

# **Deliverable 3.5.2**

# Stock structure and adaptation in the Mediterranean

# ConFish

 Connectivity among Mediterranean fishery stakeholders and scientists resolves connectivity of fishery populations –

WP3 – STUDYING

ACTIVITY 3.5 - EVOLUTIONARY ANALYSES OF STOCK STRUCTURE AND ADAPTATION

PARTNER IN CHARGE: UNIVERSITY OF ZAGREB

PARTNERS INVOLVED: CSIC, ISPRA

OCTOBER 2018



## Contents:

1.	Introduction	1							
2.	. Additional data filtering								
3.	Genomic diversity	4							
4.	Genomic differentiation	5							
5.	Mantel test	7							
6.	Populations structure	8							
7.	Signs of selection	10							
8.	Genotype-environment associations	12							
8.	1. Collection of environmental data	13							
8.	.2. Bayenv 2 analysis	13							
	8.2.1. Bayescan – Bayenv associations	15							
9.	Discussion and conclusions	16							
10.	Literature	18							



## 1. Introduction

Facing the extensive global changes experienced by marine environments in recent times, traditional fisheries are increasingly urged to incorporate the perspective of fishery stocks as evolutionary units into their management (Laugen et al. 2014, Waples & Hendry 2008). This evolutionary perspective implies the understanding that populations are inherently able to cope with changes in selective pressures by means of acclimatization or adaptation. Accordingly, the evolutionary potential of population refers to its intrinsic ability to respond positively when faced with shifts in environmental or anthropogenic influence (Eizaguirre & Baltazar-Soares 2014).

In the core of every population's evolutionary potential is genetic diversity. Populations that exhibit high levels of genetic diversity have a large pool of standing genetic variation on which selection can act upon, effectively giving the population higher probability to adapt to different environments than those with depauperate diversity. In other words, the more genetic variation is present in a population, the better the chance that population will be able to adapt and survive (Frankham 1996). Population's genetic diversity will in part depend on its connectivity to other populations (Gagnaire et al. 2015). The more connected the two populations are, and the higher the number of individuals they exchange, the bigger the likelihood that exchange of genetic material between them will occur. Hence, connectivity to other populations increases genetic variability and play a crucial role in maintaining adaptive potential of natural populations.

The adaptation to different environments is the product of population's genetic diversity, their connectivity to other populations and the external pressures acting upon them. However, populations that are well adapted to their specific environment tend to be less diverse within themselves, making them highly susceptible to selective pressures. Individuals facing an environment for which they are not adapted have a lower chance of survival and reduced number of offspring. Investigating the adaptive genetic variation is therefore extremely important for fisheries management, as it is critical for assessing the stocks resilience to the selective pressure exerted by climate change and fisheries (Valenzuela-Quiñonez 2016).



In this document, we present results of evolutionary genomics study on seventeen common octopus (*Octopus vulgaris*) and twelve red shrimp (*Aristeus antennatus*) locations (Figures 1 and 2), giving a high resolution perspective into stock structure, connectivity and local adaptation of these two species in the Mediterranean.



Figure 1 Geographic locations of analysed Octopus vulgaris populations





Figure 2 Geographic locations of analysed Aristeus antennatus populations

## 2. Additional data filtering

The genomic markers analysed in this study were obtained using *gstacks* program within Stacks 2.0 (Catchen et al. 2013). Total number of discovered single nucleotide polymorphisms (SNPs) was 79134 for octopus, and 35552 for shrimp. The protocol followed to obtain the genomic markers is detailed in Deliverable 3.5.1. For shrimp, genomic diversity and differentiation was estimated from the entire dataset of 12 populations, while the population structure, signatures of selection and genotype-environment associations were inferred from the smaller dataset containing a total of 9 populations and 21559 SNPs.



Additional data filtering to reduce the number of missing values in the dataset was done on a structure file produced by Stacks 2.0 populations program, using custom made Perl scripts, in several steps by:

- 1. Trimming the samples that were empty (more than 85% of missing values)
- 2. Deleting the loci with more than 30% of missing values
- 3. Deleting samples with over 30% of missing values
- 4. Deleting morphometric loci
- Transforming the structure file to genepop through PGDSpider software (Lischer & Excoffier 2012) and imputing the most frequent genotype of the population in the few remaining missing values.
- 6. And again transforming the file to structure, bayescan or arlequin format

The resulting datasets had a total of 16647 SNPs for 17 octopus and 4434 SNPs for 11 shrimp populations (or 414 SNPs for 8 shrimp populations in the smaller dataset). The shrimp population from Santanyí (SYa) was excluded from further analysis due to high percentage of missing data. This filtered datasets were then used in further Arlequin (Excoffier & Lischer 2010), Bayescan (Foll & Gaggiotti 2008) and Bayenv (Günther & Coop 2013) analyses.

