**Introduction**

Genetic data for assessment have relied largely on putatively neutral markers such as microsatellites. However, thorough representation of the genome is particularly critical for exploited organisms with relatively high gene flow to distinguish subtle patterns of differentiation that may be associated with local adaptation. Next-generation sequencing has provided new tools to identify large numbers of informative single-nucleotide polymorphisms (SNPs) across the genome, which includes surveys of both neutral and adaptive loci used to determine levels of genetic diversity and genetic differentiation. Eulachon are anadromous smelts found in the North Pacific Ocean ranging from northern California to the southeastern Bering Sea along the Alaska coast (Gustafson et al. 2012). Females spawn during the spring, upstream of the mouth of large river systems and fertilized eggs adhere to the substrate, where they hatch in about 20–40 d depending on water temperature (Hay & McCarter 2000). Once hatched, the larvae are immediately flushed to sea where they appear to be dispersed by estuarine and ocean currents (Barraclough 1964; Hay & McCarter 2000). The objective of the analysis is to determine population differentiation in eulachon using putative adaptive SNP data.

**Methods**

For this analysis,Putatively adaptive SNP data of Eulachon was obtained from Dryad. Reference: Candy JR, Campbell NR, Beacham TD, Grinnell MH, Narum SR, Larson WA (2015) Data from: Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (Thaleichthys pacificus, Osmeridae), an anadromous Pacific smelt. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.1797v>

**Data analysis**

Statistical analysis was carried out using R packages

SNP data was loaded into a R using the function read. genpop in the package readr. Summary of the data provided the number of individuals as 494, group sizes of samples as 41 33 37 22 40 36 42 71 41 33 66 32, number of alleles per locus as 2. The function popNames in adegenet provided the different populations of samples which are SS08, BELCOL03, COW02, CR12, FRAS09, KC02, KEM01, KEN04, KLK02, SKE10, STIK06, TMR01.

Genetic diversity between populations

Genetic diversity was determined by computing the mean allelic richness and mean heterozygosity for each of the twelve populations using the package diversity with the function divBasic.

Population structure

For population structure, the test for genetic differentiation between pairs of populations carried out using the Weir and Cockerham’s estimate 1984 of genetic distance in diversity package and determine FST values between pairs of samples. Region membership for each group of sample was determined by DAPC conducted in R with the package ADEGENET to identify and describe clusters of genetically related individuals. A neighbour-joining tree plotted to visualize the distances for each of these loci sets based on the distance obtained.

**Results**

The genetic diversity measures between populations for the putative adaptive SNPs displayed a high mean number of alleles (allelic richness) in TMRO1, KENON, BELCO3, KLK02, Fraser and COW02 compared the CR12, SS08, KC02, KEM, SKEL10 and STIK06 populations. However, differences in allelic richness was not much. Mean AR ranged from 3.23 to 3.42 in the total population. Fig 1. The test for differentiation also showed an overall FST of 0.05. and pairwise FST genetic differentiation between all population. Table2. comparison was highly significant (P < 0.05)

|  |  |  |
| --- | --- | --- |
| POP | Heterozygosity | Allelic richness |
| TMR01 | 0.3969 | 3.421 |
| KEN04 | 0.3921 | 3.408 |
| STIK06 | 0.4058 | 3.336 |
| SS08 | 0.4056 | 3.289 |
| SKE10 | 0.4103 | 3.28 |
| KEM01 | 0.4064 | 3.23 |
| BELCOL03 | 0.4101 | 3.321 |
| KC02 | 0.4132 | 3.238 |
| KLK02 | 0.4078 | 3.293 |
| FRAS09 | 0.4197 | 3.209 |
| CR12 | 0.4251 | 3.082 |
| COW02 | 0.4181 | 3.174 |

**Table1**. Estimated mean heterozygosity and mean Allelic richness for each population.

****

**Fig 1.** Bar plot of allelic richness for each of eulachon populations.

**Table 2.** Pairwise FST values between populations for the putative adaptive SNPs

SS08 BELCOL03 COW02 CR12 FRAS09 KC02 KEM01 KEN04 KLK02 SKE10 STIK06

BELCOL03 0.0109

COW02 0.0502 0.0519

CR12 0.0764 0.08411 0.0319

FRAS09 0.0511 0.0482 0.0136 0.0486

KC02 0.0200 0.0095 0.0493 0.0891 0.0401

KEM01 0.0036 0.0115 0.0532 0.0812 0.0482 0.0147

KEN04 0.0807 0.0784 0.0835 0.1050 0.0951 0.0918 0.0744

KLK02 0.0129 0.0075 0.0492 0.0807 0.0437 0.0042 0.0087 0.0863

SKE10 0.0120 0.0068 0.0512 0.0843 0.0485 0.0157 0.0083 0.0768 0.0117

STIK06 0.0119 0.0087 0.0470 0.0816 0.0520 0.0181 0.0094 0.0745 0.0108 0.0059

TMR01 0.0959 0.0913 0.0966 0.1233 0.1038 0.1059 0.0914 0.0050 0.1018 0.0914 0.088



PC2

A

B

PC1

**Fig 3.** Scatterplots showing the first two principal components of the discriminant analysis of principal components (DAPC). Ovals are the inertial ellipse, dots represent individual genotypes and the lines extend to the center of each population. (A) shows the clustering of population of eulachon into three groups and (B) shows the various groups of populations clustered in each group.

Population genetic structure was examined and pairwise FST values ranged from 0.003 for the KM01-SS08 to 0.123 for TMR01 -CRI12 comparison. DAPC analysis showed that the optimal number of clusters of individual genotypes was three using the package adegenet. Fig 3. Shows the cluster of the populations into three groups. These are group1, consists of three populations (COW02, CR12,FRAS09), group2 consists of a seven populations (SS08,BELCOL03, KC02, KEM01, KLK02, SKE10, STIK06) and group3 consists of two populations (TMR01, KEN04).



**Fig 4.** Neighbor-joining tree based on genetic distances calculated using Weir and Cockerham’s estimate 1984.

The Neighbour-joining tree result also indicated a three population with respect to the clustering of each sample group which is similar to the result obtained in the DAPC analysis.

**Discussion and conclusion**

RAD sequencing data is used to discover putatively adaptive SNP variation in eulachon population. The putatively adaptive SNP data demonstrated genetic differentiation among the twelve populations of eulachon. There was more variation between different groups and less variation within groups of populations with respect to their Fst values computed. The genetic diversity observed in the three groups can be said to be related to a demographic expansion from a refuge where leading- edge colonizers following glacial retreat would have both lower allelic richness and expected heterozygosity (Hewitt 1996). The low population differentiation of eulachon in each group is low can be attributed to probably the life history of eulachon, or due to the lingering historical effects of demographic radiation from glacial refugia (Candy et al., 2015).

**References**

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a FST-outlier method. BMC Bioinformatics, 9, 323.

Candy, J. R., Campbell, N. R., Grinnell, M. H., Beacham, T. D., Larson, W. A., & Narum, S. R. (2015). Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (Thaleichthys pacificus, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources*, *15*(6), 1421–1434. https://doi.org/10.1111/1755-0998.12400

Gustafson RG, Ford MJ, Adams PB et al. (2012) Conservation status of eulachon in the California Current. Fish and Fisheries, 13, 121–138.

Hay D, McCarter PB (2000) Secretariat CSA (2000) Status of the eulachon Thaleichthys pacificus in Canada. Canadian Stock

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society, 58, 247–276.