**Genome assembly and investigation of *Hypomesus transpacificus* genome for estimates of effective population sizes and identification of sex-specific markers.**

Qualifying Exam Research Proposal

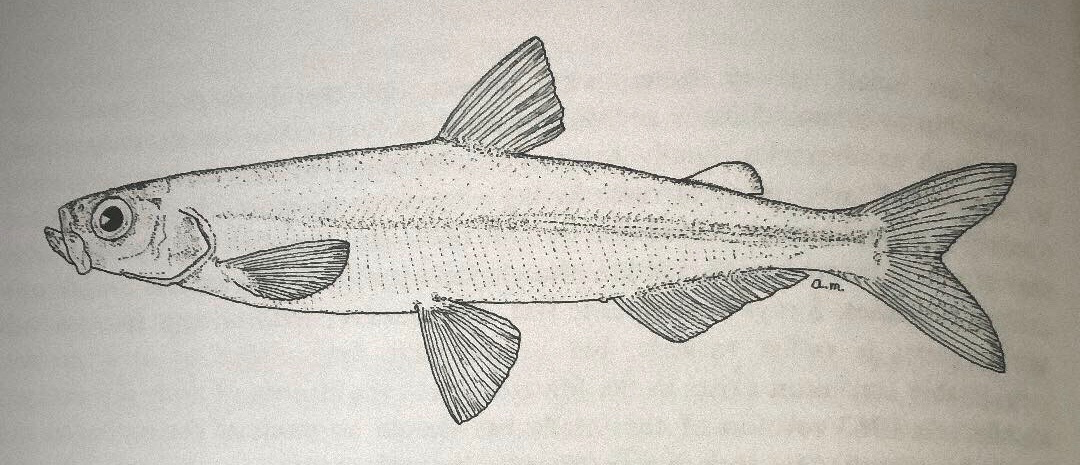


Illustration: Alan Marciochi

Shannon E.K. Joslin

[sejoslin@ucdavis.edu](mailto:sejoslin@ucdavis.edu)

Integrative Genetics and Genomics Graduate Group

PI: Dr. Mike Miller and Dr. Andrea Schreier

Tuesday **|** 30 April 2019 **|** 9:00am-12:00pm

1013 Gladys Valley Hall (first floor)

Chair: Dr. C. Titus Brown

[ctbrown@ucdavis.edu](mailto:ctbrown@ucdavis.edu)

Genomics: Dr. Graham Coop

Transmission Genetics: Dr. Thomas Gradziel

Molecular Genetics: Dr. Julin Maloof

Population & Quantitative Genetics: Dr. Benjamin Sacks

**Specific Aims**

To date, the conservation and management of listed species endemic to California, such as *Hypomesus transpacificus* (delta smelt), relies on physical surveys or limited molecular techniques to provide insights into the population health of native fishes. These methods provide highly variable estimates and are relatively low throughput. On the other hand, next generation sequencing (NGS) approaches are superior at investigating the evolution and genetic status of a species to inform conservation policies of wild populations.1 Despite numerous studies on the distribution, ecology, physiology and population dynamics of wild delta smelt, genetic findings have produced vague results in elucidating the demography of the species.2,3 My work aims to expand genetic resources for delta smelt by providing genomic management tools for this imperiled species. To do this, I propose the following:

***AIM #1:*** *Produce a high-quality reference genome for delta smelt.*

***AIM #2:*** *Develop a reproducible framework to infer contemporary effective population sizes of delta smelt.*

***AIM #3:*** *Discover sex-specific markers for sex identification of delta smelt.*

**Background**

In the San Francisco Estuary (SFE), many native pelagic fishes that were once abundant, such as delta smelt, have undergone a broad decline in population size.4 The SFE is a dynamic ecosystem encompassing 1,000 square miles of open water and wetlands in Northern California (Figure 1). Since the Sacramento-San Joaquin Delta (the Delta) became the primary distribution hub of California’s water supply in the 1960’s, the SFE has been heavily altered by anthropogenic activity. Agriculture, water delivery, shipping, and urban development have spurred changes in the way water is distributed throughout the estuarine environment.

Delta smelt is a native species to the SFE that has undergone a population collapse associated with drought and anthropogenic effects. It is now believed stochastic processes may push the species to extinction.3,5 They are a small (6 - 9cm), translucent, diadromous species that migrates between fresh and saline water and reproduces annually.6 Delta smelt are part of the Osmeridae family which represent a prosperous food source for human consumption in Japan, Europe, and North America, and have declining populations worldwide.5,7,8 Because of their annual life cycle and rapid response to conditions of their ecosystem, delta smelt are considered an indicator of the overall health of the SFE ecosystem. Once abundant and widely distributed throughout the SFE, the delta smelt population has been declining since the 1980s.9 Delta smelt have been listed as threatened under the Federal Endangered Species Act (ESA) since 1993 and endangered under the California ESA since 2009. As a result of their continued decline, resource management agencies actively monitor abundance in the wild population.

