

Population genomic analysis on the greater amberjack (*Seriola dumerili*) in the Mediterranean and Eastern Atlantic, based on single nucleotide polymorphism (SNP) markers

Aikaterini Katirtzoglou
Supervisor : Costas TSIGENOPOULOS

Abstract

This study was carried out to investigate the population structure of the greater amberjack (*Seriola dumerili*) in the Mediterranean and Eastern Atlantic from sequencing data. The bioinformatic analysis of the reads yielded 1,051 SNP markers. Two structure scenarios arise from our data, both suggest that the species forms two genetically distinct groups, one in the Mediterranean and one in the Atlantic, which is consistent with previous findings. The difference in the two cases, lies on whether an admixture zone between the Mediterranean and Atlantic stock exists. Fifteen candidate outlier loci were identified in the dataset, one of which is positioned in a genomic region that is potentially involved in temperature acclimation. This study confirms that greater amberjack Mediterranean and Atlantic populations should be managed as separate stocks, and integrates neutral and evolutionary signs for a better understanding of the species' distribution patterns.

Introduction

- The greater amberjack (*Seriola dumerili*) (Figure 1) is a member of the Carangidae family that has a circumglobal distribution in subtropical and temperate waters (Figure 2) and a prominent aquaculture candidate due to rapid growth rate, excellent flesh quality, big size and high worldwide demand [1,2]
- Understanding the population structure of a species, is pivotal to the successful management of both wild and captured stocks. Population structure, can be studied through the prism of local adaptation, the evolutionary process that leads to higher fitness of a population to its local environment .



Figure 1 : *Seriola dumerili* downloaded from <https://fishider.org/en/guide/osteichthyes/carangidae/seriola/seriola-dumerili>

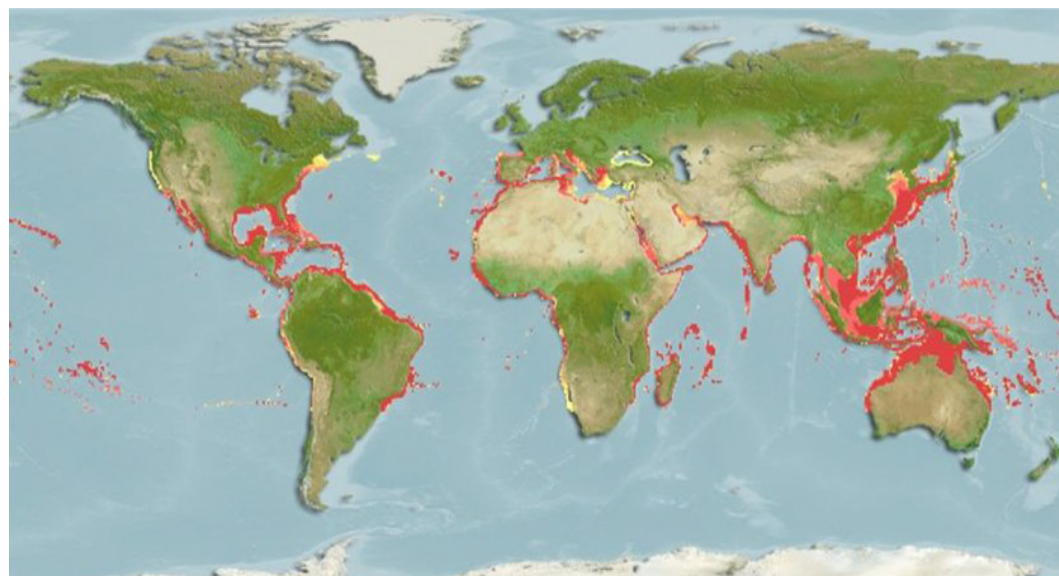


Figure 2 : Current *Seriola dumerili* distribution AquaMaps (2019, October)