## 3. Genomic diversity

Genomic diversity, in the form of observed heterozygosity, was estimated from the dataset of 79134 SNPs for octopus and 35552 SNPs for shrimp populations, using *arlecore*, console version of population genetics data analysis software Arlequin 3.5.2.2 (Excoffier & Lischer 2010). The snps.vcf file obtained from Stacks 2.0 populations program was only filtered for the samples with over 30% of missing values (which excluded the SYa population), and transformed to Arlequin format using PGDSpider software. The setting file was created beforehand using the desktop version of Arlequin, and the parameters of analysis were set as: Molecular Distance = Pairwise difference and Gamma Value = 0.



The obtained values for observed heterozygosity are presented in tables below (Table 1 and 2). Octopus populations that showed lowest level of genetic diversity are Puglia Ionica (PI) and Mola di Bari (ML) from the South Italian Adriatic Sea. For shrimp, the lowest observed heterozygosity was measured for populations from Santa Pola (SAa) and Castelsardo (CSa).

**Table 1** Mean values obtained for observedheterozygosity (Ho) in analysed octopuspopulations

Population	Но
AL	0.24467
DN	0.18786
FM	0.18669
GH	0.21022
IG	0.19174
LL	0.22318
LP	0.18486
ML	0.17273
ОН	0.19515
PI	0.17841
PL	0.21965
PT	0.19297
SP	0.19573
SY	0.20517
TR	0.20323
VR	0.18376
VS	0.20466

**Table 2** Mean values obtained for observedheterozygosity (Ho) in analysed shrimppopulations

Population	Но						
ALa	0.31226						
CSa	0.21856						
DNa	0.22505						
GRa	0.30138						
ОТа	0.27378						
PLa	0.23316						
PSa	0.24333						
РТа	0.24287						
SAa	0.21517						
SLa	0.22928						
TRa	0.22993						

## 4. Genomic differentiation

Pairwise population differentiation was estimated in *arlecore*, console version of Arlequin 3.5.2.2 software (Excoffier & Lischer 2010) on the filtered dataset of 16647 SNPs for octopus, and 4434 SNPs for shrimp. The setting file was created beforehand using the desktop version of Arlequin, and it defined parameters of analysis as: No. of permutations for significance = 100; No. of permutations for Mantel test = 1000; Gamma value for distance matrix = 0; No. of steps in



Markov chain = 100000; No. of Dememorisation Steps = 10000; Required precision on Probability = 0; Significance level = 0.05.

The obtained results showed highest F<sub>ST</sub> values (standard measure of population differentiation) for octopus' population from Atlantic - Olhão (OH), followed by two Adriatic populations – Vir (VR) and Vis (VS), and Alcúdia (AL) – the octopus population from Eastern Balearic Islands (Table 3). For shrimp, the sites that showed highest differentiation from others were also Alcúdia (ALa) in the Eastern Balearic, as well as Garrucha (GRa) from the Southern part of Spanish Mediterranean coast (Table 4). High F<sub>ST</sub> values found on these sites suggest genomic differentiation of octopus and shrimp populations in those areas.

**Table 3** Pairwise  $F_{ST}$  values obtained for Mediterranean octopus populations using Arlequin software. A colour scale is used to better visualise the difference in recorded values (dark red – high vales, dark blue – low values).

