Project co-financed by the European Regional Development Fund

Deliverable 3.5.1

Evolutionary genomics data collection

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

In today's rapidly changing marine ecosystems, exploited species are in increasing risk of environmental and anthropogenic disturbances. Being selective pressures, it is only natural to expect such disturbances to trigger evolutionary responses. In fisheries, the most prominent example of this is the observed reduction in body size as a response to size-selective fishing pressure, which translate in earlier age of maturation recorded in several heavily exploited species (Conover & Munch 2002, Olsen et al. 2004). Hence, exploring the evolutionary potential, or the capacity to respond positively to shifts in selective pressures (Eizaguirre & Baltazar-Soares 2014) is critical to successfully manage exploited fish stocks.

The most straightforward way to implement evolutionary perspective to stock management is through the use of genetic tools in stock assessments. Over the decades, a variety of genetic markers have been used to provide information on connectivity and boundaries between stocks of the same species or estimate overall genetic diversity - an indicator of population health. However, with the proliferation of next generation sequencing (NGS) methods that allow extensive genome-wide collection of several thousands of genomic markers, high-resolution insight into neutral and adaptive evolutionary signatures can now be obtained from natural populations. These genomic markers, called single nucleotide polymorphisms or SNPs, represent differences in a single nucleotide at a specific position in the DNA among individuals.

As an example, genomic studies were successful in showing that herring populations from the Atlantic stock might be locally adapted (Lamichhaney et al. 2012) and have revealed cryptic population structure in Australian stocks of mullet (Krück et al. 2013) and European hake (Milano et al. 2014). Others have shown correlations between specific genetic variants and environmental characteristics such as temperature and salinity, in the American lobster and herring, respectively (Limborg et al. 2012, Benestan et al. 2016). These and other studies laid a solid groundwork towards an evolutionary-based fisheries management, as they provided deep insights into both the connectivity and adaptation of exploited stocks.

Here, we employ NGS genotyping-by-sequencing (GBS) approach in order to study the genomic signature and stock structure of two exploited marine species in the Mediterranean Sea: the common octopus (*Octopus vulgaris*) and the red shrimp (*Aristeus antennatus*).

Across the Mediterranean, 17 sites where common octopus is targeted and 12 sites where red shrimp is frequently exploited were sampled for the analysis of evolutionary genomic endpoints.

Table 1 *Octopus vulgaris* sampling sites

No	Site	Code
1.	Alcúdia	AL
2.	Santanyí	SY
3.	Palamós	PL
4.	Tarragona	TR
5.	Llançà	LL
6.	Olhão	OH
7.	Garrucha	GR
8.	Dénia	DN
9.	Puglia Ionica	PI
10.	Mola di Bari	ML
11.	Lampedusa	LP
12.	Porto Torres	PT
13.	Su Pallosu	SP
14.	Fiumicino	FM
15.	Isola del Giglio	IG
16.	Vir	VR
17.	Vis	VS

Table 2 *Aristeus antennatus* sampling sites

No	Site	Code
1.	Alcúdia	ALa
2.	Sóller	SLa
3.	Santanyí	SYa
4.	Palamós	PLa
5.	Port de la Selva	PSa
6.	Garrucha	GRa
7.	Tarragona	TRa
8.	Patti	PTa
9.	Otranto	OTa
10.	Dénia	DNa
11.	Santa Pola	SAa
12.	Castelsardo	CSa

