**Task 4: Sex Marker**

***Background***

Sex determination in fish is a highly variable trait62 and understanding its mechanisms is crucial for not only for understanding the biology of the individual species of fish but for gaining insight into the evolution of sex chromosomes and genetic mechanisms underlying sex determination.63 Fish represent the most diverse group of vertebrates with over 30,000 described species.64 With this diversity and constant exposure to variable environments comes a vast array of morphological, physiological, behavioral, developmental and sexual mechanisms.65–68 In teleost fishes, sex determination can be genetic or environmental and varies between closely related species.62,69–71 Delta smelt are a unisexual species that do not appear to have environmental regulation of sex determination which suggests sex may be determined genetically. Endogenous genetic sex determination mechanisms can occur at the chromosomal level where heterogametic males (XY) or females (ZW) have been observed, or they can be at the genic level where single or multiple genes influence the sex determination.70 Thus, identifying sex-associated markers is of use for increasing biological knowledge and practical management.

The ability to non-invasively identify sex in delta smelt will assist in management of the captive colony and develop knowledge of the biology of wild delta smelt. Currently, wild fish can only be sexed using the expression of gametes in ripe adult fish or through dissection, both sacrifice the lifeof the fish or gamates. When sex is identified by expression pressure is put on the the abdomen of fish until eggs are excreted females or running milt in males.72 Because sexual identification of wild fish primarily relies on the physiological status of an individual fish, only about two-thirds of wild adult delta smelt sampled can be sexed (Hammock pers. comm.). Knowledge of the genetic underpinnings of sex determination in fishes is a vital asset to the better management of captive populations, basic knowledge of life history characteristics of the species, ecological surveys and management regarding population metrics, species modeling, demographic inference, and sex-based survival.73

In order to non-invasively identify the sex of wild and captive delta smelt, we sought to identify potential candidate allele(s) which could be used as genetic diagnostics for classifications of sex.

***DNA sampling & sequencing***

*Methods*

To identify a sex specific marker(s) for delta smelt, we sampled adipose fin clips from 24 female and 24 male captive bred individuals taken from the FCCL and sexually identified through either dissection or gametic expression. DNA was extracted using the Qiagen DNEasy 96 Blood & Tissue Kit with a modification of elution in100uL of H2O rather than the AE Buffer included with the kit.

Prior analyses that attempted to determine sex markers in delta smelt used the *Sbf1* restriction enzyme, which cuts DNA approximately every 65,000 base pairs, but no sex markers were identified. For this library preparation we sought to maximize the number of restriction enzyme cut sites and acquire reads from more locations throughout the genome. To do this we digested extracted DNA using the *Pst1* restriction enzyme, which shears DNA sixteen times as often as the *Sbf1* restriction enzyme, or roughly once every 4,100 base pairs. RAD sequencing libraries were prepared at the GVL according to Ali et al (2016)<CITE>, and sequenced at the UC Davis Sequencing Center with 150 bp paired-end reads on an Illumina HiSeq.

***Genome wide association study***

*Methods*

We performed two rounds of a genome wide association study (GWAS) using the male and female reference genomes. For each GWAS, we first aligned raw RAD sequencing data from 24 male and 24 female sexed fish to each assembly. Next, we performed a dominant model case-control GWAS using males as controls (0) and female as cases (1) with all of the loci spread across the corresponding reference genome using the program ANGSD12. The association of a particular allele with sex category was reported as likelihood ratio test (LRT) statistic and is chi square distributed with one degree of freedom. We applied a conservative significance cutoff with a Bonferroni corrected p-value of 0.05 using the formula where is the number of loci analyzed, is the desired p-value, and is the adjusted p-value given the number of loci used in the analysis.

*Results*

We analyzed 848,444 and 922,975 loci spread across the male and female genome assemblies, respectively. These loci correspond to a Bonferroni corrected p-value of 5.893141e-08 in loci found within the male reference genome and 5.417265e-08 in loci analyzed within the female reference genome. No significant association in sex was found using the female reference genome. Two loci located in the male reference genome on Chromosome 5 were significantly associated with sex in delta smelt (Figure 7). The SNP most associated with sex in delta smelt had an LRT score of 21.352698 which corresponds to p-value of 0.000003821.

We analyzed 260,256 loci spread across the genome which corresponds to a Bonferroni corrected p-value of 0.000002 or less. The SNP most associated with sex in delta smelt had an LRT score of 21.352698 which corresponds to p-value of 0.000003821.

***Depth analysis***

*Methods*

We looked for male and female differences in the presence and depth of RAD tags spread throughout the genome. To do this we performed two experiments, one using the male genome as a reference and another using the female genome as a reference. Each experiment used the 24 male and 24 female alignment files from the previous GWAS. First, we acquired the depth of aligned reads at each location in the reference genome using samtools depth. Next, using custom bash and perl scripts we threw out all locations where no male or female RAD data aligned, and we totaled the number of male alignments and gathered the total coverage for each sex (<https://github.com/shannonekj/DS_sex-marker/blob/master/analysis/>, see 04.analyze\_B\_compareDepths.\*Ref\_1\_\*.slurm where \* can take the place of any text). Finally, we sorted the coverage difference files and looked for locations in the genome where one sex had high and consistent coverage and the other sex had shallow or no coverage.

*Results*

We found no large areas that corresponded to an inflation of any one particular sex having higher depth of coverage compared to the sex and could not identify any sex specific loci in this analysis.

***K-mer analysis***

*Methods*

We blahblahblahslkhd.

*Results*

tktktktktkktktkt