UNIVERSITY OF CRETE

SCHOOL OF SCIENCES & ENGINEERING

BIOLOGY DEPARTMENT

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

Master's Thesis

Aikaterini Katirtzoglou

1058

Supervisor : Constantinos Tsigenopoulos

Researcher, HCMR

Heraklion 2023

Prologue

This Master's Thesis was conducted in the Genomics and Bioinformatics group, of the Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC) of the Hellenic Centre for Marine Research (HCMR). Many great people contributed to its successful completion.

Abstract

The greater amberjack, *Seriola dumerili* (Risso, 1810), is an oceanodromous, pelagic species that belongs to the Carangidae family. It has attracted considerable economic interest since the 1990s, because it possesses traits that distinguish it as a promising candidate in aquaculture, such as good-quality fillet and high growth rate. However, the challenges of its successful rearing combined with the need for development of management practices, underline the importance of research into the genetic diversity of wild and farmed stocks. Single Nucleotide Polymorphisms (SNPs) are the most utilized markers in population genetics research due to their abundance in the genome and their easier detection due to the rapid development of sequencing technologies. In the present study, a bioinformatic analysis of DNA sequencing data, after digestion with two restriction enzymes (double digest restriction-site associated DNA sequencing, ddRADseq) of 254 individuals from nine greater amberjack populations was conducted, with the aim of finding single nucleotide polymorphisms. Then, population analyses of the genetic structure of the species in the Mediterranean and the Eastern Atlantic were performed. The pipeline yielded 1,051 SNPs, and two structure scenarios were identified. The first suggests that the species forms two genetically distinct groups, one in the Mediterranean and one in the Atlantic, which is consistent with previous findings, and the second scenario indicates the existence of an admixture zone between the two stocks. 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 work enriches our knowledge on the genetic diversity of wild populations of the greater amberjack in the Mediterranean and the Eastern Atlantic, and attempts to investigate signs of local adaptation towards a better understanding of the species' distribution patterns.

Key words

Greater amberjack, ddRAD, sequencing, SNPs, Atlantic, populations, structure, outliers

Περίληψη

Το μαγιάτικο, *Seriola dumerili* (Risso, 1810), είναι ένα ωκεανόδρομο, πελαγικό είδος που ανήκει στην οικογένεια Carangidae και κατανέμεται σε εύκρατα και υποτροπικά νερά παγκοσμίως. Το μαγιάτικο έχει προσελκύσει σημαντικό ενδιαφέρον από τη δεκαετία του 1990 επειδή διαθέτει χαρακτηριστικά που το διακρίνουν ως πολλά υποσχόμενο υποψήφιο είδος ιχθυοκαλλιέργειας, όπως η καλή ποιότητα φιλέτου και ο γρήγορος ρυθμός αύξησης. Οι προκλήσεις όμως της επιτυχημένης μαζικής παραγωγής σε συνδυασμό με την ανάγκη για ορθή διαχείριση του είδους, υπογραμμίζουν τη σημασία έρευνας στη γενετική ποικιλότητα άγριων και εκτρεφόμενων αποθεμάτων, διότι Τα τελευταία χρόνια, οι μονονουκλεοτιδικοί πολυμορφισμοί (*Single Nucleotide Polymorphisms, SNPs*) αποτελούν τους πιο διαδεδομένους γενετικούς δείκτες στις πληθυσμιακές έρευνες λόγω αφθονίας στο γονιδίωμα και ευκολότερης πλέον ανίχνευσης τους, εξαιτίας της ραγδαίας εξέλιξης των τεχνολογιών αλληλούχισης. Στην παρούσα μελέτη πραγματοποιήθηκε αρχικά, η βιοπληροφορική ανάλυση δεδομένων αλληλούχισης DNA μετά από πέψη με δύο περιοριστικά ένζυμα (*double digest restriction-site associated DNA sequencing, ddRAD seq*) 254 ατόμων από 9 πληθυσμούς μαγιάτικου, με σκοπό την εύρεση μονονουκλετιδικών πολυμορφισμών, και στη συνέχεια, η πληθυσμιακή μελέτη της γενετικής δομής του είδους στη Μεσόγειο και στον Ανατολικό Ατλαντικό. Βρέθηκαν συνολικά 1.051 SNPs βάσει των οποίων προκύπτουν δύο πιθανά σενάρια ομαδοποίησης των πληθυσμών.

Σύμφωνα με το πρώτο σενάριο, το είδος σχηματίζει δύο γενετικά διακριτές ομάδες, μία στη Μεσόγειο και μία στον Ατλαντικό, κάτι που είναι σύμφωνο με προηγούμενα ευρήματα, ενώ το δεύτερο σενάριο υποδηλώνει την ύπαρξη υβριδικής ζώνης μεταξύ των δύο αποθεμάτων. Είχε αποδειχθεί για πολλά θαλάσσια είδη, μια σαφής γενετική ασυνέχεια μεταξύ της Μεσογείου και του Ατλαντικού που σε πολλές περιπτώσεις σχετίζεται με βιολογικές και φυσικές διαδικασίες του λεγόμενου φραγμού Αλμερία-Οράν (στη Θάλασσα του Αλμποράν στη Δυτική Μεσόγειο). Αυτό, παρότι για πολλά χρόνια το βιογεωγραφικό όριο μεταξύ των δύο θαλάσσιων λεκανών που υποτίθεται ότι λειτουργούσε ως φραγμός γονιδιακής ροής θεωρούνταν τα στενά του Γιβραλτάρ.

Τα ευρήματα μας έδειξαν ότι στην Κεντρική και Ανατολική Μεσόγειο παρατηρείται ένα ομοιογενές γενετικό γκρουπ. Αυτό μπορεί να εξηγηθεί εν μέρει από βιολογικά χαρακτηριστικά του μαγιάτικου όπως οι μεγάλες κολυμβητικές του ικανότητες και η εκτεταμένη παθητική μεταφορά των λαρβών και νεαρών ατόμων, που τείνουν να παραμένουν κοντά σε στρώματα φυκών (algae mats). Στο σύνολο των δεδομένων, επίσης, εντοπίστηκαν 15 υποψήφιοι αποκλίνοντες (*outliers*) γενετικοί τόποι, δηλαδή τόποι που δυνητικά εμπλέκονται σε διαδικασίες προσαρμογής ενός πληθυσμού στο περιβάλλον του (*local adaptation*). Ένας από

τους outliers που βρέθηκαν είναι πιθανώς συνδεδεμένος με ένα γονίδιο που έχει προταθεί ότι εμπλέκεται σε μηχανισμούς προσαρμογής στη θερμοκρασία.

Η εργασία αυτή εμπλουτίζει τη γνώση μας σχετικά με τη γενετική ποικιλότητα άγριων πληθυσμών μαγιάτικου στη Μεσόγειο και στον Ανατολικό Ατλαντικό και ενισχύει τις βάσεις για την ανάπτυξη προγραμμάτων γενετικής βελτίωσης ενός είδους μεγάλης οικονομικής σημασίας στην ιχθυοκαλλιέργεια. Τέλος, μελετώντας γενετικούς δείκτες που εμπλέκονται σε εξελικτικούς μηχανισμούς (outliers) μπορεί να εξετασθεί πιο εμπεριστατωμένα η δομή ενός πληθυσμού και τα πρότυπα κατανομής του σε διαφορετικά περιβάλλοντα, και έτσι να προστατευθεί η γενετική ποικιλότητα του είδους.

Λέξεις κλειδιά

Μαγιάτικο, ddRAD, αλληλούχιση, SNPs, Μεσόγειος, Ατλαντικός, πληθυσμοί , δομή, outliers

Table of contents

I. Introduction

1.1 The greater amberjack (*Seriola dumerili*)

The greater amberjack (*Seriola dumerili*) is a member of the Carangidae family, lives in pelagic waters at depths of 20–70 meters, and is often found near reefs (Manooch & Potts, 1997; Mazzola et al., 2000).

The greater amberjack is an oceanodromous species that was initially known as a popular catch for recreational fishers due to its size, endurance and especially the demonstration of fighting capabilities when hooked (Harris et al., 2007; Manooch & Potts, 1997). It has a circumglobal distribution in subtropical and temperate waters (Wells & Rooker, 2004) and it is common in the Mediterranean (Thompson et al., 1999) (Figure 1).
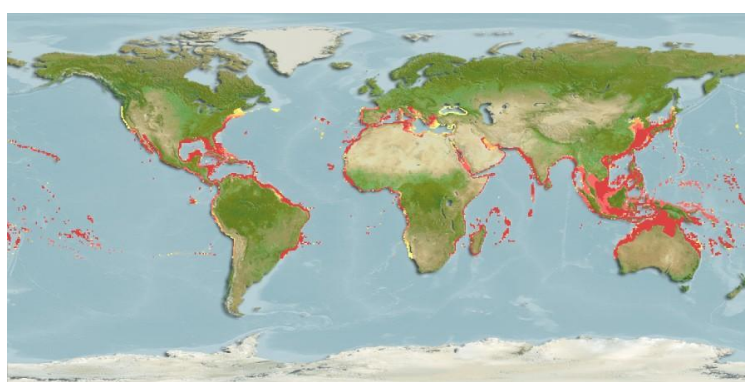
Figure 1 : Current *Seriola dumerili* distribution Aqua Maps (2019, October). Computer generated distribution maps for *Seriola dumerili* (Greater amberjack)

Other members of the *Seriola* genus include the Japanese yellowtail (*Seriola quinqueradiata*) in Japan, and the yellowtail kingfish (*Seriola lalandi*) which is found in genetically distinct populations in Australia - New Zealand, Japan and California, often considered as separate species (*S.lalandi, S.aureovittata and S. dorsalis*, respectively) (Martinez-Takeshita et al., 2015). *Seriola rivoliana* is another economically important carangid, widely distributed in the Eastern and Western Pacific (Corriero et al., 2021; Fernández-Palacios et al., 2015) that has also, recently been reported in the Mediterranean basin (Šegvić-Bubić et al., 2016).

1.2 Importance for the aquaculture

The greater amberjack has attracted strong interest in Europe and Japan since the 1990s, because it possesses many qualities classifying it as an aquaculture candidate, like a rapid growth rate, excellent flesh quality, big size and high worldwide demand (Papandroulakis et al., 2005).

Nevertheless, the challenging reproduction of the species in captivity (even if the fish were taken from the sea at a really early stage in their development) has prevented its rapid and extensive commercial production. Now, reproduction can

successfully be induced hormonally in wild-caught greater amberjack reared in captivity (Mylonas et al., 2004), demonstrating the potential of the species in aquaculture. The need to diversify the aquaculture worldwide (Cahiers Options Méditerranéennes,1995) led to a rekindled interest for greater amberjack, especially in developing methods to control egg production in captivity and studying the reproductive biology of the species (Ottolenghi & Food and Agriculture Organization of the United Nations., 2004; Pousis et al., 2018; Zupa et al., 2017) .

Other than that, bottlenecks for the further expansion of *Seriola* spp. farming appear to be disease control, lack of genetic selection programs and incomplete knowledge of nutrient requirements for these species (Sicuro & Luzzana, 2016). These difficulties occurring for the successful rearing of a new organism in aquaculture, highlight the importance of exploring the genetic background of new marine candidates. In addition, genetic studies constitute the baseline for breeding programs, to trace the origin of breeders and can be utilized to assess the potential genetic risk of intra-specific breeding among wild and farmed fish (Šegvić-Bubić et al., 2016).

1.3 Mediterranean and Eastern Atlantic phylogeography of *Seriola dumerili* and other pelagic species

The understanding of the distribution of a species' genetic diversity, plays a pivotal role in its successful rearing and the management of both wild and captured stocks(Araki & Schmid, 2010). Genetic diversity is the raw material for natural selection allowing species to adapt to new environmental conditions, thus its loss will decrease species adaptive potentials (Conover et al.,2006).

It was recently shown that many marine species display a clear genetic discontinuity between the Mediterranean Sea and the Atlantic Ocean, related to the Almeria-Oran front(Bargelloni et al., 2003; Cimmaruta et al., 2005; Lemaire et al., 2005; Magoulas et al., 2006; Viñas et al., 2004). On the contrary, for many years the biogeographical boundary between the two basins assumed to act as the barrier of gene flow was considered to be the Straits of Gibraltar. However, more recently it has been suggested that the interaction of the Atlantic water with the denser and more saline Mediterranean water in the Alboran Sea, results in the formation of the hydrographic front of Almería-Oran (Figure 2) which is the transition point for many species (Reuschel et al., 2010).

Figure 2 : Straits of Gibraltar (GB) and Almeria-Oran front (AOF). The breakpoint between Atlantic and Mediterranean basin for many species is the Almeria – Oran zone, a hydrographic front close to the Straits of Gibraltar.

In recent population genetics analyses, a more holistic approach is applied with use of spatial, oceanographic and genomic data that influence the gene flow between groups. Studies are numerous, and below few examples are given. The Mediterranean rainbow wrasse (*Coris julis*) is divided into two genetically distinct populations (Mediterranean –Atlantic) with an intermediate group that shares the "Atlantic" and "Mediterranean" genetic background, located in the Alboran Sea (Fruciano et al., 2011). An important member of the sea bass family, Serranidae, *Serranus cabrilla* is characterized by further division within the Mediterranean, associated with less studied fronts, such as the Ibiza Channel and the Balearic Front, that dictate the population structure within the basin (Schunter et al., 2011). A clear genetic break between the two basins was also reported for European hake (*Merluccius merluccius*) with a weak differentiation within them. The use of SNP markers, however, gave insight to potential signs of local adaptation and revealed more complex structure patterns within the two seas (Milano et al., 2014). A whole genome assembly in European sea bass (*Dicentrarchus labrax)* also showed a Mediterranean - Atlantic divergence and associated gene expansions with adaptation to salinity (Tine et al., 2014)

For the greater amberjack, there are still numerous gaps concerning the population structure and the ecology of the species. Two genetic stocks were recognized in the United States of America coast, one in the Gulf of Mexico and one in the southeast Atlantic (Gold & Richardson, 1998). A hypothesis tested by emerging data, indicated that there is a high level of admixture between them, and the genetic structure pattern of the species is more complex than was originally assumed (Hargrove et al., 2018). There is also evidence for two groups of populations of greater amberjack in Asia, one in the East China Sea and one in the South (Araki et al., 2018; Hasegawa et al., 2020; Nugroho et al., 2000)

The genetic structure of Eastern Atlantic and Mediterranean populations indicates the existence of a strong breakpoint between the two basins, with Mediterranean populations being genetically distinct from the Atlantic one (Šegvić-Bubić et al., 2016). Moreover, the Mediterranean group is characterized by very low divergence between the populations sampled. The lack of population subdivision of the greater amberjack in the Mediterranean was confirmed again in a later study (Šegvić-Bubić et al., 2022) and has been mainly attributed to its biological features.

