Deliverable Report

Deliverable No:	D2.4		**Delivery Month:**	36
Deliverable Title	Construction of a genetic linkage map in meagre			
WP No:	2		**WP Lead beneficiary:**	P3. IRTA
WP Title:	Reproduction and Genetic-meagre			
Task No:	2.5		**Task Lead beneficiary:**	P1. HCMR
Task Title:	Development of Single Nucleotide Polymorphisms (SNP) marker tools for the genetic characterization of fast and slow growers			
Other beneficiaries:				
Status:	Delivered		**Expected month:**	36
………..				

Lead Scientist preparing the Deliverable: Tsigenopoulos, C.S. (HCMR).

Other Scientists participating: Manousaki, T. (HCMR), Tsakogiannis, A. (HCMR), Lagnel, J. (HCMR), Duncan, N. (IRTA).

Objective: The objective of this Deliverable was to construct a genetic linkage map in meagre (*Argyrosomus regius*): Single Nucleotide Polymorphisms (SNPs) were identified using ddRAD (double-digest Restriction site Associated DNA) sequencing and were used in a genetic linkage mapping analysis to build the first linkage map of the species, and further investigate whether they have significant association with growth (see ***D2.5 identification of genetic markers related to growth for use in marker assisted breeding programs for meagre through QTL map***).

Description: Genetic maps in fish are the result of estimations that take into account the inheritance patterns of genetic markers and are used for a range of evolutionary, ecological, aquaculture and biomedical applications (Rexroad et al., 2008). In the past, the construction of genetic linkage maps was initiated with markers initially developed for population genetic analyses such as microsatellites, and more recently with the addition of markers such as AFLPs (Amplified Fragment Length Polymorphism) and SNPs.

Linkage maps are valuable tools for the investigation of the genetic basis of complex traits in farmed animal species. For several decades, linkage maps have enabled the mapping of quantitative trait loci (QTL), and formed the basis of attempts at positional cloning of these QTL in both terrestrial (Goddard and Hayes, 2009) and aquatic farmed species (Danzmann and Gharbi, 2001). High throughput sequencing technologies have now expedited the discovery of millions of SNP markers (Liu, 2010). These SNPs form the basis of modern, high-resolution genetics studies, and underpin genomic selection for faster genetic improvement in terrestrial livestock and laterally for aquaculture breeding programs (Meuwissen et al., 2001; Goddard et al., 2010; Sonesson, 2010; Yáñez et al., 2014, 2015). Scoring of genome-wide SNPs in large populations is achieved either through genotyping by sequencing (Davey et al., 2011), or by creation and application of SNP arrays (e.g., Houston et al., 2014; Tsai et al., 2016). High density linkage maps based on these SNP

datasets can aid in high resolution mapping of loci underpinning complex traits in farmed animals, improvements in assembly of reference sequences, and knowledge of the recombination landscape of the genome.

In this deliverable, we report the construction of the first linkage map for the meagre, based on SNPs. Our primary objective was to provide a powerful tool and facilitate future selective breeding for traits associated with aquaculture production efficiency (growth, sex, food conversion ratio, disease resistance, etc), which is considered as a key to reduce production costs and to promote continued aquaculture industry expansion (Garber and Sullivan, 2006). However, traditional selection is time-consuming and labor-intensive. Marker-assisted selection (MAS) is becoming popular in crop and aquaculture breeding programs due to its high efficiency. Construction of a linkage map and mapping of trait-related quantitative trait loci (QTL) is required as a foundation for implementation of MAS.

Material and Methods

Biological material

On January 20th 2016, 400 meagre fish were sampled from a large fish-cage that formed part of a commercial farm site on the Spanish coast in the community of Valencia. The juveniles were from the largest grade of fish that came from the same group of spawns collected from a broodstock that contained 19 breeders (8 females and 11 males) that were injected with GnRHa to induce spawning. Total length and weight was measured for all 400 sampled fish.

The DNA was extracted from all fish using standard protocols and genotyped for the 10 loci multiplex (Soula et al., 2011) used in Task 2.1 (UBA005, UBA006, UBA042, UBA050, UBA053, UBA054, UBA853, Soc405, Soc431, and Cacmic14) using the Qiagen multiplex PCR kit. Results were evaluated with FAP software (Taggart, 2007) to infer parentage of those 400 fish based on the parental genotypes of the 19 breeders.