	ОН	DN	TR	PL	LL	GH	SY	AL	SP	РТ	IG	FM	LP	PI	ML	VS	VR
он		0.388	0.451	0.447	0.449	0.377	0.430	0.436	0.416	0.427	0.433	0.445	0.436	0.471	0.462	0.475	0.511
DN	0.388		0.067	0.089	0.090	0.067	0.058	0.108	0.048	0.060	0.067	0.056	0.072	0.085	0.075	0.095	0.153
TR	0.451	0.067		0.098	0.104	0.096	0.074	0.140	0.061	0.073	0.078	0.063	0.079	0.083	0.070	0.102	0.165
PL	0.447	0.089	0.098		0.131	0.114	0.103	0.160	0.083	0.102	0.108	0.093	0.103	0.118	0.102	0.137	0.200
ш	0.449	0.090	0.104	0.131		0.120	0.103	0.159	0.087	0.107	0.101	0.086	0.102	0.113	0.095	0.128	0.197
GH	0.377	0.067	0.096	0.114	0.120		0.090	0.138	0.076	0.096	0.105	0.091	0.102	0.122	0.110	0.133	0.192
SY	0.430	0.058	0.074	0.103	0.103	0.090		0.134	0.058	0.072	0.079	0.065	0.077	0.086	0.077	0.109	0.170
AL	0.436	0.108	0.140	0.160	0.159	0.138	0.134		0.109	0.134	0.132	0.122	0.128	0.146	0.135	0.167	0.233
SP	0.416	0.048	0.061	0.083	0.087	0.076	0.058	0.109		0.059	0.063	0.049	0.062	0.076	0.061	0.088	0.146
РТ	0.427	0.060	0.073	0.102	0.107	0.096	0.072	0.134	0.059		0.076	0.063	0.080	0.089	0.077	0.106	0.163
IG	0.433	0.067	0.078	0.108	0.101	0.105	0.079	0.132	0.063	0.076		0.064	0.073	0.081	0.070	0.099	0.158
FM	0.445	0.056	0.063	0.093	0.086	0.091	0.065	0.122	0.049	0.063	0.064		0.065	0.070	0.057	0.085	0.145
LP	0.436	0.072	0.079	0.103	0.102	0.102	0.077	0.128	0.062	0.080	0.073	0.065		0.078	0.063	0.088	0.147
PI	0.471	0.085	0.083	0.118	0.113	0.122	0.086	0.146	0.076	0.089	0.081	0.070	0.078		0.049	0.073	0.126
ML	0.462	0.075	0.070	0.102	0.095	0.110	0.077	0.135	0.061	0.077	0.070	0.057	0.063	0.049		0.062	0.112
vs	0.475	0.095	0.102	0.137	0.128	0.133	0.109	0.167	0.088	0.106	0.099	0.085	0.088	0.073	0.062		0.150
VR	0.511	0.153	0.165	0.200	0.197	0.192	0.170	0.233	0.146	0.163	0.158	0.145	0.147	0.126	0.112	0.150	



**Table 4** Pairwise F<sub>ST</sub> values obtained for Mediterranean red shrimp populations using Stacks 2.0 population program. A colour scale is used to better visualise the difference in recorded values (dark red – high vales, yellow – low values).



### 5. Mantel test

Mantel test is one of the most popular approaches used to evaluate spatial processes driving population structure. This test, based on comparison of genetic distance (estimated by pairwise F<sub>ST</sub> values) with geographical distances between populations of interest, was performed in R (3.5.0) using *ade4*, and *reshape2* packages.

In the case of octopus, a correlation between geographical distances and genomic differentiation between sites could not be excluded (p-value = 0.002; Figure 3A), suggesting that geographic separation could be influencing species genomic divergence. For shrimp populations, however, there was no significant correlation between spatial distances and measured values of genomic differentiation within the Western Mediterranean basin (p-value = 0.279; Figure 3B), implying that most of the observed genomic differences are environmentally driven.



**Figure 3** Mantel correlogram obtained for A) analysed octopus populations and B) analysed shrimp populations

### 6. Populations structure

Population structure of both species was inferred from unfiltered data of 79134 SNPs for 17 populations of octopus, and 21056 SNPs for a total of 9 populations of shrimp, using fastStructure 1.0 program, a clustering software based on a variational Bayesian framework for posterior inference (Raj et al. 2014). Populations structure files obtained from Stacks 2.0 populations program (Catchen et al. 2013), were edited to correspond to fastStructure input format using custom made Perl scripts. Settings used for the analysis were prior=logistics, and K set to 17 for octopus and to 10 for red shrimp populations.