Currently, agencies estimate local abundance based on the number of fish captured in targeted surveys throughout the SFE.10,11 However, estimating the abundance of delta smelt provides neither insight into the genetic diversity nor key aspects of basic biological knowledge, such as sex ratios, which are important for understanding the population dynamics of a species. Two previous studies used microsatellite markers () to estimate the contemporary effective population size () of delta smelt, but found variable results.2,3 NGS technology presents more power to make precise estimates by increasing the number of putatively neutral markers.12–14

In addition to an increase in loci, better understanding sex distribution of delta smelt throughout their life cycle will inform demographic inference, ecological modeling, and management. We do not know the underlying genetic composition for sex determination in delta smelt. Gametic expression is the only method available to identify sex in individuals caught in surveys. However, many fish are not sexually mature at the time of capture, resulting in missing data regarding sex ratios of the delta smelt population. Contrasting demographic results and a lack of knowledge in the foundational biology of delta smelt provide motivation to better understand the genetic status of the species through more refined and current genetic techniques.

The purpose of this proposal is to gain better understanding of delta smelt genetics by 1) assembling a high-quality reference genome; 2) inferring contemporary effective population sizes of delta smelt in the wild; and 3) identifying sex-specific genetic markers to be used for sex identification.

**AIM #1: Produce a high-quality reference genome for delta smelt.**

***Rationale***

A high-quality reference genome provides a useful tool in many areas of modern biology. For my thesis work, it will allow for fine scale resolution of population genetic analysis, and assist in elucidating genetic and genomic sex-specific differences in delta smelt. Extending one step beyond my work, providing a reference genome will immediately assist in multiple studies currently being conducted at UC Davis and management agencies. Current proposed studies outside the scope of my thesis range from adaptation of thermal tolerance (Whitehead & Fangue Labs), to observing genomic differences in domestication (Schreier/Miller Lab), to testing whether rearing environment affects phenotype through epigenetic alterations (California Department of Fish and Wildlife/Schreier lab). More broadly, a chromosome-scale scaffolded delta smelt genome will be the first Osmerid sequenced and assembled beyond 15x coverage and the first and only *Hypomesus* reference-quality genome available15. My research will make a new genus available for large-scale genomic studies and add to the collaborative efforts of documenting the world’s diversity.16,17

***Experimental Design***

In order to construct a *de novo* reference genome, high molecular weight (HMW) genomic DNA (gDNA) was extracted from sampled specimens. Since sex determination in delta smelt is unknown, one male and one female were sampled for extraction and subsequent sequencing from the captive population located at the Fish Conservation and Culture Laboratory (FCCL) in Discovery Bay, CA. After collection, samples were brought to the UC Davis DNA Technologies Core where DNA was extracted and sequenced. HMW gDNA was extracted following the protocol described in Wasko et al.18 and the range of sizes of extracted DNA fragments were determined using a pulse field gel (Figure 2). A whole genome sequencing library was prepared using a 10X Genomics Chromium Genome Library prep and the library was sequenced to an estimated 120x coverage on an Illumina NovoSeq6000 sequencer with 150 base pair (bp) paired-end reads.

To produce an initial 10X assembly, I used the software Supernova to assemble raw reads into contigs and scaffolds.19 Currently, Supernova is optimized to assemble human genomes at 38-56x coverage with higher coverage occasionally being advantageous in non-human species. As such, I will run multiple assemblies ranging from 30-70x coverage to find a level that is best suited to assemble the delta smelt genome. I will assess the qualities of the resulting 10X assemblies using statistics corresponding to the completeness and contiguity of the Supernova assembly results. First, I will look at the contig N50 and scaffold N50 which are the shortest sequence length where 50% of the total genome size is contained in contigs or scaffolds larger than that size, respectively.20 Next, I will assess the 10X assembly completeness through looking at the proportion of expected single-copy orthologs contained in the assembly using BUSCO.21

After 10X assembly, I will cluster contigs into chromosome-scale scaffolds by using chromatin conformation capture. Hi-C based chromatin interaction maps provide a powerful tool to make all-by-all interactions of nucleotide sequence data that can be used to scaffold *de novo* genome assemblies.22–25 Hi-C libraries will be prepared according to Burton et al.26 at the UC Davis DNA Technologies Core. Paired-end reads resulting from sequencing libraries will be scaffolded with my 10X assembly using a Hi-C proximity guided assembly with the LACHESIS software.27 I will then repeat the calculation of N50 and run BUSCO to assess the completeness of my reference genome.

***Initial Results***

Initial HMW gDNA extraction was far superior on the female sample, so an initial 10X library and assembly was only carried out on the female fish. Assembly results show estimated genome size of 0.9Gb which is 0.3Gb smaller than expected from genome size prediction based on *Sbf1* RAD-derived loci. This discrepancy resulted in an excessive read depth for the first assembly with a raw read coverage of 91.06x. Despite having a sequencing depth outside of that recommended by Supernova, the initial assembly produced scaffold N50 of 1.59MB.