Construction and sequencing of ddRAD library

The ddRAD library preparation protocol was based on the methodology originally reported by Peterson et al. (2012). The modified protocol used here is essentially the one described in Manousaki et al. (2016), with some alterations indicated below. Briefly, each of the 144 DNA samples in each of the libraries (1st Library: 2 parents in triplicates, 2 in quadruplicates and 130 offspring DNA samples, 2nd Library: 5 parents in triplicates and 129 offspring DNA samples; 21ng DNA per sample) was simultaneously digested by two high fidelity restriction enzymes (RE): *Sbf*I (CCTGCA|GG recognition site), and *Sph*I (GCATG|C recognition site), both sourced from New England Biolabs, (NEB) UK. Digestions were incubated at 37ºC for 50 min, using 20 U of each enzyme per microgram DNA in 1x CutSmart Buffer (NEB), in a 6 µl total reaction volume. Barcoded adapters were designed such that adapter– genomic DNA ligations did not reconstitute RE sites, while residual RE activity limited concatemerization of genomic fragments. After cooling the reactions to room temperature, 3 µl of a premade adapter mix was added to the digested DNA, and incubated at room temperature for 10 min. This adapter mix comprised individual-specific combinations of P1 (*Sbf*I-compatible) and P2 (*Sph*I-compatible) adapters at 6 nM and 72 nM concentrations respectively (1:12 P1 to P2 adapter ratio), in 1x reaction buffer 2 (NEB). The P1 and P2 adapters included an inline five- or seven- base barcode for sample identification. Ligation was performed over 3 hr at 22ºC by addition of a further 3 µl of a ligation mix comprising 4 mM rATP (Promega, UK) and 2000 cohesive-end units per microgram of T4 ligase (NEB) in 1x CutSmart buffer (NEB). The ligated samples were then heat denatured

at 65ºC for 20 min, cooled and combined into a single pool. The pooled sample was column-purified (MinElute PCR Purification Kit, Qiagen, UK) and eluted in 70 µl EB buffer (Qiagen, UK). Size selection of fragments ranged from approximately 400 bp to 700 bp. Following gel purification (MinElute Gel Extraction Kit, Qiagen, UK), the eluted size-selected template DNA (68 µl in EB buffer) was PCR amplified (15 cycles PCR; 36 separate 12.5 µl reactions, each with 1 µl template DNA) using a high fidelity Taq polymerase (Q5 Hot Start High-Fidelity DNA Polymerase, NEB). The PCR reactions were combined (450 µl total), and column-purified (MinElute PCR Purification Kit) and eluted in EB buffer, in a total volume of 33 µl for the first library and 31 µl for the second.

Finally, each ddRAD library was first test-sequenced at the Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC) of HCMR in Crete using two runs of an Illumina MiSeq (v2 chemistry, 300 cycle kit, 162 bp paired end reads). The two libraries were finally shipped to NSC in Norway (subcontract) to get sequenced on 2 lanes of HiSeq4000 and the 150bp PE option.

Data analysis of ddRAD library

The raw sequence data produced were analysed in STACKS v1.40 (Catchen et al., 2013). Each library was analysed separately. Quality control, filtering for ambiguous barcodes and restriction sites, and demultiplexing took place using the script process_radtags (options -c -q -r). The unique RAD loci and SNPs of each individual were identified with the script *ustacks*. Secondary reads were not used for genotype calling to reduce possible genotypic errors (option -H). Then, the four parental samples were used to build a catalogue of loci with the script *cstacks* using default parameters. The progeny loci were afterwards matched to the catalogue loci with the script *sstacks*. A MySQL database was built and the output files were loaded with the script *load_radtags.pl*. The database was indexed with the script *index_radtags.pl*. Finally, the genotypes of the parental loci with more than 50 progeny genotyped were selected and extracted using the script *genotypes*.

Building the linkage map

The genotype data extracted from STACKS, were edited and transformed to *linkage* format with custom perl scripts. The linkage mapping software LepMap2 (Rastas et al., 2016) was selected for building the map, as it is designed to accommodate data from multiple full-sib families such as the data of this deliverable. First, the *linkage* files were filtered for loci with segregation distortion of the Mendelian ratios (p - value 0.01) using the module *Filtering*. The surviving markers were assigned to different linkage groups (LG) with the module *SeparateChromosomes*. The assignment was done by computing all pairwise LOD (logarithm of the odds) scores between markers using a LOD score limit 7. Then, the markers within each LG were ordered using the module *OrderMarkers* by maximizing the likelihood of the data given the order. The constructed map was visualised in MapChart (Voorrips, 2002).

