**Tables & Figures**

**Table 1**. List and description of deliverables with the status of each task.

|  |  |  |
| --- | --- | --- |
| **Deliverable** | **Description** | **Status** |
| **Task 1. Sequence and assemble a draft delta smelt genome + linkage map** | | |
| **Task 1.1 Linked-read sequencing (10X Genomics)** | | |
| a. Sample acquisition | Sample tissue of male (M) and female (F) fish. | Completed |
| b. Sequencing | 10X Chromium Library Prep | F Completed,  M Completed |
| Illumina NovaSeq Sequencing | F Completed,  M completed |
| c. Assembly | Supernova assembly | F Completed,  M Completed |
| Assembly QC | F Completed,  M Completed |
| tigmint assembly correction | In progress |
| **Task 1.2 Long-read sequencing (PacBio)** | |  |
| a. Sample preparation | Sample tissue and organs of male and female fish. | F completed,  M completed |
| b. Sequencing | PacBio Prep | F completed,  M completed |
| HiFi Sequencing | F in progress,  M completed |
| c. Assembly | Canu assembly | F in progress,  M completed |
| Assembly polishing | In progress |
| **Task 1.3 Hi-C Chromatin Conformation Capture & Sequencing (Phase Genomics)** | | |
| a. Sample preparation | Sample tissue and organs of male and female fish. | F completed,  M completed |
| b. Sequencing | Hi-C Library Prep | F completed,  M in progress |
| Deep Sequencing | F completed,  M in progress |
| c. Assembly | Salsa2 scaffolding | In progress |
| Assembly curation | In progress |
| **Task 1.4 High-density linkage map** | | |
| a. SNP Discovery | Sample tissue and organs of male and female fish. | In progress |
| b. Construction | Construction of a linkage map using HighMap | In progress |
| **Task 1.5 Map linked regions to reference genome** | | |
| a. Map regions | Align linked regions to genome | In progress |
| **Task 2. Estimate contemporary Ne using RAD-seq data & draft genome** | | |
| **Task 2.1 DNA collection and extraction** | | |
| a. Sample Collection | Collect samples from BY2017, BY2018 and BY2019 cohorts | Completed |
| b. DNA Extraction | Extract DNA from BY2017, BY2018 and BY2019 cohorts | BY2017 completed  BY2018 & BY2019 in progress |
| c. RAD-sequencing | Sequence BY2017, BY2018 and BY2019 cohorts | BY2017 completed  BY2018 & BY2019 in progress |
| **Task 2.2 Estimate Ne in BY2017, BY2018 and BY2019 cohorts** | | |
| a. RAD-sequencing data processing | Split, align and quality control all sequencing data | In progress |
| b. Estimation of Ne | Temporal Ne  LD Ne | In progress |
| **Task 3. Domestication selection** | | |
| **Task 3.1 Select individuals for RAD-seq analysis** | | |
| a. Determine domestication index (DI) | Calculate domestication index for each individual from each generation | Completed |
| b. Group individuals based on their DI | Assign individuals to three DI group: Low, medium and high | Completed |
| c. Determine recovery rate of each individual | Calculate recovery rate of each individual | Completed |
| d. Group individuals based on their recovery rate | Assign individuals to two groups: Low and high recovery rate | Completed |
| **Task 3.2 Preparation and sequencing of individuals** | | |
| a. Sample Collection | Collect samples from BY2008 to BY2015 cohorts | Completed |
| b. DNA Extraction | Extract DNA from BY2008 to BY2015 cohorts | Completed |
| c. RAD-sequencing | Sequence BY2008 to BY2015 cohorts | Completed |
| **Task 3.3 Population Genetic Analysis** | | |
| a. RAD-sequencing data processing | De-multiplexing  Quality filtering | Completed |
| a. Identifying SNPs of domestication selection | Genotype calling  Fst analysis between groups with completed reference genome assembly | In progress |
| d. Association study | GWAS between DI and recovery rate | In progress |
| **Task 4. Identification of sex specific markers** | | |
| **Task 4.1 Investigation of delta smelt genome** | | |
| a. RAD-sequencing data processing | De-multiplex | Completed |
| QC Sequencing data | Completed |
| b. Identification of sex marker(s) | GWAS in M vs F | Completed |
| Depth comparison | Completed |
| **Task 4.2 Fluidigm assay creation** | | |
| a. Submit sequences | Submit sex specific sequences to Fluidigm for SNP Type Assay design | In progress |
| b. Validation | Validation of SNP Assay | In progress |