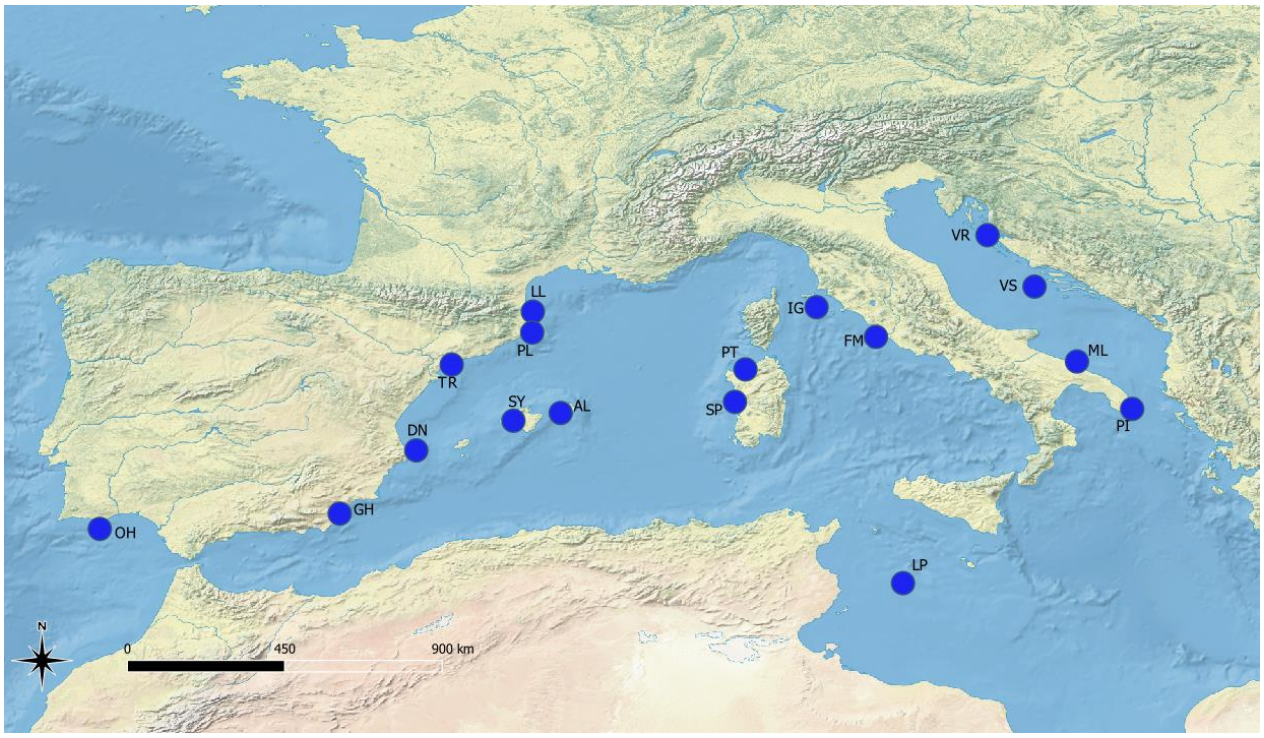

Figure 1 Geographic locations of analysed *Octopus vulgaris* sampling sites

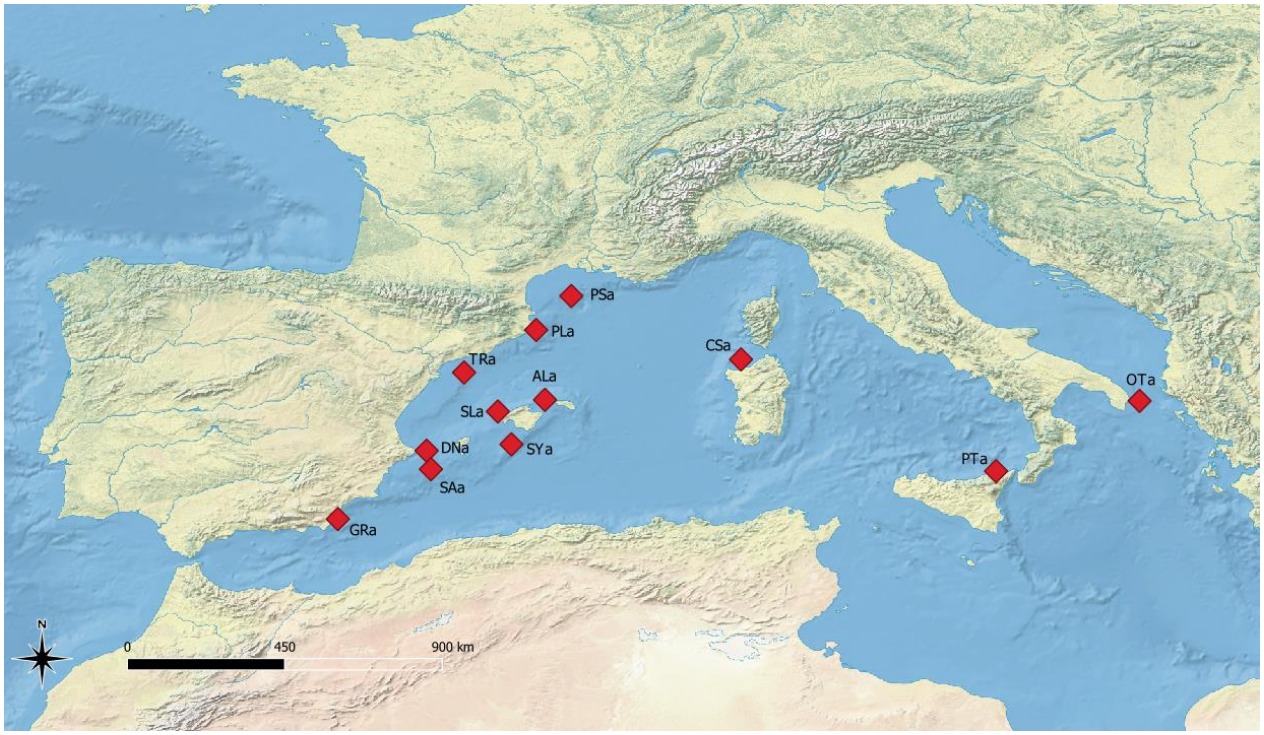

Figure 2 Geographic locations of analysed *Aristeus antennatus* sampling sites

2. Collection and analysis of phenotypic data

3.1. Morphometric measurements

A total of six morphometric measurements were taken for octopus and eight for shrimp individuals. Measurements were taken by researchers on the field, in the lab and in some cases by fisherman, with measuring ruler or calliper, and digital scales. Measurements considered in this study are directly related to individual´s overall metabolic processes, rapid locomotion or feeding behaviour, and are described in earlier morphological and genetic studies (Kapiris & Kavvadas 2009, Leite et al. 2008, Roper & Voss 1983, Sardà et al. 1995). Sex, as the most common cause of phenotypic variation among individuals in natural populations, was also determined for all sampled individuals.

Table 3 Morphometric measurements taken for *Octopus vulgaris* individuals

No	Abbreviation	Measurement	Definition
1	dML	Dorsal Mantle Length	Measured from midpoint between eyes to posterior end of mantle (0.1 cm)
2	vML	Ventral Mantle Length	Measured from midpoint between eyes to the anterior border of the mantle ventral midline (0.1 cm)
3	MW	Mantle Width	Greatest straight-line (dorsal) width of mantle (0.1 cm)
4	HW	Head Width	Greatest width of head at level of eyes (0.1cm)
5	FuL	Funnel Length	The length of the funnel from the anterior funnel opening to the posterior border measured along the ventral midline (0.1 cm)
6	TW	Total Weight	Total weight in g

Table 4 Morphometric measurements taken for *Aristeus antennatus* individuals

No	Abbreviation	Measurement	Definition
1	CL	Carapace Length	From the posterior margin of the orbit to the posterior margin of the carapace, parallel to the axis of the body (0.1 mm)
2	AL	Abdominal Length	Latero-dorsally from the anterior margin of the first abdominal segment to the posterior margin of the last abdominal segment (0.1 mm)
