

Population genomic analysis of the European anchovy (*Engraulis encrasicolus*) in the Western and Central Mediterranean

Δαμιανός Αλεξανδρίδης
Επιβλέπων: Κωνσταντίνος Τσιγγενόπουλος

Abstract

Genomic studies have recently become a powerful tool to gain compelling insights into population history, patterns and genetic connectivity. The European anchovy (*Engraulis encrasicolus*), one of the most important pelagic fish in the Mediterranean Sea, constitutes a keystone species within the marine ecosystem due to its role in regulating trophic interactions, while also contributing significantly to the region's fishing industry. The aim of the study was to assess the European anchovy's population genetic structure in the Western and Central Mediterranean using a genomic approach. To identify a great number of polymorphic loci, we first assembled a draft reference genome of 1.69 Gb and 79.8% BUSCO completeness serving as a powerful mean for our downstream analysis. Then, we generated ddRAD-sequencing data from 398 anchovy individuals across twelve Western and Central Mediterranean sites, which led to the successful genotyping and further study based on 9,497 single nucleotide polymorphisms (SNPs). Results showed that our samples were divided into two clusters: one grouping the individuals from the Atlantic and the Alboran Sea, and the other including individuals from the Northwestern and the Central Mediterranean Seas ($F_{ST}=0.09$). This differentiation highlights the presence of two distinct genetic pools, which is in alignment with prior research on the European anchovy, indicating once again the Almeria–Oran front as a potential dispersal barrier for gene flow in the species. Consequently, our research sets the ground for delineating distinct stocks and implementing targeted conservation strategies to safeguard the species from over-exploitation and habitat degradation.

Introduction

The Mediterranean anchovy (*Engraulis encrasicolus*), one of the most important pelagic fish in the Mediterranean Sea, constitutes a keystone species within the marine ecosystem due to its role in regulating trophic interactions while also contributing significantly to the region's fishing industry. Population studies have shown a particular genetic structure in the species throughout its distribution area. Inside the Mediterranean basin, morphological and genetic data have shown population substructure in the Eastern Mediterranean but no genetic differentiation in the Western Mediterranean. In particular, only the Alboran population was shown to be genetically distinct from the rest of Western Mediterranean, whereas is genetically close to adjacent Atlantic regions.

Aim

We aim at providing a better understanding of the current population structure of the European anchovy in the Western and Central Mediterranean by analyzing genome-wide single nucleotide polymorphisms (SNPs). To better achieve this, we first attempted to generate the draft whole genome assembly of the species. This decision was motivated by the need to have a high number of loci resulting in high resolution in our dataset.