***Potential Problems***

Due to the physical size difference in male and female delta smelt, it may be challenging to extract a sufficient quantity of HMW DNA to carry out Hi-C in male samples. Depending on the genomic architecture that differentiates male from female this may present an issue. If it does not appear male delta smelt have novel sex chromosomes, then I will carry out Hi-C using only female delta smelt. If it appears male delta smelt have novel sex chromosome(s) and I cannot extract sufficient DNA from one male fish, I will work with the FCCL to produce an inbred line and pool DNA from two male full-siblings together.

**AIM #2: Develop a reproducible framework to infer contemporary effective population sizes of delta smelt.**

***Rationale***

The U.S. Fish and Wildlife Service has increased monitoring delta smelt through estimating local abundance in their year-round weekly sampling throughout the Delta since November 2016. Abundance of delta smelt is calculated from the product of the number of fish physically caught, and the total sampled volume of water.28 This method gives highly variable abundance estimates depending on a difference of relatively few fish. Recent abundance estimates range from over 18,000 fish in January 2019 where nine fish were caught, followed by zero fish in March 2019 where no fish were caught.29 Additionally, this surveillance method does not indicate genetic health or diversity of the species. A complementary method of quantifying levels of delta smelt can be achieved through monitoring the fish by estimating the contemporary of the species through time.

is a useful tool in monitoring endangered populations as it can inform how likely alleles in the population are to be lost or fixed.30,31 The of a given species is defined as the size of a Wright Fisher population that would have the same rate of change of a genetic parameter as the population under study.32 A number of methods exist to estimate both historical for evolutionary purposes, as well as contemporary, or short-term, to quantify modern day genetic diversity.33 Contemporary methods use different genetic parameters such as linkage disequilibrium (LD) between two allele frequencies (inbreeding), allele frequency variance between generations (genetic drift), the difference in the number of heterozygotes with respect of Hardy-Weinberg equilibrium (heterozygote excess) and molecular co-ancestry (coalescent).34–39 Today, LD and temporal methods are commonly used in conservation genetic studies and management as they can detect population declines within one generation post-decline40 and reliably detect population declines ten generations post decline.14 Early and reliable detection of diminished genetic diversity is important when protecting endangered species because conservation programs seek to increase or maintain genetic diversity, as it is the raw component for natural selection to act on. 41

Two previous studies estimated contemporary in the wild population of delta smelt using 12-15 markers. Fisch et al. 20113 found the to be decreasing from 2003 to 2009 study period. In contrast, Finger et al. 20172 found the could not be accurately estimated from 2011 to 2014 due to infinite confidence intervals. Multiple factors may have contributed to this discrepancy: 1) different versions of NeEstimator42,43 were used between the two studies; 2) differences in the number of informative loci analyzed; or 3) the number of generations factored into the analysis. By using more generations and NGS, I will increase the power to estimate contemporary through the use of thousands of potentially informative loci.

***Experimental Design***

We sequenced archived samples collected in state and federal trawls from 1993-2014. Contemporary samples were collected as fin clips by Interagency Ecological Program surveys from 2015-2017. Samples from 2018 and 2019 will be collected and transferred into the custody of the Genomic Variation Lab by December 2019. Genomic DNA has been extracted using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer’s protocol for a total of 2,605 samples to date. In order to produce a large number of loci in a cost-effective manner, restriction site associated DNA (RAD) sequencing was carried out for all individuals. RAD libraries were prepared using the *Sbf1* restriction enzyme according to the ‘new RAD protocol’ described in Ali et al., 44 and sequenced 100 bp paired-end reads on an Illumina HiSeq 4000.

In order to estimate contemporary of delta smelt I will need to generate a list of high-quality polymorphic loci to make demographic inferences with. First, I will align all sequences to the reference genome created in Aim #1 using samtools45 or, if I do not complete Aim #1 I will align to a RAD-seq derived contig assembly I created using PRICE, BWA and samtools.45–47 Second, I will recalibrate base quality calls using GATK to eliminate sequencer biases in base quality scores and improve downstream SNP discovery.48–50 Finally, I will call genotypes using site allele frequencies as priors in ANGSD51 and filter sequencing data in the following ways: remove individuals with insufficient read coverage and depth, remove paralogous loci using the site frequency spectrum as a prior, remove reads that do not meet mapping and base call qualities, remove low frequency SNPs, SNPs not in Hardy-Weinberg equilibrium and hybridized individuals. I will then test for population structure in my dataset by carrying out a principal component analysis in ANGSD51 and running STRUCTURE52 for each year.