3	U.endo.	Uropodal Endopodite Lenght	From the margin of the commissure of the articulation to the end of the terminal spine on the dorsal ridge (0.1 mm)
4	U.exo.	Uropodal Exopodite Length	From the margin of the commissure of the articulation to the end of the terminal spine on the dorsal ridge (0.1 mm)
5	SL	Scaphocerite Length	From the margin of the commissure of the articulation to the end of the spine on the inner ridge (0.1 mm)
6	ArL	Article Lenght	The length of the *ischium*, *merus*, *carpus*, and *propodus* on the third right *pereiopod* (0.1 mm)
7	TL	Telson Lenght	From the insertion with the sixth abdominal somite to the distal tip (0.1 mm)
8	TW	Total Weight	Total body weight in g

3.2. Phenotypic analysis

The recorded morphometric values were corrected for total body size using dorsal mantle length as *a proxy* in octopus, and the sum of carapace, abdominal and *telson* length for shrimps. The values were then log transformed (scale of 10) to reduce skewness of the data before the analysis. Since the dataset showed non-normal distribution non-parametric tests were used for the analysis, Kruskal-Wallis test was employed to test for significance of morphological

differences between sites and sexes. Linear discriminant analysis (LDA) was performed to investigate the patterns of phenotypic variations among the populations. The analysis was conducted using MASS package in RStudio (R 3.4.3).

3.2.1. Octopus vulgaris

The analysis of octopus populations showed separation of several sites due to differences in individuals' morphological features. Predominantly, those sites were Vir (VR), Mola di Bari (ML) and Puglia Ionica (PI) – all sampled in the Adriatic Sea (Figure 3). The observed separation is mostly due to significant differences in mantle, head and funnel length in the individuals from the abovementioned sites. Curiously, there was no evidence for sexual dimorphism in any of the measurements taken (Kruskal-Wallis; p-value > 0.05).