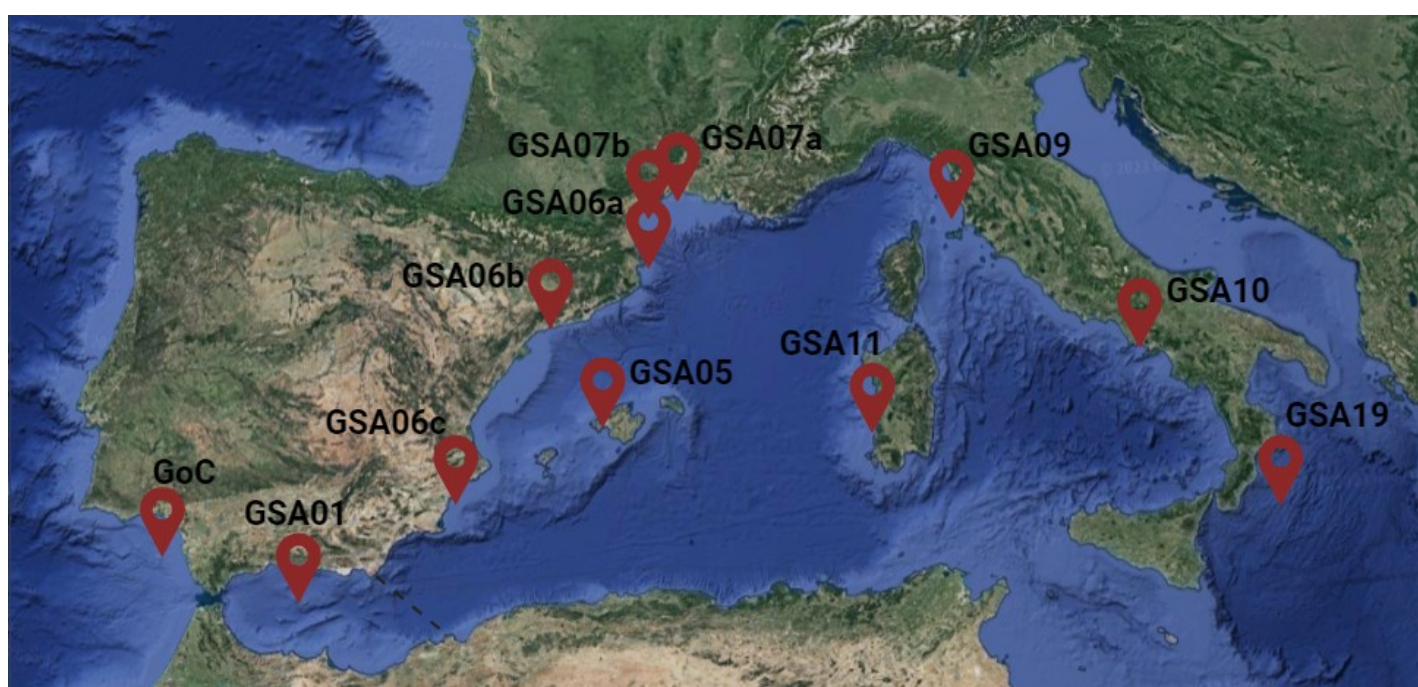


Figure 1. Geographical distribution of sampling sites. The Almeria-Oran Front (AOF) is indicated by a black dashed line.

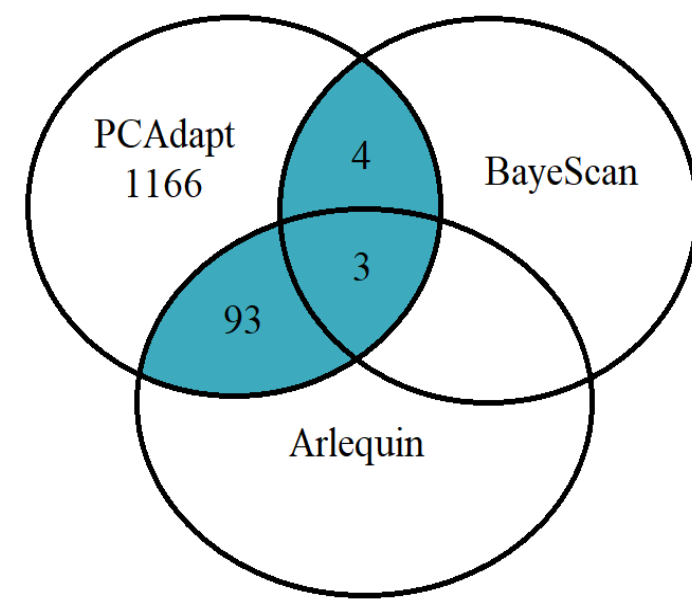


Figure 2. Venn diagram representing the potential outliers that occurred by three approaches and the Overlapping SNP loci.

Materials and Methods

Sampling, DNA extraction and ddRAD sequencing

408 anchovy individuals were collected from 12 different locations across the Western and Central Mediterranean Sea, and the adjacent Atlantic waters (Fig. 1). Genomic DNA was extracted and libraries were generated for ddRAD sequencing.

Sampling, DNA extraction, Nanopore sequencing, Genome assembly and evaluation

One anchovy individual was collected from the Euboean Gulf for constructing the European anchovy draft genome. DNA extraction, size selection and library preparation were conducted. Then, three MinION flowcells were run (Nanopore sequencing) and the following basecalling was done with Guppy v6.5.7 and Dorado v0.4.1. Flye v2.9 was used for the genome assembly. The assembly was polished with flye and racon v1.4.20. The generated assembly was evaluated in terms of expected genes content using BUSCO.

Population genomic analyses

ddRAD data were analyzed both with *de novo* and reference-based approaches to check the number of polymorphic loci increment when using the reference genome. STACKS v2.64 pipeline was used for cleaning, quality control the reads and genotype each individual. 20% missingness cut-off was applied to the dataset and highly linked loci ($r^2=0.5$) were excluded using PLINK v1.9. BayeScan v2.1, PCAdapt R package and Arlequin v3.5 was used for outlier detection (fig 2). Clustering analyses were conducted with STRUCTURE, AMOVA and DAPC. F-statistics were calculated using Arlequin and hierfstat R package.