Additionally, a recent project on greater amberjack populations in the two basins, with the use of mitochondrial and microsatellite markers provided further insights to the distribution of the species (Kolios 2017, personal communication). The nuclear loci provided evidence of two separated clusters (Mediterranean-Atlantic) with the suggestion of a hybrid group between them. The mtDNA data showed two co-existing lineages within the two basins, supporting a hypothesis of the concomitant entrance of two already genetically differentiated groups of greater amberjack, during abrupt paleoclimatic events (Šegvić-Bubić et al., 2016).

The species shows high dispersal abilities and an extended pelagic larval stage. The survival of the juveniles that congregate in schools at this stage, is closely connected with algae mats or other flowing objects, and it has been reported that they can be passively transported at long distances by oceanic currents (Hasegawa et al., 2017, 2020). Even slight changes in survival rates of juveniles, could generate order-of-magnitude differences in the supply of new recruits (Houde, 2008). In addition, its reproductive traits (multiple spawnings) that take place from May to

July in the Mediterranean, and from April to September in the Eastern Atlantic (Harris et al., 2007; Sley et al., 2014)could play a central role in the species' seasonal movements. Spawning areas can be determined by various criteria, such as abundance of predators, water temperature or seawater circulations, and can seriously affect levels of gene flow between individuals (Ottmann et al., 2021). In the Mediterranean basin, a spawning site of the greater amberjack has been reported in the Pelagie islands area (Central Mediterranean) by Andaloro et al.(1992).

1.4 RAD techniques

The genetic markers that have recently prevailed to study the population structure of a species, are microsatellites and single nucleotide polymorphisms (SNPs). Modern advances in sequencing technologies make the discovery of large numbers of SNPs possible in the population genomics 'era'. However, whole genome sequencing is still relatively expensive and fortunately in many studies, gathering complete genomic sequence data is not necessary (Etter et al., 2011)

Restriction-site associated DNA sequencing (RAD sequencing or RAD-Seq) combines the use of genome complexity reduction with restriction enzymes (REs), and the high sequencing output of modern Next Generation Sequencing platforms (Andrews et al., 2016). The initial protocol for RAD-Seq was described by Baird et al.(2008). Three modifications of this methodology have been mainly used in aquaculture genetics research. Aside from the original RAD-Seq (Baird et al., 2008) that digests genomic DNA with one restriction enzyme, the 2b-RAD method (Wang et al., 2012) uses type IIB restriction enzymes, which cleave DNA upstream and downstream of the recognition site, and ddRAD (Peterson et al., 2012) uses two restriction enzymes, with adaptors specific to each enzyme (Figure 3) (Andrews et al., 2016).
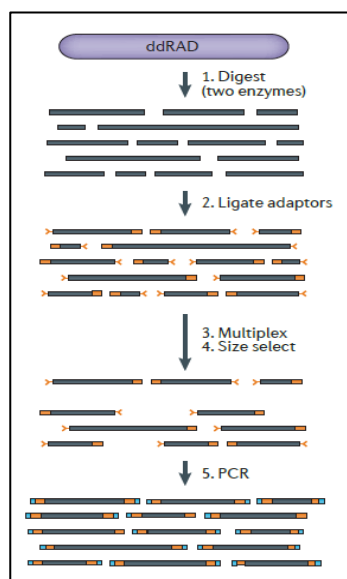
Figure 3 : Double Digest Restriction enzyme Associated DNA sequencing protocol. The main difference of ddRAD methodology is the use of two Restriction Enzymes and size selection of fragments with a specific length for sequencing (Andrews et al.,2016)

The applications of RAD sequencing in aquaculture (Andrews et al., 2016; Robledo et al., 2018) are numerous and have expanded in many directions such as construction of genetic maps (Recknagel et al., 2013), comparative genomics in non-model organisms (Manousaki et al., 2016), genome assemblies for valuable aquaculture species (Tine et al., 2014), and mapping QTL associated with traits of economic importance in salmonid species (Houston et al., 2008). RAD-seq techniques have been extensively applied to generate population-level SNP genotype data and infer population structure and evolutionary patterns in many marine organisms, such as the blind cavefish (Bradic et al., 2013), the threespine Stickleback *Gasterosteus aculeatus* (Hohenlohe et al., 2010), the Atlantic mackerel *Scomber scombrus* (Rodríguez-Ezpeleta et al., 2016), the Nassau grouper *Epinephelus striatus* (Sherman et al., 2020), the redlip croaker *Larimichthys polyactis* (Zhang et al., 2019), the Red drum (Hollenbeck et al., 2019) , the American lobster *Homarus americanus* (Benestan et al., 2015), the Chilean blue mussels *Mytilus chilensis* (Araneda et al., 2016) and freshwater fish like brook trout *Salvelinus fontinalis* (Létourneau et al., 2018), the scaly yellowfish *Labeobarbus natalensis* (Stobie et al., 2018) and many others.

1.5 Advances in genomic studies of *Seriola* genus

The availability of genomic resources for the *Seriola spp.* has begun to expand only the last few years, as a consequence of their increasing economic value in aquaculture. In Japanese yellowtail, a genome wide analysis was implemented that led to the discovery of quantitative trait loci (QTL) associated with mechanisms of resistance to parasites (Ozaki et al., 2013) and economically important traits (Aoki et al., 2015). An annotated, high-quality draft genome was recently constructed for the same species (Yasuike et al., 2018). Also, linkage mapping and identification of markers associated with valuable traits have been reported in yellowtail kingfish (Nguyen et al., 2018).A very important milestone in genomic related analyses of the *Seriola* genus, is the chromosome-level assembly in Japanese yellowtail jack (*S. aureovittata*) along with the discovery of a putative sex chromosome (Li et al., 2022). As mass production of the greater amberjack is still in its infancy, genomic research on the species has been limited. Two attempts of whole genome sequencing (WGS) exist at present for the species, both at the scaffold level (Araki et al., 2018; Sarropoulou et al., 2017)

II. Aim of the thesis

The objective of this thesis was to infer the population structure of the greater amberjack in the Mediterranean and Eastern Atlantic based on SNP markers generated from ddRAD sequencing data. In addition, neutral and outlier loci were identified and used to explore the potential signatures of local adaption. Results were also compared to those previously reported by (Kolios 2017, personal communication), from microsatellite and mitochondrial DNA sequencing of the present dataset.

This project can set the ground for the development of future management strategies and genetic improvement programs in wild greater amberjack stocks. Stock assessment plays an essential role in protecting the biodiversity of a species in the wild, and can determine the success rate of a captured organism in mass production. The selection of breeders in aquaculture has a fundamental effect on stock quality. The main goal for successful breeding practices, is the maintenance of the species genetic variability. Inbreeding can lead to greater vulnerability to parasites, lower tolerance to environmental stress, and a lower degree of reproductive success, thereby reducing the fitness of individuals (Oliveira et al., 2018). Moreover, this reduced genetic variability, can pose a threat for wild populations in case of escape events.

Lastly, in this thesis we investigated markers involved in evolutionary mechanisms that can lead to an in-depth inspection of the genetic diversity of a species. They can potentially provide fine-scale patterns of population structure not detected with neutral markers, and a better insight to the species future geographical distribution as a consequence of the ongoing fluctuation of environmental variables, due to constant climate changes.

III. Materials and methods[1]

Sampling[1]

In order to study the genetic structure of greater amberjack in the Mediterranean and Eastern Atlantic, 254 samples from different aquaculture companies were used in this study. The information provided by the companies indicate that the individuals are wild and they were captured in nine different locations across the Mediterranean Sea and the Eastern Atlantic Ocean. The dataset consists of 3 populations originating from the Canary Islands (1-3) and 6 populations from Central and Eastern parts of the Mediterranean (4-9), as shown in Table 1 and Figure 4.

Table 1 : *Seriola dumerili* sampling details. Population ID (1-9), fishing area and number of captured individuals

Population ID	Fishing area of greater amberjack	Number of samples
1	Gran Canaria, Spain	25
2	North Tenerife, Canary Islands (caught in 2015)	13
3	Tenerife, Canary Islands (caught in 2013-14)	22
4	Lampedusa, Italy	45
5	Astakos, Greece	54
6	Chalkidiki, Greece	18
7	North Crete, Greece	4
8	Cyprus	56
9	Mersin, Turkey	17
	Total	**254**

Figure 4 : Geographical distribution of greater amberjack populations capturing sites

---

[1] Sampling and library preparation were conducted in previous experiments. The present research project begun with the Bioinformatic analysis of the reads (3.1)

Double Digestion RAD library preparation and sequencing[1]

Total DNA from the 254 greater amberjack samples was extracted following a modified salt-based extraction protocol described by Manousaki et al. (2016). The DNA samples were separately digested by two restriction enzymes (RE): *Sbf*I (CCTGCA|GG recognition site), and *Sph*I (GCATG|C recognition site). Barcoded adapters were designed in order to meet specific criteria. The first is that restriction sites should not be reconstructed by adapter-genomic DNA ligations. Second, the adapter mix consists of P1 (*Sbf*1 compatible) and P2 (*Sph*I compatible) adapters that include a five to seven base inline barcode each. The number of different short sequences used as barcodes for sample identification is such, so that the unique combination of two, is individual-specific. Third, the P1 and P2 were added with ratio 1:12, as this was expected to more accurately reflect the abundance of *Sbf*1 and *Sph*I cutting sites. The ligated samples were then denatured, and mixed into a single DNA pool. Size selection of the DNA fragments was performed by agarose gel separation, with a 300-600 base pairs (bp) range, followed by PCR amplification of the DNA template. The last step of the ddRAD library preparation included some further clean up and size selection with the use of magnetic beads. Lastly, paired end sequencing of the library was performed in Novaseq Illumina platform.

3.1 Bioinformatic Data Analysis

3.1.1 Quality check and demultiplexing

The raw data were included in two compressed FASTQ files: the *R1.fastq.gz* representing the forward reads, and the *R2.fastq.gz* with the reverse reads, with read length of 151 BP. Raw reads were analyzed in Stacks v2.59 (Rochette et al., 2019)the software assembles large numbers of short read sequences from a set of individuals, identifies and then genotypes loci within them, in search of single nucleotide polymorphisms (SNPs). Before running the Stacks pipeline (Figure 5), a quality control check was performed in fastQC 0.11.9 (Andrews and Babraham Bioformatics, 2010).

Then, the program *process_radtags* from Stacks was used in order to recover, clean and demultiplex the raw data into the individual samples. A barcode text file was supplied to the program, where each individual is assigned to the combination of barcodes used for its identification. The settings in *process_radtags* specify the position of the barcodes in the sequence, in this case - - inline_inline, as well the restriction enzymes used in library construction, --renz_1 sbfI --renz_2 sphI. The parameters set were - c , - q , - r . The – c parameter is used to clean the data and remove reads with uncalled base, – q for discarding reads with low quality scores (Phred score < 10) and - r corrects sequence errors in the RAD cut site and barcodes that would otherwise make them undetectable for the analysis

---

[1] Sampling and library preparation were conducted in previous experiments. The present research project begun with the Bioinformatic analysis of the reads (3.1)

3.1.2 Data alignment against greater amberjack reference genome

The reference genome of greater amberjack used in alignments is from Araki et al. (2018) (RefSeq assembly accession: GCF_002260705.1). To align our samples against the reference genome, the Burrows-Wheeler Alignment (bwa) software package was employed (Li & Durbin, 2009). BWA first constructs an index of the reference genome. Then, it proceeds to align the paired end reads of each sample on it, via BWA MEM algorithm. The default options were used plus the –M parameter which flags additional hits of a read on the reference genome as secondary. The SAM output files were converted to BAM files with SAMtools 1.9 (Li et al., 2009) with – samtools_ sort command. Last, in order to assess the mapping rate of the reads – samtools flagstat was used, which generates a report on alignment statistics for each individual.
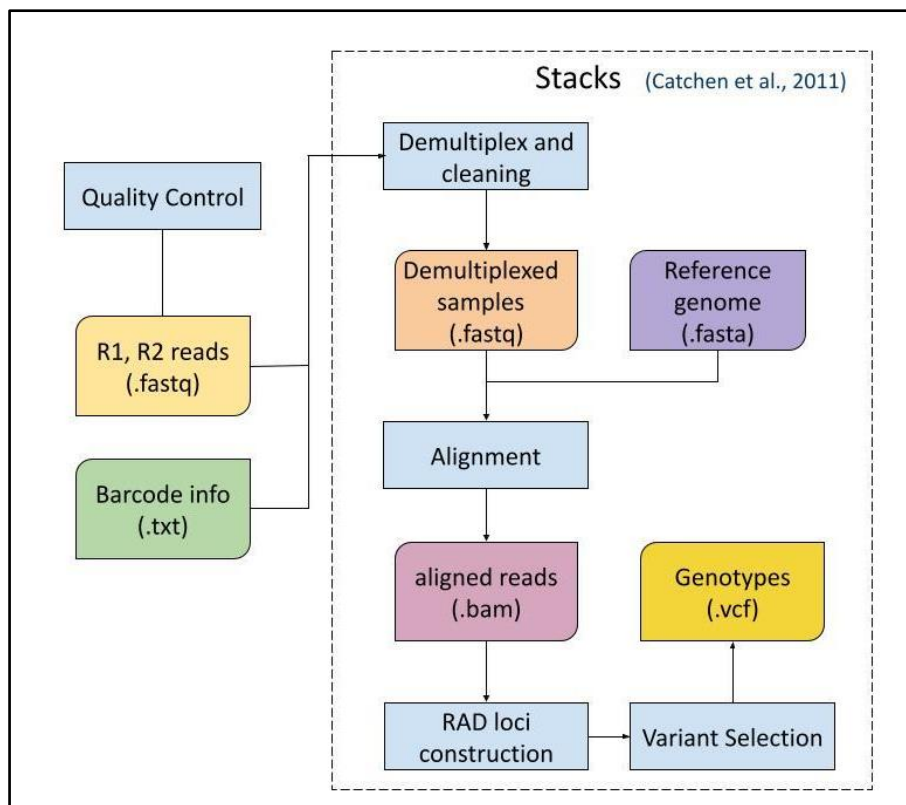
Figure 5 : SNP calling pipeline. After quality control reads are demultiplexed and aligned on the reference genome. The RAD loci are then constructed and in the last step variants are selected based on custom filtering. The pipeline results in a file with genotypes that is used in downstream analyses. Implemented in Stacks (Rochette et al., 2019)

3.1.3 Genotyping RAD loci

The following step in the pipeline (Figure 5), *gstacks*, examines a RAD data set one locus at a time, looking at all individuals for that locus and proceeds to identify SNPs within it. The BAM files generated by SAMtools are passed to *gstacks*, along with a population map (Table 1, Appendix). All the genotyped loci across the total of individuals are stored in the *Catalog* FASTA file.