**Table 2.** Table of tissue type and storage method of sampled delta smelt from three different trips. Included are the names referred to in the text. T= trip, F= female, M=male, BM = back muscle, SC = scales, IO = internal organ, FF = flash frozen, and PG = propylene glycol.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trip Number** | **Sex** | **Sample Name** | **Tissue Type** | **Storage Method** | **Used for Sequencing** |
| 1 | F | T1F01 | Back muscle | Flash frozen | - |
| T1F02 | yes |
| T1F03 | - |
| M | T1M01 | Back muscle | Flash frozen | - |
| T1M02 | - |
| T1M03 | - |
| 2 | M | T2M01 | Back muscle | Flash frozen | - |
| T2M02 | - |
| T2M03 | - |
| 3 | F | T3F01\_BM\_PG | back muscle | propylene glycol | - |
| T3F01\_BM\_FF | back muscle | flash frozen | - |
| T3F01\_IO\_FF | internal organ | flash frozen | - |
| T3F01\_SC\_FF | scales | flash frozen | - |
| T3F02\_BM\_PG | back muscle | propylene glycol | - |
| T3F02\_BM\_FF | back muscle | flash frozen | - |
| T3F02\_IO\_FF | internal organ | flash frozen | - |
| T3F02\_SC\_FF | scales | flash frozen | yes |
| M | T3M01\_BM\_PG | back muscle | propylene glycol | - |
| T3M01\_BM\_FF | back muscle | flash frozen | - |
| T3M01\_SC\_FF | scales | flash frozen | - |
| T3M02\_BM\_PG | back muscle | propylene glycol | - |
| T3M02\_BM\_FF | back muscle | flash frozen | yes |
| T3M02\_SC\_FF | scales | flash frozen | - |
| 4 | M | T4M01\_BM\_FF | back muscle | flash frozen | - |
| T4M01\_IO\_FF | internal organ | flash frozen | yes |
| T4M01\_SC\_FF | scales | flash frozen | - |

**Table 3.** Table of raw data sequencing statistics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sequencing Technology** | **Sequencing Company** | **Sex** | **Number of Runs** | **Coverage** |
| Linked Reads | 10X | M | 1 | 120 |
| Linked Reads | 10X | F | 1 | 80 |
| Long Reads | PacBio HiFi | M | 2 |  |
| Long Reads | PacBio HiFi | F | 3 |  |
| Hi-C | Phase | F | 1 |  |
| Hi-C | Arima | M | 1 |  |

Diagram

Description automatically generated

**Figure 1.** Workflow diagram for completion of assembling reference genome for delta smelt (Task 1). Both male and female delta smelt will undergo the pipeline separately to produce two reference genomes––one male and one female. All work was completed at the GVL unless otherwise noted. A.) Tissue from male and female delta smelt are sampled from the captive colony at the FCCL. B.) HMW DNA is extracted from the fish tissue samples. C.) Extracted DNA of sufficient length undergoes a library prep, subsequent sequencing and quality control for each of the three NGS technologies (linked-read, long-read and hi-c). D.) Linked- and long-read sequencing data are assembled into individual draft genomes. The draft genomes are quality assessed by the software BUSCO13. E.) The linked- and long-read assemblies are scaffolded together to produce a consensus assembly. The quality of the consensus assembly is then assessed by the software BUSCO13. F.) Hi-c data connects long-range gaps in the linked- & long-read consensus assembly. The quality of the consensus assembly is then assessed by the software BUSCO13. G.) Final polishing to produce a reference genome.

A picture containing text, electronics, computer

Description automatically generated

**Figure 2.** Pulse field gel images of extracted HMW gDNA from three sampling trips (“T”) and six rounds of extractions (“E”). Green boxes surround lanes from extracted samples usable for long-read and linked-read sequencing (extraction distribution centered ~ 50kb), yellow boxes surround lanes from samples with insufficient extraction lengths. A) T1, E1: one usable female sample (T1\_F02) B) T2, E2: no usable samples C) T2, E3: no usable samples D) T2, E4: no usable samples (E) T3, E5: one usable male sample (T3M02\_BM\_FF) F) T3, E6: one usable female (T3F02\_SC\_FF).