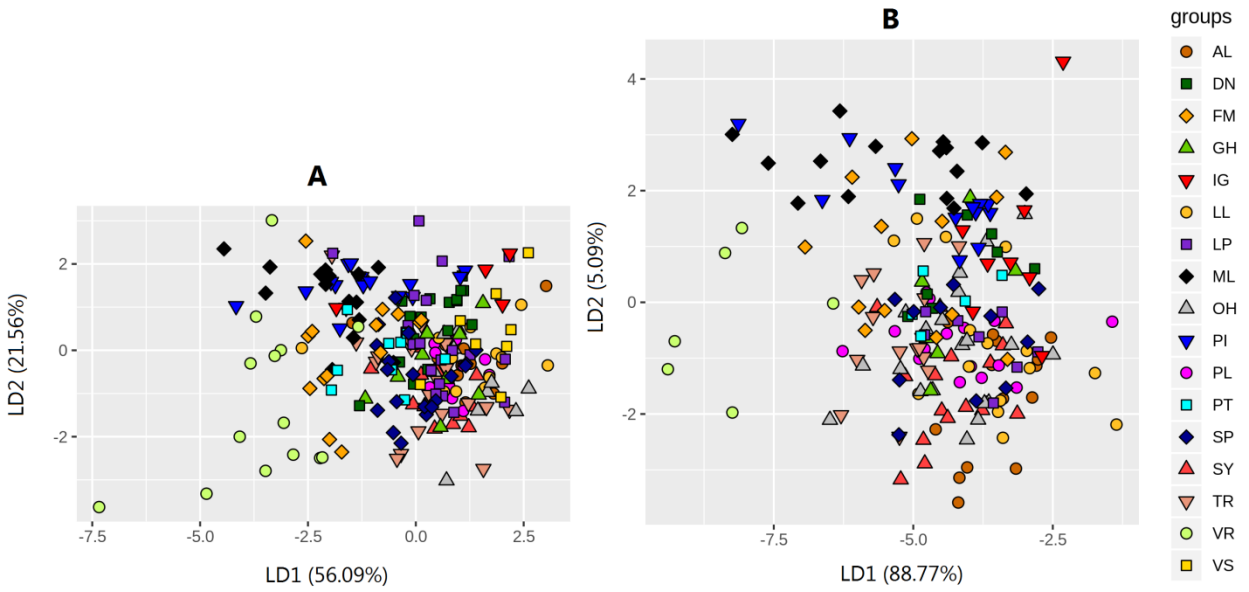

Figure 3 LDA analyses of morphometric measurements taken separately for A) female and B) male *Octopus vulgaris* individuals

3.2.2. Aristeus antennatus

For shrimp, the analysis of phenotypic diversity was performed on measurements recorded for the individuals from nine, out of twelve sampled sites. The shrimp locations showed significant differences between sexes for all measured variables (Kruskal-Wallis; p-value < 0.05). Additionally some locations were separated from the rest based on phenotypic differences in the LDA analysis (Figure 4). Both male and female individuals from Port de la Selva (PSa) were not grouped with the rest of the sites, as well as male shrimp sampled from Otranto (OTa), differences mostly stemming from significant disparity in carapace size between sites.

