ATTACHMENT C **EVOSTC Annual Project Report Form**

Form Rev. 8.30.18

1. Program Number:

18170115

2. Project Title:

Genomic mechanisms that underlie lack of recovery of Prince William Sound herring following the 1990's collapse

3. Principal Investigator(s) Names:

Andrew Whitehead

4. Time Period Covered by the Report:

February 1, 2018-January 31, 2019

5. Date of Report:

April 2019

6. Project Website (if applicable):

http://pwssc.org/herring-research-and-monitoring/

7. Summary of Work Performed:

Overview: The causes of the collapse of the Prince William Sound (PWS) Pacific herring stock are controversial, and the reasons for the lack of recovery remain a mystery. In this series of projects we interrogate the genome structure and genome function of PWS fish to test hypotheses about the causes and consequences of the collapse, by revealing ecological, evolutionary, and genetic mechanisms governing the demographic trajectory of PWS fish over the past ~30 years. Conspicuous events that coincided with the dramatic PWS collapse include the *Exxon Valdez* oil spill four years previous, and the emergence of disease. We test hypotheses concerning the effects of oil exposure, the effects of disease challenge, and the potential interactive effects of oil exposure and disease challenge, on herring health and fitness. Physiological measurements and patterns of genome-wide gene expression will serve to reveal similarities and differences in mechanisms of response to these stressors between PWS and reference population fish. These studies should provide novel insights into the causes and consequences of recent dramatic demographic changes in PWS fish, potentially inform novel intervention strategies, and provide modern genomic resources for management and conservation of Pacific herring. We have performed work on three aspects of the project during this second year of the research program.

1) Animal exposure experiments: This year saw the completion of a massive and complex experiment, which would not have been possible without the hard and highly collaborative work between Tony Gill (Ph.D. student in the Whitehead lab), all members of collaborator Paul Hershberger's lab, partners at the National Oceanographic and Atmospheric Administration (NOAA; Nat Scholz and John Incardona's group), and folks at the Sitka Science Center (including Angie Bowers in particular) and at the Alaska Department of Fish and Game (ADF&G; including Eric Coonradt in particular). This experiment involved exposing three different populations of Pacific herring to a broad range of low concentrations of

oil during development, including populations from Sitka Sound, Prince William Sound, and Puget Sound. These experiments were difficult because we had to secure gametes from spawning fish from each of these regions, transport them back to exposure facilities at the Marrowstone Marine Field Station Fish Health Laboratory, complete fertilizations, and then conduct exposures during development. Since fish mature at different times in each of these regions, these experiments required much coordination. And since a core component of experiments is to assess population differences to oil exposure, we needed exposures to be highly reproducible between time-staggered experiments between populations that were spawning during different weeks in the early spring. Because recovering and shipping gametes live is difficult, and because the Sitka population mysteriously disappeared for several weeks, this required several trips to Alaska. Since experiments with different populations were staggered in time because of spawning time differences, we needed an exposure system that was highly reproducible. This led to our decision to buy and contract the assembly of the SINTEF Exposure System (SINTEF, Norway). This is a highly sophisticated state-of-the-art apparatus that can carefully control dose through a computer-controlled solenoid based dosing system. This was expensive and time consuming to order and build, but I think it was a crucial investment that will pay dividends in the reproducibility of experiments, and ultimately will enable our ability to rigorously compare populations in their developmental response to oil exposure. The animals from these exposure experiments hatched, were grown out to late juvenile stage (at which time their immune system has matured), and challenged with pathogens in October/November 2018. We then initiated a second round of experiments, in which remaining animals have been maintained on a low calorie diet, and will be re-challenged with pathogen in March 2019.

2) Genome sequencing and assembly: I am very pleased to report that we have completed the first draft of a reference genome assembly for Pacific herring. There is some additional work to be completed before we have a final assembly, but in the current state the assembly is sufficient to serve as a mapping reference to support the population genomics read mapping and RNA-seq read mapping. We prepared libraries compatible with the 10X Genomics technology. We received the sequencing data in early spring. We have been working through iterative assemblies, and the current assembly is of very high quality. N50 scaffold sizes are greater than 2 Gb, which is above average contiguity compared to other fish genomes. We anticipate even higher contiguity once we complete additional tasks, including scaffolding from Hi-C libraries, and ordering of scaffolds through recombination mapping. Data have been collected for Hi-C assemblies. Our first pass at incorporating those data into the assembly has provided extremely promising results: our 10X genomics assembly has been ordered into 26 large scaffolds by using the Hi-C data. Since 26 is the expected number of chromosomes, this is a very good sign. There are still some very small and very large scaffolds, which suggests that optimization steps are necessary. The optimization algorithms are currently running, and we should have results by the end of March or beginning of April. Data for the recombination map have been collected, and we will incorporate those data once we have a polished Hi-C assembly. We will be continuing with the genome annotation effort once the Hi-C assembly is complete. Sequencing of a reference transcriptome will contribute to this annotation. Reference transcriptome sequencing has been completed. We used one of the latest technologies – Pacific Biosciences IsoSeq – to generate full-length RNA sequences from a library of mixed adult and early life stage tissues. We have added to this ~1 gigabases of additional RNA sequencing using Illumina reads. The IsoSeq assembly is complete, and will be hybridized with the Illumina data soon. This will be a crucial resource for genome annotation as a mapping reference for the quantitative transcriptomics that we will perform on samples generated from FY19 animal experiments outlined in the previous section. A description of the reference genome, including assembly strategy with the reference transcriptome, and annotation, will be packaged into a manuscript. We anticipate that this will be submitted for publication by the end of this year.

