belonging to 263 families. The resultant data has been used to construct 29 male (1806cM) and female (2519cM) linkage groups containing 5443 SNPs. These linkage maps will be presented along with preliminary results from the International Salmon HapMap project. This activity has involved the collection and genotyping of 1517 samples from 52 river/lake populations located throughout Europe and Eastern North America and should provide information on genetic diversity, population structure, LD estimation, signatures of selection etc. in Atlantic Salmon.

---

## SNPs detection and genotyping in oysters

**Sylvie Lapègue**<sup>1</sup>, Serge Heurtebise<sup>2</sup>, Estelle Harrang<sup>3</sup>  
Benjamin Morga, Emilie Flahauw, Christopher Sauvage, Pierre Boudry  
Ifremer France

During the last five years, a great effort has been made to develop genomic tools in oyster species all over the world. In Europe, the Pacific cupped oyster, *Crassostrea gigas*, and the European flat oyster, *Ostrea edulis* received special attention as they are the key species for local aquaculture. The Pacific oyster in particular is a marine bivalve of major economical and ecological importance worldwide. Better knowledge of its physiology and its genetic diversity contributes to the optimization of its aquaculture production using selective breeding, or to an understanding of invasive behavior in some coastal areas. Currently, international development of genomic resources include EST databases, BAC libraries, genome mapping and sequencing. The Genoscope (Evry, France) and the Max Plank Institute (Berlin, Germany) have notably performed sequencing projects, as part of the collaborative work developed within the EU funded projects ("Marine Genomics Europe": [www.marine-genomics-europe.org](http://www.marine-genomics-europe.org) and "Aquafirst": <http://aquafirst.vitamib.com>). These efforts led to the release of a novel dedicated EST database ([www.siginae.org.aquafirst](http://www.siginae.org.aquafirst)), allowing the development of additional markers (microsatellites and SNPs), and transcriptome studies. SNPs were detected in candidate ESTs that were differentially expressed between lines selected for high or low survival to summer mortality in microarray studies. A very high level of polymorphism was observed. Moreover, a consensus linkage map was built using SNPs and microsatellites markers and allowed the detection of quantitative trait loci for better survival to summer mortality and Ostreid Herpes virus type 1 load, a virus involved in summer mortality. This kind of approach is of particular interest as the function of candidate genes can be indirectly investigated by performing association studies between polymorphic candidate genes and phenotypic traits. Furthermore, at the population level, genome scans are performed to reveal the footprints of natural selection by identifying loci showing higher genetic differentiation than neutral ones. SNPs have been similarly developed in the European flat oyster. Thanks to the INTERREG funded projects "Arc Atlantic Aquaculture Group" ([www.arcaqua.org](http://www.arcaqua.org)) a genetic map is now available, and QTL for resistance to the parasite *Bonamia ostreae* were identified. SNPs detected in candidate genes differentially expressed in SSH libraries (obtained in experimentes where oysters were challenged by *Bonamia ostreae*) will be used to obtain higher density linkage maps and more precision in QTL mapping. Although SNP detection is becoming easier and more efficient - thanks to the availability of numerous sequences - the challenge remains in the high throughput genotyping of markers in very polymorphic species.

---

## Using custom-designed capture arrays and next-generation sequencing to streamline SNP discovery: An example from the study of global population structure of fin whales (*Balaenoptera physalus*)

**Matthew Leslie**<sup>1</sup>, Amy Frey<sup>2</sup>, Brittany Hancock<sup>2</sup>, Kelly M. Robertson<sup>2</sup>, Phillip A. Morin<sup>2</sup>  
Scripps Institution of Oceanography, Southwest Fisheries Science Center, NOAA Fisheries<sup>1</sup>  
Southwest Fisheries Science Center, NOAA Fisheries<sup>2</sup>

We describe an approach that combines custom micro-array capture technology and next-generation sequencing to ascertain SNPs for use in population structure analyses. The example provided is an ongoing investigation of population genetic structure to determine distinct population segments and subspecies taxonomy within a globally distributed species, the fin whale (*Balaenoptera physalus*). Nuclear locus sequences were initially characterized from a single fin whale DNA sample using conserved mammalian comparative anchor-tag sequence (CATS) primers in a targeted-gene approach. Tiled complimentary sequence probes were designed for a custom micro-array (Agilent) for hybridization-based genomic library enrichment. For fin whales we designed probes that covered the entire mitochondrial genome and fifty nuclear loci. In order to multiplex (N=50 whales per array), individual whale genomes were fragmented and labeled with short unique index tags, creating indexed libraries that were pooled in equal concentrations prior to hybridization capture. Finally, the pooled library was hybridized to the capture array and subsequently eluted and amplified for sequencing on an Illumina Genome Analyzer. This procedure is a streamlined method for collecting large amounts of nuclear and mitochondrial sequence data for SNP discovery and population structure analyses of wild populations.

---

## SNP discovery by 454 sequencing in non-model organisms

**Morten T. Limborg**, FishPopTrace Consortium  
National Institute of Aquatic Resources, Technical University of Denmark

Based on the 454 Next-Generation-Sequencing technology (Roche) a high throughput screening method was devised in order to generate novel genetic markers (SNPs). SNP discovery was performed for three target species of marine fish: hake (*Merluccius merluccius*), herring (*Clupea harengus*) and sole (*Solea solea*). For each species, a total of 7-8 individuals from 4-5 populations, that in most cases spanned major parts of the species' distributions, were individually tagged and subsequently pooled for cDNA library construction. After 454 sequencing reads were de-multiplexed, cleaned, repeat-masked and assembled into high quality contig sequences using the Genomics Workbench ([www.clcbio.com](http://www.clcbio.com)) and Cap3 (Huang and Madan, Genome Res. 1999) softwares. Since no genome reference information was available for any of these species, SNP detection was performed by mapping reads to the contig sequences as a reference using GigaBayes (<http://bioinformatics.bc.edu/marthlab/GigaBayes>). From the predicted polymorphic sites potential SNPs were filtered for intron-exon boundaries and visually inspected. We selected 1,536 candidate SNPs which were then used to design probes for validation by large-scale genotyping assays (Illumina Golden Gate platform). Successfully genotyped SNPs will be used for studying genetic population structures, with special focus on detecting signatures of divergent selection at functionally important genomic regions. An aim is to compile the most informative markers into a cost-effective tool for monitoring fish populations and tracing fish samples and products.