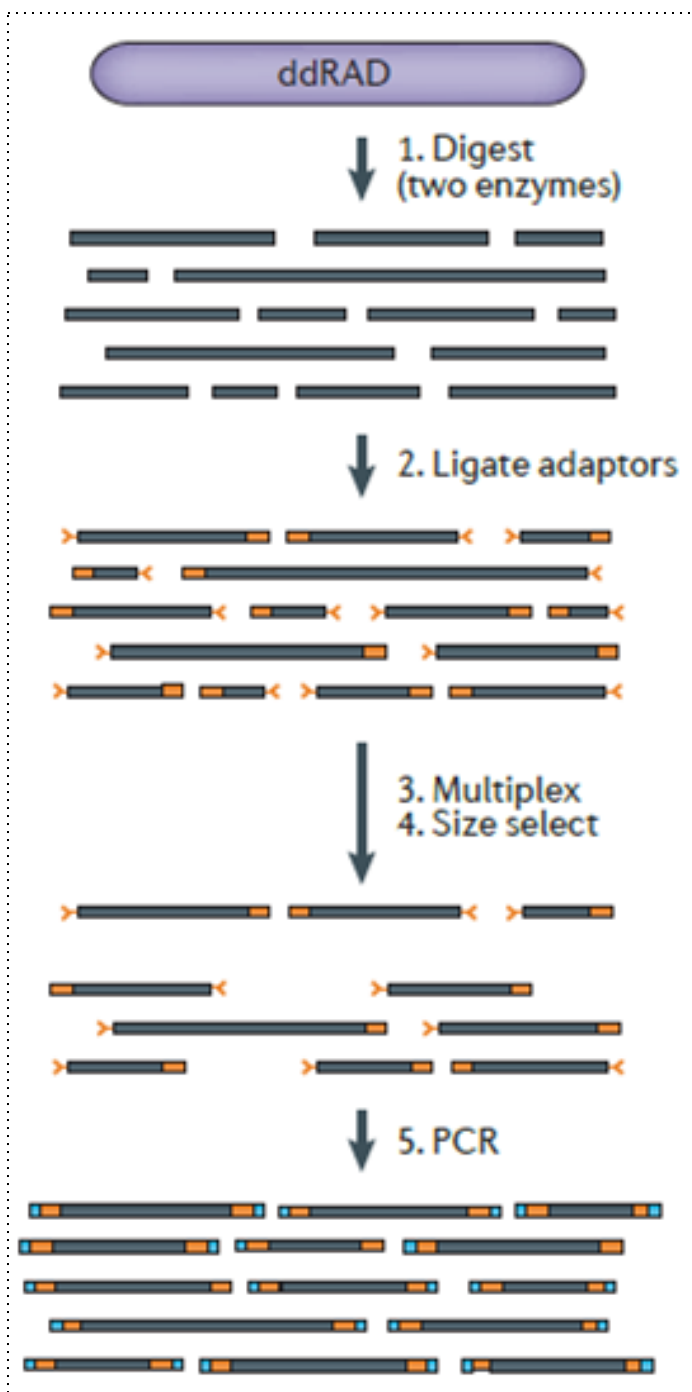


Figure 3 : ddRAD sequencing protocol [4]

Materials and Methods

- Sampling from nine localities (Figure 8) and double digest restriction-site associated DNA sequencing or ddRADseq [3] (Figure 3)*
- SNP calling workflow [5] (Figure 4)