Based on results from comparative analysis by Gilbert and Whitlock (2015)53 and Wang (2016),54 I will use both temporal and LD methods to estimate contemporary in delta smelt through the programs MLNe and NeEstimator v2.01.43,55 MLNe is a pseudo-likelihood method for estimating from temporally-spaced samples. NeEstimator v2.01 is a software that implements various updates to the original LDNe method for estimating from single-sample data.13,43,55,57,58 Temporal methods use allele frequency change between generations at individual loci to quantify the level of genetic drift in a population and estimate the harmonic mean .35,56 LD methods use the non-random association of allele frequencies at different loci to quantify the level of inbreeding within the population of the parental generation.34,56 Thus, the two methods give estimations of at different timepoints. In order to reduce bias and increase precision I will incorporate a linkage map into my LD estimates of and use more precise confidence intervals and estimate bias according to Waples et al (2016).59,60

***Expected Results***

I expect to see a decline in the contemporary of delta smelt from 1993 to 2019. Initial results indicate a consistent decline from 1994 – 2016 in when measures of genetic diversity ( and ) are input into . While this equation gives a long-term estimate of effective population size, the consistent decline in indicates genetic diversity continues to decrease in the population. I expect that contemporary estimates will show a similar pattern of decreasing through time but with lower estimates of contemporary .

***Potential Problems***

Both temporal and LD methods of estimating account for the expected contribution from random sampling error () to the observed genetic characteristic of or , respectively. As such, estimates of can be negative with infinite confidence intervals due to the expected contribution of random sampling being larger than the actual contribution 57. If I observe negative estimates of , I will run replicate subsets of samples from each year to estimate and obtain an overall estimate of by taking the harmonic mean of all the estimates as prescribed in Waples and Do (2010).12 Additionally, a new method of jointly estimating and pedigrees as a single-sample method shows promise to give accurate estimates of contemporary .61 However, this method is under review.

**AIM #3: Discover sex-specific markers for sex identification of delta smelt.**

***Rationale***

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) and 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 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. Sex is identified by observing egg excretion through pressure on the abdomen of females or through running milt in males.72 Because identification 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 assists with 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

***Experimental Design***

I proposed to discover sex-specific SNPs for the identification of male and female delta smelt by identifying SNPs associated with each corresponding sex using RAD-seq data.74–79 To discover markers, I sequenced 24 female and 24 male samples from a captive colony of delta smelt. Due to the possibility of having only a single gene influence sex determination in delta smelt, it is essential to survey as many sites across the genome as possible. To maximize the number of cut sites, we digested DNA using the *Pst1* restriction enzyme, which shears DNA sixteen times as often as the *Sbf1* restriction enzyme used in Aim #2. We prepared RAD libraries and sequenced with 150 bp paired-end reads on an Illumina HiSeq.

After sequencing, I will align, filter and test for population structure as I did in Aim #2, 45–52 and carry out a case-control genome-wide association study using PLINK.80 I will correct for multiple testing by determining significance using a Benjamini-Hochberg false discovery rate81 and permutation testing through PLINK.80

Sequences containing sex associated SNPs will be submitted to Fluidigm for SNP Type assay design as described in Benjamin et al. (2018).82 SNPs will be validated using an additional 24 females and 24 males sexed through either dissection or gametic expression, which were not included in the initial dataset.

***Expected Results***

If sex determination is determined by sex-specific chromosomes or a small number of genes I expect to find SNPs that differentiate female from male fish from RAD-seq data.74–79 Determining sex in wild and captive populations is an important parameter, as it can assist in demographic estimates and ecological modeling. This study will produce a SNP assay which will contribute a valuable resource for identifying sex in delta smelt.

***Potential Problems***

It is possible, that RAD-seq alone does not have high enough resolution to find sex associated markers if sex determination is a complex trait in delta smelt.83,84 If RAD-seq is not sufficient, whole-genome sequencing data of male and female fish will provide higher resolution and increase the probability of detecting linked SNPs associated with sex.85 If it appears sex determination is a polygenic trait, a GWAS may result in nonsignificant associations of loci that are indeed linked to sex determining genes. To increase the GWAS power, I can use the reference genome created in Aim #1 with the software FINDOR to leverage functional enrichment in my analysis.86 If this method does not work, I can perform QTL mapping analysis to identify sex-specific markers using RAD data87–91 or find sex specifically expressed genes using RNA-seq.92,93