---

## High-throughput melt curve analysis for verification and genotyping of SNPs in 4 Alaska populations of chum salmon

**Ken Livak**<sup>1</sup>, Lisa Seeb<sup>2</sup>, Jim Seeb<sup>2</sup>, Jun Wang<sup>1</sup>, Krishna Datta<sup>1</sup>, Carita Pascal<sup>2</sup>  
Fluidigm Corporation<sup>1</sup>, University of Washington<sup>2</sup>, USA

Next generation sequencing has dramatically increased the availability of SNPs for studies of population genetics. Cost effective laboratory procedures at sufficient throughput for SNP verification and genotyping have not kept pace. At the same time, the use of SNPs for tackling very difficult population problems has become very appealing for fish management and conservation. North Pacific Anadromous Fish Commission nations are currently faced with the daunting task of identifying populations of chum salmon as they migrate through a changing and patchy marine environment. We describe newly developed melt curve analyses used to rapidly and inexpensively verify putative SNPs identified through next generation sequencing. Using the power of ascertainment bias, we focus these efforts on four major and very closely related sets of populations inhabiting the eastern Bering Sea. For initial verification, heteroduplex analysis of PCR products is performed to estimate the minor allele frequency for putative SNPs in each of the four populations using 96 fish from each population. This is done to eliminate putative SNPs that either failed PCR or have a low minor allele frequency. For the reduced number of SNPs, genotyping is performed on the same set of individuals using a  $T_m$  shift method that combines allele-specific PCR and melt curve analysis. This eliminates putative SNPs derived from sequencing results of homeologous loci. The result is a set of verified SNPs that can now be tested across all chum populations to assemble a robust set of identification SNPs.

---

## Power of SNPs for estimation of population genetic parameters

**Gordon Luikart**<sup>1</sup>, Tiago Antao<sup>2</sup>, Steve Amish Amish<sup>1</sup>, Matthew Gruber<sup>1</sup>  
University of Montana<sup>1</sup>, Liverpool School of Tropical Medicine<sup>2</sup>, USA

SNPs will replace microsatellites for many applications in fish and wildlife research. However we lack information on the number of SNPs needed to achieve the same power as microsatellites for many research questions. We assess the relative usefulness of SNPs and microsatellites using literature reviews, empirical data, and computer simulations for applications such as detecting selection, identifying hybrids, estimating effective population size ( $N_e$ ), and correlating landscape features with gene flow in landscape genetics. For detecting selection, several SNPs should be genotyped per locus because use of only one SNP often will not detect selection signatures (e.g.  $F_{st}$ -outliers). For  $N_e$  estimation, approximately 100 SNPs gives power similar to 20-40 microsatellites (8 alleles per locus) using either the linkage disequilibrium method or the temporal method. We also present power analyses for identification of F1 hybrids (in Yellowstone wolves), and for detecting genetic discontinuities using landscape genetic approaches and computer simulations.

---

## The distribution of genetic variation among Chinook salmon lineages in the Columbia River Basin: a landscape genetics perspective

**Andrew Matala**, Shawn Narum

Columbia River Inter-Tribal Fish Commission, USA

Chinook salmon (*Oncorhynchus tshawytscha*) life history types in the Columbia River Basin (CRB) in the Pacific Northwest, USA, exhibit a high degree of variability. Three primary Chinook salmon lineages persist in the CRB, including one form occupying the lower Columbia River, and two sympatric interior forms with distinct biological attributes (ocean- and stream-type). Ocean- and stream-types likely originated from separate glacial refuges and have experienced long-term geographic/reproductive isolation ( $F_{ST} = 0.100$ ). Low level variation and evidence of population bottlenecks within the stream-type lineage suggests substantial genetic drift has occurred among populations. We genotyped 52 Chinook salmon populations sampled among lineages, ranging in location from the upper Salmon River to near the Columbia River estuary, using a panel of 96 single nucleotide polymorphism (SNP) assays. Our goal is to investigate the basis of genetic variation beyond the limit of inference possible with neutral marker data. Because SNPs represent potential sites (or are identified from sequenced regions) within functional genes, they retain an advantage as candidate markers for detecting positive selection. This feature provides the added evolutionary perspective of evaluating natural populations based on local adaptation. Our current efforts focus on identifying the correlations between multilocus or locus-specific SNP genotypes and environmental factors associated with Chinook salmon habitat and life histories. Differences in the distribution of SNP allele frequencies across landscapes may help delineate patterns of local and regional adaptation that influence genetic differentiation. Preliminary analyses have identified at least six candidate loci for positive selection among stream-type Chinook salmon, and allele frequencies appear most highly correlated with temperature, elevation, and migratory distance. We did not observe a significant association of isolation by distance (restricted geneflow as a function of population pairwise distances) with our SNP panel.

---

## Incorporating uncertainty into inference of haplotypes and population analysis of linked SNP genotypes

**Phillip Morin**, Eric Archer

Southwest Fisheries Science Center, USA

Using sets of linked SNPs for population analyses can increase statistical power while decreasing the effort needed to identify many SNPs. Because the process of inferring haplotypes from linked SNP genotypes has some degree of uncertainty, there is the potential for errors or bias in analyses if haplotypes are excluded based on low linkage probability. Analysis of three populations of bowhead whales using 12 sets of 2-3 linked SNPs indicated that, as the level of probability used as a cutoff for accepting SNP haplotypes was increased from 0.65 to 1, the heterozygosity decreased in 8-11 of the loci per population, 2-3 loci went from being in Hardy-Weinberg equilibrium (HWE) to being significantly out of HWE, and  $F_{ST}$  and  $G'_{ST}$  changed by 20-30% and 27-43%, respectively. The change in  $F_{ST}$  and  $G'_{ST}$  was not consistent across cutoffs and comparisons, likely a result of differences in linkage probability distributions among populations. In one pairwise comparison, an increasing cutoff level converted a result of significant differentiation to non-significant. We have developed a method of incorporating the uncertainty inherent in inference of linked SNP haplotypes into standard population analyses. Rather than setting arbitrary cutoffs for accepting inferred haplotypes, we sampled haplotypes repeatedly in proportion to their probability (as determined by the program PHASE 2.1) to obtain a distribution of data sets representing the probable distribution of linked-SNP haplotypes in the population, and then calculated statistics of population subdivision from all of the replicates. This results in distributions for each metric, as well as p-values generated from permutation tests, which properly reflect the uncertainty inherent in the SNP haplotype inference.