3) Population genomics: Last year we chose a different sequencing strategy than we had originally proposed. Rather than pooled-sample sequencing, we decided to individually index and sequence samples. This is because the cost-per-sequence versus sequence data coverage math had come out in favor of individually-indexed whole genome sequencing with the publication of a new library preparation protocol. We decided that this strategy would offer higher quality data for the same cost as originally proposed. This required us to get a new protocol up and running in the lab, which involved three months of troubleshooting. We finally got this up and running reliably in June 2018. We then created libraries for all ~1,200 samples within one month. The libraries were submitted to the core facility for sequencing in July 2018. Sequencing was completed and data available in October 2018. The result is a massive dataset that includes 20 lanes of Illumina Hi-Seq 4000 paired-end 150 base sequence and yields nearly 10 terrabytes of sequence data. Sequence trimming and QC is completed. Read mapping to the preliminary reference genome is completed. Sequence variant calling is in progress. Sequence data will then be made public by uploading to the National Center for Biotechnology Information (NCBI). The population genomics data analysis will then be started.

8. Coordination/Collaboration:

A. Projects Within a Trustee Council-funded program

1. Within the Program

This project is a formal collaboration with my research group at University of California Davis and that of Dr. Paul Hershberger at the U.S. Geological Survey, Marrowstone Marine Field Station. Animal experiments are being conducted by Dr. Hershberger's group (Project 18120111-E) at the Marrowstone facility. A Ph.D. student from my research group travelled to Marrowstone to participate in animal experiments in spring 2018. Members of Dr. Hershberger's group collaborated in collecting and shipping herring gametes from Alaska for animal experiments. They also collected tissues from a single herring family (offspring from in vitro fertilization of gametes from one male and one female) to be used for generating a linkage map for the reference genome. In the past year I have travelled to the Seattle area for face-to-face logistics and project coordination meetings with Dr. Hershberger and his group. My Ph.D. student and postdoctoral scholar also joined us for those meetings.

2. Across Programs

a. Gulf Watch Alaska

N/A

b. Data Management

We have nothing yet to report, as no data have yet been made public. We plan to make our data publicly available once quality control is completed. The reference genome sequence will be uploaded to NCBI. Population genomics data will be also uploaded to NCBI, as will RNA-seq data. We will then create a project page at EBI BioStudies that will include links to raw data at NCBI, and will also house variant call files for the population genomics data and matrices of read counts for the RNA-seq data. The EBI site will also house data from animal challenge experiments. All custom bioinformatics scripts will be archived at GitHub, and will be linked through the EBI BioStudies project site. Publications will also eventually be linked through the BioStudies project site. The databases described above are designed to accommodate the types of data that we need to make public, and they are durable. Once we have started these data uploads, we will create links to them through the Gulf of Alaska Data Portal.