Comparative genome mapping

The mapped RAD loci in meagre were used in a comparative analysis with the genome from the European seabass (*Dicentrarchus labrax*, http://seabass.mpipz.de) to confirm that the mapping process was successful. Meagre RAD loci sequences were extracted from the MySQL database built through Stacks pipeline and were used in BLASTN sequence similarity searches against the genome of the European sea bass (version dicLab v1.0c, *e*-value threshold 10^{-9}). The top hit per sequence was kept and considered homologous to the RAD locus. Then, we determined for each meagre's LG the chromosome of the reference species that

corresponded to that LG based on the similarity search. If the majority of the loci of a meagre LG were homologous to loci from a single chromosome of a reference species, we considered them homologous chromosomes.

Results and Discussion

Parentage assignment

The fish were genotyped in order to infer parentage allocations. Parentage was based on nine loci since locus UBA053 was excluded from the analysis. Single parentage assignment (match) was successfully described for 345 of those fish (86.25%) and the rest had multiple matches. Fish belonged to 17 families (out of the 88 theoretically expected). Only 5 out of the 8 females were identified as probable parents of the offspring; females No 403F and 404F had participated the most, and to a smaller extend females 391F and 406F. Likewise, six out of the 11 males were identified as probable parents of the offspring; Male No 405M seems to be responsible for nearly half of the offspring followed by males 397M, No 388M and No 402M (see **Table 1**).

Table 1. Parentage assignment in the meagre (*Argyrosomus regius*) stock.

		Males						
		388M	397M	398M	401M	402M	405M	Total
Females	391F	10	3	3	7	10	25	58
	394F						1	1
	403F		77			3	24	104
	404F	29				6	95	130
	406F	1	8			3	40	52
	Total	40	88	3	7	22	185	

Sequencing outcome for the ddRAD library

Sequencing of the constructed ddRAD libraries yielded 765,712,194 and 788,654,246 total raw reads. Following demultiplexing and quality control we obtained 542,447,568 and 566,148,132 reads for the two libraries respectively. For parental samples, we obtained ~6 million reads on average, while for progeny samples we got ~4 million reads on average (see **Supplementary Table 1** for details on all samples). Following STACKS analysis, we obtained in total 87,522 ddRAD loci in the parental samples which constituted the ddRAD catalogue. The progeny contained on average 19,402 ddRAD loci, and the number of SNPs identified seems to be proportional to the sequencing depth (reads sequenced per fish, see **Figure 1**). For details, see **Supplementary Table 2** for summary on the ddRAD loci and SNPs discovered.