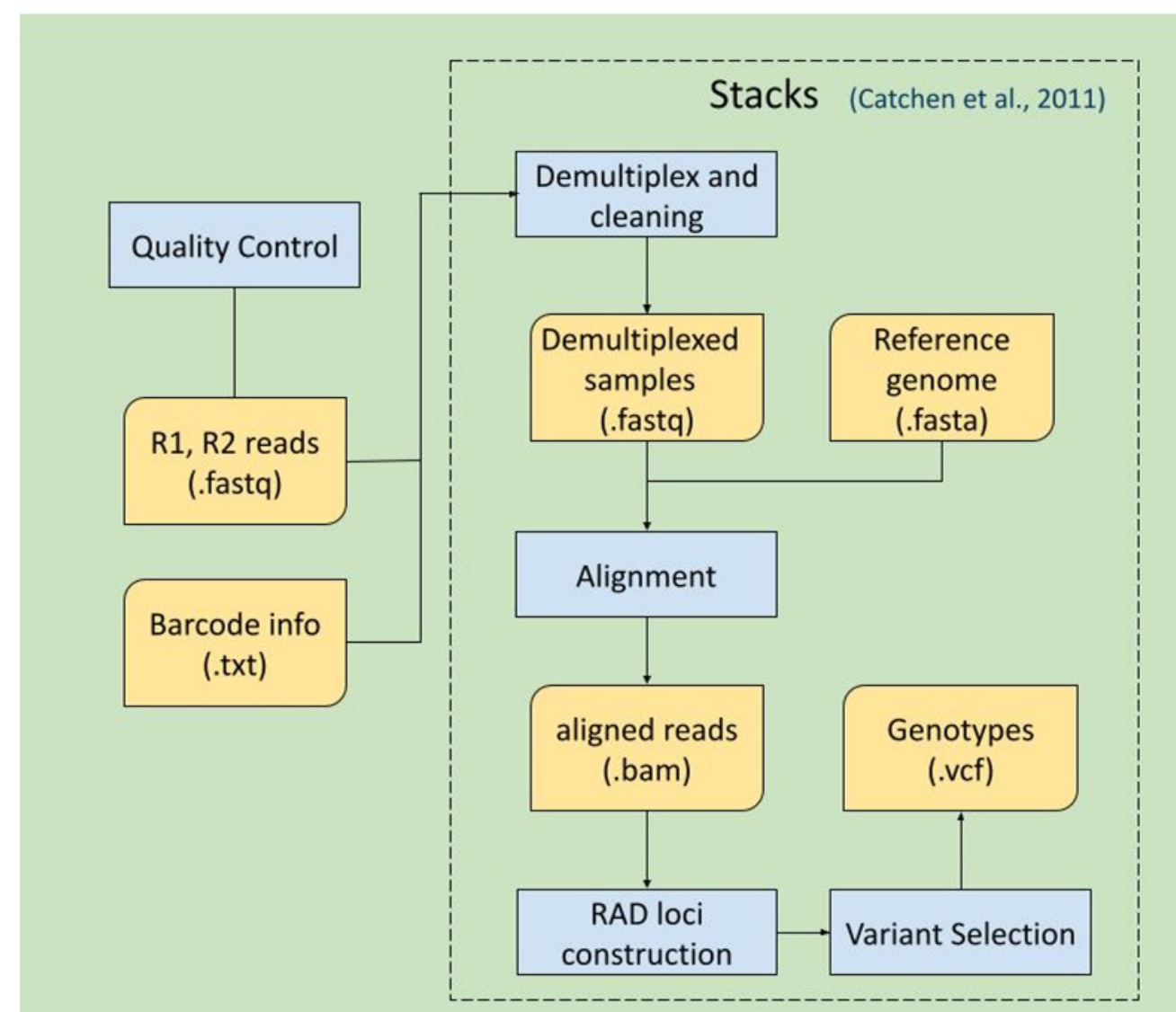


Figure 4 : SNP calling bioinformatic pipeline, Stacks [5]

* Sampling and ddRADSeq were implemented in previous lab projects

- Calculation of genetic distances between pairs of populations (pairwise F_{ST})
- Cluster analyses (Ancestry [6], Bayesian [7], Multivariate[8] and Analysis of Molecular Variance [9])
- Detection of signatures of local adaptation (outlier loci) , with two software, and identification of their closest genes [9,10,11]