In order to conclude to a specific number of polymorphic loci, the program *populations* was used as the final step of the Stacks pipeline with the following settings: -R = -0.75 which indicates that a locus needs to be genotyped in at least 75% of individuals across all populations, --min-maf=0.05 and --max-obs-het 0.7 which are the minimum minor allele frequency and the maximum heterozygosity required to process a nucleotide site at a locus, respectively. The --write-random-snp option restricts data analysis to select one random SNP per locus. The options –genepop, -vcf and –structure were used, providing different formats of storing the data for downstream analyses. The data set was reformed by excluding individuals with a high number of missing data. This was implemented by counting the missing data occurrences per individual and excluding the individuals that contain > 20% of missing data.

3.1.4 *In silico* digestion

The gstacks program identified a relatively low number of RAD loci in the catalog file, compared to some other ddRAD studies in teleosts (Maroso et al., 2021; Nousias et al., 2022), where the libraries were constructed with the same restriction enzymes and reads were analyzed in Stacks (Rochette et al., 2019). Consequently, we wanted to confirm if the expected number of sites is similar to the observed one.

SimRAD R package (Lepais & Weir, 2014)simulates the construction of a restriction site associated DNA sequencing (RAD) library, by performing *in silico* restriction enzyme digests and fragment size selection on a given DNA sequence.

In order to run the SimRAD program, the reference genome of greater amberjack (Araki et al., 2018)was provided along with recognition sites of the enzymes A:*Sbf*I and B:*Sph*I. The digestion method was determined (double digest) and lastly, the type of DNA fragments (AB) along with the size selection window (300-600 bp).

3.2 Measures of genetic differentiation

3.2.1. Pairwise F_{ST}s

A very reliable measure of gene flow and genetic differentiation between two populations is pairwise F_{ST}, that calculates the distance as a result of genetic drift (Reynolds et al., 1983). This was implemented by the population comparison test in

Arlequin v 3.5 (Excoffier & Lischer, 2010). The input file used in all Arlequin analyses is an *.arp* file created by the .structure file from Stacks (Rochette et al., 2019) via PGDSpider (Lischer & Excoffier, 2012). The number of permutations performed was 10,000 and the significance level was set to 0.05. The allowed level of missing data was set to 1 (all loci included), as the data filtering took place in the previous stage of the analysis.

3.2.2 Inbreeding coefficient (F_{IS}) and Hardy-Weinberg exact tests

The F_{IS} index or inbreeding coefficient, first described by Wright 1965, explains the deviation of the observed heterozygosity from the expected heterozygosity within a subpopulation as a result of non-random breeding of individuals.

The inbreeding coefficient F_{IS} (Weir and Cockerham , 1984) for each population (1-9) was estimated with the R package *genepo*p v. 4.7.5 (Rousset, 2020). Also, Hardy Weinberg exact tests (Probability test, Fisher's method) were conducted to investigate whether each population is in equilibrium. The F_{IS} estimate of each population was accepted based on the significance of population-specific HWE tests.

3.3 Cluster Analysis

3.3.1 Analysis of Molecular Variance (AMOVA)

Analysis of Molecular Variance is one of the approaches used in order to estimate the population genetic structure. AMOVA is based on the F-statistics (Wright, 1965), which are calculated from a set of covariance components, corresponding to the different hierarchical levels assumed to be present in the population structure: individuals, populations, and groups (Weir & Cockerham, 1984). F_{ct} index describes the variance explained by groups of populations, F_{sc} describes the variance among populations (within groups) and F_{st} index describes the variance within groups among all samples.

A locus-by-locus AMOVA (separate AMOVA performed for each locus), was implemented on our data using Arlequin (Excoffier&Lischer, 2010) as it is favored in the case of missing data. The number of permutations was set to 99,999 and the level of missing data/SNP to 1. In AMOVA, the genetic structure (groups) must be determined before the analysis by the user. In this study, four different clustering scenarios were tested, based mainly on geographical criteria.

In the first scenario, populations were divided into two groups: Eastern Atlantic (1-3)– Mediterranean (4-9); in the second, into three groups Gran Canarias (1)-North Tenerife, Tenerife (2,3) –Mediterranean (4-9); in the third case into four: Gran Canarias(1) – North Tenerife, Tenerife (2,3) –Lampedusa, Astakos, Chalkidiki, North Crete, (pop. 4-7)–Cyprus, Mersin (pop. 8-9); and, in the fourth scenario the possible groups tested were Gran Canaria (1)-Tenerife (2,3) , Lampedusa, Astakos (4,5), Chalkidiki, North Crete, Cyprus, Mersin (6-9).

3.3.2 Population structure estimation using Bayesian Analysis

Another method to infer population structure by determining a number of clusters (groups) observed without prior knowledge, is using STRUCTURE(Pritchard et al., 2000). STRUCTURE (Pritchard et al., 2000) analyzes differences in the distribution of genetic variants amongst populations with a Bayesian iterative algorithm by placing samples into groups whose members share similar patterns of variation. In this Bayesian approach, Markov Chain Monte Carlo (MCMC) estimation is applied to the dataset.

STRUCTURE (Pritchard et al., 2000) applies a model to the data of K assumed genetic groups (clusters), each characterized by a subset of allele frequencies identified in the data. For each number of predetermined K groups, the program calculates a posterior probability Pr(X/K) (Porras-Hurtado et al., 2013). The number K with the maximum likelihood value is assumed to be correlated with the existing genetic structure of the populations (Tsaparis, 2011).

In our data, the admixture model was used, according to which, individuals have mixed ancestry inherited by more than one K groups. This membership ratio of different ancestries in each individual's genome is calculated by the membership coefficient Q. The final Q plot represents each individual in the data set by a single vertical line, partitioned into K colored segments that show the individual's estimated membership fraction in each of the K inferred clusters.

The Bayesian analysis was implemented in Structure_threader (Pina-Martins et al., 2017) on the *.structure* output file by *populations*, with the following settings: –K = 10 and -R = 10 indicate that ten replicate runs will be performed testing the posterior probability for K=1-10 genetic groups, with a burning length of 100,000 and 500,000 MCMC iterations. The clustering algorithm used did not consider the geographic origin of the individuals. Additional parameters are set in the *extraparams* (concerning the model of the analysis) and *mainparams* (concerning the information provided in the genotyping data) files. The best K is chosen according to the Evanno et al., 2005 method, based on the ad hoc statistic ΔK , as besides the posterior probability of K, it also takes into account the distribution of it between successive values of K.

STRUCTURE HARVESTER (Earl & vonHoldt, 2011) was used for the implementation of the Evanno method and a summary of the Structure results (Evanno et al.,2005). The graphical representation of Q-plots for the chosen K value, was produced by CLUMPAK (Kopelman et al., 2015).

3.3.3 Ancestry Analysis

A similar clustering approach to STRUCTURE (Pritchard et al., 2000) was tested on the data with ADMIXTURE (Alexander et al., 2009). ADMIXTURE (Alexander et al., 2009) estimates ancestry of a given set of unrelated individuals from autosomal SNP genotype data sets and determines the nest value of K using a cross-validation procedure.

The program outputs the proportion of ancestral population for each individual. To run the program, a prior belief number of ancestral populations (K) must be provided (Kao et al., 2015).

We used a plink file as the primary input (.bed) file, with the associated .bim (binary marker information) file and .bam (pedigree stub)files and run a fivefold cross-validation procedure for K=1-5, as according to our previous analyses, the number of groups fluctuates between lower K values. The optimal K values are selected according to the cross-validation error that is also reported in the output. The final Q-plot was generated in R, from the ADMIXTURE output .Q file for the best value for K.

3.3.4 Multivariate Analysis

The last method used to infer the genetic structure of our data is DAPC – Discriminant Analysis of Principal Components (Jombart et al., 2010). Unlike STRUCTURE-like approaches, DAPC does not assume that markers are linked and populations are panmictic. In this multivariate statistical approach, variance in the sample is partitioned into a between-group and within-group component, in an effort to maximize discrimination between groups.

In DAPC, a principal component analysis (PCA) is firstly applied on the data, with the selection of a number of principal components (PCs) that best describe it. Then, a k-means clustering algorithm is applied that runs sequentially with increasing values of K and different clustering solutions are compared using Bayesian Information Criterion (BIC). A Discriminant Analysis is performed on the retained PCs and a number of Discriminant Analysis eigenvalues (DAs) can be selected. The final result of DAPC is a scatter plot visualizing the clusters of the dataset (Jombart & Collins, 2015).

Here, DAPC was implemented by *adegenet 2.0.0* (Jombart& Collins, 2015) and *ade4* (Dray & Dufour, 2007) R packages.

3.4 Outlier loci

3.4.1 Detection of Outlier loci

Outlier loci are markers which point to genomic locations that show behavior or patterns of variation that are very divergent from the rest of the genome (Wang et al., 2005). Outliers can be involved in local adaptation, a term referring to the process and resulting patterns of populations evolving traits after interaction with their local habitat (Feng et al., 2015).

In order to detect potential outlier loci in a reliable way, two methods were used. The first one is an F_{ST}-based method proposed by Excoffier et al., 2009 and was applied via the Arlequin program (Excoffier&Lischer, 2010). According to this approach, the outliers are identified as those being present in the tails of the

generated distribution of the F_{ST} values across loci, as a function of heterozygosity between populations, by carrying out simulations under the hierarchical Island model (Slatkin & Voelm, 1991).

The number of simulations was set to 100,000, for 100 simulated demes in 10 simulated groups, with the allowed level of missing data/SNP set to 1. The genetic structure was defined *a priori* for K=3, and loci with a p-value≤0,001 were considered outliers.

Pcadapt (Luu et al., 2017) is the second approach used to identify candidate outliers in the ddRAD dataset, based on Principal Component Analysis (PCA). First, PCA is performed on the genotype matrix (.bed file) and then the Mahalanobis distance is computed for a given SNP, based on the z-scores obtained when regressing SNPs with the K principal components. The outliers were selected with an adjusted p-value = 0,01 using the Benjamini-Hochberg correction.

3.4.2 Identification of closest genomic features to the outlier loci

The functional annotation of the outlier loci on the reference genome, was implemented with the aim of exploring the biological identity of adjacent genes potentially under selection (as they are linked with outliers), using the function closestBed under the BEDTools (Quinlan & Hall, 2010)software.

Firstly, in both software used in outlier detection, different ID numbers are assigned to the loci, therefore, custom scripts were used to match the outliers to their Stacks IDs and then find their genomic intervals.

The closestBed function (Quinlan & Hall, 2010)) searches for overlapping features in A and B. In the event that no feature in B overlaps the current feature in A, closest will report the nearest (that is, least genomic distance from the start or end of A) feature in B (Figure 6).

As the A genomic feature a *.bed* file containing information of one outlier per line was created, consisting of three columns that represent the name of the scaffold, the starting and ending position of the locus respectively. As a B feature a .bed file of the GFF file (from the FTP directory of the NCBI GCF_002260705.1_Sdu_1.0 genome assembly). Additionally, the parameter –d was used, that reports the distance between A and B features (Figure 6).
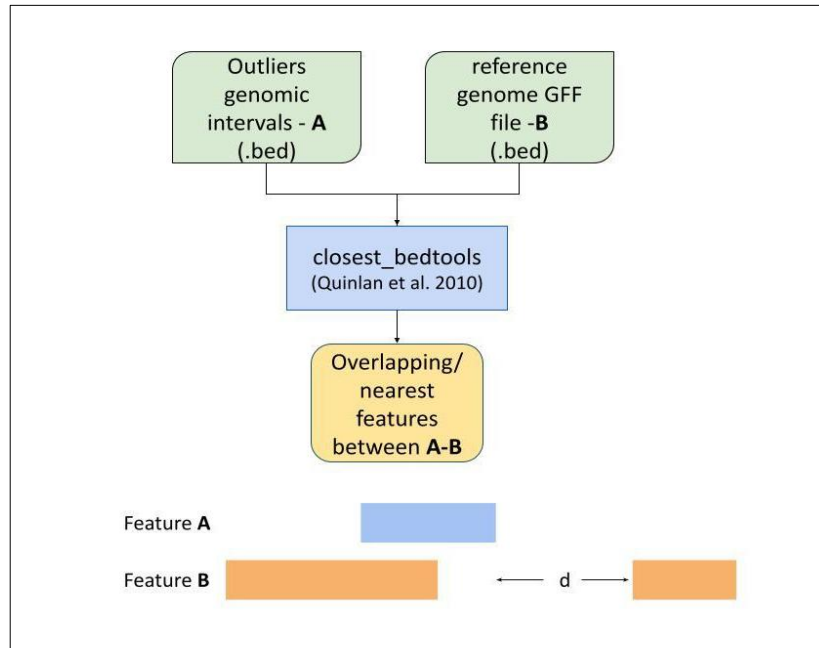
Figure 6 : Exploration of the closest genomic features to the outlier loci. The software finds overlaps between feature A, here genomic intervals of the outlier loci, and feature B, here the mapped genomic features of the reference genome (Quinlan & Hall, 2010)

IV. Results

4.1 SNP calling

Sequencing yielded 665,877,326 raw reads (332,938,663 per FASTQ file) which passed the initial quality control filtering. After *process_radtags*, approximately 4% of the reads were discarded; 550,230 due to low quality (Phred score < 10), 24,855,776 and 2,049,539 for barcode or RAD cut site not found, respectively. The rest of the reds were demultiplexed successfully, by being assigned in one of the 254 individuals of the dataset (Table 1, Appendix).

After alignment against the genome, the percentage of properly paired reads (reads that have the correct orientation towards each other and expected distance in between) was very high for 251 Individuals in the analysis (86% or higher) whereas samples (ITALY_21 (49.75 %), ITALY_31 (43,94 %) and FCPCT_02 43,7 %) presented a low alignment rate and were therefore excluded from the dataset in downstream analysis (Table 1, Appendix).

The *gstacks* program produced a catalog consisting of 11,706 loci with a mean insert length of 402,4 bp. The effective mean coverage per sample in gstacks.log.distribs file, shows the number of times the RAD loci identified in a sample are covered by the reads. The majority of the individuals (~98%) project 20x or higher coverage ((Table 1, Appendix)).

The final dataset consists of 197 individuals that have a maximum 20% of missing data per individual and are genotyped for 1,051 loci with a single nucleotide polymorphism each, by the population program. This final result of polymorphic sites is similar to the number of loci/fragments (2,531) produced by the *in silico* digestion method. The difference can be attributed to the lack of loci filtering in the *in silico* method, as fragments are chosen based on the cut sites and fragment length. The 1,051 SNPs are used in the following analyses.