---

## SNP markers development in polyploid fish species

### **Nikolai Mugue**

Russian Institut of Fisheries and Oceanography (VNIRO) , Russia

SNPs are powerful tools for the assessment of genetic diversity of natural populations, genetic identification of species and for QTL analysis and linkage map construction. However, applicability of SNPs in polyploid species is complicated, because of abundance of duplicated loci often prevents proper assignment of allelic composition. Results of two projects based on development of SNP markers in polyploids will be presented. First, development of species- and population-specific SNP markers for a number of sturgeon species (tetra- and octoploids) will be described, and second, the search for SNP markers for Carp (*Cyprinus carpio*) linkage map construction will be explored.

Intra-individual variation of two regions (exon 1 and intron 4) of a conservative gene coded for vimentin, presented as a single-copy gene in most vertebrate genomes, revealed presence up to 4 copies in low-chromosome numbers ( $2n \sim 120$ ) and up to 8 copies in high chromosome numbers ( $2n \sim 250$ ) in each individual. Comparing allelic composition from few individuals from each of ten species (over 1000 sequences of cloned PCR products) permits the design of SNP markers for identification of several commercially important sturgeon species. However, no fixed species-specific alleles have been found for Siberian (*Acipenser baerii*), Russian (*A. gueldenstaedtii*) and Persian (*A. persicus*) sturgeons. Presence of a common *A.baerii*-specific allele in the Caspian sturgeon species and absence of this allele in the Black/Azov Sea populations supports the hypothesis of pre-glacial hybridization of *A. gueldenstaedtii* with *A. baerii* within the Caspian basin.

Unlike efforts in the sturgeon species, development of SNP markers in common carp (young allotetraploid species) does not require laborious and expensive cloning approaches. Analysis of existing EST databases as well as 454 mRNA sequence data (CarpBase, UoL, UK) reveals paralog-specific clusters for the majority of highly expressed genes. Design of paralog specific primers and subsequent sequencing of each paralogous gene separately allows unambiguous detection of polymorphic sites on each gene copy of the duplicated genome.

---

## Testing for selection and local adaptation with SNP markers

### **Shawn Narum, Jon Hess**

CRITFC, USA

Genome scans with many genetic markers provide the opportunity to investigate local adaptation in natural populations and identify candidate genes under selection. In particular, SNPs are dense throughout the genome of most organisms, and are commonly observed in both coding and non-coding regions of functional genes, making them ideal markers to study adaptive molecular variation. This has become a commonly employed approach in ecological and population genetics studies to detect outlier loci that are putatively under selection. However, there are several challenges to address with outlier approaches including genotyping errors, population stratification and false positives, variation in mutation rate, and limited sensitivity (false negatives). In this study, we evaluate a wide variety of outlier tests to detect candidate loci under selection in both simulated and empirical data sets. Comparisons are made among several approaches for testing outliers and their effectiveness for studies in adaptive variation.

---

## Application of candidate gene SNP's for detection of adaptive population divergence in high gene flow marine species: Insights from targeted genome scans in Atlantic cod (*Gadus morhua*)

**Einar Nielsen**, Jakob Hemmer-Hansen, Nina Therkildsen  
Technical University of Denmark

Classical marine species commonly display low levels of genetic structuring which could suggest limited or lacking adaptive divergence among local populations. Population genomic methods such as “targeted genome scans”, using samples of genetic markers enriched with candidate genes, can provide an efficient shortcut for demonstrating divergent directional selection in local populations, by allowing the identification of outlier loci with inflated levels of genetic differentiation. We have used combinations of random gene associated SNPs and specifically selected candidate gene SNPs for elucidating potential adaptive population divergence in Atlantic cod (*Gadus morhua*) on various geographical levels across the species global distribution. In general, signatures of differential selection have been widespread and apparent for proteins with highly divergent functions. Distributions of allele frequencies at loci allegedly under selection have been complex but correlations with environmental variables such as temperature and salinity have been identified. A particular effort has recently been diverted to SNPs in candidate genes for growth and maturation as these traits show large divergence across the species distribution and have been claimed to be subject to fisheries induced evolution. We present our most recent results from this particular interesting class of genes including temporal genome scans using DNA from historical otoliths. Finally, we point to new avenues of research which are expected to shed further light on natural and human induced evolution in natural populations of high gene flow organisms.

---

## Unlocking the potential of genomic technologies for conservation management and wildlife forensics

**Rob Ogden**  
TRACE Wildlife Forensics Network, UK

Conservation management and wildlife crime enforcement are increasingly relying on genetic techniques to provide data for directing management decisions and to deliver forensic evidence to investigators. Both of these applications require robust molecular markers, informative at the level of the population and the individual, in a wide range of species. SNPs have long been recognized as offering many potential advantages over microsatellite markers; however, the best methods for their discovery, validation and genotyping in non-model species remains an area of novel research and much discussion. The potential availability to wildlife geneticists of deep sequencing platforms and high density genotyping arrays appears to promise an almost infinite source of variable markers for investigating issues ranging from *ex situ* population management to forensic geographic origin identification. This paper will examine the drivers for developing SNP genotyping panels and their potential as applied tools, before examining a range of strategies that are being employed to try to unlock this potential and address current questions in wildlife conservation and enforcement.