**Figures**

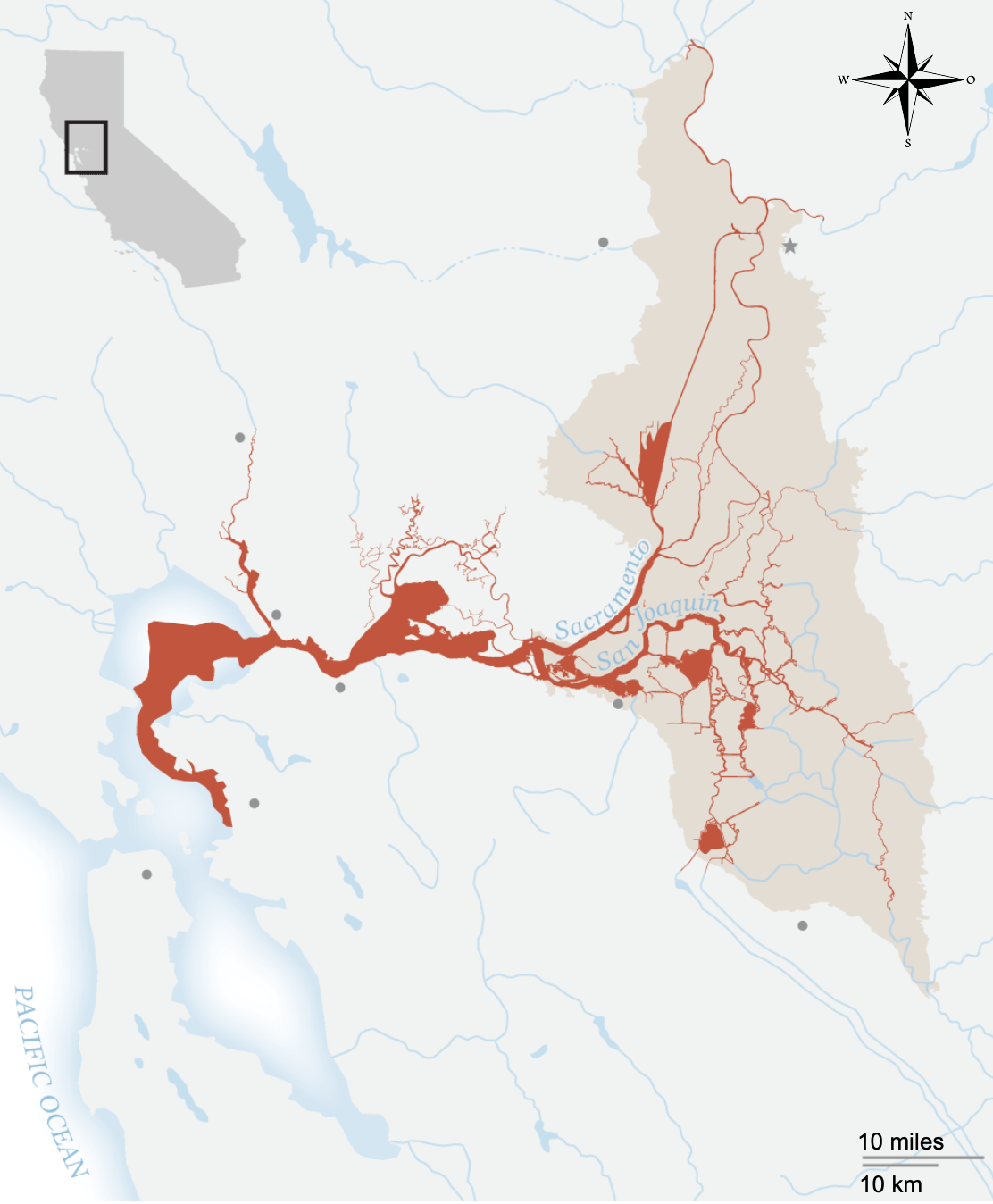


Figure 1. A map of the habitat range of delta smelt in the San Francisco Estuary in Northern California. Source:<https://news.nationalgeographic.com/2015/04/150403-smelt-california-bay-delta-extinction-endangered-species-drought-fish/>

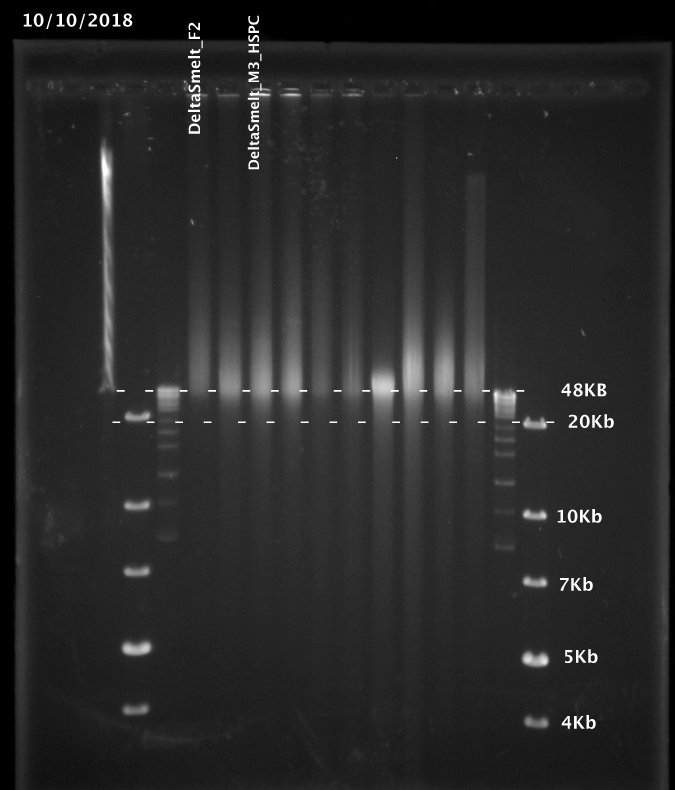


Figure 2. Image of a pulse field gel containing extracted high molecular weight genomic DNA from female (DeltaSmelt\_F2) and male (DeltaSmelt\_M3\_HSPC). The female lane shows more extracted gDNA larger than the 48kb size band.