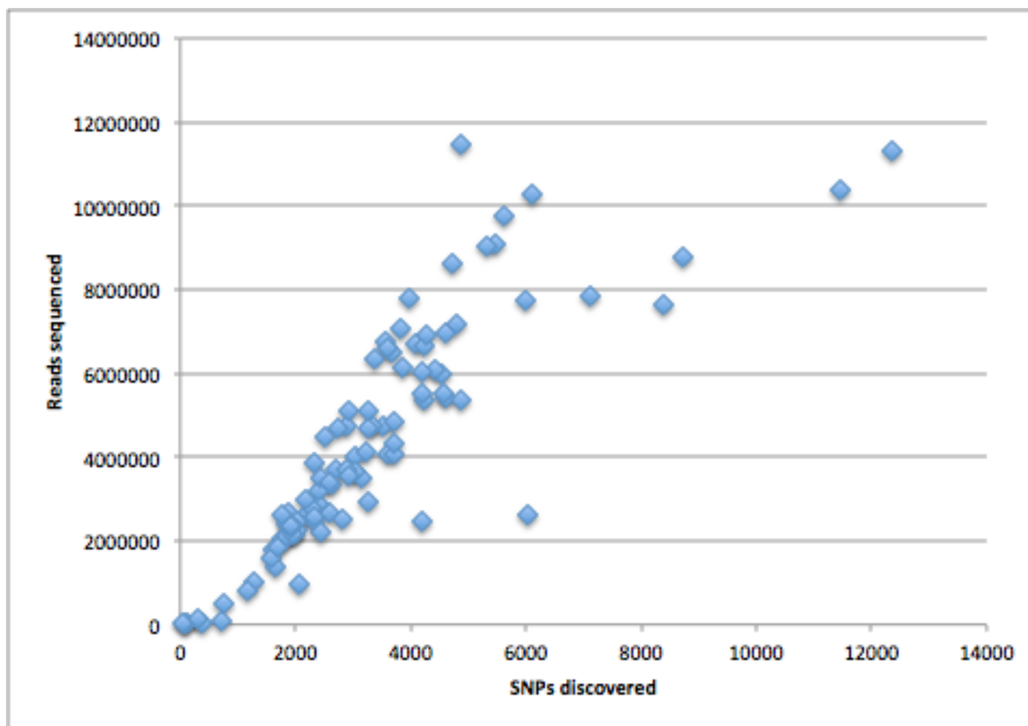

Figure 1. The discovered SNPs relative to the sequenced depth.

The first genetic linkage map of meagre

Following the ddRAD analysis in STACKS, the two largest families were chosen for building the linkage map. From both families, 1,013 markers survived the analysis criteria and were exported for the construction of the first meagre linkage map. Following filtering of the markers that deviated significantly from the Mendelian ratios, 61 markers were excluded from family 403F&397M and 69 from family 404F&405M. In total, 950 markers were maternally informative, while 929 were paternally informative summing up to 1,008 total markers used to build a sex-averaged linkage map.

For the linkage map construction, a range of LOD score 5-8 was used. An LOD score 7 was selected based on the number of LGs produced and the accuracy of the linkage mapping estimated through the comparative genomic analysis with European seabass. When LOD score was lowered, some large meagre LGs were fused but not the smaller ones expected to be fused when additional data will be available (see below).

Karyotype analysis of meagre indicated a haploid set of 24 chromosomes (Soares et al., 2012; Merlo et al., 2013). The constructed linkage map included 731 markers organised in 27 linkage groups (**Figure 2**), that means 3 LGs-chromosomes more than the haploid number determined in the karyotype of this species; it is anticipated that when more markers are added, some LGs will merge into larger ones, and the number of LGs should eventually match the number of chromosomes. The LGs range from 14,001 to 67,333 cM in length and the whole map spans 1,158.299 cM. The loci included in the LGs range from 11 to 54 ddRAD loci per LG (see **Supplementary Table 3** for further details).

Figure 2. The genetic linkage map of meagre (*Argyrosomus regius*).

The mapped loci were used for a comparative genomic analysis against the genome of European seabass. The similarity search revealed that more than one third (36.66%) and 268 out of 731 loci have a homologous region in the genome of European seabass. The homologous loci revealed extensive conservation of synteny (regions or fragments are conserved of order within two sets of chromosomes that are being compared with each other) between the two species (**Table 2**). Out of 268 markers, 13 were mapped on un-identified (UN) European seabass regions. From the remaining 255 markers, 14 more did not follow a one-to-one pattern against the European seabass LGs/chromosomes.

Table 2. Comparative mapping results of the meagre (*Argyrosomus regius*) linkage map against the European sea bass (*Dicentrarchus labrax*) genome. The meagre markers with significant BLAST hits in European sea bass are presented, and the putative systemic pairs are shown in light grey along the diagonal.

	***Dicentrarchus labrax* LGs**																									
***Argyrosomus regius* LGs**	5	7	10	20	22/25	13	24	19	18/21	4	11	9	12	X	15	16	3	1A	14	8	1B	6	2	17	UN	Sum
1	16																	1								17
2		14																							2	16
3			7									1						1							1	10
4				15																					1	16
5					9																				2	11
6						9																				9
7					1		10																			11
8	1							8																	1	10
9									6				1												2	9
10										13																13
11											13															13
12												14														14
13													11													11
14														12												12
15															13											13
16																7						1				8
17																	10									10
18																		8							1	9
19			2																8							10
20																				9					1	10
21		1	1																		6					8
22																						3				3
23																							4		1	5
24																				6						6
25			1													1									1	3
26																						6				6
27															1									4		5
Sum	17	15	11	15	10	9	10	8	6	13	13	15	12	12	14	8	10	10	8	15	6	10	4	4	13	268

Based on the comparison of the meagre linkage map against the European seabass genome map (Tine et al., 2014), a reduction of the number of LGs to 24 is possible. Fusion of LG20 with LG24 is suggested since both map to seabass LG8. Likewise, LG22 should join LG26, as markers from both LGs are found on the same seabass LG6. Lastly, the LG25 might not be supported since it comprises markers from different seabass LGs and could potentially in the future disintegrate when linkage map are enriched.