In order to choose the appropriate number of model components that explain structure in the dataset (i.e the likely number of clusters), chooseK.py script provided with the program was run on fastStructure output. For octopus the model complexity that maximizes marginal likelihood was 2; and model components used to explain structure in data value was 1. For shrimp the model complexity that maximizes marginal likelihood was 10; and model components used



to explain structure in data value was 9. To explore the data further the output was plotted using *distruct2.2.py* script from the same program.

fastStructure analysis showed clear differentiation between Atlantic (OH) and Adriatic octopus populations (PI, ML, VS, and VR), with Western Mediterranean populations appearing as a genetic admixture between those two clusters (Figure 4). Red shrimp populations showed lower degree of population structuring within the Western Mediterranean basin. Particularly, Eastern Balearic populations (ALa and SYa), as well as Garrucha (GRa) from Southern Spain, showed more similarity with Italian shrimp populations than from those off the Spanish coast (Figure 5), that are geographically much closer.



**Figure 4** Structure barplot (K=7) showing the genetic background of analysed octopus populations in the Mediterranean, obtained by fastStructure software. On the x-axis, individuals are grouped by populations and on the y-axis is represented the admixture level. Colours represent the likelihood to belong to determined K.





**Figure 5** Structure barplot (K=9) showing the genetic background of analysed red shrimp populations in the Mediterranean, obtained by fastStructure software. On the x-axis, individuals are grouped by populations and on the y-axis is represented the admixture level. Colours represent the likelihood to belong to determined K.

## 7. Signs of selection

To determine if any of analysed SNPs show signs of selection, Bayescan 2.0 program was run (Foll & Gaggiotti 2008). This program identifies candidate loci under natural selection from genetic data using differences in allele frequencies between populations. The analysis was run together on nine populations of shrimp, and separately for Adriatic and Western Mediterranean populations on octopus, in order to eliminate the false positives that may arise from joint analysis of structured populations. The results were explored and plotted in R using their provided plot\_R.r script (R 3.4.3). In total, 196 SNPs showed signs of directional and/or balancing selection in Western Mediterranean populations of octopus (Figure 6), as did the 21 SNPs identified as candidate loci under selection in Adriatic octopus populations (Figure 7). Only one of these SNPs was significant in both Mediterranean and Adriatic populations. For shrimp, only 3 SNPs emerged as candidate loci (Figure 8).





Figure 6 Bayescan analysis of Western Mediterranean octopus populations.



Figure 7 Bayescan analysis of Adriatic octopus populations.





Figure 8 Bayescan analysis of Mediterranean shrimp populations

## 8. Genotype-environment associations

In order to test if any of the identified genomic markers are further involved in local adaptation, the dataset was tested for unusual correlation with the differences in environmental factors between the sites using Bayenv 2 program (Günther & Coop 2013). This program uses a Bayesian method to estimate the empirical pattern of covariance in allele frequencies between populations from a set of markers, and then employs this as a null model for a test at individual SNPs.



#### 8.1. Collection of environmental data

Environmental variables used in the analysis were extracted from Bio-ORACLE 2.0 online database - a set of raster layers providing geophysical, biotic and environmental data for surface and benthic marine realms at a spatial resolution of 5 arcmin (Assis et al. 2017). Mean values of 12 environmental variables measured in 2014 at maximum depth were extrapolated for each site using their R package:

- Chlorophyll concentration
- Current velocity
- Dissolved oxygen concentration
- Iron concentration
- Phosphate concentration
- Nitrate concentration
- Sea water temperature
- Carbon phytoplankton biomass
- Primary production
- Sea water salinity
- Silicate concentration
- Sea floor depth

Variables were standardized by subtracting the mean and dividing with standard deviation across the populations. The resulting database was used as an input file for Bayenv 2 analysis.

#### 8.2. Bayenv 2 analysis

First, a correlation matrix was estimated by running the program on a subset of randomly chosen SNPs for octopus and all available SNPs for shrimp. The last matrix from the estimate matrix output was used as matrix input file for Bayenv 2. To execute the program, the provided calc\_bf.sh script was run with -t - f - X - c flags. The loci that were considered significant in the



analysis were in top 1% of Bayes factor (BF) and top 1% of p (Spearman's coefficient) values. All analysed environmental variables correlated with at least some of the genomic markers, but the ones showing the highest level of correlation for both species were, as expected, differences in biologically important factors such as temperature, salinity, dissolved oxygen concentration and seafloor depth among sites. The values for these factors are shown in the figures below (Figure 9 and 10).



**Figure 9** Bayenv results for correlation testing between genomic markers and differences in environmental factors affecting octopus populations. SNPs considered significant are in the ones with highest Bayes factor and Spearman's coefficient values (in the top right).