**Works Cited**

1. Zhang, J. The impact of next-generation on genomics. *J Genet Genomics* **38,** 95–109 (2011).

2. Finger, A. J., Schumer, G., Benjamin, A. & Blankenship, S. Evaluation and Interpretation of Genetic Effective Population Size of Delta Smelt from 2011–2014. *San Fr. Estuary Watershed Sci.* **15,** (2017).

3. Fisch, K. M., Henderson, J. M., Burton, R. S. & May, B. Population genetics and conservation implications for the endangered delta smelt in the San Francisco Bay-Delta. *Conserv. Genet.* **12,** 1421–1434 (2011).

4. Moyle, P. B. *Inland Fishes of California*. (University of California Press, 2002).

5. Moyle, Peter B., Brown, Larry R., Durand, John R., Hobbs, J. A. Delta Smelt: LIfe History and Decline of a Once-Adundant Species in the San Francisco Estuary. *San Fr. Estuary Watershed Sci.* **14,** 1–40 (2016).

6. Sommer, T., Mejia, F., Nobriga, M., Feyrer, F. & Grimaldo, L. The Spawning Migration of Delta Smelt in the Upper San Francisco Estuary. *San Fr. Estuary Watershed Sci.* **9,** 1–44 (2011).

7. McAllister, D. E. *Revision of the smelt family, Osmeridae.* (National Museum of Canada, Biological Series 191, 1963).

8. Rosenfield, J. A. & Baxter, R. D. Population Dynamics and Distribution Patterns of Longfin Smelt in the San Francisco Estuary. *Trans. Am. Fish. Soc.* **136,** 1577–1592 (2007).

9. Moyle, P. B., Herbold, B., Stevens, D. E. & Miller, L. W. Transactions of the American Fisheries Society Life History and Status of Delta Smelt in the Sacramento-San Joaquin Estuary , California. *Trans. Am. Fish. Soc.* **121,** 67–77 (1992).

10. Newman, K. B., Mitchell, L. & Barnard, D. *Summary of Enhanced Delta Smelt Monitoring for LOBO Independent Review Panel*. (2017).

11. Newman, K. B. Sample design-based methodology for estimating delta smelt abundance. *San Fr. Estuary Watershed Sci.* **6,** 1–18 (2008).

12. Waples, R. S. & Do, C. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evol. Appl.* **3,** 244–262 (2010).

13. Waples, R. S. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121,** 379–391 (1989).

14. Nunziata, S. O. & Weisrock, D. W. Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity (Edinb).* **120,** 196–207 (2018).

15. Malmstrøm, M., Matschiner, M., Tørresen, O. K., Jakobsen, K. S. & Jentoft, S. Data descriptor: Whole genome sequencing data and de novo draft assemblies for 66 teleost species. *Sci. Data* **4,** 1–13 (2017).

16. Lewin, H. A. *et al.* Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci.* **115,** 4325–4333 (2018).

17. Koepfli, K.-P., Paten, B., Scientists, the G. 10k C. of & O’Brien, S. J. The Genome 10K Project: A Way Forward. *Annu. Rev. Anim. Biosci.* **3,** 57–111 (2015).

18. Wasko, A. P., Martins, C., Oliveira, C. & Foresti, F. Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. *Hereditas* **138,** 161–165 (2003).

19. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Corrigendum: Direct determination of diploid genome sequences. *Genome Res.* **27,** 757–767 (2017).

20. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).

21. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).

22. Korbel, J. O. & Lee, C. Genome assembly and haplotyping with Hi-C. *Nat. Biotechnol.* **31,** 1099–1101 (2013).

23. Kaplan, N. & Dekker, J. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat. Biotechnol.* **31,** 1143–1147 (2013).

24. Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker’s guide to Hi-C analysis: Practical guidelines. *Methods* **72,** 65–75 (2015).

25. Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* **546,** 524–527 (2017).

26. Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps. *G3&amp;#58; Genes|Genomes|Genetics* **4,** 1339–1346 (2014).

27. Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31,** 1119–1125 (2013).

28. Rose, K. A., Kimmerer, W. J., Edwards, K. P. & Bennett, W. A. Individual-based modeling of delta smelt population dynamics in the upper san francisco estuary: I. model description and baseline results. *Trans. Am. Fish. Soc.* **142,** 1238–1259 (2013).