Deviations: Due to recent technological and research advances concerning molecular methods to genotype and genetically characterize fish (or organisms in general) and some constraints that are related to the biology of the meagre (*Argyrosomus regius*), we proposed (Amendment 3, Nov 2016) to modify the DOW for Task 2.5, in terms of (a) the method to be used and (a) the time-schedule. Because it is the first time that these methods were going to be applied in meagre we believed that we needed a 6-month extension of the associated deliverables, from Mo 30 to 36). No change in budget allocation, staff effort or the number of deliverables has resulted from this modification.

References:

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait, 2011 Stacks: building and genotyping loci de novo from short-read sequences. G3-Genes Genomes Genetics 1: 171-182.

Danzmann, R. G., and K. Gharbi, 2001 Gene mapping in fishes: a means to an end. Genetica 111: 3–23.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen et al., 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12: 499–510.

Goddard, M. E., and B. J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381–391.

Liu, Z. J., 2010 Next Generation Sequencing and Whole Genome Selection in Aquaculture, Blackwell Publishing Ltd., Des Moines, IA., 10.1002/9780470958964.

Manousaki, T., A. Tsakogiannis, J. B. Taggart, C. Palaiokostas, D. Tsaparis, J. Lagnel, D. Chatziplis, A. Magoulas, N. Papandroulakis, C. C. Mylonas, and C. S. Tsigenopoulos. 2016. Exploring a Nonmodel Teleost Genome Through RAD Sequencing—Linkage Mapping in Common Pandora, *Pagellus erythrinus* and Comparative Genomic Analysis. G3: Genes|Genomes|Genetics 6:509-519.

Merlo, M., Cross, I., Rodríguez-Rúa, A., Manchado, M., Rebordinos, L. 2013 First approach to studying the genetics of the meagre (*Argyrosomus regius*; Asso, 1801) using three multigene families. Aquac Res. 2013; 44: 974–984.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7:e37135.

Rastas, P., F. C. F. Calboli, B. Guo, T. Shikano, and J. Merilä. 2016. Construction of Ultradense Linkage Maps with Lep-MAP2: Stickleback F2 Recombinant Crosses as an Example. Genome Biology and Evolution 8:78-93.

Soares, F., A. Leitão, M. Moreira, J. T. de Sousa, A. C. Almeida, M. Barata, S. W. Feist, P. Pousão-Ferreira, and L. Ribeiro. 2012. Sarcoma in the thymus of juvenile meagre *Argyrosomus regius* reared in an intensive system. Diseases of Aquatic Organisms 102:119-127.

Sonesson, A. K., 2010 Genomic selection for aquaculture: principles and procedures, pp. 151–163 in Next Generation Sequencing and Whole Genome Selection in Aquaculture, edited by Z. (J.) Liu. Blackwell Publishing, Des Moines, IA.

Soula, M., Zamorano, M. J., Navarro, A., Sánchez, J.J., Neil, D., Alejandro, G., Afonso, J. M. 2011. Diseño de dos nuevas PCRs múltiplex para corvina (*Argyrosomus regius*), Congreso National de Acuicultura, Barcelona España 2011.

Taggart, J. B. 2007. FAP: an exclusion-based parental assignment program with enhanced predictive functions. Molecular Ecology Notes 7:412-415.

Tine, M., H. Kuhl, P. A. Gagnaire, B. Louro, E. Desmarais et al., 2014 European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5: 5770.

Tsai, H. Y., D. Robledo, N. R. Lowe, M. Bekaert, J. B. Taggart, J. E. Bron, and R. D. Houston. 2016. Construction and Annotation of a High Density SNP Linkage Map of the Atlantic Salmon (*Salmo salar*) Genome. G3: Genes|Genomes|Genetics 6:2173-2179.

Voorrips, R. E., 2002 MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93: 77-78.

Yáñez, J. M., R. D. Houston, and S. Newman, 2014 Genetics and genomics of disease resistance in salmonid species. Front. Genet. 5: 415.

Yáñez, J. M., S. Newman, and R. D. Houston, 2015 Genomics in aquaculture to better understand species biology and accelerate genetic progress. Front. Genet. 6: 128.