**Figure 10** Bayenv results for correlation testing between genomic markers and differences in environmental factors affecting shrimp populations. SNPs considered significant are in the ones with highest Bayes factor and Spearman's coefficient values (in the top right).

#### 8.2.1. Bayescan – Bayenv associations

The SNPs that were considered significantly correlated with differences in any of the analysed environmental variables among sites were compared to those previously identified as candidate loci under selection using Bayescan software. However, there was no overlap between the SNPs considered significant in those two analyses, neither for octopus nor for shrimp populations.



### 9. Discussion and conclusions

The genomic study of common octopus and red shrimp among sampled locations provided deep insight of each species connectivity within the Mediterranean area. For octopus, clear genomic differentiation was found between populations from the Adriatic Sea and the population from Atlantic Ocean, with the Western Mediterranean populations exhibiting a mix of genomic signatures from those two groups. This highlights the previously unreported reduced connectivity and population structuring between octopus populations from Western Mediterranean basin and Adriatic Sea. Within the Western Mediterranean basin, both octopus and red shrimp populations showed varied levels of differentiation. Notably, high levels of genomic differentiation were found for shrimp populations from Eastern Balearic Islands and South of Spain. Interestingly, those populations are genetically similar to those of the Italian coast. Indeed, both octopus and shrimp populations along Spanish coast are genetically distinct from one another, despite their apparent geographic proximity.

While the large-scale population structure for octopus seems to be in part influenced by geographical distance (Mantel test p-value < 0.05), the observed patterns of genomic differences indicate that the large part of differentiation is also stemming from environmental differences. Indeed, both species investigated in this study showed correlations between patterns of genomic differentiation, and differences in environmental factors such as temperature, salinity and dissolved oxygen among sites. Octopus populations in the Adriatic and shrimp in Balearic Islands especially appear to be locally adapted to those specific environments. Further data analysis will be necessary to determine exactly how much of this difference between the populations stems from either geographical obstacles to migration, or local adaptations arising due to different environment they inhabit.

Yet, populations that are genetically isolated and well adapted to their specific environment tend to be less diverse. Indeed several octopus populations from Adriatic showed lower levels of genetic diversity, as did the shrimp populations from Eastern Balearic and South of Spain that also exhibited higher population differentiation. These populations should be taken special care of as they exhibit limited connectivity, high local adaptation, and, under fast shifting



environmental scenarios, lower evolutionary potential. In theory, those populations would not easily recover and are particularly endangered from changes in external stressors. This also means that in the case of population collapse it would be harder for migrants from other populations to replace them, as they are not specifically adapted to those environments.

The disparity in connectivity levels exhibited by octopus and shrimp in the Mediterranean appears to be due to species different life histories and juvenile development. These findings emphasises the need for stock identification based on each species particular characteristics, and not necessarily geopolitical boundaries. Recognizing these populations as separate and implementing stock specific management into Mediterranean fisheries should be the first step towards their effective management.



## 10. Literature

- Assis, J. et al., 2017. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27(3), pp.277–284.
- Catchen, J. et al., 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), pp.3124–3140.
- Eizaguirre, C. & Baltazar-Soares, M., 2014. Evolutionary conservation-evaluating the adaptive potential of species. *Evolutionary Applications*, 7(9), pp.963–967.
- 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), pp.564–567.
- Foll, M. & Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), pp.977– 993.
- Frankham, R., 1996. Relationship of Genetic Variation to Population Size in Wildlife. *Conservation Biology*, 10(6), pp.1500–1508.
- Gagnaire, P.A. et al., 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, 8(8), pp.769–786.
- Günther, T. & Coop, G., 2013. Robust identification of local adaptation from allele frequencies. *Genetics*, 195(1), pp.205–220.
- Laugen, A.T. et al., 2014. Evolutionary impact assessment: Accounting for evolutionary consequences of fishing in an ecosystem approach to fisheries management. *Fish and Fisheries*, 15(1), pp.65–96.
- Lischer, H.E.L. & Excoffier, L., 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), pp.298–299.
- Raj, A., Stephens, M. & Pritchard, J.K., 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), pp.573–589.
- Valenzuela-Quiñonez, F., 2016. How fisheries management can benefit from genomics? *Briefings in Functional Genomics*, 15(5), pp.352–357.
- Waples, R.S. & Hendry, A.P., 2008. EDITORIAL: Special Issue: Evolutionary perspectives on salmonid conservation and management. *Evolutionary Applications*, 1(2), pp.183–188.