29. Fish, U. S. & Service, W. *Enhanced Delta Smelt Monitoring 2018-2019 Phase 1 Sampling Preliminary Analysis: DRAFT*. (2019).

30. Frankham, R. Genetics and extinction. *Biol. Conserv.* **126,** 131–140 (2005).

31. Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K. & Allendorf, F. W. Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. *Conserv. Genet.* **11,** 355–373 (2010).

32. Wright, S. Evolution in Mendelian Populations. *Genetics* **16,** 97–159 (1931).

33. Wang, J. Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. B Biol. Sci.* **360,** 1395–1409 (2005).

34. Hill, W. G. Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.* **38,** 209 (1981).

35. Krimbas, C. B. & Tsakas, S. THE GENETICS OF DACUS OLEAE. V. CHANGES OF ESTERASE POLYMORPHISM IN A NATURAL POPULA- TION FOLLOWING INSECTICIDE CONTROL-SELECTION OR DRIFT? *Evolution (N. Y).* 454–460 (1971).

36. Watterson, G. A. On the Number of Segragating Sites in Genetic Models without Recombination. *Theor. Popul. Biol.* **276,** 256–276 (1975).

37. Caballero, A. Developments in the prediction of effective population size. *Heredity (Edinb).* **73,** 657–679 (1994).

38. Robertson, A. The interpretation of genotypic ratios in domestic animal populations. *Anim. Prod.* **7,** 319–324 (1965).

39. Nei, M. & Tajima, F. Genetic drift and estimation of effective population size. *Genetics* **98,** 625–640 (1981).

40. Antao, T., Pérez-Figueroa, A. & Luikart, G. Early detection of population declines: High power of genetic monitoring using effective population size estimators. *Evol. Appl.* **4,** 144–154 (2011).

41. Griffiths, A., Gelbart, W., Miller, J. & Lewontin, R. *Modern Genetic Analysis*. (W. H. Freeman, 1999).

42. Peel, D., Ovenden, J. R. & Peel, S. L. NeEstimator: software for estimating effective population size (version 1.3). *Dep. Prim. Ind. Fish. Queensl. Gov. Brisbane* (2004).

43. Do, C. *et al.* Ne ESTIMATOR v2 : re - implementation of software for the estimation of contemporary effective population size ( N e ) from genetic data. *Mol. Ecol. Resour.* **14,** 209–214 (2014).

44. Ali, O. A. *et al.* Rad capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics* **202,** 389–400 (2016).

45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

46. Ruby, J. G., Bellare, P. & DeRisi, J. L. PRICE: Software for the Targeted Assembly of Components of (Meta) Genomic Sequence Data. *G3&amp;#58; Genes|Genomes|Genetics* **3,** 865–880 (2013).

47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

48. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–501 (2011).

49. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 254–260 (2010).

50. Ni, S. & Stoneking, M. Improvement in detection of minor alleles in next generation sequencing by base quality recalibration. *BMC Genomics* **17,** 1–12 (2016).

51. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15,** 1–13 (2014).

52. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).

53. Gilbert, K. J. & Whitlock, M. C. Evaluating methods for estimating local effective population size with and without migration. *Evolution (N. Y).* **69,** 2154–2166 (2015).

54. Wang, J. A comparison of single-sample estimators of effective population sizes from genetic marker data. *Mol. Ecol.* **25,** 4692–4711 (2016).

55. Wang, J. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genet. Res. (Camb).* **78,** 243–257 (2001).

56. Waples, R. S. Genetic estimates of contemporary effective population size: To what time periods do the estimates apply? *Mol. Ecol.* **14,** 3335–3352 (2005).

57. Waples, R. S. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.* **7,** 167–184 (2006).

58. Jones, A. T., Ovenden, J. R. & Wang, Y. Improved conf dence intervals for the linkage disequilibrium method for estimating effective population size. *Nat. Publ. Gr.* **117,** 217–223 (2016).

59. Lew, R. M. *et al.* Using Next-Generation Sequencing to Assist a Conservation Hatchery: a Single-Nucleotide Polymorphism Panel for the Genetic Management of Endangered Delta Smelt. *Trans. Am. Fish. Soc.* **144,** 767–779 (2015).

60. Waples, R. K., Larson, W. A. & Waples, R. S. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity (Edinb).* **117,** 233–240 (2016).

61. Ko, A. & Nielsen, R. Joint Estimation of Pedigrees and Effective Population Size Using Markov Chain Monte Carlo. *bioRxiv* 1–6 (2018). doi:10.1534/genetics.XXX.XXXXXX

62. Volff, J. N. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb).* **94,** 280–294 (2005).

63. Mei, J. & Gui, J. F. Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. *Sci. China Life Sci.* **58,** 124–136 (2015).

64. Nelson, J. S., Grande, T. C. & Wilson, M. V. H. *Fishes of the World*. (John Wiley & Sons, 2016).

65. Nagahama, Y. Molecular mechanisms of sex determination and gonadal sex differentiation in fish. *Fish Physiol. Biochem.* **31,** 105–109 (2005).