---

## Searching for evidence of selection in the genome of a marine fish: hoki (*Macruronus novaezelandiae*)

**Sebastien Rioux Paquette**, Fred Allendorf, Peter Ritchie  
Victoria University of Wellington, New Zealand

Hoki (*Macruronus novaezelandiae*) is New Zealand's most important commercial fish species. Although management of this species assumes a two-stock hypothesis based on documented differences in morphology and life history traits, previous studies with neutral markers have shown little evidence of genetic differentiation among hoki spawning populations. Our microsatellite results support this notion of extensive gene flow between stocks ( $F_{ST}$  near zero). However, one out of nine microsatellite loci exhibits differentiation significantly greater ( $F_{ST} = 0.06$ ) than expected under neutral models; hierarchical Bayesian analyses also indicate strong evidence of a locus-specific effect. In order to further investigate signatures of selection among hoki stocks, we will develop a panel of genome-wide SNPs. We will initially obtain 6x coverage of the hoki genome from Solexa (Illumina) sequencing with 75 bp paired-end reads. These reads will be assembled into short contigs through de novo assembly and alignment with the recently completed cod genome. Mapped contigs of interest (e.g. excluding repetitive elements) will be employed to develop a targeted resequencing array (SureSelect). This array will be used to sequence multiplexed genomic DNA from several individuals, ensuring deep coverage of target regions and thus allowing for SNP discovery and validation.

---

## Detecting green turtle population structure in the Pacific using single nucleotide polymorphisms (SNPs)

**Suzanne Roden**<sup>1</sup>, Peter Dutton<sup>1</sup>, Phillip Morin<sup>1</sup>, George Balazs<sup>2</sup>, Patricia Zarate<sup>3</sup>, I.J. Cheng<sup>4</sup>  
National Marine Fisheries Service, Southwest Fisheries Science Center<sup>1</sup>, National Marine Fisheries Service, Pacific Islands Science Center<sup>2</sup>, Charles Darwin Foundation<sup>3</sup>, National Taiwan Ocean University<sup>4</sup>

We developed and applied a set of nuclear single nucleotide polymorphisms (SNPs) to detect genetic stock structure among Pacific *Chelonia mydas* nesting populations. Sampled populations included Galapagos n=56, Mexico n=74, Hawaii n=136, and Taiwan n=12, to represent eastern, central, and western Pacific regions. A combination of single independent loci and linked loci combined as haplotypes were used for a total of 19 independent markers. Our nuclear markers confirmed significant differentiation between populations in the three Pacific regions.

---

## SNP discovery in Pacific herring using 454 transcriptome sequencing

**Steven Roberts**, Jim Seeb, Lisa Seeb, William Templin, Lorenz Hauser  
University of Washington, USA

Despite two decades of research and extensive restoration efforts since the *Exxon Valdez* oil spill in 1989, Pacific herring populations in Prince William Sound in Alaska, USA, have not fully recovered. While the causes for the initial collapse of Prince William Sound herring are not completely clear, it is essential to identify current population structure to be able to better understand dynamics of recolonization. This information is critical for Prince William Sound restoration efforts. The challenge is that current genetic information does not allow for the increased resolution necessary to distinguish populations in the region. However, the advent of new genetic technologies allowing the sequencing of the entire expressed genome (transcriptome) of non-model species provides the opportunity for large scale genetic marker discovery. The objective for this project is to use single nucleotide polymorphisms (SNPs) discovered by high density sequencing, to identify fine-scale population structure in Pacific Herring. Results of 454 pyrosequencing of Pacific herring liver and gonad libraries will be presented. Reads were *de novo* assembled using CLC Genomics Workbench and over 16,000 putative SNPs were identified. Of particular interest are genes that are known to be involved in pollution and disease response, as it is possible these genes could be under variable selection across populations and could therefore be valuable in identifying specific populations.

---

## An application of SNP markers to stock identification of chum salmon in the Bering Sea during the summer of 2007

**Shunpei Sato**<sup>1</sup>, William D. Templin<sup>2</sup>, James E. Seeb<sup>3</sup>, Lisa W. Seeb<sup>3</sup>, Hiroyuki Nagoya<sup>1</sup>, Shigehiko Urawa<sup>1</sup>  
National Salmon Resources Center, Fisheries Research Agency<sup>1</sup>, Alaska Department of Fish and Game<sup>2</sup>,  
University of Washington<sup>3</sup>, Japan and USA

Populations of chum salmon from Asia and North America co-migrate in the North Pacific Ocean and the Bering Sea. The Bering Sea is a major feeding habitat for chum salmon during the summer season. Accurate identification of the composition of these migrating mixtures in the summer Bering Sea can provide important monitoring information for management, stock assessment, and conservation. Our objective was to apply single nucleotide polymorphisms (SNP) markers to estimate stock origins. Chum salmon (n=3,849) were collected from 22 stations in the Bering Sea (52.63N-59.38N, 174.97E-170.18W) aboard the Japanese research vessel R/V Hokko-maru between 22 July and 3 August, 2007. Pectoral fin samples were collected and fixed in 100% ethanol. Each sample was assayed for 33 SNP loci by TaqMan chemistry. Stock contributions (Japan, Russia, and North America) of immature and maturing chum salmon were estimated by a conditional maximum likelihood algorithm using SNP baseline dataset from 147 populations from the Pacific Rim. The percentages of immature and maturing chum salmon were 86.5% and 13.5%, respectively. The genetic stock identification (GSI) and GSI-estimated CPUE (catch per unit effort) suggested that most of immature and maturing chum salmon originated from Japan and Russia, and they were widely distributed in the survey areas of the Bering Sea. The estimated abundance of immature Japanese chum salmon was similar to that of immature Russian chum salmon in the central and eastern Bering Sea; however, immature Russian chum salmon were predominant in the southern and western Bering Sea. On the other hand, maturing Japanese chum salmon were dominant in the all survey areas, particularly the eastern and western Bering Sea. Immature and maturing North American chum salmon were generally less abundant than Asian chum salmon in the summer Bering Sea. These results support previous estimates made with allozyme and mtDNA data and indicate that chum salmon stocks from Asia continue to be dominant in the Bering Sea during summer.