Results

Genome assembly

Flye pipeline assembled a genome of total length 1.69 Gb, which contains 76,551 contigs (contig N50: 57,970). BUSCO assessment found 2,905 (79.8%) completed, 257 fragmented (7.1%) and 478 (13.1%) missing BUSCOs.

Population genomic analyses

398 individuals and 9,497 polymorphic loci passed the quality control. In comparison, the *de novo* approach yielded 952 polymorphic loci. One hundred outlier loci were found from at least 2 approaches (Fig. 2). The best clustering scenario was K=2 as determined by STRUCTURE (Fig. 3), DAPC, AMOVA.

The first cluster (ATL) contains the Cadiz population (E. Atlantic) and the Alboran population. The second cluster (MED) consists of the Central and Western anchovy populations (except Alboran). Pairwise F_{ST} between the clusters was 0.097 (statistically significant). No substructure was observed in the MED cluster (Fig. 4). Only a few individuals from the GSA19 stock (Western Ionian Sea) might be genetically differentiated from the MED cluster.

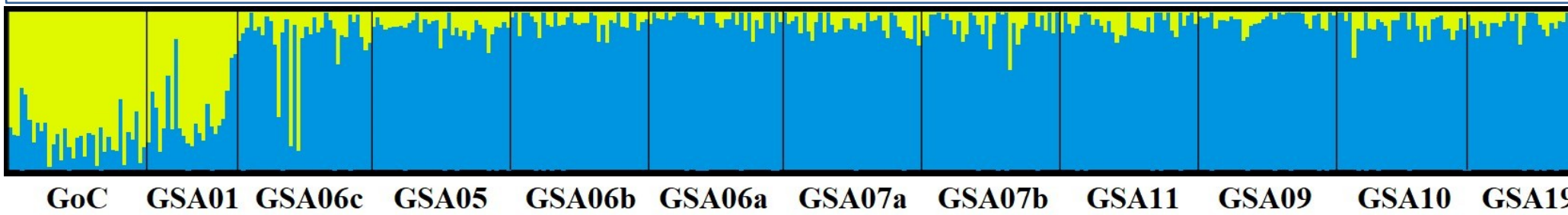


Figure 3. Population structure of European anchovy in the Eastern Atlantic and Western and Central Mediterranean Sea as inferred by STRUCTURE for K = 2.