4.2 Measures of genetic differentiation

4.2.1 Pairwise F_{ST}s

From the matrix of the genetic distances (pairwise F_{ST}s), population 1 (Gran Canaria) appears to be relatively distant from the Mediterranean populations (populations 4-9), especially population 7 (Crete), as well as the Eastern Atlantic populations 2 (North Tenerife) and 3 (Tenerife) despite them being geographically close. Populations 2 and 3 appear to be divergent from populations 4 to 9, but less than population 1. Lastly, the genetic distances between the Mediterranean populations are small and close and thus a homogenous color appears in the F_{ST} matrix. Statistically significant F_{ST} values (p < 0.05) are presented with an asterisk (*) in the pairwise F_{ST} matrix (Figure 7, Appendix Table 2).
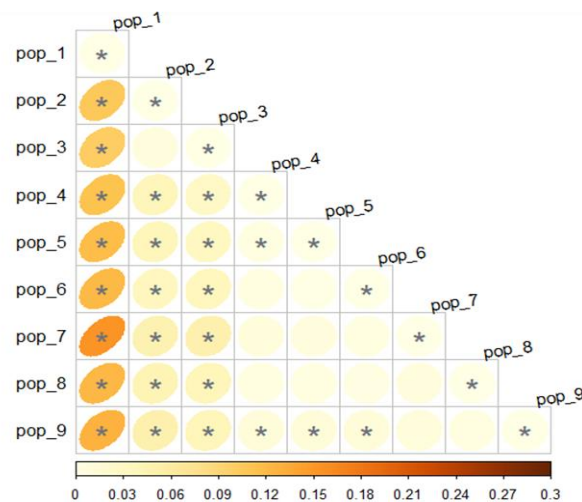
Figure 7 : Heatmap of pairwise F_{ST} matrix with significance level of 0.05 and 10000 permutations. The scale of F_{ST} values is represented with color intensity and statistically significant F_{st}s values (p < 0.05) are displayed with an asterisk (*). Populations : 1 – Gran Canaria, 2 – North Tenerife ,3 – Tenerife, 4 – Lampedusa, 5 – Astakos, 6 – Chalkidiki, 7 – Crete, 8 – Cyprus, 9 – Mersin

4.2.2 Inbreeding coefficient (F_{IS})

Measures of inbreeding showed an excess of homozygotes in every population, since the analysis yielded positive values, from 0.1531 to 0.3067. These results indicate that all populations apart from North Crete (7) significantly deviate from the Hardy-Weinberg equilibrium and suggest non-random mating among individuals especially in North Tenerife, Tenerife and Lampedusa (2,3,4) populations, with the higher F_{IS} values (Table 2). The F_{IS} values encountered are relatively much higher than the ones observed in Mediterranean and Atlantic greater amberjack populations in a previous study (from −0.210 to 0.041 in Šegvić-Bubić et al., 2016).

Table 2 : inbreeding coeffient (F_{IS}- Weir and Cockerham, 1984) estimates. Hardy-Weinberg exact significance tests were conducted for each population marked with an asterisk (*) for p-value <0,01, thus all populations apart from North Crete (7) deviate from HWE

Pop ID	Location	F_{IS} (WC)
1	Gran Canaria	0.1531*
2	North Tenerife	0.3017
3	Tenerife	0.3067*
4	Lampedusa	0.2831*
5	Astakos	0.2771*
6	Chalkidiki	0.1930*
7	North Crete	0.0862
8	Cyprus	0.2765*
9	Mersin	0.2467*

4.3 Cluster Analysis

4.3.1 Analysis of Molecular Variance

Locus-by-locus AMOVA was implemented in four clustering scenarios as presented in Table 3. The tested scenarios were based on the geographical barrier between the two seas (Mediterranean – Atlantic division) and the previous pairwise FST analysis that showed that population 1 is likely to be separate from populations 2 and 3 despite their limited distance.

Table 3 : Population structure cases tested in locus-by-locus AMOVA. Four different scenarios of population structure were tested based on geographical criteria and pairwise F_{ST} results

Groups	Case 1	Case 2	Case 3	Case 4
1	Gran Canaria, North Tenerife, Tenerife (pop. 1-3)	Gran Canaria (pop 1)	Gran Canaria (pop 1)	Gran Canaria (pop 1)
2	Lampedusa, Astakos, Chalkidiki, North Crete, North Crete, Cyprus, Mersin (pop. 4-9)	North Tenerife, Tenerife (pop. 2,3)	North Tenerife, Tenerife (pop. 2,3)	North Tenerife, Tenerife (pop. 2,3)
3		Lampedusa, Astakos, Chalkidiki, North Crete, North Crete, Cyprus, Mersin (pop. 4-9)	Lampedusa, Astakos, Chalkidiki, North Crete, North Crete (pop. 4-7)	Lampedusa, Astakos, (pop. 4,5)
4			Cyprus, Mersin (pop. 8-9)	Chalkidiki, North Crete, North Crete, Cyprus, Mersin (pop. 6-9)

The source of variation (groups, populations or individuals) from each AMOVA analysis is presented in Table 4. The genetic variation among groups (F_{CT}) is maximum in the second scenario. That means that among the tested cases, the structure "Gran Canaria (1) - North Tenerife, Tenerife (2,3) -Chalkidiki, North Crete, North Crete, Cyprus, Mersin (4-9)" explains best the existing variation in our data.

Table 4 : F-statistics of locus-by-locus AMOVA in different population structure scenarios. F_{CT}, F_{SC} and F_{ST} are genetic structure indices, used to a explain the source of the variations. Statistically significant values (p<0.05) are represented with an asterisk (*). The number of permutations was 99999.

Variation (%)	Case 1	Case 2	Case 3	Case 4
Among groups	4.04910	6.55765	4.18665	3.77080
Among populations within groups	2.06998	0.77985	0.71684	0.82581
Within populations	93.88091	92.66250	95.09651	95.40339

Fixation indices				
F_{CT}	0.04049*	0.06558*	0.04187*	0.03771*
F_{SC}	0.02157*	0.00835	0.00748	0.00858*
F_{ST}	0.06119*	0.07337*	0.04903*	0.04597*

4.3.2 Population structure estimation using Bayesian Analysis

STRUCTURE was implemented on three datasets: the total number of loci, candidate outlier loci, and neutral loci (total loci minus outliers). The number of clusters (K) for the studied populations was determined with the use of the Evanno method (Evanno et al., 2005) for the highest value of Delta K. For the total number of 1,051 loci, STRUCTURE revealed evidence for two clusters (K=2) (Appendix, Figure 1). The Gran Canaria population is the only one that belongs into the first genetic cluster, and all the Mediterranean populations (4-9) form a homogeneous group. Populations 2 and 3 (North Tenerife and Tenerife) show a high level of admixture of the existing clusters, but seem more closely related to the Mediterranean group (Figure 8b). The same pattern was shown by the neutral loci (Figure 8c). The outliers showed a similar distribution (Figure 8a), with many individuals in populations 2, 3 and 4 having a highly admixed profile of the two clusters.
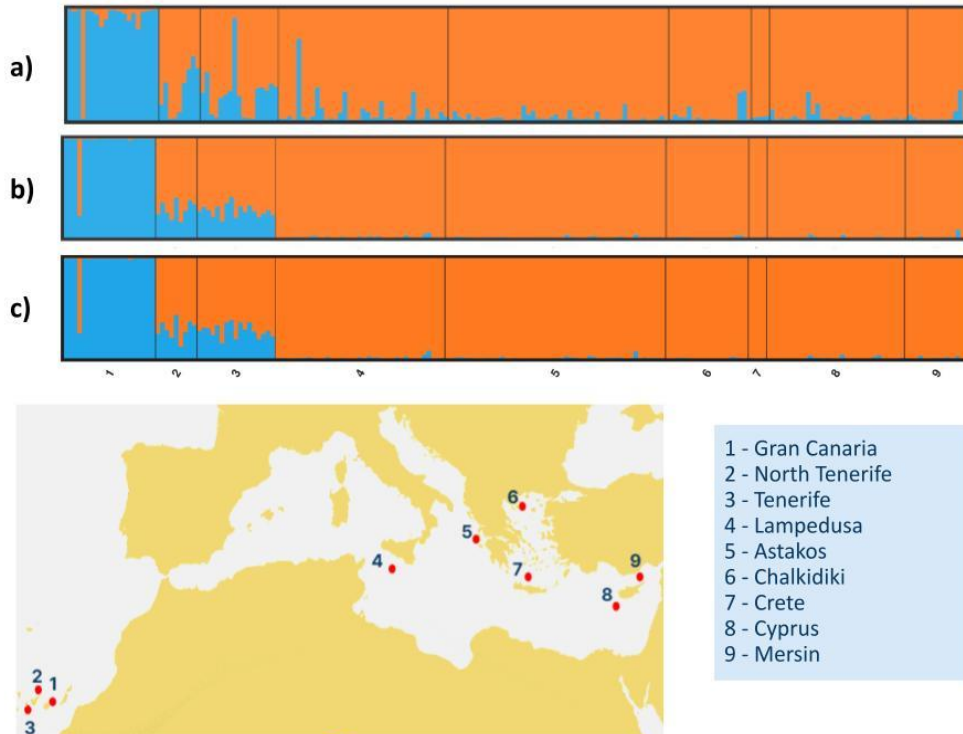
Figure 8 : STRUCTURE Q-plots for a) outlier loci b) total loci c) neutral loci, along with origin locations of the specimen . Membership coefficient plots (Q- plots) show strong evidence for two genetic clusters in the data in a, b and c structure analysis for outliers, total number of loci and neutral loci respectively. Populations 2 and 3 have a highly admixed profile.

4.3.3 Ancestry Analysis

The ADMIXTURE analysis for all loci (Alexander et al., 2009) provided an almost identical output with Structure, suggesting the same clustering model (Figure 9). The most fit number of clusters is 2, as the cross-validation error is the minimum for K=2 (Appendix, Figure 2)
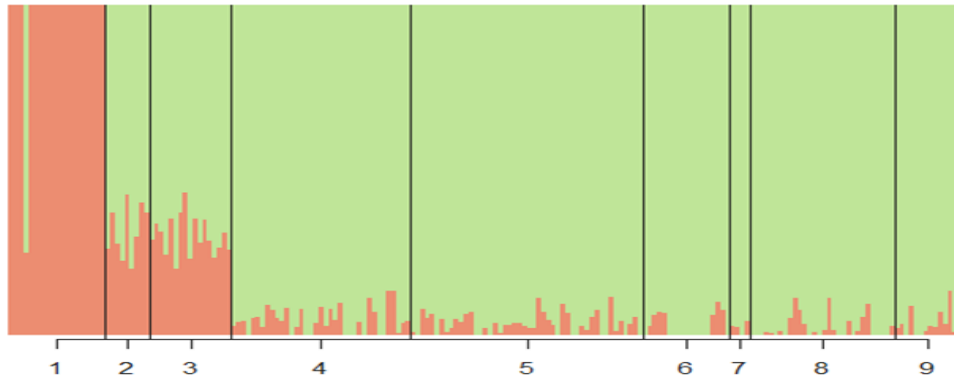
Figure 9 : Ancestry analysis Q-plot for all loci. Membership coefficient plot (Q-plot) produced by ADMIXTURE (Alexander et al., 2009), provides additional evidence on the existence of 2 genetic groups in the data. The clustering scenario of the populations is the same produce by STRUCTURE

4.3.4 Multivariate Analysis

In the multivariate analysis (DAPC), the data was optimally explained by three clusters for 20 PCA and 2 DA eigenvalues, respectively. Tenerife populations 2 and 3 form a single group, which is closer to the Mediterranean group than to Gran Canaria population (Figure 10 a,b)
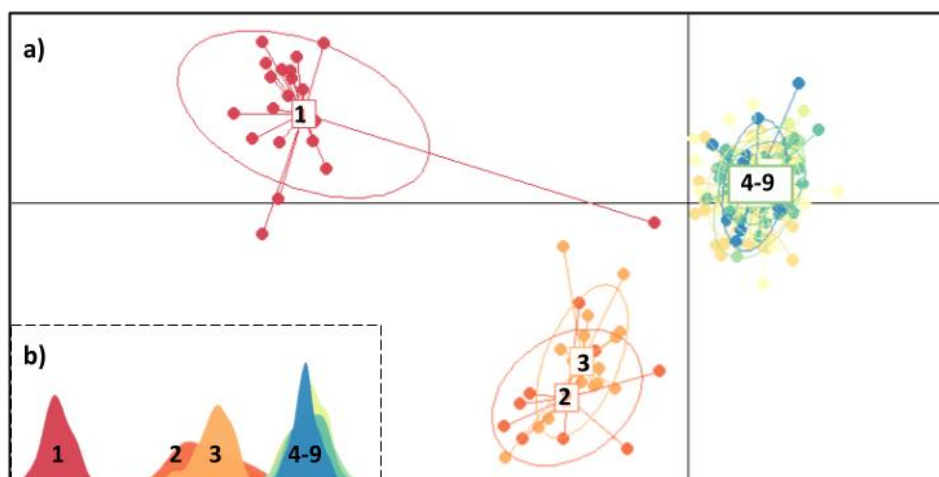
Figure 10 : DAPC analysis: a) The genetic structure of the data is explained by three clusters, for 20 Principal Components and 2 Discriminant Analysis components b) The three clusters shown as peaks in one dimension

4.4 Outlier loci

4.4.1 Detection of outlier loci

The number of potential outlier loci identified from Arlequin (Excoffier&Lischer, 2010) and pcadapt (Luu et al., 2017) were 30 and 32, respectively. Fifteen were shared by both methods and were used in the downstream analysis such as Bayesian cluster analysis and exploration of the biological identity of their closest genes on the reference genome (Figure 11).
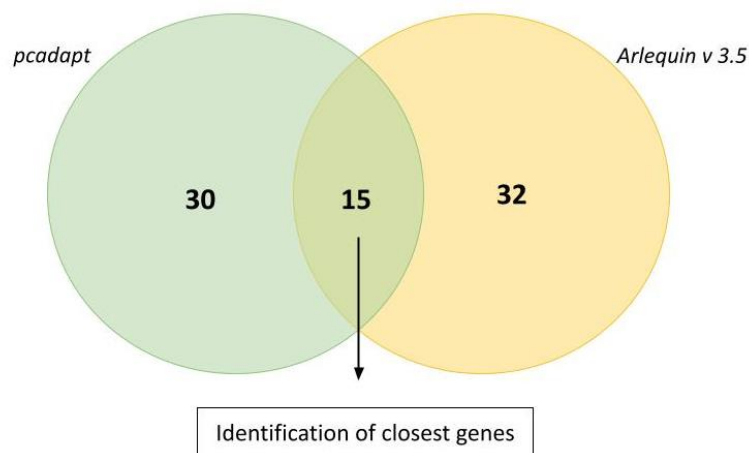
Figure 11 : Venn diagram representing the potential outliers that occurred by two approaches and the overlapping sites. Thirty-two loci were identified by pcadapt (Luu et al., 2017) and thirty by Arlequin (Excoffier&Lischer, 2010). Fifteen loci in total were identified by both detection methods

4.4.2 Understanding the functional background of greater amberjack's adaptation loci

Searching genomic regions adjacent to the outlier loci, can lead to identification of candidate genes for environmental adaptation. The results of closestBedtools (Quinlan & Hall, 2010) showed that 6 out of 15 outliers are located within gene regions (Figure 12). The closest genes identified are associated with immune response, peptide hormone activity, cytoskeleton organization, neurogenic and skeletal development, protein transport – metabolism, pH homeostasis. Lastly, tef-1 gene (transcription enhancer factor-1) potentially related with local adaptation was detected. In medaka (*Oryzias latipes*) muscle tissue this gene produces two splicing variants, TEF-1A and TEF-1B mRNAs. During cold acclimation, the mRNA level of TEF-1A decreased, whereas that of TEF-1B increased. The same study also showed that three putative downstream genes of tef-1 are also transcribed in a temperature-dependent pattern (Yamasaki et al., 2006) (Figure 12).