Results

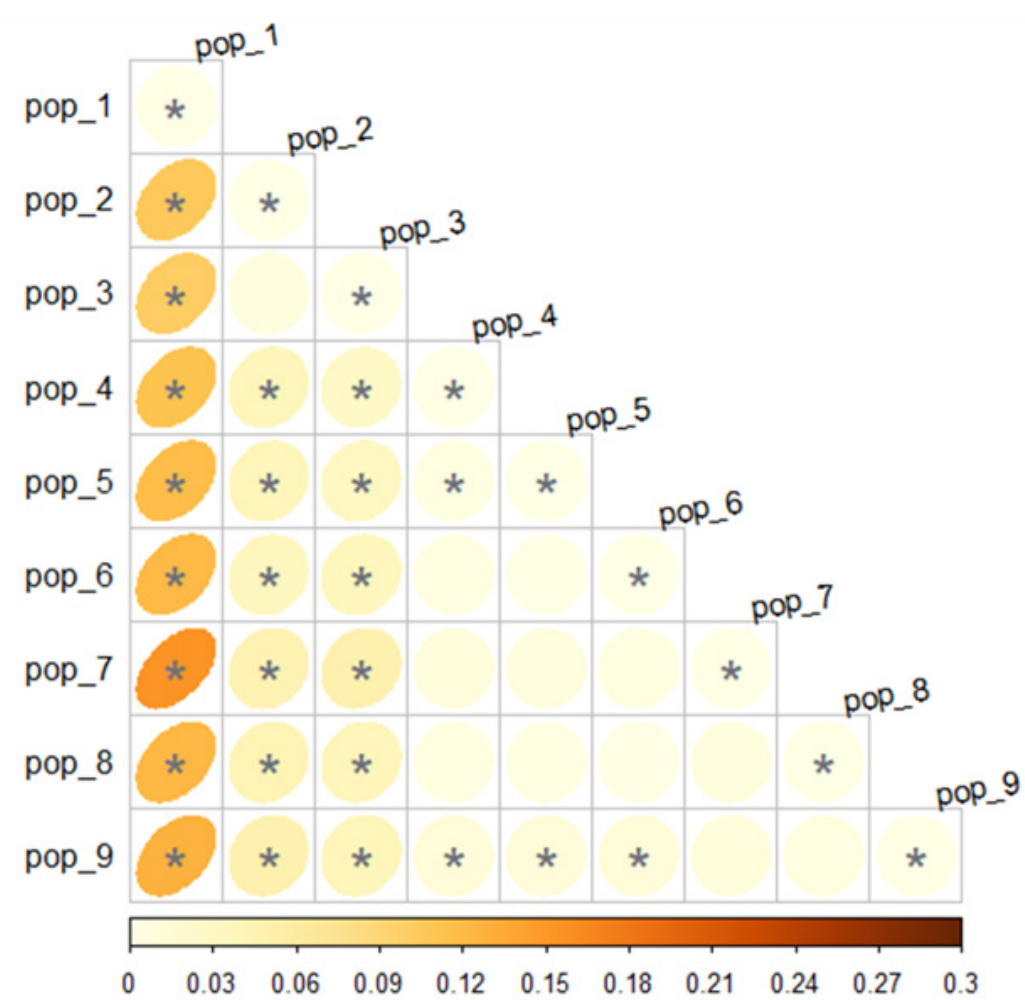


Figure 5: Pairwise F_{ST} matrix implemented in Arlequin ver 3.5 [9]

- The pipeline (Figure 4) yielded **1,051 SNP** markers used in downstream genetic analyses

- The pairwise F_{ST} matrix showed that Gran Canaria population (1) is genetically distant from the Mediterranean populations (4-9) and from the rest of the Atlantic ones (2, 3). Tenerife populations (2,3) are also differentiated from the Mediterranean ones (4-9) but relatively less than Gran Canaria (1). Lastly, a homogenous pattern is observed along the Mediterranean group (4-9) as the populations have low pairwise F_{ST} values (Figure 5)