---

## Congeneric species identification using TaqMan<sup>®</sup> SNP genotyping assays

**Piper Schwenke**, Jon Baker, Linda Park  
Northwest Fisheries Science Center, USA

Species identification has been an important part of fisheries law enforcement specifically related to endangered salmonids and mislabeled salmon fraud. Currently, sequencing mtDNA, PCR-RFLP, and allele specific assays are most commonly used for identifying species. The TaqMan<sup>®</sup> SNP genotyping assays are a robust and rapid way to identify biallelic genotypes for large numbers of loci and individuals. MtDNA loci accumulate sufficient amounts of diagnostic site mutations between species yet intraspecific polymorphisms pose a challenge for finding conserved regions for primer/probe locations. Intraspecific polymorphism data are required to design robust and reliable species identification assays for congeneric species. Baseline sequence was collected at mtDNA COIII/ND3 in 276 samples across 7 species of *Oncorhynchus* spp. Several samples from closely related genera, *Salmo* and *Salvelinus* were also sequenced to identify *Oncorhynchus* diagnostic sites. A suite of 8 TaqMan<sup>®</sup> SNP custom assays were designed and tested for accurate species identification of known *Oncorhynchus* samples. The benefit of running 8 SNPs over sequencing a single locus is time saved during analysis, especially for large sample sizes. Transforming SNP allele calls to final species identification can be easily automated. Results from these trials are presented here along with design and implementation challenges encountered.

---

## Rapid Advances in SNP Discovery and Genotyping in Non-model Organisms

**Jim Seeb**

University of Washington, USA

Molecular genetic studies provide exceptional insight into relationships, movement and migration, and evolution of natural populations. In the 1970s it became clear that techniques such as allozyme electrophoresis would provide a basic framework for understanding adaptation and conserving natural genetic variability. Shortly thereafter innovators began to dream of potential applications for conservation and management economically exploited species that included using molecular markers to determine the population-of-origin of migrating animals. The decades since were punctuated by improvements in molecular and statistical techniques that produced a torrent of assignment tests, structure and kinship analyses, and now genome scans based upon an ever increasing resolution of individuals and populations. A jump in progress occurred with the marriage of next-generation sequencing for SNP discovery and the development of high-throughput genotyping platforms for SNP analyses (two primary sponsors of SNP Workshop III). In this workshop we will interact and exchange ideas to further fuel advances with SNP discovery and high-throughput genotyping of non-model organisms.

---

## Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon

**Lisa W. Seeb**<sup>1</sup>, William. D. Templin<sup>2</sup>, Shunpei Sato<sup>3</sup>, Syuiti Abe<sup>4</sup>, Kenneth Warheit<sup>5</sup>, and Jim E. Seeb<sup>1</sup>  
<sup>1</sup> School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Seattle, Washington, USA 98112 <sup>2</sup>Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, Alaska, USA 99518 <sup>3</sup>National Salmon Resources Center, Fisheries Research Agency, Sapporo, Hokkaido 062-0922, Japan <sup>4</sup>Division of Marine Biosciences, Graduate School of Fisheries Sciences, Hokkaido University, Hakodate 041-8611, Japan <sup>5</sup>Washington Department of Fish and Wildlife, PO Box 600, Olympia, Washington USA 98501

Studies of the oceanic and near shore distributions of Pacific salmon whose migrations typically span thousands of kilometers have become increasingly valuable as we enter periods of climate change, increasing hatchery production from Pacific Rim nations, and potentially high rates of bycatch in offshore fisheries. However, these types of studies require extensive databases of spawning populations originating from across the species' range. Single nucleotide polymorphisms (SNPs) have been particularly amenable for these multi-national applications because they are easily shared, require little inter-laboratory standardization, and can be assayed through increasingly efficient technologies. Here we discuss the development and standardization of an extensive SNP baseline for chum salmon. The collaboration among United States and Japanese researchers, termed "PACSNP", is focused on three major areas: 1) SNP discovery and genotyping, 2) joint projects on distribution on the high seas, and 3) development of shared and online databases. We also review various issues associated with SNP development across a species' range including ascertainment bias, linkage relationships, and outlier loci.

---

## Single feature polymorphisms, transcript presence/absence and differential expression between strains of *Danio rerio* (Zebrafish).

**Matthew Settles**, Barrie Robison, Terence Soule  
University of Idaho, USA

High-throughput microarray experiments often contain more biological information than required to answer the question for which the experiment was designed. Most microarray experimental analyses focus solely on differential expression. However, when the dataset contains multiple genotypes, additional genetic information can be gleaned from array data. When experiment samples harbor genetic polymorphisms, a subset of these may be detected using new techniques for detection of single feature polymorphisms (SFP). We have developed an improved technique for the identification of single feature polymorphisms in array data. In this study, we use this technique to extract single feature polymorphisms from an Affymetrix zebrafish microarray experiment, comprised of four strains (two wild, two domesticated). To validate identification of SFPs, each of the four strains' transcriptomes was also sequenced using 454 pyrosequencing technology. Analysis of the relationship between SFP and expression reveals insights into the relationship between genetics, transcription and phenotype. Further cis-acting and tran-acting SFP are identified and associated with Gene Ontologies and KEGG pathways and compared to expression.