Immune response
- *nox1* (NADPH oxidase 1) *
- *nclr3* (NOD-like receptor family CARD domain containing 3)

Peptide hormone activity
- *ecel1* (Endothelin Converting Enzyme Like 1) *

Cytoskeleton organization
- *arhgap10* (Rho GTPase Activating Protein 10) *

pH homeostasis
- *gpr4* (G-protein coupled receptor 4-like) *

Temperature acclimation (*Oryzias latipes*)
(Yamasaki et al., 2006)
- *tef-1* (transcription enhancer factor-1) *

Neurogenic development
- *grm8* (metabotropic glutamate receptor 8) *
- *nrxn2* (Neurexin 2) *
- *pdgfa* (Platelet Derived Growth Factor Subunit A)
- *nfia* (Nuclear Factor I)

Skeletal development
- *fras1* Fraser extracellular matrix protein 2-like *
- *col27a1* (Collagen Type XXVII Alpha 1 Chain) *

Protein metabolism/transport
- *wac* (WW Domain Containing Adaptor With Coiled-Coil)
- *etfbkmt* (Electron Transfer Flavoprotein Subunit Beta Lysine Methyltransferase)
- *sec23ip* (sec23 Interacting Protein) *

Figure 12 : **Closest genes to the outlier loci organized based on their biological function.** These genes seem to participate in development, cell cycle and immunity. Genes that are potentially associated with local adaptation are marked with yellow and outliers located inside gene regions are displayed with an asterisk.

[28]

V.Discussion

We found two separate genetic clusters of this highly migratory pelagic teleost species in the Mediterranean and Eastern Atlantic, with the possibility of a hybrid zone between them and compared our results with previous population studies.

5.1 Inbreeding within populations

In our analysis very high F_{IS} values were reported for every population , apart from North Crete (population 7), were the F_{IS} value was not statistically significant, compared to the ones by (Šegvić-Bubić et al., 2016).This could be due to the different type and number of markers used as there are reports of microsatellite and SNP markers leading to different conclusions concerning genetic diversity and population structure (Defaveri et al., 2013). Also, the sampling of the individuals, sequencing errors and missing data especially in SNP markers, could really affect whether different types of markers are representative of population variability. High rate of homozygotes could suggest that individuals are reproduced in a non-random way and consequently, further analyses on inbreeding rate and relations between individuals need to be conducted in order to enrich our knowledge on the genetic variability of the wild populations and ensure that these individuals have a better performance as potential breeders in aquaculture.

5.2 Structure Analysis in Mediterranean and Eastern Atlantic

Our results yielded two possible clustering scenarios for the species in the area of study (Figure 13). The two-cluster scenario (K=2) is supported by the ancestry and Bayesian analysis (Figure 13a). According to this hypothesis, the Atlantic populations of Tenerife (2,3) are clustered with the Mediterranean group despite their distant geographic origin. The three-cluster scenario (K=3) on the other hand, supported by multivariate analysis and analysis of molecular variance (Figure 13b), suggests that Atlantic populations are divided into two groups. In addition, pairwise F_{ST} values (Figure 7) provide further evidence that the two populations from Tenerife are genetically differentiated from the Mediterranean populations, considerably less, however, than the Gran Canaria one (population 1).
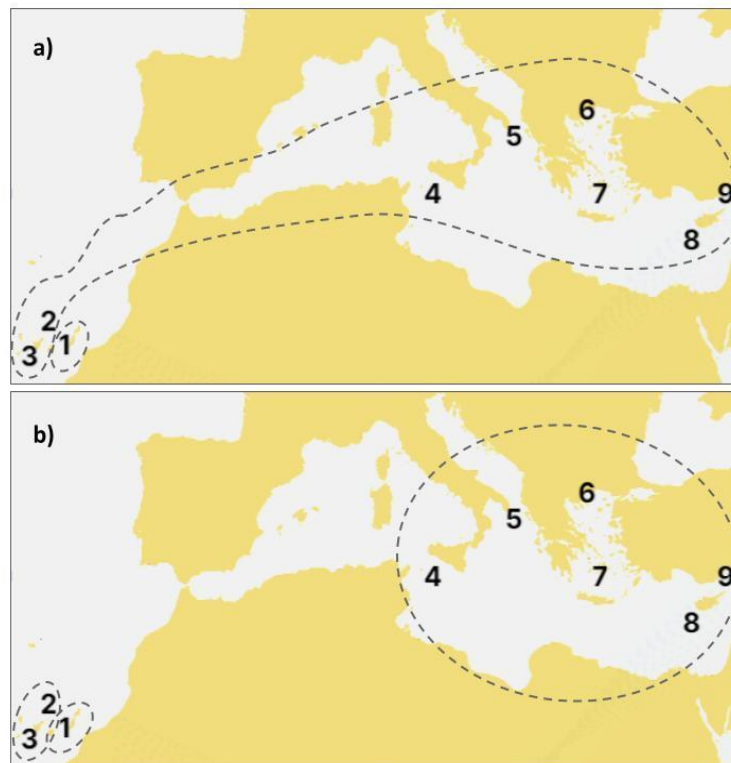
Figure 13 : Population structure scenarios from different clustering approaches a) Two-cluster scenario (K=2) suggested by Ancestry and Bayesian analysis b) Three-cluster scenario (K=3) suggested by multivariate analysis and analysis of molecular variance

A clear separation between geographically adjacent populations could potentially be explained by the existence of a phylogeographic break between them. In the Atlantic-Mediterranean area, several oceanographic discontinuities, mostly on the Spanish coast, have been reported to act as potential barriers of gene flow in marine organisms, like the Balearic front (BF) and the Ibiza Channel (Galarza et al., 2009; García-Merchán et al., 2012), the Sicily and Sardinian channel (Serra et al., 2010; Zitari-Chatti et al., 2009). The most well-studied oceanographic barrier is the Almeria-Oran front with strong evidence by studies in several species (1.3 Introduction, Schunter et al., 2011). However, there has not been any reports of genetic breaks in Macronesia waters (African Atlantic coasts) and consequently, the existence of a front that could explain the limited gene flow between the Atlantic populations (Gran Canaria-Tenerife populations) is not likely.

This is also strongly supported by the recent population genetic analyses conducted in the greater amberjack in the Atlantic-Mediterranean area with microsatellite markers (Šegvić-Bubić et al., 2016, 2022), where two genetically differentiated clusters occurred, a Mediterranean and an Atlantic one, with no sign of a third or intermediate group as indicated in our current results. Despite the clear genetic structure that occurred in their study, the authors implemented a barrier analysis that revealed genetic discontinuities within the Mediterranean basin, located at the Almeria-Oran and Siculo-Tunisian areas (Šegvić-Bubić et al., 2016).

In a previous study conducted on the present dataset, microsatellite and mtDNA

markers were investigated (Kolios 2017, personal communication), while integrating findings from Šegvić-Bubić et al., 2016. Two co-existing phylogroups (phylogroup A, B) were detected for the greater amberjack in the Eastern Atlantic and Mediterranean with the use of mtDNA markers. Despite the existence of private haplotypes (found in the sample from one particular population, but absent in the samples from other populations), approximately 60% similarity of haplotype composition between the two basins was found. This pattern of two clearly separated lineages, being similarly distributed in the Mediterranean, without a detectable geographical partitioning, has been observed again in swordfish and suggests the hypothesis of a simultaneous colonization of the two phylogroups in the Mediterranean basin that had already been differentiated elsewhere(Alvarado Bremer et al., 2005; Šegvić-Bubić et al., 2016). This is additionally supported by the paleoceanographic events that have shaped the Mediterranean basin and its biodiversity. In the late Miocene, the Messinian Salinity Crisis (MSC) caused a massive extinction of the Mediterranean marine ichthyofauna which was later restored, by the re-opening of the Strait of Gibraltar, allowing a rapid migration of Atlantic species in the Mediterranean basin, possibly with greater amberjack being among them (Domingues et al., 2005).

The nuclear markers (eleven microsatellite loci) yielded results that align with the ones of the present study (1051 SNPs), in terms of population clustering. Microsatellite analysis indicated that Gran Canaria population was genetically distant from the rest of the populations (North Tenerife, Tenerife and Mediterranean ones) whereas, Tenerife populations (2,3) displayed an intermediate zone, between the Mediterranean populations (4-9) and Gran Canaria (1) (Kolios 2017, personal communication).

However, considering the type of genetic markers used is also an important aspect in comparing population studies. The use of SNP loci over microsatellites has proved to allow a more precise estimation of population-level diversity, higher power to identify groups in clustering methods, and the ability to consider local adaptation (Zimmerman et al., 2020). Taking the above into account, and since the sampling of the current study did not include individuals from the Western Mediterranean, we cannot exclude with certainty the existence of a second stock in the Mediterranean.

Consequently, one possible scenario is that the two Tenerife populations (2 and 3) were captured on a migratory route and they belong to the assumed Western Mediterranean group that is characterized by a high rate of gene flow with the Mediterranean populations and to a relatively lower degree with the Atlantic ones, explaining Bayesian and Ancestry analyses' hybrid zone. A hybrid zone, with intermediate allelic frequencies, has been reported again for Mediterranean rainbow wrasse (*Coris julis*) in the Alboran Sea (Fruciano et al., 2011). In this light, another plausible structural model that could decipher our data is if individuals from Gran Canaria (1) originate from the Central Atlantic and were consequently caught around Gran Canaria on a dispersal route since they appear genetically distant from the rest of the samples.

The above hypotheses are supported by the highly migratory nature of this marine teleost that has a cosmopolitan distribution, despite the fact that its life cycle movements are still seriously understudied (Hasegawa et al., 2020). Finally, our study confirms the consistent and homogeneous pattern of genetic connectivity in

the Central and Eastern Mediterranean that are connected to the species' biological features, such as extended pelagic larval phase with passive dispersal patterns associated with seaweed and multiple spawning (Part 1.3 Introduction).

In conclusion, extended studies centered on the ecology and life history traits of the greater amberjack are required, as they would lead to a better understanding of its distribution patterns. Very importantly, a thorough sampling of wild individuals along the Mediterranean basin, with special attention to locations that have been considered as potential barriers to gene flow is vital for the elucidation of the species' structure and the identification of its transition area between the two basins.

5.3 Neutral and outlier markers

An increasing number of population genomic studies searches for the potential genetic basis of local adaptation, the evolutionary process that leads to higher fitness of a population to its local environment (Milano et al., 2014; Rodríguez-Ezpeleta et al., 2016; Sherman et al., 2020; Zhao et al., 2018; Zimmerman et al., 2020). Considering directional selection while inferring population structure, local adaptation is very informative in terms of recently diverged stocks and has potential applications for designating management units in fisheries and aquaculture (Russello et al., 2012).

Separate investigation of outlier markers has unveiled fine-scale patterns of genetic structure which could not be detected with the use of neutral markers, in European hake's Atlantic and Mediterranean populations (Milano et al., 2014). However, comparing our Bayesian STRUCTURE analysis results for the outlier, neutral and all loci, it did not reveal any different genetic profile of the populations (Figure 8). The outliers' Q-plot displays an introgression of the Atlantic genetic component into the Mediterranean one, which is more gradual compared to the "sharp" genetic composition, which is identical, in neutral and total loci Q-plots. This might suggest ongoing or recent gene flow between the basins but, other than that no cryptic stock was detected.

5.4 Markers potentially involved in local adaptation

Investigating the footprints of local adaptation can provide a clear insight into the patterns that lead to the differentiation of a population, and the traits that promote its survival. However environmental pressure can disrupt the fitness that an organism has developed to its habitat.

In marine ecosystems sea temperature increases as a consequence of global climate change and can lead to radical changes, such as invasion of non-native species that could completely alter the habitat and thus, the distribution of the endemic organisms (Gentilucci et al., 2021). The Mediterranean Sea was possibly the first of the world seas where changes related to temperature were attributed to climate change (Bethoux et al., 1990). As a consequence, identification of potential genetic markers like *tef-1* that are expressed in a temperature-dependent pattern,

could give insight for this shift in marine biodiversity composition and species distribution in the Mediterranean basin, as a result of fluctuation of environmental variables. Exploring markers that could contribute in monitoring climate crisis and protection of biodiversity will definitely be a priority for researchers in the future.

VI. Concluding remarks

This is the first project to date, that uses SNP markers from a reduced-complexity sequencing approach (ddRAD sequencing) to study the population structure of the greater amberjack in the Mediterranean and Eastern Atlantic, while exploring the performance of neutral/outlier loci to elucidate the patterns of genetic connectivity among wild populations. This study confirms the presence of two separate stocks in the Mediterranean and the Atlantic, by taking into consideration results from previous analyses on the species. However, further research is needed to determine the phylogeographic fronts of the greater amberjack in the Western Mediterranean and the possibility of an intermediate group with a ''shallow'' structure. Lastly, a more thorough investigation of local adaptation signatures is required, as they could be connected with important environmental parameters that can potentially provide a more detailed insight to the species present and future distribution.