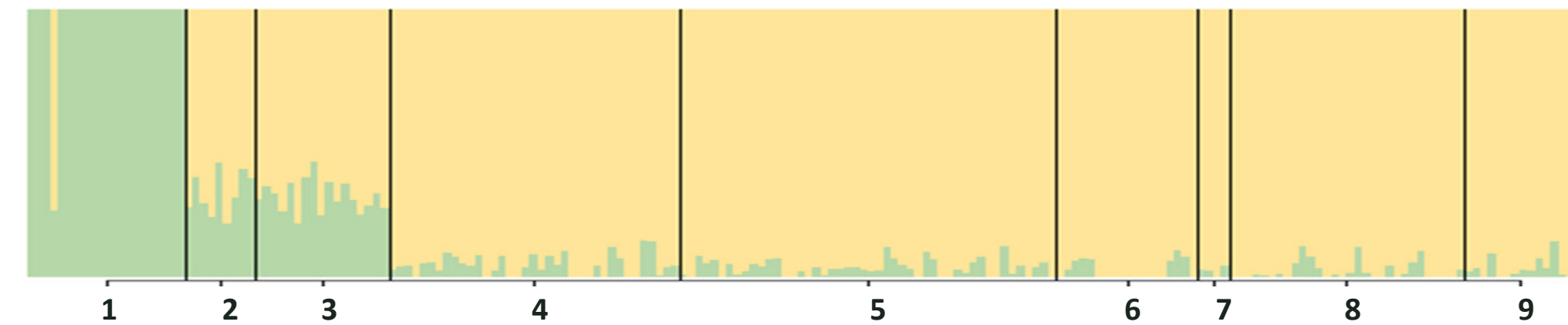


Figure 6 : Ancestry Analysis Q-plot for K=2 for nine populations (1-9) [6]

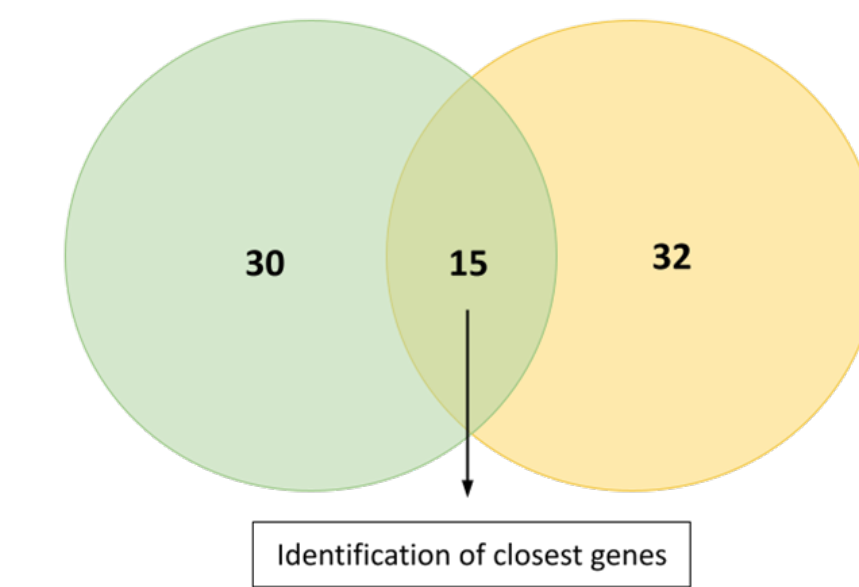


Figure 7 : Venn diagram for identifying the overlapping outlier loci

- Ancestry and Bayesian analysis identified a two cluster (K=2) model (Figure 6, Figure 8a) whereas Analysis of Molecular variance and multivariate analysis a three cluster model (K=3) (Figure 8b)

- Fifteen outlier loci where identified by both detection methods and their genomic locations were investigated

Discussion

A. POPULATION STRUCTURE HYPOTHESES AND OUTLIER INVESTIGATION

Existence of one Mediterranean stock and one in the Eastern Atlantic. The divergent Gran Canaria population (1) could be caught in a migratory route from a more distant location in the Atlantic ocean (Figure 8b) or Tenerife populations (2,3) could potentially belong in a hybrid zone between the two stocks (Figure 8a).