---

## A 40-year retrospective of catch compositions of sockeye salmon (*Oncorhynchus nerka*) in Bristol Bay

**Matt Smith**<sup>1</sup>, Carita Pascal<sup>1</sup>, Zac Grauvogel<sup>2</sup>, Chris Habicht<sup>2</sup>, Lisa Seeb<sup>1</sup>, Jim Seeb<sup>1</sup>  
University of Washington<sup>1</sup>, Alaska Department of Fish and Game<sup>2</sup>

We are using single nucleotide polymorphisms (SNPs) in archived sockeye salmon scales up to forty-five years old to provide historical catch composition estimates in the fishing districts of Bristol Bay, Alaska, USA. While the application of SNPs to historical samples can provide a unique window into the past by extending the temporal scale of a wide array of population studies, there are several challenges associated with using historical, low-quality tissues. Genetic studies using historical scale collections have typically been limited by low DNA yields, high levels of degradation, and tissue-to-tissue contamination. SNPs offer advantages when using historical, degraded samples because the amplified fragments can be very short. Here we present an efficient, cost effective method to accurately genotype low-quality scale samples on a high-throughput single nucleotide polymorphism (SNP) genotyping platform. We demonstrate that multiplex preamplification of forty-five SNP loci can generate accurate, reproducible genotypes with low drop-out rates. Furthermore, we describe how to screen samples to detect contamination between samples using a suite of four preamplified highly polymorphic microsatellite loci. These methods have resulted in the successful genotyping of over 18,000 individual sockeye salmon scales and have made it possible to examine catch compositions in a fishery that took place up to forty-five years ago.

---

## Searching for SNPs: Mining the sockeye salmon, *Oncorhynchus nerka*, transcriptome for high-resolution molecular markers

**Caroline Storer**, Carita Pascal, Eric Grau, Steven Roberts, Lisa Seeb, Jim Seeb  
University of Washington USA

Advances in molecular techniques and decreasing technology costs have enabled a growing use of molecular markers for the study of non-model organisms. Recently, single nucleotide polymorphisms (SNPs) have begun to replace microsatellites as the molecular marker workhorse of population genetics. SNPs provide varying degrees of resolution across a range of temporal and spatial scales. In order to effectively use SNPs for fisheries population genetics there is a need for novel SNPs with high-resolution, and temporal and spatial stability. Here we describe the use of next generation transcriptome sequence data and high resolution melting techniques to discover and develop new SNP assays for sockeye salmon that are spatially and temporally stable. Over seventy new SNP markers have been developed using these techniques thus far and the resolution of these new SNPs is currently being assessed for a range of populations spanning the Pacific. This SNP discovery workflow can be adopted for the development of high-resolution SNP markers for many non-model organisms.

---

## High-resolution stock identification for migratory studies of Chinook salmon

**William Templin**<sup>1</sup>, Lisa Seeb<sup>2</sup>, James Murphy<sup>3</sup>, James Seeb<sup>2</sup>  
Alaska Department of Fish and Game<sup>1</sup>, International Program for Salmon Ecological Genetics<sup>2</sup>, NOAA - Fisheries<sup>3</sup>, USA

While the life history and ecology of Chinook salmon in freshwater is well known, less is known of the oceanic migration patterns and relative survival of individual stocks in the marine environment. Until recently, investigation of the effects of fluctuating marine conditions on the abundance and distribution of Chinook salmon has only been approachable through the sporadic collection of tagged individuals and analysis of scale patterns. The Alaska Department of Fish and Game has developed a baseline of 53 markers based on single nucleotide polymorphisms (SNPs) surveyed in 175 populations across the species range in the North Pacific. This baseline provides the foundation for the application of genetic stock identification to the high-resolution exploration of the distribution of Chinook salmon in marine waters. We compare this method to the resolution possible from previous methods to demonstrate its utility for mixed stock analysis of high seas samples. Our results show that this baseline provides a rapid and cost effective approach to analysis of samples from complex mixtures encountered in multi-national research and fishery monitoring efforts.

---

## Inbreeding effective size and parentage analysis without parents

**Robin Waples<sup>1</sup>**, Ryan Waples<sup>2</sup>  
NOAA Fisheries, Seattle<sup>1</sup>, Casa Azul, Seattle<sup>2</sup>, USA

In a typical parentage analysis, multilocus genotypes are scored in both parents and progeny, and these data are used to 'assign' progeny to parents. This type of study, made possible only recently by highly polymorphic molecular markers, has produced many novel insights, including the ability to directly calculate the number of genes contributed to the next generation by each parent ( $k_i$ ). From this information, and knowing the total number of parents ( $N$ ), it is possible to use standard formulas to calculate effective population size ( $N_e$ ). But what if your sample of parents is incomplete or missing entirely? In theory, it should be possible to reconstruct parental genotypes based entirely on a sample of progeny. This is not quite feasible yet but might be in the near future if large enough numbers of SNPs become available. If the two parents for each individual in a sample could be identified, it would be possible to construct a vector of parental  $k_i$  values, but this would provide information only about parents that actually contributed offspring to your sample. What about the 'null' parents that produced no offspring in your sample, and how would they affect an estimate of  $N_e$ ? The surprising answer is that null parents have no effect at all. We show the following: 1) It is possible to rewrite the standard formula for inbreeding effective size so that it is only a function of  $\sum(k_i)$  and  $\sum(k_i^2)$ . That is, it is not necessary to know  $N$  to calculate inbreeding  $N_e$ . This same relationship does not hold for variance  $N_e$ . 2) It is easy to prove #1 assuming a complete sample of all progeny. However, even if only a sample of progeny is taken, simulated data show that the estimate of  $N_e$  using the simple formula that ignores null parents is unbiased. This means that parentage analysis without parents can be used to provide an unbiased, single-sample estimate of inbreeding  $N_e$ , provided parental contributions to the offspring sample can be resolved. 3) It is not necessary to actually reconstruct parental genotypes; from a matrix of pairwise relationships (as can be estimated by some current software programs) it is possible to construct the vector of  $k_i$  values and estimate  $N_e$ . 4) Accuracy and precision of the new method based on parentage analysis without parents compares favorably with single-sample estimators of  $N_e$  currently in use.