Bibliography

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

Andaloro, F., Potoschi, A., 1992. Contribution to the knowledge of growth of greateramberjack, Seriola dumerili Cuv. (1817) in the Sicilian Channel (MediterraneanSea). Rapport Comm. Int. Explor. Mer Méditerr. 33, 282.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Alvarado Bremer, J. R., Viñas, J., Mejuto, J., Ely, B., & Pla, C. (2005). Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular Phylogenetics and Evolution*, *36*(1), 169–187. https://doi.org/10.1016/J.YMPEV.2004.12.011

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

Aoki, J. Y., Kai, W., Kawabata, Y., Ozaki, A., Yoshida, K., Koyama, T., Sakamoto, T., & Araki, K. (2015). Second generation physical and linkage maps of yellowtail (Seriola quinqueradiata) and comparison of synteny with four model fish. *BMC Genomics*, *16*(1), 1–11. https://doi.org/10.1186/S12864-015-1600-7/FIGURES/4

Araki, H., & Schmid, C. (2010). Is hatchery stocking a help or harm?: Evidence, limitations and future directions in ecological and genetic surveys. *Aquaculture*, *308*(SUPPL.1), S2–S11. https://doi.org/10.1016/J.AQUACULTURE.2010.05.036

Araki, K., Aokic, J. Y., Kawase, J., Hamada, K., Ozaki, A., Fujimoto, H., Yamamoto, I., & Usuki, H. (2018). Whole Genome Sequencing of Greater Amberjack (Seriola dumerili) for SNP Identification on Aligned Scaffolds and Genome Structural Variation Analysis Using Parallel Resequencing. *International Journal of Genomics*, *2018*. https://doi.org/10.1155/2018/7984292

Araneda, C., Larraín, M. A., Hecht, B., & Narum, S. (2016). Adaptive genetic variation distinguishes Chilean blue mussels (Mytilus chilensis) from different marine environments. *Ecology and Evolution*, *6*(11), 3632–3644. https://doi.org/10.1002/ECE3.2110

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*(10), 1–7. https://doi.org/10.1371/journal.pone.0003376

Bargelloni, L., Alarcon, J. A., Alvarez, M. C., Penzo, E., Magoulas, A., Reis, C., & Patarnello, T. (2003). Discord in the family Sparidae (Teleostei): Divergent phylogeographical patterns across the Atlantic-Mediterranean divide. *Journal of Evolutionary Biology*, *16*(6), 1149–1158. https://doi.org/10.1046/j.1420-9101.2003.00620.x

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). *Molecular Ecology*, *24*(13), 3299–3315. https://doi.org/10.1111/MEC.13245

Bethoux, J. P., Gentili, B., Raunet, J., & Tailliez, D. (1990). Warming trend in the western Mediterranean deep water. *Nature 1990 347:6294*, *347*(6294), 660–662. https://doi.org/10.1038/347660a0

Bradic, M., Teotónio, H., & Borowsky, R. L. (2013). The population genomics of repeated evolution in the blind cavefish Astyanax mexicanus. *Molecular Biology and Evolution*, *30*(11), 2383–2400. https://doi.org/10.1093/MOLBEV/MST136

Cimmaruta, R., Bondanelli, P., & Nascetti, G. (2005). Genetic structure and environmental heterogeneity in the European hake (Merluccius merluccius). *Molecular Ecology*, *14*(8), 2577–2591. https://doi.org/10.1111/J.1365-294X.2005.02595.X

Corriero, A., Wylie, M. J., Nyuji, M., Zupa, R., & Mylonas, C. C. (2021). Reproduction of greater amberjack (Seriola dumerili) and other members of the family Carangidae. *Reviews in Aquaculture*, *13*(4), 1781–1815. https://doi.org/10.1111/raq.12544

Conover D. O., Clarke L. M., Munch S. B., Wagner G. N. (2006). Spatial and Temporal Scales of Adaptive Divergence in Marine Fishes and the Implications for Conservation. *J. Fish Biol.* 69, 21–

47. doi: 10.1111/j.1095-8649.2006.01274.x

de Oliveira, R. C., Santos, M. da C. F., Bernardino, G., Hrbek, T., & Farias, I. P. (2018). From river to farm: an evaluation of genetic diversity in wild and aquaculture stocks of Brycon amazonicus (Spix & Agassiz, 1829), Characidae, Bryconinae. *Hydrobiologia*, *805*(1), 75–88. https://doi.org/10.1007/S10750-017-3278-0/TABLES/4

Defaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNPs. *Molecular Ecology Resources*, *13*(3), 377–392. https://doi.org/10.1111/1755-0998.12071

Domingues, V. S., Bucciarelli, G., Almada, V. C., & Bernardi, G. (2005). Historical colonization and demography of the Mediterranean damselfish, Chromis chromis. *Molecular Ecology*, *14*(13), 4051–4063. https://doi.org/10.1111/J.1365-294X.2005.02723.X

Dray, S., & Dufour, A. B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/JSS.V022.I04

Earl, D. A., & vonHoldt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources 2011 4:2*, *4*(2), 359–361. https://doi.org/10.1007/S12686-011-9548-7

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. *Methods in Molecular Biology*, *772*, 157–178. https://doi.org/10.1007/978-1-61779-228-1_9/COVER

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/J.1365-294X.2005.02553.X

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity 2009 103:4*, *103*(4), 285–298. https://doi.org/10.1038/hdy.2009.74

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

Feng, X. J., Jiang, G. F., & Fan, Z. (2015). Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, Ceracris kiangsu. *Scientific Reports 2015 5:1*, *5*(1), 1–11. https://doi.org/10.1038/srep13758

Fernández-Palacios, H., Schuchardt, D., Roo, J., Hernández-Cruz, C. M., & Izquierdo, M. (2015). Multiple GnRHa injections to induce successful spawning of wild caught greater amberjack (Seriola dumerili) matured in captivity. *Aquaculture Research*, *46*(7), 1748–1759. https://doi.org/10.1111/ARE.12330

Fruciano, C., Hanel, R., Debes, P. V., Tigano, C., & Ferrito, V. (2011). Atlantic-Mediterranean and within-Mediterranean molecular variation in Coris julis (L. 1758) (Teleostei, Labridae). *Marine Biology*, *158*(6), 1271–1286. https://doi.org/10.1007/s00227-011-1647-1

Galarza, J. A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G. F., & Rico, C. (2009). The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1473–1478. https://doi.org/10.1073/PNAS.0806804106

García-Merchán, V. H., Robainas-Barcia, A., Abelló, P., Macpherson, E., Palero, F., García-Rodríguez, M., Gil de Sola, L., & Pascual, M. (2012). Phylogeographic patterns of decapod crustaceans at the Atlantic–Mediterranean transition. *Molecular Phylogenetics and Evolution*, *62*(2), 664–672. https://doi.org/10.1016/J.YMPEV.2011.11.009

Gentilucci, M., Parisi, C., Coppola, M. R., Majdoubi, F. Z., Madonna, A., & Guerriero, G. (2021). Influence of Mediterranean Sea Temperature Increase on Gaeta Gulf (Tyrrhenian Sea) Biodiversity. *Proceedings of the Zoological Society*, *74*(1), 91–103. https://doi.org/10.1007/S12595-020-00334-6

Gold, J. R., & Richardson, L. R. (1998). Population structure in greater amberjack, Seriola dumerili, from the Gulf of Mexico and the western Atlantic Ocean. *Fishery Bulletin*, *96*(4), 767–778.

Hargrove, J. S., Murie, D. J., Parkyn, D. C., Saarinen, E. V., & Austin, J. D. (2018). Mixing rates in weakly differentiated stocks of greater amberjack (Seriola dumerili) in the Gulf of Mexico. *Genetica*, *146*(4–5), 393–402. https://doi.org/10.1007/s10709-018-0031-1

Harris, P. J., Wyanski, D. M., White, D. B., Mikell, P. P., & Eyo, P. B. (2007). Age, Growth, and Reproduction of Greater Amberjack off the Southeastern U.S. Atlantic Coast. *Transactions of the American Fisheries Society*, *136*(6), 1534–1545. https://doi.org/10.1577/t06-113.1

[35]

Hasegawa, T., Lu, C. P., Hsiao, S. T., Uchino, T., Yeh, H. M., Chiang, W. C., Chen, J. R., Sassa, C., Komeyama, K., Kawabe, R., Sakamoto, T., Masumi, S., Uchida, J., Aoshima, T., & Sakakura, Y. (2020). Distribution and genetic variability of young-of-the-year greater amberjack (Seriola dumerili) in the East China Sea. *Environmental Biology of Fishes*, *103*(7), 833–846. https://doi.org/10.1007/s10641-020-00985-6

Hasegawa, T., Takatsuki, N., Kawabata, Y., Kawabe, R., Nishihara, G. N., Ishimatsu, A., Soyano, K., Okamura, K., Furukawa, S., Yamada, M., Shimoda, M., Kinoshita, T., Yamawaki, N., Morii, Y., & Sakakura, Y. (2017). Continuous behavioral observation reveals the function of drifting seaweeds for Seriola spp. Juveniles. *Marine Ecology Progress Series*, *573*, 101–115. https://doi.org/10.3354/meps12154

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, *6*(2), 1000862. https://doi.org/10.1371/JOURNAL.PGEN.1000862

Hollenbeck, C. M., Portnoy, D. S., & Gold, J. R. (2019). Evolution of population structure in an estuarine-dependent marine fish. *Ecology and Evolution*, *9*(6), 3141–3152. https://doi.org/10.1002/ECE3.4936

Houde, E. D. (2008). Emerging from Hjort's shadow. *Journal of Northwest Atlantic Fishery Science*, *41*, 53–70. https://doi.org/10.2960/J.V41.M634

Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., McAndrew, B. J., & Bishop, S. C. (2008). Major Quantitative Trait Loci Affect Resistance to Infectious Pancreatic Necrosis in Atlantic Salmon (Salmo salar). *Genetics*, *178*(2), 1109. https://doi.org/10.1534/GENETICS.107.082974

Jombart, T., & Collins, C. (2015). *A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0*.

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

Kao, J. Y., Zubair, A., Salomon, M. P., Nuzhdin, S. V., & Campo, D. (2015). Population genomic analysis uncovers African and European admixture in Drosophila melanogaster populations from the south-eastern United States and Caribbean Islands. *Molecular Ecology*, *24*(7), 1499–1509. https://doi.org/10.1111/MEC.13137

Kolios V. (2017) Genetic structufe of *Seriola dumerili* Risso,1810 (Perciformes, Carangidae) in the Mediterranean and Eastern Atlantic (Unpublished master's thesis), HCMR, Heraklion

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Lemaire, C., Versini, J. J., & Bonhomme, F. (2005). Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology*, *18*(1), 70–80. https://doi.org/10.1111/J.1420-9101.2004.00828.X

Lepais, O., & Weir, J. T. (2014). SimRAD: An R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources*, *14*(6), 1314–1321. https://doi.org/10.1111/1755-0998.12273

Létourneau, J., Ferchaud, A. L., Le Luyer, J., Laporte, M., Garant, D., & Bernatchez, L. (2018). Predicting the genetic impact of stocking in Brook Charr (Salvelinus fontinalis) by combining RAD sequencing and modeling of explanatory variables. *Evolutionary Applications*, *11*(5), 577–592. https://doi.org/10.1111/EVA.12566

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Project, G., & Subgroup, D. P. (2009). The Sequence Alignment/Map format and SAMtools. *BIOINFORMATICS APPLICATIONS NOTE*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, S., Liu, K., Cui, A., Hao, X., Wang, B., Wang, H. Y., Jiang, Y., Wang, Q., Feng, B., Xu, Y., Shao, C., & Liu, X. (2022). A Chromosome-Level Genome Assembly of Yellowtail Kingfish (Seriola lalandi). *Frontiers in Genetics*, *12*, 2807. https://doi.org/10.3389/FGENE.2021.825742/BIBTEX

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting

population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/BIOINFORMATICS/BTR642

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

Magoulas, A., Castilho, R., Caetano, S., Marcato, S., & Patarnello, T. (2006). Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (Engraulis encrasicolus). *Molecular Phylogenetics and Evolution*, *39*(3), 734–746. https://doi.org/10.1016/J.YMPEV.2006.01.016

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

Manousaki, T., Tsakogiannis, A., Taggart, J. B., Palaiokostas, C., Tsaparis, D., Lagnel, J., Chatziplis, D., Magoulas, A., Papandroulakis, N., Mylonas, C. C., & Tsigenopoulos, C. S. (2016). Exploring a nonmodel teleost genome through rad sequencing-linkage mapping in common pandora, Pagellus erythrinus and comparative genomic analysis. *G3: Genes, Genomes, Genetics*, *6*(3), 509–519. https://doi.org/10.1534/g3.115.023432

Maroso, F., Gkagkavouzis, K., de Innocentiis, S., Hillen, J., do Prado, F., Karaiskou, N., Taggart, J. B., Carr, A., Nielsen, E., Triantafyllidis, A., Bargelloni, L., Bekkevold, D., Frank-Gopolos, T., Franch, R., Ferraresso, S., Babbucci, M., Greco, C., Crosetti, D., Simionati, B., … Pagelson, G. (2021). Genome-wide analysis clarifies the population genetic structure of wild gilthead sea bream (Sparus aurata). *PLoS ONE*, *16*(1 January), 1–16. https://doi.org/10.1371/journal.pone.0236230

Martinez-Takeshita, N., Purcell, C. M., Chabot, C. L., Craig, M. T., Paterson, C. N., Hyde, J. R., & Allen, L. G. (2015). A Tale of Three Tails: Cryptic Speciation in a Globally Distributed Marine Fish of the Genus Seriola. *Copeia*, *103*(2), 357–368. https://doi.org/10.1643/CI-124-224

Mazzola, A., Favaloro, E., & Sarà, G. (2000). Cultivation of the Mediterranean amberjack, Seriola dumerili (Risso, 1810), in submerged cages in the Western Mediterranean Sea. *Aquaculture*, *181*(3–4), 257–268. https://doi.org/10.1016/S0044-8486(99)00243-4

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A. J., Hansen, J. H., Helyar, S. J., Nielsen, E. E., Ogden, R., Patarnello, T., Stagioni, M., Tinti, F., & Bargelloni, L. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (Merluccius merluccius). *Molecular Ecology*, *23*(1), 118–135. https://doi.org/10.1111/mec.12568

Mylonas, C. C., Papandroulakis, N., Smboukis, A., Papadaki, M., & Divanach, P. (2004). Induction of spawning of cultured greater amberjack (Seriola dumerili) using GnRHa implants. *Aquaculture*, *237*(1–4), 141–154. https://doi.org/10.1016/J.AQUACULTURE.2004.04.015

Nguyen, N. H., Rastas, P. M. A., Premachandra, H. K. A., & Knibb, W. (2018). First high-density linkage map and single nucleotide polymorphisms significantly associated with traits of economic importance in yellowtail kingfish Seriola lalandi. *Frontiers in Genetics*, *9*(APR), 127. https://doi.org/10.3389/FGENE.2018.00127/BIBTEX

Nousias, O., Oikonomou, S., Manousaki, T., Papadogiannis, V., Angelova, N., Tsaparis, D., Tsakogiannis, A., Duncan, N., Estevez, A., Tzokas, K., Pavlidis, M., Chatziplis, D., & Tsigenopoulos, C. S. (2022). Linkage mapping, comparative genome analysis, and QTL detection for growth in a non-model teleost, the meagre Argyrosomus regius, using ddRAD sequencing. *Scientific Reports*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-09289-4

NUGROHO, E., TANIGUCHI, N., KATO, K., & MIYASHITA, S. (2000). Genetic Difference among Seed Populations of Greater Amberjack Used in Aquaculture Farm of Japan. *Aquaculture Science*, *48*(4), 665–674. https://doi.org/10.11233/AQUACULTURESCI1953.48.665

Ottmann, D., Fiksen, Ø., Martín, M., Alemany, F., Prieto, L., Álvarez-Berastegui, D., & Reglero, P. (2021). Spawning site distribution of a bluefin tuna reduces jellyfish predation on early life stages. *Limnology and Oceanography*, *66*(10), 3669–3681. https://doi.org/10.1002/LNO.11908

Ottolenghi, F., & Food and Agriculture Organization of the United Nations. (2004). *Capture-based aquaculture : the fattening of eels, groupers, tunas, and yellowtails*. 308.