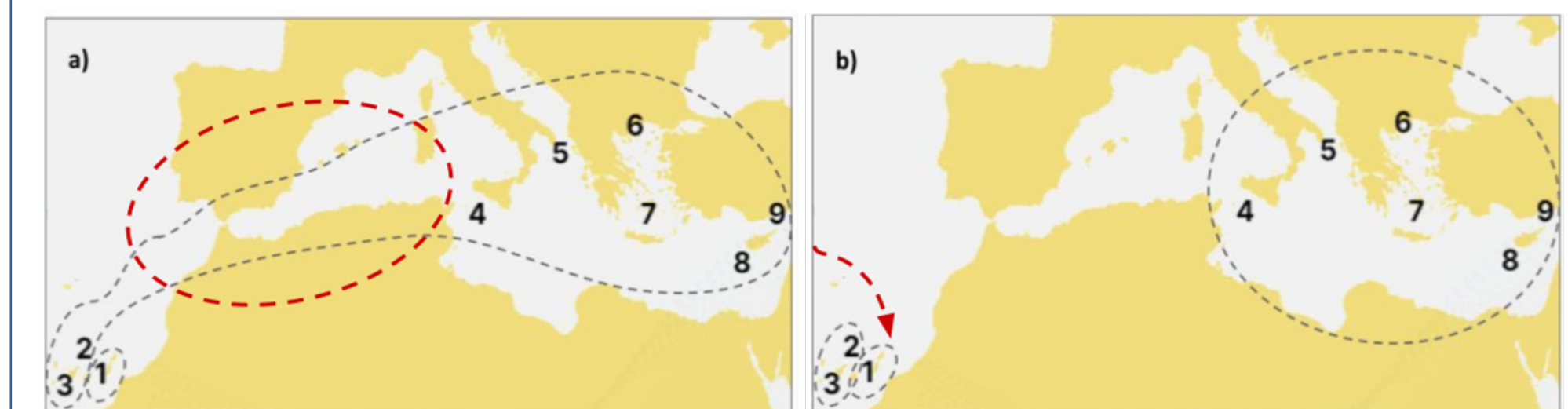


Figure 8 : Hypotheses for the distribution of the greater amberjack

- One of the fifteen is potentially linked with **tef-1** gene (transcription enhancer factor-1) that has been associated with temperature acclimation in medaka [12]

B. CONCLUDING REMARKS

- Atlantic and Mediterranean populations should be considered as separate stocks for future management practices, however more research is needed, centered on the transition point between the two basins
- Investigating markers involved in local adaption can potentially contribute in protecting biodiversity and monitoring a species' movements as a response to the environmental pressure by climate changes

Reference

- Manooch, C. S., & Potts, J. C. (1997). Age, growth and mortality of greater amberjack from the southeastern United States. *Fisheries Research*, 30(3), 229–240. [https://doi.org/10.1016/S0165-7836\(96\)00554-1](https://doi.org/10.1016/S0165-7836(96)00554-1)
- Papandroulakis, N., Mylonas, C. C., Maingot, E., & Divanach, P. (2005). First results of greater amberjack (*Seriola dumerili*) larval rearing in mesocosm. *Aquaculture*, 250(1–2), 155–161. <https://doi.org/10.1016/j.aquaculture.2005.02.036>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5)
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92. <https://doi.org/10.1038/nrg.2015.28>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postelwait, J. H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. 63
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 157(2), 945–959. <https://doi.org/10.1093/GENETICS/155.2.945>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 1–15. <https://doi.org/10.1186/1471-2156-11-94/FIGURES/9>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67–77. <https://doi.org/10.1111/1755-0998.12592>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/BIOINFORMATICS/BTQ033>
- Yamasaki, Y., Komoike, Y., & Higashinakagawa, T. (2006). Adaptive Changes in TEF-1 Gene Expression During Cold Acclimation in the Medaka. <https://doi.org/10.2108/Zsj.23.903>