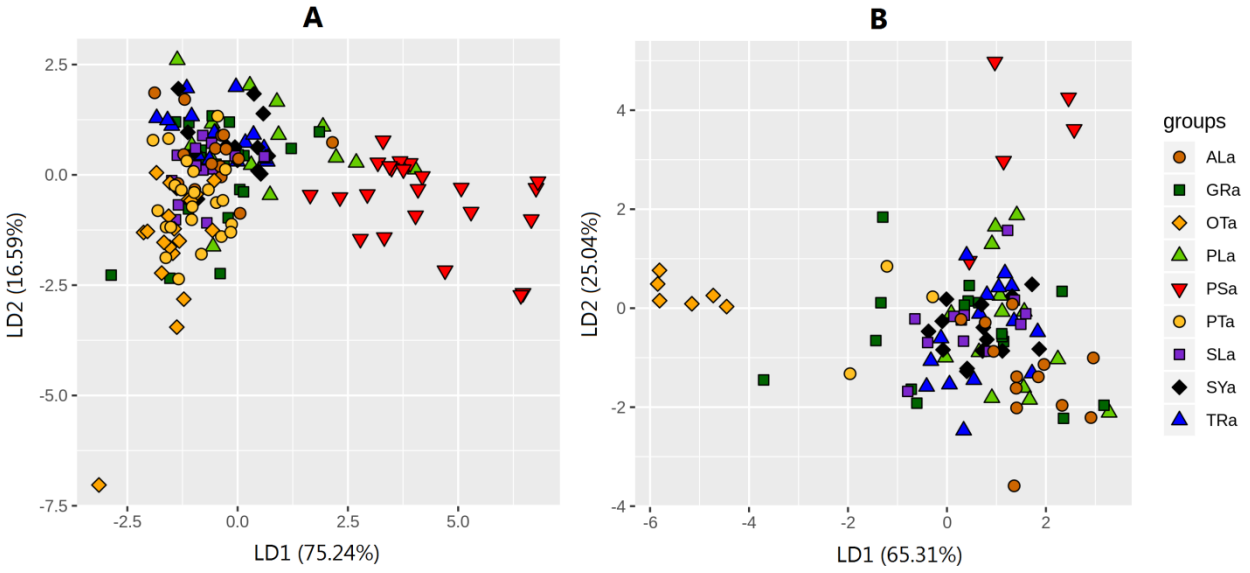

Figure 4 LDA analyses of morphometric measurements taken separately for A) female and B) male *Aristeus antennatus* individuals

4. Collection of genomic data

DNA samples for ~20 individuals per population were taken on the field or in the lab. Approximately 100 mg of arm tissue from octopus, and abdomen muscle tissue from shrimp individuals were harvested for genomic analysis. The tissue was stored in 96% ethanol and shipped to the Faculty of Science in Zagreb for further processing.

4.1. DNA extraction

A total of 384 samples from 17 locations were processed for octopus and 240 samples from 12 locations for shrimp. 15 - 20 mg of sampled tissue was flash frozen in liquid nitrogen and minced with scissors. Genomic DNA was then extracted with Sigma Aldrich-GenElute Mammalian Genomic DNA Miniprep Kit, using their provided protocol. The quality of extracted DNA was checked by agarose gel electrophoresis and the quantity of DNA was measured on a nanodrop spectrophotometer.

4.2. Library preparation

Library preparations were made according to modified GBS protocol from Parchman et al. (2012) and ddRAD protocol by Peterson et al. (2012).

Briefly, the extracted DNA was first digested with EcoR1 and MseI restriction enzymes, by mixing the samples with restriction mix (Table 5) and incubating them at 37 °C for 8 hours. Second, custom made EcoR1 adaptors containing 8 - 10 bp long barcodes that differed by a minimum of 4 bases, and a Y-shaped MseI adaptor (Table 7), were ligated on the digested DNA. In order to get the annealed, double-stranded adaptors, 100 µM stocks of single-stranded oligos were first mixed with nuclease free water, heated to 95 °C for 5 minutes and slowly cooled down to room temperature. So prepared, EcoR1 (final concentration 1 µM) and MseI (final concentration 10 µM) adaptor working stocks were added to ligation mix (Table 6), mixed with digested DNA and incubated at 16 °C for 6 hours. Lastly, the digest-ligation products were diluted up to 100 µL with 0.1X TE, and amplified in the PCR using Illumina PCR compatible primers (Table 7). For the reaction, 4 µL of digest-ligation products were mixed with the PCR mix (Table 8). PCR conditions included 98 °C for 30 s, followed by 16 PCR cycles (98 °C for 20 s; 60 °C for 30 s; 72 °C for 40 s) and a final extension at 72 °C for 10 min. The quality of PCR products was checked with an aliquot on an agarose gel, after which the samples were pooled together.

Table 5 Restriction mix used in library preparation

Chemical	1 sample (µL)
10X T4 Buffer	1.15
1 M NaCl	0.60
1 mg/mL BSA	0.60
H_2O	0.25
Mse1 enzyme	0.12
EcoR1 enzyme	0.28

Table 6 Ligation mix used in library preparation

Chemical	1 sample (µL)
EcoR1 adaptor	1
Mse1 adaptor	1
H_2O	0.072
10X T4 buffer	0.1
1M NaCl	0.05
1mg/µL BSA	0.05
T4 ligase	0.1675

Table 7 Sequences of adaptors and primers used in the library preparations. Barcodes imbedded in EcoR1 adaptors are marked with red X.

Oligo name	Sequence 5' 3'
EcoR1_1	AATTGXXXXXXXXXAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
EcoR1_2	CTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXXXXXC
Mse1_1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Mse1_2	TAAGATCGGAAGAGCGAGAACAA
PCR1	A*A*TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PCR2	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC

Table 8 PCR mix used in library preparation

Chemical	1 sample (µL)
H_2O	9.67
5X Iproof buffer	4.0
dNTP (10 mM)	0.4
$MgCl_2$ (50 mM)	0.4
Primers (2.5 µM each)	1.33
Iproof taq	0.2
DMSO	0.15

4.3. Sequencing

Obtained PCR products were sent to BGI Tech Solutions company in Hong Kong for pair-end genotyping by sequencing (GBS) on Illumina HiSeq X 10 platform. The company provided quality control, 250 - 450 bp gel size selection and sequencing with 40% PhiX, along with first data filtering and demultiplexing of produced reads.