Ozaki, A., Yoshida, K., Fuji, K., Kubota, S., Kai, W., Aoki, J. ya, Kawabata, Y., Suzuki, J., Akita, K., Koyama, T., Nakagawa, M., Hotta, T., Tsuzaki, T., Okamoto, N., Araki, K., & Sakamoto, T. (2013). Quantitative Trait Loci (QTL) Associated with Resistance to a Monogenean Parasite (Benedenia seriolae) in Yellowtail (Seriola quinqueradiata) through Genome Wide Analysis. *PLOS ONE*, *8*(6),

e64987. https://doi.org/10.1371/JOURNAL.PONE.0064987

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). https://doi.org/10.1371/journal.pone.0037135

Pina-Martins, F., Silva, D. N., Fino, J., & Paulo, O. S. (2017). Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. *Molecular Ecology Resources*, *17*(6), e268–e274. https://doi.org/10.1111/1755-0998.12702

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4*(MAY), 1–13. https://doi.org/10.3389/fgene.2013.00098

Pousis, C., Mylonas, C. C., De Virgilio, C., Gadaleta, G., Santamaria, N., Passantino, L., Zupa, R., Papadaki, M., Fakriadis, I., Ferreri, R., & Corriero, A. (2018). The observed oogenesis impairment in greater amberjack Seriola dumerili (Risso, 1810) reared in captivity is not related to an insufficient liver transcription or oocyte uptake of vitellogenin. *Aquaculture Research*, *49*(1), 243–252. https://doi.org/10.1111/ARE.13453

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/GENETICS/155.2.945

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

Recknagel, H., Elmer, K. R., & Meyer, A. (2013). A hybrid genetic linkage map of two ecologically and morphologically divergent midas cichlid fishes (amphilophus spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3: Genes, Genomes, Genetics*, *3*(1), 65–74. https://doi.org/10.1534/G3.112.003897/-/DC1/TABLES5.PDF

Reuschel, S., Cuesta, J. A., & Schubart, C. D. (2010). Marine biogeographic boundaries and human introduction along the European coast revealed by phylogeography of the prawn Palaemon elegans. *Molecular Phylogenetics and Evolution*, *55*(3), 765–775. https://doi.org/10.1016/J.YMPEV.2010.03.021

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). ESTIMATION OF THE COANCESTRY COEFFICIENT: BASIS FOR A SHORT-TERM GENETIC DISTANCE. *Genetics*, *105*(3), 767–779. https://doi.org/10.1093/GENETICS/105.3.767

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., & Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, *10*(3), 670–682. https://doi.org/10.1111/raq.12193

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/MEC.15253

Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources*, *16*(4), 991–1001. https://doi.org/10.1111/1755-0998.12518

Rousset, F. (n.d.). *Genepop version 4.7.5*.

Russello, M. A., Kirk, S. L., Frazer, K. K., & Askey, P. J. (2012). Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, *5*(1), 39–52. https://doi.org/10.1111/J.1752-4571.2011.00206.X

Sarropoulou, E., Sundaram, A. Y. M., Kaitetzidou, E., Kotoulas, G., Gilfillan, G. D., Papandroulakis, N., Mylonas, C. C., & Magoulas, A. (2017). Full genome survey and dynamics of gene expression in the greater amberjack Seriola dumerili. *GigaScience*, *6*(12), 1–13. https://doi.org/10.1093/GIGASCIENCE/GIX108

Schunter, C., Carreras-Carbonell, J., MacPherson, E., TintorÉ, J., Vidal-Vijande, E., Pascual, A., Guidetti, P., & Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. https://doi.org/10.1111/j.1365-294X.2011.05355.x

Šegvić-Bubić, T., Marrone, F., Grubišić, L., Izquierdo-Gomez, D., Katavić, I., Arculeo, M., & Lo Brutto, S. (2016). Two seas, two lineages: How genetic diversity is structured in Atlantic and Mediterranean greater amberjack Seriola dumerili Risso, 1810 (Perciformes, Carangidae). *Fisheries Research*, *179*, 271–279. https://doi.org/10.1016/j.fishres.2016.03.018

Šegvić-Bubić, T., Talijančić, I., Žužul, I., Žuvić, L., Grubišić, L., & Izquierdo-Gomez, D. (2022). Culture of Seriola dumerili in a marine ecosystem: Insights from genetic and morphometric fish traits and implications of escape events. *Estuarine, Coastal and Shelf Science*, *278*(May). https://doi.org/10.1016/j.ecss.2022.108115

Serra, I. A., Innocenti, A. M., Di Maida, G., Calvo, S., Migliaccio, M., Zambianchi, E., Pizzigalli, C., Arnaud-Haond, S., Duarte, C. M., Serrao, E. A., & Procaccini, G. (2010). Genetic structure in the Mediterranean seagrass Posidonia oceanica: disentangling past vicariance events from contemporary patterns of gene flow. *Molecular Ecology*, *19*(3), 557–568. https://doi.org/10.1111/J.1365-294X.2009.04462.X

Sherman, K. D., Paris, J. R., King, R. A., Moore, K. A., Dahlgren, C. P., Knowles, L. C., Stump, K., Tyler, C. R., & Stevens, J. R. (2020). RAD-Seq Analysis and in situ Monitoring of Nassau Grouper Reveal Fine-Scale Population Structure and Origins of Aggregating Fish. *Frontiers in Marine Science*, *7*, 157. https://doi.org/10.3389/FMARS.2020.00157/BIBTEX

Sicuro, B., & Luzzana, U. (2016). The State of Seriola spp. Other Than Yellowtail (S. quinqueradiata) Farming in the World. *Reviews in Fisheries Science and Aquaculture*, *24*(4), 314–325. https://doi.org/10.1080/23308249.2016.1187583

Slatkin, M., & Voelm, L. (1991). FST in a hierarchical island model. *Genetics*, *127*(3), 627–629. https://doi.org/10.1093/GENETICS/127.3.627

Sley, A., Hadj Taeib, A., Jarboui, O., Ghorbel, M., & Bouain, A. (2014). Reproductive biology of greater amberjack Seriola dumerili (Risso, 1810) from the Eastern Mediterranean Sea (Tunisia, Gulf of Gabes). *Cahiers de Biologie Marine*, *55*(4), 421–430.

Stobie, C. S., Oosthuizen, C. J., Cunningham, M. J., & Bloomer, P. (2018). Exploring the phylogeography of a hexaploid freshwater fish by RAD sequencing. *Ecology and Evolution*, *8*(4), 2326–2342. https://doi.org/10.1002/ECE3.3821

Thompson, B. A., Beasley, M., & Wilson, C. A. (1999). Age distribution and growth of greater amberjack, Seriola dumerili, from the north-central Gulf of Mexico. *Fishery Bulletin*, *97*(2), 362–371.

Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., … Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*, 5770–5770. https://doi.org/10.1038/NCOMMS6770

Tsaparis, D., &Τσαπάρης, Δ. (2011). *Γενετική ποικιλότητα και στοιχεία οικολογίας των πληθυσμών του ζαρκαδιού (Capreolus capreolus) στην Ελλάδα*. https://doi.org/10.12681/EADD/28125

Viñas, J., Alvarado Bremer, J. R., & Pla, C. (2004). Inter-oceanic genetic differentiation among albacore (Thunnus alalunga) populations. *Marine Biology*, *145*(2), 225–232. https://doi.org/10.1007/S00227-004-1319-5

Wang, K. S., Liu, M., & Paterson, A. D. (2005). Evaluating outlier loci and their effect on the identification of pedigree errors. *BMC Genetics*, *6*(SUPPL.1), 1–5. https://doi.org/10.1186/1471-2156-6-S1-S155/TABLES/3

Wang, S., Mckay, J. K., & Matz, M. V. (2012). *2b-RAD: A simple and flexible method for genome-wide genotyping Establishing the cutting-edge genomic techniques and databases View project*. https://doi.org/10.1038/nmeth.2023

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358. https://doi.org/10.2307/2408641

Wells, R. J. D., & Rooker, J. R. (2004). Distribution, age, and growth of young-of-the-year greater amberjack (Seriola dumerili) associated with pelagic Sargassum. *Fishery Bulletin*, *102*(3), 545–554.

Wright, S. (1965). The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution*, *19*(3), 395. https://doi.org/10.2307/2406450

Yasuike, M., Iwasaki, Y., Nishiki, I., Nakamura, Y., Matsuura, A., Yoshida, K., Noda, T., Andoh, T., & Fujiwara, A. (2018). The yellowtail (Seriola quinqueradiata) genome and transcriptome atlas of the digestive tract. *DNA Research*, *25*(5), 547–560. https://doi.org/10.1093/dnares/dsy024

[39]

Zhang, B. D., Xue, D. X., Li, Y. L., & Liu, J. X. (2019). RAD genotyping reveals fine-scale population structure and provides evidence for adaptive divergence in a commercially important fish from the northwestern Pacific Ocean. *PeerJ*, *2019*(7), e7242. https://doi.org/10.7717/PEERJ.7242/SUPP-6

Zhao, Y., Peng, W., Guo, H., Chen, B., Zhou, Z., Xu, J., Zhang, D., & Xu, P. (2018). Population Genomics Reveals Genetic Divergence and Adaptive Differentiation of Chinese Sea Bass (Lateolabrax maculatus). *Marine Biotechnology*, *20*(1), 45–59. https://doi.org/10.1007/S10126-017-9786-0/FIGURES/9

Zimmerman, S. J., Aldridge, C. L., & Oyler-Mccance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics*, *21*(1), 1–16. https://doi.org/10.1186/S12864-020-06783-9/FIGURES/6

Zitari-Chatti, R., Chatti, N., Fulgione, D., Caiazza, I., Aprea, G., Elouaer, A., Said, K., & Capriglione, T. (2009). Mitochondrial DNA variation in the caramote prawn Penaeus (Melicertus) kerathurus across a transition zone in the Mediterranean Sea. *Genetica*, *136*(3), 439–447. https://doi.org/10.1007/S10709-008-9344-9/FIGURES/3

Zupa, R., Fauvel, C., Mylonas, C. C., Pousis, C., Santamaria, N., Papadaki, M., Fakriadis, I., Cicirelli, V., Mangano, S., Passantino, L., Lacalandra, G. M., & Corriero, A. (2017). Rearing in captivity affects spermatogenesis and sperm quality in greater amberjack, Seriola dumerili (Risso, 1810). *Journal of Animal Science*, *95*(9), 4085. https://doi.org/10.2527/JAS2017.1708

Sample	PopID	Retained reads	Properly paired (%)	Mean_coverage	Sample	PopID	Retained reads	Properly paired (%)	Mean_coverage
FCPCT_01	1	493001	95.63	97.305	ARGO_23	5	792092	92.56	146.008
FCPCT_02	1	8066	43.70		ARGO_24	5	908757	97.10	189.499
FCPCT_03	1	512671	93.98	93.460	ARGO_25	5	2838905	95.51	396.338
FCPCT_04	1	1404285	95.13	203.058	HCMR_01	5	2406917	89.45	309.336
FCPCT_05	1	917982	94.49	145.965	HCMR_02	5	1616870	91.65	230.540
FCPCT_06	1	1430290	96.35	203.558	HCMR_03	5	6961613	94.78	811.972
FCPCT_07	1	118180	94.45	18.592	HCMR_04	5	209388	90.98	61.102
FCPCT_08	1	3722929	95.84	476.490	HCMR_05	5	1237045	92.40	179.546
FCPCT_09	1	980671	88.65	153.643	HCMR_06	5	1036582	87.43	161.548
FCPCT_10	1	3364219	92.16	404.618	HCMR_07	5	3153606	91.34	402.292
FCPCT_11	1	1305834	88.87	186.556	HCMR_08	5	128856	89.24	57.846
FCPCT_12	1	1176576	89.71	170.051	HCMR_09	5	1500072	91.04	217.790
FCPCT_13	1	1563584	92.13	223.312	HCMR_10	5	5707239	94.11	687.080
FCPCT_14	1	5129332	91.87	591.305	HCMR_11	5	7397652	92.42	854.618
FCPCT_15	1	3762095	89.99	449.130	HCMR_12	5	3105606	90.96	405.769
FCPCT_16	1	3426707	91.76	420.748	HCMR_13	5	2056857	92.61	290.488
FCPCT_17	1	2665972	91.45	340.247	HCMR_14	5	5788318	94.56	683.148
FCPCT_18	1	1920395	93.25	278.201	HCMR_16	5	7513712	93.26	851.643
FCPCT_19	1	609733	93.50	117.553	HCMR_17	5	529768	90.44	92.483
FCPCT_20	1	4518888	86.70	512.327	HCMR_18	5	3250909	94.58	425.887
FCPCT_21	1	1612816	86.52	215.845	HCMR_19	5	473554	93.10	98.773
FCPCT_22	1	5082094	86.42	566.167	HCMR_20	5	1107863	92.78	159.677
FCPCT_23	1	55451	86.16	31.830	HCMR_21	5	1237562	90.61	182.989
FCPCT_24	1	2151266	89.90	291.047	HCMR_22	5	2288305	92.84	303.417
FCPCT_25	1	1149703	93.56	178.412	HCMR_24	5	1791923	95.57	262.326
SPAIN_23	2	233483	96.53	73.501	HCMR_25	5	781001	94.28	135.468
SPAIN_24	2	1633123	91.34	227.308	HCMR_26	5	3505218	96.15	460.933
SPAIN_25	2	3803121	94.36	459.354	HCMR_28	5	5902678	96.01	733.841
SPAIN_26	2	2818289	93.25	359.786	HCMR_30	5	3599906	95.24	465.095
SPAIN_27	2	575735	91.90	96.028	HCMR_31	5	128122	89.42	58.062
SPAIN_28	2	196478	91.10	51.730	HCMR_32	5	1503874	93.64	224.305
SPAIN_29	2	4461198	93.39	529.894	HCMR_33	5	2613961	94.11	344.777
SPAIN_30	2	863612	93.15	132.297	SITEIA_01	6	1583174	91.99	222.630
SPAIN_32	2	170780	92.02	49.096	SITEIA_02	6	17076668	94.72	1.731.951
SPAIN_33	2	2301470	92.85	301.540	SITEIA_03	6	3159178	92.12	399.353
SPAIN_34	2	1173983	95.74	168.045	SITEIA_04	6	2152054	95.94	320.690
SPAIN_35	2	2368132	92.56	311.197	SITEIA_05	6	4060743	94.29	492.161
SPAIN_31	2	317471	94.29	68.169	SITEIA_06	6	7090867	95.18	827.502
SPAIN_01	3	4905006	95.31	612.564	SITEIA_07	6	3533472	90.95	446.891
SPAIN_02	3	7667862	95.68	863.710	SITEIA_08	6	6260302	89.57	705.549
SPAIN_03	3	491384	93.33	92.722	SITEIA_09	6	7255413	91.39	817.713
SPAIN_04	3	1871458	93.71	261.777	SITEIA_10	6	4175753	89.97	489.836

