**DNA Sequencing & Otolith Microchemistry Data Set**

**Pacific cod, South Korea coastal waters**

**Dataset Description**

This data set consists of genetic and otolith microchemistry data for Pacific cod, to be used to determine stock structure of Pacific cod around the Korean peninsula. Mature Pacific cod were sampled from winter spawning aggregates in coastal South Korean waters over the course of several spawning seasons. For each individual, a fin clip or tissue sample was taken and preserved in 100% denatured ethanol, and either one or both otoliths (ear bones) collected and stored dried. Samples were collected by Dr. Sukyung Kang of the National Institute of Fisheries Science (spawning seasons 2014 – 2016), and Dr. WooSeok Gwak of Gyeongsang University (2007-2008 spawning season). Tissue samples from the 2007-2008 spawning season are subsamples from Gyeongsang National University.

The spawning aggregates span three regions of South Korean coastal waters: the western coast (2 sites), southern coast (4 sites), and eastern coast (1 site). The data also includes temporal replicates; one site was sampled during both the 2013-2014 and 2014-2015 spawning seasons to test between-year differences, and one site sampled at the beginning and end of the 2007-2008 spawning season to test within-year differences.

After data filtering and processing, genetic data is available for 253 individuals. Of the 253 individuals processed for genetic analysis, 198 individuals have corresponding otolith data.

Genetic data collection consisted of restriction-site associated DNA sequencing, which produced thousands of single nucleotide polymorphisms (referred to from here on as “markers”). DNA sequencing prep was conducted by the Molecular Ecology Research Lab at the University of Washington, and sequencing completed at the University of Oregon’s Genomics & Cell Characterization Core Facility. Otoliths were processed and aged by the Alaska Fisheries Science Center’s Age & Growth Program. Otolith microchemical data collection was completed by Age & Growth Program lab technicians using the University of Oregon’s mass spectrometer.

The combination of the two datasets is an exploratory analysis of how population structure differs, or remains the same, on evolutionary versus ecological timescales.

The genetic data set can determine the population structure of Pacific cod around the Korean peninsula on evolutionary timescales. The purpose of the otolith microchemistry analysis is to determine population structure on ecological timescales. To do so, multivariate analyses will be conducted on otolith microchemical concentrations of the bone layer at (1) the core of the otolith, representative of the body of water where the fish spent time as an early juvenile, and (2) the edge of the otolith, representative of the body of water where the fish was caught. This comparison will show the extent to which fish show site fidelity to their natal spawning population, or the extent to which there is mixing between genetically distinct populations on ecological timescales.

It would be interesting to access existing environmental data (particularly water temperature data, given temperature-specific spawning behavior in Pacific cod) for the dates and locations where samples were taken. This is not readily available; I would contact the National Institute of Fisheries Science to see if they have a similar buoy system to NOAA, collecting continuous measurements of environmental variables at sea.

**Variables**

All Possible Explanatory Variables

|  |  |
| --- | --- |
| Site | Sampling Site |
| Season | Sampling Spawning Season |
| Region | Sampling Region (West / South / East) |
| TL | Total length of fish (cm) |
| W | Weight of fish (g) |
| GW | Gonad weight of fish (g) |
| Sex | Sex of fish (M/F)\* |
| Maturity | Sexual maturity of fish at time of capture\* |
| Age | Age of fish, as determined from otolith (years) |
| OW | Otolith weight (g) |

\* Low confidence that these variables were correctly measured in the field. Will likely remove from final data set.

All Possible Response Variables

1. Genetic markers
   1. 5,405 single nucleotide polymorphisms (SNPs), recorded as “Locus\_”
2. Otolith chemical markers
   1. All chemicals are expressed as a *relative* concentration to a Calcium standard

|  |  |  |
| --- | --- | --- |
| **Abbreviation** | **Element** | **Correlated with…** |
| B11 | Boron |  |
| Ba138 | Barium 138 | Salinity of water |
| Li7 | Lithium 7 |  |
| Mg25 | Magnesium 25 | Heavy metals (magnesium) in water |
| Mn55 | Manganese 55 |  |
| Pb208 | Lead 208 | Heavy metals (lead) in water |
| Si86 | Silicon 86 |  |
| Zn66 | Zinc 66 | Heavy metals (zinc) in water |

Individual Codes

Each individual sample is coded according to sampling site, sampling date. Following the code is a specific numerical identifier for that sample, and any appropriate suffixes (see next section).

|  |  |  |  |
| --- | --- | --- | --- |
| **Abbreviation** | **Sampling Site Location** | **Sampling Date** | **Analysis Group** |
| PO010715 | Pohang | 1/7/2015 | Pohang 2015 |
| PO020515 | Pohang | 2/5/2015 | Pohang 2015 |
| GE011215 | Geoje | 1/12/15 | Geoje 2015 |
| GE012315 | Geoje | 1/23/2015 | Geoje 2015 |
| GEO020414 | Geoje | 2/4/2014 | Geoje 2014 |
| YS121316 | Yellow Sea Block 161 | 12/13/2016 | Yellow Sea 2016 |
| JB121807 | Jinhae Bay | 12/18/2007 | Jinhae Bay 2007, Early Spawn |
| JB021108 | Jinhae Bay | 2/11/2008 | Jinhae Bay 2007, Late Spawn |
| JUK07 | Jukbyeon | 12/10/2007 | Jukbyeon 2007 |
| BOR07 | Boryeong | 1/24/2007 | Boryeong 2007 |

Individual Suffixes

Describe genetic data only. Suffixes on individual codes provide an additional description of how the sample was processed. This was included in the data set because the difference in processing may affect data quality, and so these samples were tracked through data filtering steps.

|  |  |
| --- | --- |
| Suffix | Processing |
| “\_2” | The sample used in the data set is a re-run of the original sample. This may have been done because of contamination, low read depth, or low DNA quality. |
| “\_c” | The sample used in the data set is a combination of two separate DNA extractions, because of low quantities of DNA in either extraction. |
| “\_rep” | The sample used in the data set was originally prepared as a replicate for an existing sample. The replicate showed higher quality or read depth, and so was maintained in the data set instead of the original sample. |
| “.1” | The sample was put through paired end sequencing. Whereas only forward sequencing reads are available for samples without this prefix, “.1” samples have both forward and reverse reads available for data analysis. The reverse reads will not be used for this project. |