5. From raw reads to SNPs

Around 60 GB of raw data were obtained by GBS sequencing. The sequencing reads were delivered in FastaQ format and saved on a hard disc and a server in the Department of Biology at the Faculty of Science in Zagreb for further processing.

5.1. Cleaning and filtering

The raw reads were trimmed of residual adaptor and/or barcode contamination and standardized to 100 bp length using custom made Perl scripts. The quality check and trimming of low quality reads was done using custom made Perl scripts and *process_radtags.pl* script in Stacks 2.2 (Catchen et al. 2013) by specifying input pair-end reads and their format, used restriction site enzymes, and *--clean*, *--quality*, and *–rescue* flags in program options.

5.2. *Octopus vulgaris*

5.2.1. Genome assembly

The assembly of octopus genome was done using *RADAssembler* pipeline (Li et al. 2018). This assembly method utilizes Stacks produced contigs and reassembles then into scaffolds. These are further embedded into 1000 artificial sequences called colloquially "pseudoscaffolds". The assembly was done in several steps: 1.- Number of mismatches within individual was estimated by running *chooseM* program within the pipeline on three randomly chosen samples with maximum mismatch set to 10 and minimum stack depth to 2; 2.- *ustacks* was run on a small subset of randomly chosen samples i.e. 3 per population, with the optimized mismatch value obtained

in the first step (m = 2) and maximum mismatch set to 10; 3.- *chooseN* program was run within the pipeline on the *ustacks* output to optimize mismatch across individuals; 4.- Denovo assembly was done by using the optimized similarity thresholds (m=2, n=2) with minimum read depth of a locus to export for assembly option set to range from 10 to 100.

5.2.2. Mapping the genome

The mapping of octopus reads was done via Bowtie2 (version 2.3.4.1), a short read aligner program that enables the aligning of large sets of short DNA sequence reads to large genomes (Langmead & Salzberg 2012). The genome was first indexed using *bowtie2,* program build-in within the software, using previously obtained pseudoscaffolds as a reference and the default settings. The reads were then mapped with Bowtie2 program using the default settings and previously obtained Bowtie index.

5.2.3. SNP calling

The mapped reads were then run through Stacks 2.2 *ref_map.pl* pipeline (Catchen et al. 2013). First, loci were assembled according to the alignment positions provided for each read, and SNPs called in each sample with *gstacks* program. Second, by enabling *populations* program, the SNPs were filtered according to their quality and position, and the population-level summary statistics were generated. The program parameters included: restricting the data analysis to one random SNP per locus; setting the minimum minor allele frequency required to process a nucleotide site at a locus to 0.05; minimum percentage of individuals in a population required to process a locus for that population set to 0.2; minimum number of populations a locus must be present in to process a locus set to 17 (100%); and kernel-smoothed enabled. 79134 single nucleotide polymorphisms (SNPs) were identified in the dataset in total.

5.3. *Aristeus antennatus*

5.3.1. De novo assembly and SNP calling

The filtered and cleaned reads from shrimp populations were assembled and mapped in *denovo_map.pl* pipeline in Stacks 2.2 (Catchen et al. 2013). First, *ustacks* program was run on all shrimp samples to build loci and call SNPs de novo. Second, a catalog of all loci across the population was created by running *cstacks* program on a small subset of samples (random three per population), and all samples were then matched against the catalog using *sstacks* program. Next, *gstacks* program was run to assemble and call SNPs in each sample. Last, *populations* program was run to filter the SNPs according to their quality and position, and generate population-level summary statistics. The program parameters included: restricting the data analysis to one random SNP per locus; setting the minimum minor allele frequency required to process a nucleotide site at a locus to 0.05; minimum percentage of individuals in a population required to process a locus for that population set to 0.2; minimum number of populations a locus must be present in to process a locus set to 9 (100%); and kernel-smoothed enabled. 35552 single nucleotide polymorphisms (SNPs) were identified in the dataset in total.