c. Lingering Oil

N/A

B. Projects not Within a Trustee Council-funded program

N/A

C. With Trustee or Management Agencies

This project is a formal collaboration with my research group at University of California Davis and that of Dr. Paul Hershberger at the U.S. Geological Survey, Marrowstone Marine Field Station. Animal experiments are being conducted by Dr. Hershberger's group (Project 18120111-E) at the Marrowstone facility. A Ph.D. student from my research group travelled to Marrowstone to participate in animal experiments in spring 2018. Members of Dr. Hershberger's group collaborated in collecting and shipping herring gametes from Alaska for animal experiments. They also collected tissues from a single herring family (offspring from in vitro fertilization of gametes from one male and one female) to be used for generating a linkage map for the reference genome. In the past year I have travelled to the Seattle area for face-to-face logistics and project coordination meetings with Dr. Hershberger and his group. My Ph.D. student and postdoctoral scholar also joined us for those meetings. NOAA Fisheries scientists John Incardona and Nathaniel Scholz (NOAA Northwest Fisheries Science Center, Seattle) are collaborating in animal exposure experiments, since they have research goals that include exposure impacts on growth and development. These measurements were easily added to tasks associated with our (Whitehead and Hershberger) 2017 and 2018 exposure experiments, so it made sense to join forces and leverage the collective experience, personnel, and equipment of our groups – the most important of these being the outstanding aquatics systems for rearing live herring at Marrowstone. A challenging portion of this year's experiments was collecting gametes from PWS and Sitka Sound populations, shipping them live back to Marrowstone, then spawning them for experiments. Particularly challenging was the collection of animals and tissues from Sitka Sound because fish mysteriously disappeared for over a week at the time of our trip. My Ph.D. student Tony Gill traveled to Alaska to await their return. During this period colleagues based in Sitka graciously assisted him. Angie Bowers (Aquaculture director at Sitka Sound Science Center) frequently took Tony out in her boat in search of spawning herring. She provided Tony room at the Center to work and assisted with dissecting and packing gametes. She lent Tony the company vehicle to make pickups and drove him to the airport. She was incredibly hospitable and generous with her time. Eric Coonradt (Fishery Biologist, ADF&G) was constantly in contact with our team, sending daily updates on spawn (or lack thereof). He is the one who spotted the spawn and invited Tony on the boat to catch them.

9. Information and Data Transfer:

A. Publications Produced During the Reporting Period

None yet

B. Dates and Locations of any Conference or Workshop Presentations where EVOSTC-funded Work was Presented

None yet

C. Data and/or Information Products Developed During the Reporting Period, if Applicable

We are currently in the final stages of reference genome assembly. Once complete, we will make the assembly publicly available through NCBI and the Ensembl genome database project. Population genomics data collection is complete, and we will upload raw reads to NCBI once quality control filtering is completed.

D. Data Sets and Associated Metadata that have been Uploaded to the Program's Data Portal

We plan to make our data publicly available once quality control is completed. The reference genome sequence will be uploaded to NCBI. Population genomics data will be also uploaded to NCBI, as will RNA-

seq data. We will then create a project page at EBI BioStudies that will include links to raw data at NCBI, and will also house variant call files for the population genomics data and matrices of read counts for the RNA-seq data. The EBI site will also house data from animal challenge experiments. All custom bioinformatics scripts will be archived at GitHub, and will be linked through the EBI BioStudies project site. Publications will also eventually be linked through the BioStudies project site. The databases described above are designed to accommodate the types of data that we need to make public, and they are durable. Once we have started these data uploads, we will create links to them through the Gulf of Alaska Data Portal.

10. Response to EVOSTC Review, Recommendations and Comments:

Science Panel Comment on FY18 Work Plan: The Panel was pleased to see the integration with Paul Hershberger's disease work, linking them to see if see if there is a genomic change in response to these different pathogens in the PWS herring population. The Panel appreciates that goals are being achieved ahead of schedule and cost-effectively, allowing for additional samples at other locations. The Panel approves the shift of funds from future years to FY18 to get the postdoc onboard to work with the data being generated. There are many great collaborations being made. The Panel is excited to have the entire genome and transcriptome for herring mapped for other studies, including the possibility of adding more value to herring stock responses in Southeast Alaska. There might be another source of archived samples in Pacific Northwest (Doug Hay - Barkley Sound?).

PI Response: Thank you.

11. Budget:

Budget Category:	Proposed FY17	Proposed FY 18	Proposed FY 19	Proposed FY 20	Proposed FY 21	TOTAL PROPOSED	ACTUAL CUMULATIVE
Personnel	$76.95	$173.89	$243.82	$175.23	$127.44	$797.3	$ 244
Travel	$0.0	$1.18	$4.81	$2.59	$2.59	$11.2	$ 5
Contractual	$0.0	$0.0	$0.0	$0.0	$0.0	$0.0	$ -
Commodities	$54.36	$95.68	$107.09	$18.33	$20.16	$295.6	$ 107
Equipment	$0.0	$0.0	$51.43	$0.0	$0.0	$51.4	$ 51
Indirect Costs (*will vary by proposer*)	$46.9	$154.0	$200.8	$99.9	$72.6	$574.1	$ 201
SUBTOTAL	$178.2	$424.7	$608.0	$296.0	$222.8	$1,729.7	$ 608
General Administration (9% of	$16.0	$38.2	$39.5	$26.6	$20.1	$140.5	N/A
PROJECT TOTAL	$194.2	$462.9	$647.5	$322.7	$242.8	$1,870.1	
Other Resources (Cost Share Funds)	$0.0	$0.0	$0.0	$0.0	$0.0	$0.0	