---

## Population differentiation of Puget Sound Chinook

**Kenneth Warheit**, Todd Kassler, Jennifer Von Bargaen, Sewall Young  
Washington Department of Fish and Wildlife, USA

We genotyped 3700 individual Chinook salmon from 24 hatchery and wild Puget Sound (Washington, USA) populations using 96 SNPs, 75 of which are part of the Genetic Analysis of Pacific Salmonids (GAPS) laboratory consortium standardized SNP Taqman assays. SNPs were genotyped using either the Fluidigm EP1 or Applied Biosystems 7900 systems. Of the 75 standardized GAPS SNPs, 62 had minor allele frequencies greater than 0.10; one assay (Ots\_SERPC1-209) was fixed in all populations. In total 19 of the 96 SNPs (13 of the 75 GAPS SNPs) were eliminated from the analysis as a result of poor data quality, and three populations were removed because all individuals from each population failed at one or more loci. For this presentation, we will use these 77 SNPs to test for genetic differentiation among the 21 populations. We will also determine the power of these SNPs for proportional and individual assignments in mixed-stock fisheries and to assist in the selection of broodstock in hatcheries. For all analyses, we will compare the results using these 77 SNPs with those obtained from analyses using the 13 standardized GAPS microsatellite loci, and with a combined dataset that includes both the SNPs and microsatellites.

---

## Russian pacific salmon SNPs database

**Daria Zelenina**, Diana Stoklitskaya, Dmitry Shchepetov  
Russian Federal Institute for Fisheries and Oceanography (VNIRO), Russia

A SNP database of Russian Pacific salmon populations has been developed. It consists of two parts: the first part contains raw data of genotyping of each specimen and the second one – allele frequencies for populations. About 400 chum specimens from seven hatcheries and one wild population from different regions of Sakhalin Island have been screened for eight SNP loci and the comparison of samples has been performed. Four samples from the eastern part of the island, three – from the south-western and one – from the south-eastern part have been analyzed. In nearest future, the results for further SNPs and additional samples of Russian chum will be added to the database. The similar project for sockeye salmon is now in preparation and will be started this year.

### SNP Workshop III Participants

Name	Organization	City State	Country	Email
Syuiti Abe	Hokkaido University	Hakodate, Hokkaido	Japan	abesyu@fish.hokudai.ac.jp
Mike Ackerman	University of Washington	Seattle WA	USA	mwa44uw.edu
Brice Adams	United States Fish & Wildlife Service	Longview WA	USA	brice_adams@fws.gov
Steve Amish	Montana Conservation Genetics Lab	Missoula MT	USA	stephen.amish@umontana.edu
Eric Anderson	Southwest Fisheries Science Center	Santa Cruz CA	USA	eric.anderson@noaa.gov
Michael Banks	Oregon State University	Newport OR	USA	michael.banks@oregonstate.edu
Renee Bellinger	Oregon State University	Newport OR	USA	renee.bellinger@oregonstate.edu
Paul Bentzen	Dalhousie University	Halifax NS	Canada	paul.bentzen@dal.ca
Elizabeth Boulding	University of Guelph	Guelph ON	Canada	boulding@uoguelph.ca
Ian Bradbury	University of Windsor	Toronto ON	Canada	ibradbur@me.com
Marine Briec	University of Washington	Seattle WA	USA	elpetrou@uw.edu
Matthew Campbell	Idaho Department of Fish and Game	Eagle ID	USA	matthew.campbell@idfg.idaho.gov
Nathan Campbell	Columbia River Inter-Tribal Fish Commission	Hagerman ID	USA	camn@critfc.org
John Candy	Pacific Biological Station	Nanaimo BC	Canada	John.Candy@dfo-mpo.gc.ca
Michael Canino	Alaska Fisheries Science Center	Seattle WA	USA	mike.canino@noaa.gov
Gary Carvalho	Bangor University	Bangor Gwynedd	United Kingdom	g.r.carvalho@bangor.ac.uk
Willie Davidson	Simon Fraser University	Burnaby BC	Canada	wdavidso@sfu.ca
Daniel Drinan	Montana State University	Bozeman MT	USA	ddrinan@montana.edu
Jennifer Eichelberger	Southern Illinois University Carbondale	Carbondale IL	USA	biojen@siu.edu
Bjorn Erickson	UC Davis	Davis CA	USA	pberickson@ucdavis.edu
Meredith Everett	University of Washington	Seattle WA	USA	mav@uw.edu
Svein-Erik Fevolden	University of Tromso	Tromso	Norway	sfe000@uit.no
Karin Fredrikson	Roche Applied Science	Ijamsville MD	USA	karin.fredrikson@roche.com
John Garza	NOAA Southwest Fisheries Science Ctr	Santa Cruz CA	USA	carlos.garza@noaa.gov
Armando Geraldes	University of British Columbia	Vancouver BC	Canada	geraldes@interchange.ubc.ca
Daniel Gomez-Uchida	University of Washington	Seattle WA	USA	dgu@uw.edu
Anders Goncalves da Silva	CSIRO Marine & Atmospheric Research	Hobart, TAS	Australia	anders.goncalvesdasilva@csiro.au
Eric Grau	University of Washington	Seattle WA	USA	grau@uw.edu
Peter Grewe	CSIRO Marine & Atmospheric Research	Hobart, Tasmania	Australia	grewepg@mac.com