6. Patterns of genomic variation across spatial scales

6.1. Genomic markers – SNPs

Tens of thousands of genomic markers - single nucleotide polymorphisms (SNPs), were identified during DNA sequence read processing for both species. Principal component analysis (PCA) of obtained SNP datasets, produced using *Adegenet* package in R (3.4.3), showed that there exists a separation of several octopus and shrimp populations based on variability of their genomic signature. For instance, there is a clear separation of Atlantic Olhão (OH), as well as a subtler separation of Adriatic Vir (VR), Vis (VS), Mola di Bari (ML) and Puglia Ionica (PI) for octopus populations (Figure 5). For shrimp populations, no separation was shown by analysing the variability of genomic markers in the PCA analysis (Figure 6).

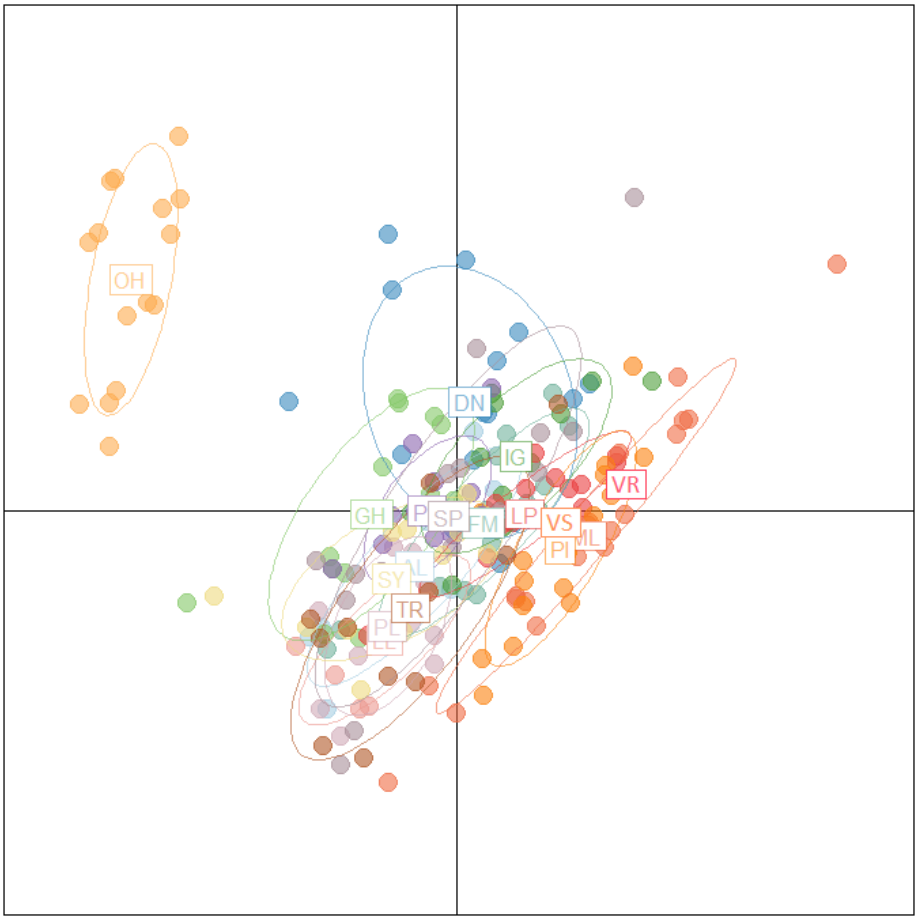

Figure 5 PCA done on genomic markers (SNPs) identified for octopus populations in the Mediterranean