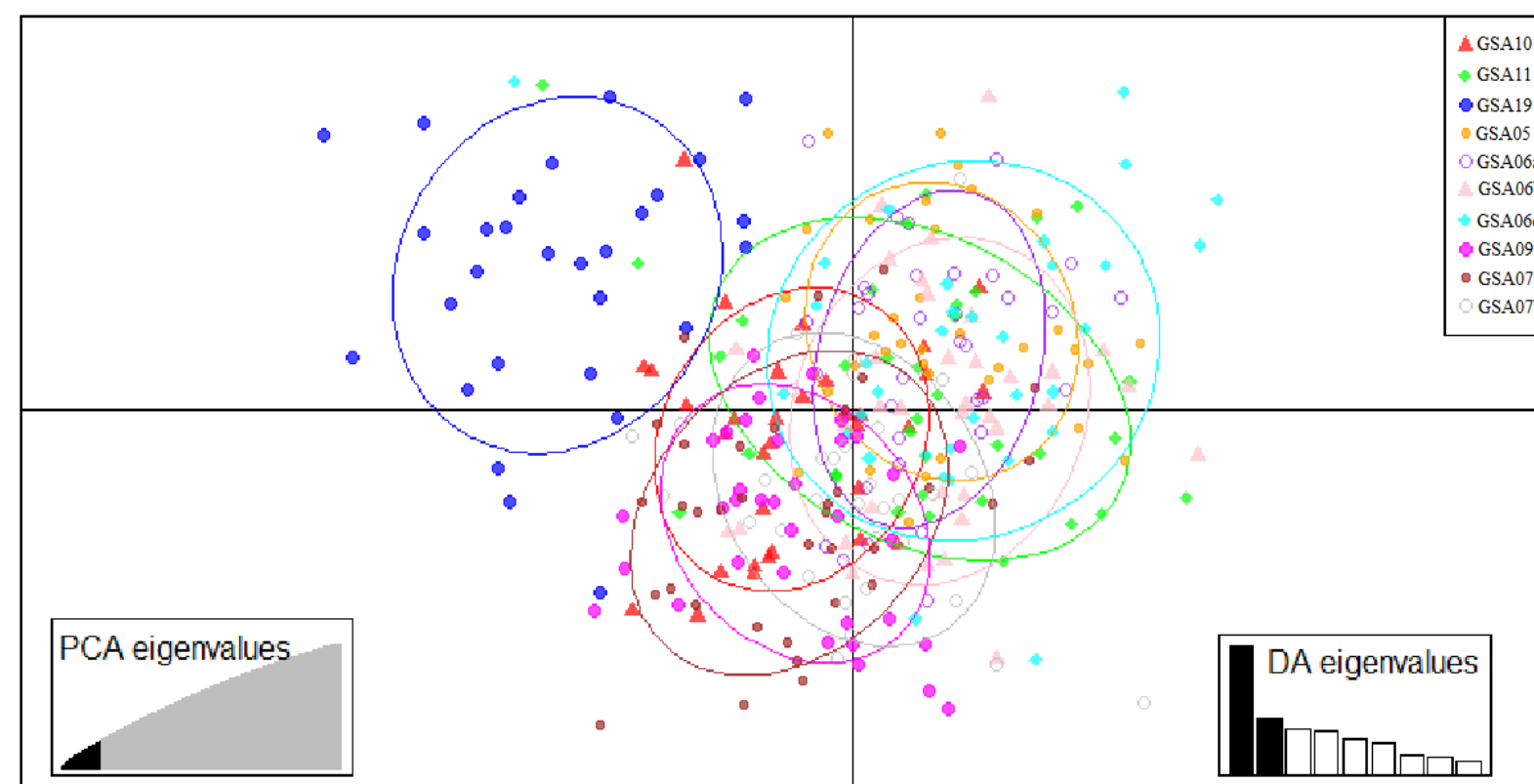


Figure 4. DAPC using all loci for the MED cluster.