66. Baroiller, J.-F., Guiguen, Y. & Fostier, A. Endocrine and environmental aspects of sex differentiation in fish. *Cell. Mol. Life Sci.* **55,** 910–931 (1999).

67. Nakamura, M., Kobayashi, T. & Chang, X. Gonadal sex differentiation in fishes.pdf. 362–372 (1998).

68. Kikuchi, K. & Hamaguchi, S. Novel sex-determining genes in fish and sex chromosome evolution. *Dev. Dyn.* **242,** 339–353 (2013).

69. Conover, D. O. & Kynard, B. E. Environmental Sex Determination : Interaction of Temperature and Genotype in a Fish Environmental Sex Determinaffon : Interaction of Temperature and Genotype in a Fish. **213,** 577–579 (2013).

70. Devlin, R. H. & Nagahama, Y. Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture* **208,** 191–364 (2002).

71. Mank, J. E. & Avise, J. C. Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex. Dev.* **3,** 60–67 (2009).

72. Lindberg, J. C. *et al.* Aquaculture methods for a genetically managed population of endangered Delta Dmelt. *N. Am. J. Aquac.* **75,** 186–196 (2013).

73. Martínez, P. *et al.* Genetic architecture of sex determination in fish: Applications to sex ratio control in aquaculture. *Front. Genet.* **5,** 1–13 (2014).

74. Liu, H. *et al.* Sex-specific markers developed by next-generation sequencing confirmed an XX/XY sex determination system in bighead carp (Hypophthalmichthys nobilis) and silver carp (Hypophthalmichthys molitrix). *DNA Res.* **25,** 257–264 (2018).

75. Carmichael, S. N. *et al.* Identification of a Sex-Linked SNP Marker in the Salmon Louse (Lepeophtheirus salmonis) Using RAD Sequencing. *PLoS One* **8,** 1–8 (2013).

76. Kafkas, S., Khodaeiaminjan, M., Güney, M. & Kafkas, E. Identification of sex-linked SNP markers using RAD sequencing suggests ZW/ZZ sex determination in Pistacia vera L. *BMC Genomics* **16,** 1–11 (2015).

77. Gamble, T. & Zarkower, D. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol. Ecol. Resour.* **14,** 902–913 (2014).

78. Fowler, B. L. S. & Buonaccorsi, V. P. Genomic characterization of sex-identification markers in Sebastes carnatus and Sebastes chrysomelas rockfishes. *Mol. Ecol.* **25,** 2165–2175 (2016).

79. Gamble, T. Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Mol. Ecol.* **25,** 2114–2116 (2016).

80. Purcell, S. *et al.* PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

81. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57,** 289–300 (2018).

82. Benjamin, A. *et al.* Use of single nucleotide polymorphisms identifies backcrossing and species misidentifications among three San Francisco estuary osmerids Collection Sites for Unknown Fish Yolo Bypass sites Lower Sacramento River site. *Conserv. Genet.* **0,** 0 (2018).

83. Loukovitis, D. *et al.* Quantitative Trait Loci involved in sex determination and body growth in the gilthead sea bream (Sparus aurata L.) through targeted genome scan. *PLoS One* **6,** (2011).

84. Vandeputte, M. & Haffray, P. Parentage assignment with genomic markers: A major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* **5,** 1–8 (2014).

85. Chen, S. *et al.* Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* **46,** 253–260 (2014).

86. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum. Genet.* **104,** 65–75 (2019).

87. Palaiokostas, C. *et al.* Mapping the sex determination locus in the Atlantic halibut (Hippoglossus hippoglossus) using RAD sequencing. *BMC Genomics* **14,** 1–12 (2013).

88. Palaiokostas, C. *et al.* Mapping and Validation of the Major Sex-Determining Region in Nile Tilapia (Oreochromis niloticus L.) Using RAD Sequencing. *PLoS One* **8,** 1–9 (2013).

89. Palaiokostas, C. *et al.* A novel sex-determining QTL in Nile tilapia (Oreochromis niloticus). *BMC Genomics* **16,** 1–10 (2015).

90. Palaiokostas, C. *et al.* A new SNP-based vision of the genetics of sex determination in European sea bass (Dicentrarchus labrax). *Genet. Sel. Evol.* **47,** 1–10 (2015).

91. Anderson, J. L. *et al.* Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One* **7,** (2012).

92. Lamatsch, D. K. *et al.* A transcriptome derived female-specific marker from the invasive western mosquitofish (Gambusia affinis). *PLoS One* **10,** 1–16 (2015).

93. Lu, J. *et al.* Transcriptomic Analyses Reveal Novel Genes with Sexually Dimorphic Expression in Yellow Catfish (Pelteobagrus fulvidraco) Brain. *Mar. Biotechnol.* **17,** 613–623 (2015).