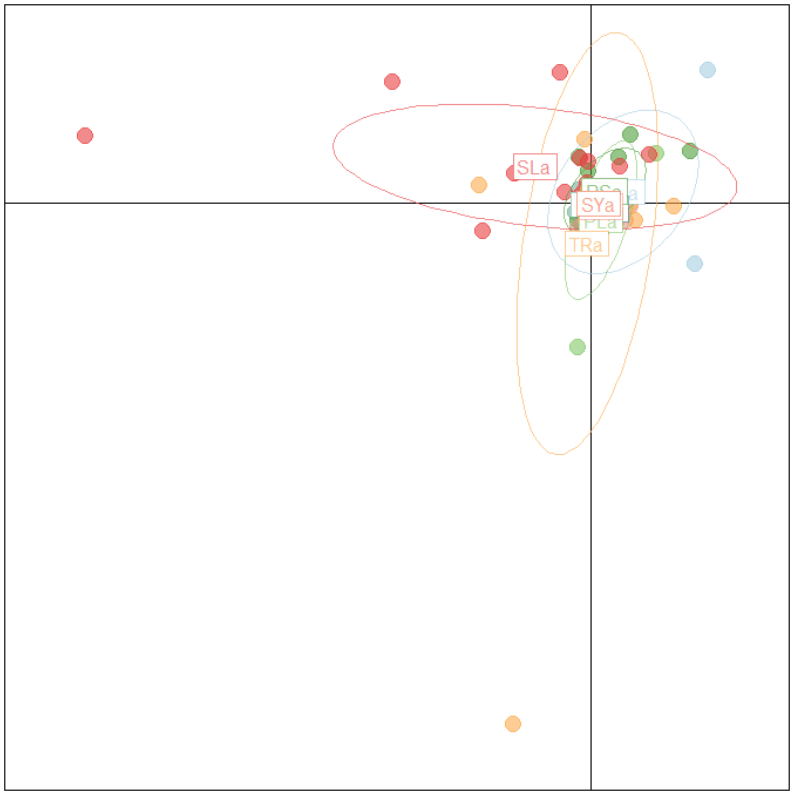

Figure 6 PCA done on genomic markers (SNPs) identified for red shrimp populations in the Mediterranean

6. Discussion and conclusions

Phenotypic analysis of both species revealed certain differences between the populations, but without any strong pattern. These differences could be the result of phenotypic plasticity and acclimatization, but may partly also be due to observed genomic variability. Additionally, some errors in the phenotypic data might have occurred on account of a large number of people with different skill levels participating in collection of morphological data itself.

However, the genomic methods used in this study proved to be fast, relatively inexpensive and efficient, as was evident by good quality of prepared libraries and obtained sequenced reads. For shrimp populations lower coverage and read depth was obtained than for octopus populations. This discrepancy in generated sequences between two species, which resulted in smaller SNP count for shrimp, could be the consequence of shrimp genome characteristics in terms of distribution of cutting sites, or due to cellular components affecting DNA extraction quantity and quality. Still, the described methods provided a large number of genomic markers for both species.

The obtained SNPs were further analysed, checked for quality and trimmed, and used as markers for population's genomics analysis (Deliverable 3.5.2).

7. Literature

Benestan, L. et al., 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). *Molecular ecology*, 25(20), pp.5073–5092.

Catchen, J. et al., 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), pp.3124–3140.

Conover, D.O. & Munch, S.B., 2002. Sustaining fisheries yields over evolutionary time scales. *Science*, 297(5578), pp.94–96.

Eizaguirre, C. & Baltazar-Soares, M., 2014. Evolutionary conservation-evaluating the adaptive potential of species. *Evolutionary Applications*, 7(9), pp.963–967.

Kapiris, K. & Kavvadas, S., 2009. Morphometric study of the red shrimp Aristeus antennatus in the Eastern Mediterranean. *Aquatic Ecology*, 43(4), pp.1061–1071.

Krück, N.C., Innes, D.I. & Ovenden, J.R., 2013. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (Mugil cephalus). *Molecular Ecology Resources*, 13(4), pp.715–725.

Lamichhaney, S. et al., 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, 109(47), pp.19345–19350.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), pp.357–359.

Leite, T.S. et al., 2008. Morphological and genetic description of Octopus insularis, a new cryptic species in the Octopus vulgaris complex (Cephalopoda: Octopodidae) from the tropical southwestern Atlantic. *Journal of Molluscan Studies*, 74(1), pp.63–74.

Li, Y.L. et al., 2018. An optimized approach for local de novo assembly of overlapping paired-end RAD reads from multiple individuals. *Royal Society Open Science*, 5(2), p.171589.

Limborg, M.T. et al., 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (Clupea harengus). *Molecular Ecology*, 21(15), pp.3686–3703.

Milano, I. et al., 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (Merluccius merluccius). *Molecular Ecology*, 23(1), pp.118–135.

Olsen, E.M. et al., 2004. Maturation trends indicative of rapid evolution preceded the collapse of northern cod. *Nature*, 428(6986), pp.932–935.

Parchman, T.L. et al., 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, 21(12), pp.2991–3005.

Peterson, B.K. et al., 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species L. Orlando, ed. *PLoS ONE*, 7(5), p.e37135.

Roper, C.F.E. & Voss, G.L., 1983. Guidelines for taxonomic descriptions of cephalopod species. *Memoirs of the National Museum of Victoria*, 44(44), pp.49–63.

Sardà, F., Bas, C. & LLeonart, J., 1995. Functional morphometry of Aristeus antennatus (Risso, 18 16) (Decapoda, Aristeidae). *Crustaceana*, 11(1981), pp.461–471.