SPAIN_05	3	646573	93.44	119.927	SITEIA_11	6	3307604	88.31	403.823
SPAIN_06	3	2098292	96.04	276.494	SITEIA_12	6	3660348	88.29	438.872
SPAIN_07	3	1160626	93.60	186.058	SITEIA_13	6	2676939	92.96	358.408
SPAIN_08	3	2177895	93.23	302.485	SITEIA_14	6	11436691	91.06	1.209.669
SPAIN_09	3	9974663	93.18	1.098.434	SITEIA_15	6	2929455	90.98	384.520
SPAIN_10	3	2516119	92.94	331.206	SITEIA_16	6	4719303	94.78	598.048
SPAIN_11	3	1736199	94.36	261.575	SITEIA_17	6	7268102	92.66	844.216
SPAIN_12	3	3038412	92.91	403.551	SITEIA_18	6	13201779	95.17	1.469.481
SPAIN_13	3	842825	94.79	144.826	HCMR_15	7	5680323	95.31	685.608
SPAIN_14	3	3920453	95.67	495.351	HCMR_23	7	4255547	95.15	535.208
SPAIN_15	3	392368	92.52	85.117	HCMR_27	7	2729709	94.50	358.348
SPAIN_16	3	1893829	95.92	282.292	HCMR_29	7	4092919	96.01	515.941
SPAIN_17	3	393115	93.25	92.671	CYPRUS_01	8	97721	96.61	42.803
SPAIN_18	3	662282	93.10	125.877	CYPRUS_02	8	1109873	97.15	197.267
SPAIN_19	3	362217	93.72	83.328	CYPRUS_03	8	552366	97.97	132.901
SPAIN_20	3	1723723	94.12	260.101	CYPRUS_04	8	95865	97.04	43.690
SPAIN_21	3	285004	93.74	65.434	CYPRUS_05	8	460431	96.89	113.186
SPAIN_22	3	566082	94.87	116.146	CYPRUS_06	8	179626	93.43	41.219
ITALY_02	4	2375966	96.80	313.960	CYPRUS_07	8	679440	96.00	118.438
ITALY_04	4	168670	93.18	49.083	CYPRUS_08	8	1645575	96.29	253.126
ITALY_05	4	1222296	94.49	162.805	CYPRUS_09	8	1825600	96.83	290.032
ITALY_06	4	3410823	96.52	448.745	CYPRUS_10	8	173144	97.01	58.803
ITALY_08	4	1089021	95.01	165.213	CYPRUS_11	8	2408841	97.02	304.950
ITALY_09	4	351382	95.38	76.580	CYPRUS_12	8	592992	97.60	130.319
ITALY_10	4	2000101	94.94	270.315	CYPRUS_13	8	679029	97.34	143.684
ITALY_11	4	1888744	93.17	250.795	CYPRUS_14	8	51518	93.64	17.406
ITALY_14	4	1055370	96.11	162.716	CYPRUS_15	8	631593	96.43	125.257
ITALY_15	4	609233	94.06	104.161	CYPRUS_16	8	179647	96.15	58.816
ITALY_16	4	1154919	93.95	174.356	CYPRUS_17	8	84764	90.75	35.985
ITALY_17	4	1551712	95.69	213.945	CYPRUS_18	8	529285	95.84	101.745
ITALY_18	4	2055664	96.14	281.205	CYPRUS_19	8	751020	96.76	138.384
ITALY_19	4	154656	95.82	52.172	CYPRUS_20	8	223620	97.05	58.613
ITALY_20	4	1571896	96.63	221.889	CYPRUS_21	8	199610	96.17	65.117
ITALY_21	4	14156	**49.75**		CYPRUS_22	8	151086	96.51	62.640
ITALY_22	4	3456646	96.41	423.825	CYPRUS_23	8	390409	97.74	101.525
ITALY_23	4	2592943	96.73	343.291	CYPRUS_24	8	80055	95.44	32.004
ITALY_24	4	495767	95.29	97.614	CYPRUS_25	8	591995	97.12	130.110
ITALY_25	4	824573	95.90	144.915	CYPRUS_26	8	27337	89.21	10.325
ITALY_26	4	736435	94.93	130.270	CYPRUS_27	8	150536	96.02	57.632
ITALY_27	4	4005420	95.98	516.584	CYPRUS_28	8	17655	94.65	19.275
ITALY_29	4	963793	95.93	147.994	CYPRUS_29	8	408985	95.20	85.414
ITALY_30	4	3517486	96.41	424.409	CYPRUS_30	8	978588	94.15	151.817
ITALY_31	4	5007	**43.94**		CYPRUS_31	8	3062840	93.82	399.395
ITALY_32	4	1005215	94.00	158.619	CYPRUS_32	8	1743545	95.36	246.218
ITALY_33	4	3313828	96.59	391.397	CYPRUS_33	8	3739194	94.82	460.850
ITALY_34	4	3373711	94.60	420.758	CYPRUS_34	8	7480486	95.91	886.160
ITALY_35	4	709631	92.62	124.956	CYPRUS_35	8	1107110	92.79	160.203

ITALY_36	4	1506904	91.55	213.572	CYPRUS_36	8	4150932	94.72	517.255
ITALY_37	4	5602451	93.05	638.239	CYPRUS_37	8	165791	93.81	49.480
ITALY_38	4	4584578	93.53	547.555	CYPRUS_38	8	3649829	95.34	463.061
ITALY_39	4	4013836	94.49	501.470	CYPRUS_39	8	1446682	95.65	217.039
ITALY_40	4	2925523	95.25	373.170	CYPRUS_40	8	1857193	94.69	265.171
ITALY_41	4	660449	94.14	119.841	CYPRUS_41	8	1180841	94.01	176.875
ITALY_42	4	4774423	94.89	560.570	CYPRUS_42	8	7974149	91.44	900.669
ITALY_43	4	417774	94.54	91.850	CYPRUS_43	8	1656321	94.74	239.067
ITALY_44	4	1761322	94.18	248.353	CYPRUS_44	8	4675413	94.75	559.508
ITALY_45	4	144928	94.13	50.647	CYPRUS_45	8	1901261	93.55	269.083
ITALY_46	4	15199501	95.54	1.617.988	CYPRUS_46	8	2131186	95.00	297.453
ITALY_47	4	648178	92.95	121.959	CYPRUS_47	8	94124	96.28	49.891
ITALY_48	4	1859366	92.36	257.871	CYPRUS_48	8	2762754	94.80	376.911
ITALY_49	4	4265026	95.77	524.676	CYPRUS_49	8	446414	91.19	90.587
ITALY_50	4	4296601	93.57	530.065	CYPRUS_50	8	6009037	94.65	718.730
ITALY_51	4	2944910	93.93	376.702	CYPRUS_51	8	816711	93.69	140.264
ARGO_01	5	1762042	91.57	248.766	CYPRUS_52	8	2651358	93.38	359.129
ARGO_02	5	663451	92.77	118.862	CYPRUS_53	8	1604807	94.63	235.929
ARGO_03	5	4041750	91.30	479.259	CYPRUS_54	8	3404003	95.42	450.141
ARGO_04	5	4952249	97.10	667.353	CYPRUS_55	8	886830	92.94	150.434
ARGO_05	5	7043572	94.09	830.266	CYPRUS_56	8	4966509	94.49	591.407
ARGO_06	5	3281782	92.41	400.964	KILIC_01	9	3466319	94.09	446.520
ARGO_07	5	1560788	94.22	221.969	KILIC_02	9	1691577	93.40	252.229
ARGO_08	5	2401177	97.16	369.982	KILIC_03	9	2065319	95.11	303.884
ARGO_09	5	2410767	94.23	321.080	KILIC_04	9	1803252	95.16	265.000
ARGO_10	5	3328013	94.66	432.811	KILIC_05	9	562237	91.86	113.817
ARGO_11	5	8087560	95.42	947.158	KILIC_06	9	9097585	93.44	1.027.328
ARGO_12	5	4639369	95.94	593.290	KILIC_07	9	18104565	93.65	1.945.245
ARGO_13	5	2401006	92.29	313.662	KILIC_08	9	2167992	91.66	287.152
ARGO_14	5	1264952	93.42	200.747	KILIC_09	9	352795	90.77	81.063
ARGO_15	5	4676611	93.69	552.053	KILIC10	9	1994629	92.86	284.084
ARGO_16	5	1890512	94.88	283.550	KILIC11	9	3459202	92.01	411.610
ARGO_17	5	995768	95.55	186.440	KILIC12	9	1726482	92.97	244.267
ARGO_18	5	857356	94.41	159.695	KILIC13	9	1246359	94.92	203.200
ARGO_19	5	1056740	93.02	180.066	KILIC14	9	5713425	95.19	700.272
ARGO_20	5	4144054	93.55	514.513	KILIC15	9	2063112	92.45	285.667
ARGO_21	5	662962	96.07	133.009	KILIC16	9	5084555	94.78	612.804
ARGO_22	5	1871763	93.09	266.096	KILIC17	9	326696	90.61	81.579

Table 1 : Population map and individual statistics from Stacks (Rochette et al., 2019). The sample name is displayed on the first column and on the second the population ID. In the third column, the number of reads assigned to each individual after cleaning and demultiplexing (*process_radtags*). In the fourth column the proportion of properly paired reads per individual is shown and individuals in yellow were removed from the analysis due to the low alignment rate, SAMtools 1.9(Li et al., 2009). Finally the last column shows, the effective mean coverage per sample. This indicates the number of times the RAD loci identified in a sample, are covered by the reads, Stacks (Rochette et al.,2019)

Pop	1	2	3	4	5	6	7	8	9
1	0.00000*								
2	0.10594*	0.00000*							
3	0.10051*	0.00675	0.00000*						
4	0.11155*	0.03800*	0.02803*	0.00000*					
5	0.11704*	0.03830*	0.03215*	0.00472*	0.00000*				
6	0.12012*	0.03530*	0.03747*	0.00414	0.00005	0.00000*			
7	0.15341*	0.04570*	0.04820*	0.00751	0.00636	0.00552	0.00000*		
8	0.12284*	0.04300*	0.03852*	0.00422	0.00435	0.00270	0.00603	0.00000*	
9	0.12840*	0.04803*	0.04043*	0.01117*	0.01051*	0.01069*	0.00987	0.00520	0.00000*

Table 2 : Pairwise F$_{ST}$ matrix. The asterisk indicates values that are statistically significant (p<0.05), Arlequin (Lischer&Excoffier, 2012)
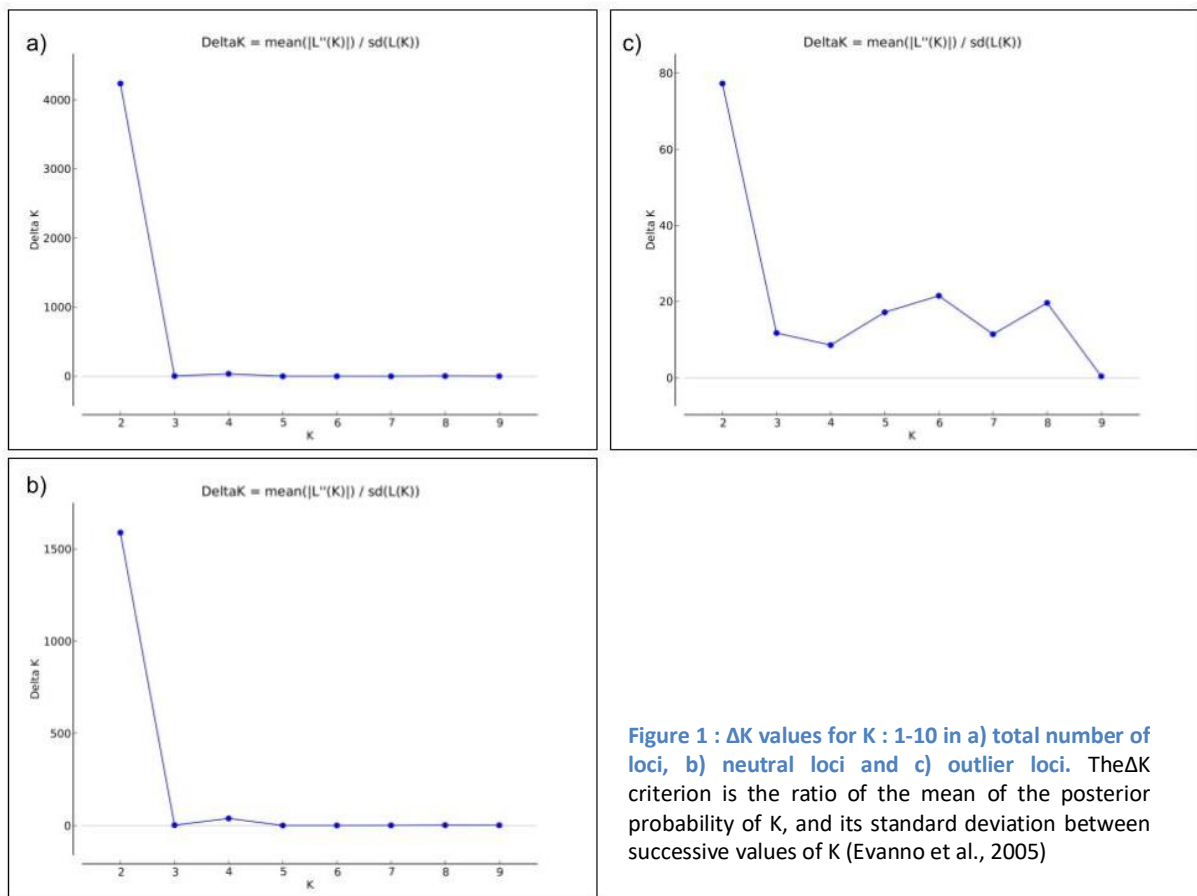

Figure 1 : ΔK values for K : 1-10 in a) total number of loci, b) neutral loci and c) outlier loci. TheΔK criterion is the ratio of the mean of the posterior probability of K, and its standard deviation between successive values of K (Evanno et al., 2005)
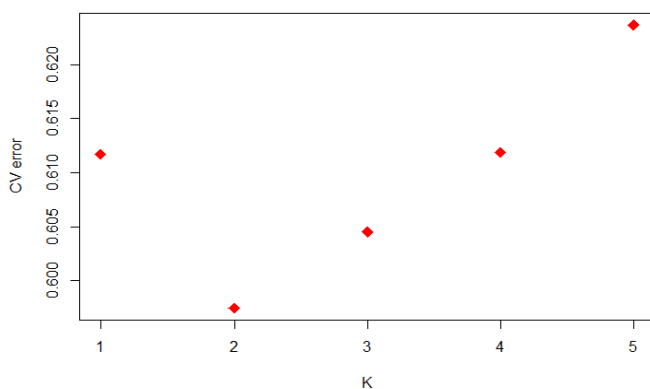

Figure 2 : Cross Validation Error for consecutive values of K in the ancestry analysis. The CV error is the criterion for choosing the best K in the ADMIXTURE (Alexander et al., 2009) software and the lowest value here is for K=2.