fodder\_doc

Intro

Genetic underpinnings may contribute to any aspect of life history, ecology and physiology. Therefore developing genomic resources and maintenance of genetic diversity within the species are key components to

genetic diversity = closely tied to the evolutionary capacity for adaptation to environmental change

phenotypic variation to allow for a broad respond to climatic factors deeply influencing the Delta. Three key components to

of basic ecology is phenotypic variation; thus, maintenance of genetic diversity is essential as it is tied to the evolutionary capacity for adaptation to environmental change. To assess genetic diversity within delta smelt, contemporary and historical effective population size estimates were made using 2318[tk] samples from 25[tk] generations.

Conservation of genetic diversity within the species will require several components, including an assembled genome to allow for fine-scale genomic change, quantifying genetic diversity through time, and identifying loci to assist in non-lethal monitoring methods of population dynamics, such as sex ratios, of the species.

genetic resources for and monitoring genetic diversity within delta smelt is a critical step to aid in research efforts for making informed decisions to manage this imperiled species. Resource development requires genomic tools for

The abundance of delta smelt (*Hypomesus transpacificus*) has been in decline for decades1,2. Genetic tools have been useful for several management concerns: monitoring overall diversity and effective population size () in the wild population, genetic management in the captive population, developing assays to detect hybrids between delta smelt, wakasagi, and longfin smelt, and developing assays for eDNA sampling or species identification3–5. Attempts at quantifying in the wild population have been hampered by several interacting factors including a lack of power from using microsatellite markers and a very large historical 6,7. In order to improve estimates of , we assembled the delta smelt genome. We then leveraged the genome to estimate contemporary and the loss of genome-wide diversity between 1993 and 2020, interrogate domestication selection, and search for a sex marker. While is useful for monitoring purposes, we strongly advice against the use of alone for making management decisions.

For this project our tasks were to: 1) assemble a high-quality reference genome for delta smelt that is publicly available for all researchers, 2) estimate contemporary effective population sizes () for the wild 2017 to 2019 birth year cohorts of delta smelt, 3) search for genetic evidence of domestication selection across the genome, and 4) search for sex-specific markers and/or chromosomes in delta smelt. In order to obtain a more comprehensive understanding of through time, we have expanded Task 2 to include analysis of historical datasets of delta smelt dating back to 1995 and samples collected from 2020. The status of each deliverable is listed in Table 1. We have divided this final report into 4 sections, each with its own background, methods, results and discussion, followed by a final conclusion integrating our findings.

GENOME

Genome Intro: Why?

* innovation (theoretical or medical)
* preserve biodiversity
  + species are going extinct faster than ever
  + 5th great extinction?
  + need to know the genetic composition of the speices that sustain ecological communities
  + “the genome is the core entitiy to life”
  + previously focused on a single gene approach but genes do not exist in isolation,
  + genetics have influenced
  + “genomics starts from a reference genome”
  + model genomic research has led to innumerous medical advances for humans
  + humans are radically altering the natural world
    - we need to understand the effects of this
  + can understand what genetics lead to making a mammal a mammal and a reptile a reptile, or bats live for fucking ever
  + from just one mammal that currently doesn’t ha
    - know about blindness,
  + disease susceptibility

What could happen?

* scientific innovation
* theories stemming from observations intangible to the naked eye
* contributing to a larger and ever growing resource
* can better understand things like aging (insert bats),

Genome Intro: old

The use of next generation sequencing (NGS) technologies in conjunction with a highly contiguous and accurate reference genome increases the power and precision of inferences made in population genetic studies (e.g., analyses of population structure, genetic diversity, and local adaptation)8. Without a reference genome, DNA sequences captured by the sequencer but not aligned to a reference genome fail to account for how each piece of sequenced DNA interacts with all other sequences (i.e., linkage patterns). However, by aligning to a reference genome we know where each segment of sequenced DNA lies within the genome and relative to other sequenced DNA. An assembled genome vastly increases the power to answer questions concerning demography, adaptation, fitness, and disease susceptibility.9,10 For example, previous work in Chinook salmon and steelhead found a small number of markers that were associated with run-timing in each species. In 2014, these markers were located on five different scaffolds and next to gaps in the highly fractured genome assembly. Reanalysis using long-reads to span gaps and reorient and link fragmented scaffolds demonstrated that all the associated markers were from a single locus (i.e., the GREB1L region)11. Since the 2017 study, continued improvements to salmonid reference genomes have enabled more detailed mapping of life-history variation and phenotypic traits in the same locus12, which may lead to significant changes in management. This example underscores how reference genomes can propel conservation research forward.

Recently, genome sequencing technologies have become both more cost effective and efficient. “Hybrid assemblies” (assemblies that use multiple NGS technologies) are a reliable way to achieve a chromosome-scale high-quality genome assembly13,14. The Vertebrate Genomes Project, a consortium aimed towards developing an assembly pipeline and quality standards for genome assemblies of all vertebrates, established quality goal metrics for the continuity, completeness, and accuracy of reference genomes14. Therefore, a main goal of this project is to develop a highly accurate chromosome-scale reference genome, hereafter called “reference genome”, using linked-reads (Box 1), long-reads (Box 2), hi-c chromatin confirmation capture (Box 3), a genetic linkage map15 and various assembly software programs (Figure 1).