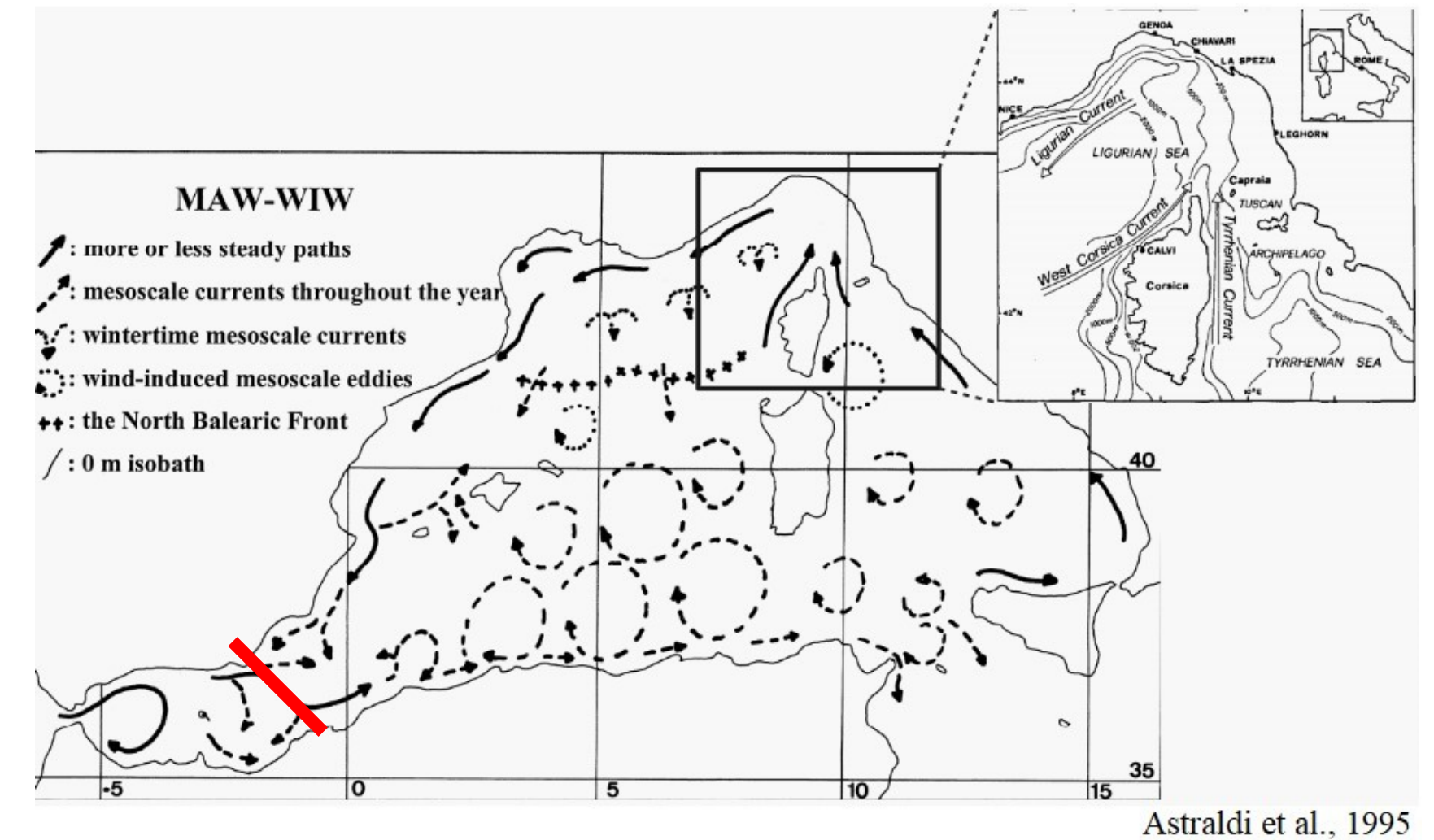


Figure 5. Circulation of the Modified Atlantic Water and the Winter Intermediate Water (Millot, 1999). Figure in top right: area in the north-western corner of the Mediterranean, showing the boundary between the Ligurian and Tyrrhenian seas (Tuscan Archipelago). Arrows illustrate the major current systems (graph from Astraldi et al., 1995). The red line is where the Almeria-Oran front is located.

Discussion

Our genomic dataset revealed two distinct genetic clusters in the Western and Central Mediterranean and Eastern Atlantic. The Almeria-Oran front is suggested to be a potential mechanism contributing to the observed clustering (Fig. 5). Furthermore, the genetic homogeneity observed in anchovy stocks within the Western Mediterranean basin is likely due to specific water current systems in the region. The low genetic differentiation in the Central Mediterranean implies potential connectivity via the Siculo-Tunisian Strait.

Conclusion

- Population genomics analyses based on a reference genome usually provide a much larger number of loci compared to *de novo* ones. Here, the reference-based approach increased ten times the number of polymorphic loci.
- Our findings are in alignment with prior research.
- The Almeria-Oran Front is suggested, once again, as potential barrier to gene flow for the European anchovy.
- The Siculo-Tunisian Strait might not be a biogeographic boundary for the species.

References

- Astraldi, M., Bianchi, C. N., Gasparini, G. P., Morri, C., & Leric, D. (1995). OCEANOLOGICA ACTA-VOL. 18-N°2 ~ Climatic fluctuations, current variability and marine species distribution: Variabilité du climat Espèces thermophiles Dispersion Epibenthos Biogéographie marine ABSTRACT RÉSUMÉ (Vol. 18).
- Bakun, A. (2006). Wasp-waist populations and marine ecosystem dynamics: Navigating the "predator pit" topographies. Progress in Oceanography, 68(2–4), 271–288. <https://doi.org/10.1016/j.pocan.2006.02.004>
- Grant, W. S. (2005). A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): Assessing models of population structure and the Black Sea isolation hypothesis. Genetics, 125(2–3), 293–309. <https://doi.org/10.1007/s10709-005-0717-z>
- Ruggeri, P., Sileadani, A., Occhipinti, G., Fioravanti, T., Sant'anni, A., Leonori, I., De Felice, A., Ameri, E., Proccacci, G., Catanesi, G., Ticina, V., Bonanno, A., Cerioni, P. N., Giovannotti, M., Grant, W. S., & Barucchi, V. C. (2016). Biocomplexity in populations of european anchovy in the adriatic sea. PLoS ONE, 11(4), <https://doi.org/10.1371/journal.pone.0153061>
- Zarronadinda, I., Iriando, M., Albaina, A., Pardo, M. A., Manzano, C., Grant, W. S., Irigoien, X., & Estomba, A. (2012). Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in european anchovy (*Engraulis encrasicolus* L.). PLoS ONE, 7(7), <https://doi.org/10.1371/journal.pone.0042201>