### SNP Workshop III Participants

Name	Organization	City State	Country	Email
Chris Habicht	Alaska Department of Fish and Game	Anchorage AK	USA	chris.habicht@alaska.gov
Matthew Hale	Purdue University	West Lafayette IN	USA	mchale@purdue.edu
Jeffrey Hard	NOAA Northwest Fisheries Science Center	Seattle WA	USA	jeff.hard@noaa.gov
Mette Hansen	Aarhus University	Tjele	Denmark	metteh.hansen@agrsci.dk
Lorenz Hauser	University of Washington	Seattle WA	USA	lhauser@uw.edu
Denise Hawkins	United States Fish & Wildlife Service	Longview WA	USA	denise_hawkins@fws.gov
Sarah Helyar	Bangor University	Bangor Gwynedd	United Kingdom	s.helyar@bangor.ac.uk
Jon Hess	Columbia River Inter-Tribal Fish Commission	Hagerman ID	USA	hesj@critfc.org
Maureen Hess	Columbia River Inter-Tribal Fish Commission	Hagerman ID	USA	hesm@critfc.org
Dave Jacobson	Oregon State University	Newport OR	USA	dave.jacobson@oregonstate.edu
Matthew Kent	CIGENE	Aas	Norway	matthew.peter.kent@umb.no
Christine Kozfkay	Idaho Department of Fish and Game	Eagle ID	USA	christine.kozfkay@idfg.idaho.gov
Sylvie Lapègue	IFREMER	La Tremblade	France	slapegue@ifremer.fr
Griffen Lauer	Fluidigm Corporation	South San Francisco CA	USA	griffen.lauer@fluidigm.com
Matthew Leslie	Scripps Institute of Oceanography	La Jolla CA	USA	mleslie@ucsd.edu
Ken Livak	Fluidigm Corporation	South San Francisco CA	USA	ken.livak@fluidigm.com
Patrick Martin		Bozeman MT	USA	pcmartin@montana.net
Andrew Matala	Columbia River Inter-Tribal Fish Commission	Hagerman ID	USA	mata@critfc.org
Molly McGlaufflin	University of Washington	Seattle WA	USA	mollymcg@u.washington.edu
Megan McPhee	University of Alaska Fairbanks	Juneau AK	USA	mvmcphee@alaska.edu
Paul Moran	NOAA Northwest Fisheries Science Center	Seattle WA	USA	paul.moran@noaa.gov
Phillip Morin	National Marine Fisheries Service	La Jolla CA	USA	phillip.morin@noaa.gov
Nikolai Muge	VNIRO	Moscow	Russia	muge@mail.ru
Kerry Naish	University of Washington	Seattle WA	USA	knaish@uw.edu
Shawn Narum	Columbia River Inter-Tribal Fish Commission	Hagerman ID	USA	nars@critfc.org
Einar Nielsen	Technical University of Denmark	Silkeborg	Denmark	een@aqua.dtu.dk
Kathleen O'Malley	Oregon State University	Newport OR	USA	kathleen.omalley@oregonstate.edu
Jenny Ovenden	Molecular Fisheries Laboratory	Brisbane, Queensland	Australia	jennifer.ovenden@deedi.qld.gov.au
Linda Park	NOAA Northwest Fisheries Science Center	Seattle, WA	USA	linda.park@noaa.gov

### SNP Workshop III Participants

Name	Organization	City State	Country	Email
Carita Pascal	University of Washington	Seattle WA	USA	cpascal@uw.edu
Eleni Petrou	University of Washington	Seattle WA	USA	elpetrou@uw.edu
Ruth Phillips	Washington State University	Vancouver WA	USA	phillipsr@vancouver.wsu.edu
Kim Praebel	University of Tromsø	Tromsø	Norway	kim.praebel@uit.no
Clint Rhode	Stellenbosch University	Stellenbosch, Western Cape	South Africa	14313103@sun.ac.za
Jamie Rice	Fluidigm Corporation	South San Francisco CA	USA	jamie.rice@fluidigm.com
Sebastien Rioux Paquette	School of Biological Sciences	Wellington	New Zealand	sebastien.riouxpaquette@vuw.ac.nz
Steven Roberts	University of Washington	Seattle WA	USA	sr320@uw.edu
Suzanne Roden	NOAA-NMFS	La Jolla CA	USA	roden@noaa.gov
Kristen Ruegg	Stanford University	Pacific Grove CA	USA	kruegg@stanford.edu
Michael Russello	University of British Columbia	Kelowna BC	Canada	michael.russello@ubc.ca
Daniel Ruzzante	Dalhousie University	Halifax NS	Canada	daniel.ruzzante@dal.ca
Shunpei Sato	National Salmon Resources Center	Sapporo, Hokkaido	Japan	sato.shunpei@fra.affrc.go.jp
Piper Schwenke	NOAA/NMFS	Seattle WA	USA	piper.schwenke@noaa.gov
Jim Seeb	University of Washington	Seattle WA	USA	jseeb@uw.edu
Lisa Seeb	University of Washington	Seattle WA	USA	lseeb@uw.edu
Matthew Settles	University of Idaho	Pullman ID	USA	msettles@uidaho.edu
Rachel Simmons	UC Davis	Davis CA	USA	resimmons@ucdavis.edu
Christian Smith	United States Fish & Wildlife Service	Longview WA	USA	christian_smith@fws.gov
Matt Smith	University of Washington	Seattle WA	USA	smithm8@u.washington.edu
Caroline Storer	University of Washington	Seattle WA	USA	cgs5@uw.edu
William Templin	Alaska Department of Fish and Game	Anchorage AK	USA	bill.templin@alaska.gov
Morten Limborg Tonsberg	Technical University of Denmark	Silkeborg	Denmark	mol@aqua.dtu.dk
Jesse Tsai	University of Washington	Seattle WA	USA	jes37235@u.washington.edu
Wendy Tymchuk	Fisheries and Oceans Canada	West Vancouver BC	Canada	wendy.tymchuk@dfo-mpo.gc.ca
Fred Utter	University of Washington	Seattle WA	USA	fmutter@uw.edu
Brent Vadopalas	University of Washington	Seattle WA	USA	brentv@uw.edu
Tim Vines	Molecular Ecology Resources	Vancouver BC	Canada	managing.editor@molecol.com

**SNP Workshop III Participants**

<b>Name</b>	<b>Organization</b>	<b>City State</b>	<b>Country</b>	<b>Email</b>
Robin Waples	NOAA Fisheries Service	Seattle WA	USA	robin.waples@noaa.gov
Ryan Waples	Casa Azul	Seattle WA	USA	ryan.waples@gmail.com
Kenneth Warheit	Washington State Dept. of Fish & Wildlife	Olympia WA	USA	kenneth.warheit@dfw.wa.gov
Gajus Worthington	Fluidigm Corporation	South San Francisco CA	USA	gyne.riser@fluidigm.com
Yong Yi	Fluidigm Corporation	South San Francisco CA	USA	yong.yi@fluidigm.com
Daria Zelenina	VNIRO	Moscow	Russia	dz67@mail.ru