Muscle and liver transcriptome characterization and genetic marker discovery in the farmed meagre, *Argyrosomus regius*

Manousaki, T.[1], Tsakogiannis, A.[1,2], Lagnel, J.[3], Kyriakis D.[1,4], Duncan, N.[5], Estevez, A.[5], Tsigenopoulos, C. S.[1*]

[1] *Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Greece*

[2] *Department of Biology, University of Crete, Heraklion, Greece*

[3]*INRA PACA, UR 1052 GAFL, 67 Allée des Chênes, CS 60 094, 84143 Montfavet Cedex, France*

[4] *School of Medicine, University of Crete, Heraklion, Greece*

[5] *IRTA, Sant Carles de la Rapita, Tarragona, Spain*

**Author for Correspondence: Costas S. Tsigenopoulos, Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Greece, tel +30 2810337854, fax +30 2810337870, tsigeno@hcmr.gr*

Abstract

Meagre (*Argyrosomus regius*), a teleost fish of the family Sciaenidae, is part of a group of marine fish species considered new for Mediterranean aquaculture representing the larger fish cultured in the region. Meagre aquaculture started ~25 years ago in West Mediterranean, and the supply of juveniles has been dominated by few hatcheries. This fact has raised concerns on possible inbreeding, urging the need for genetic information on the species and for an assessment of the polymorphisms found in the genome. To that end we characterized the muscle and liver transcriptome of a pool of meagre individuals, from different families and

phenotypic size, to obtain a backbone that can support future studies regarding physiology, immunology and genetics of the species. The assembled transcripts were assigned to a wide range of biological processes including growth, reproduction, metabolism, development, stress and behavior. Then, to infer its genetic diversity and provide a catalogue of markers for future use, we scanned the reconstructed transcripts for polymorphic genetic markers. Our search revealed a total of 42,933 high quality SNP and 20,581 STR markers. We found a relatively low rate of polymorphism in the transcriptome that may indicate that inbreeding has taken place. This study has led to a catalogue of genetic markers at the expressed part of the genome and has set the ground for understanding growth and other traits of interest in meagre.

Keywords: aquaculture, RNASeq, SNPs, STRs

Introduction

The meagre, *Argyrosomus regius* (Asso y del Rio 1801) is a teleost fish that belongs to the family Sciaenidae and is widely distributed along the eastern Atlantic Ocean coast and the entire Mediterranean Sea (Chao, 1986). Throughout the distribution, meagre holds an important role in fisheries and now represents one of the newly emerging and promising aquaculture species across the Mediterranean region. There appears to be few fast growing large aquaculture species in the Mediterranean region and meagre together with greater amberjack (*Seriola dumerili*) fill this niche. Meagre aquaculture started in late nineties in France and Italy and since then has expanded in other European countries (FAO, 2015). Interestingly, meagre fry production has been for years carried out through a single hatchery in France (Monfort 2010), a fact that raises concerns regarding the genetic diversity of the European aquaculture stocks and requires evaluation.

Coupled with the increasing interest in the aquaculture industry, meagre is being explored in various fields, such as reproduction and broodstock management (DUNCAN *et al.* 2012; MYLONAS *et al.* 2015) and spawning with (MYLONAS *et al.* 2013b; FERNÁNDEZ *et al.* 2014) and without (MYLONAS *et al.* 2013a; SOARES *et al.* 2015) hormones, larval rearing conditions (ESTEVEZ et al. 2007; ROO *et al.* 2010; VALLÉS AND ESTÉVEZ 2013)), larval nutritional requirements (CAMPOVERDE AND ESTEVEZ 2017; EL KERTAOUI *et al.* 2017), skeletal development (CARDEIRA *et al.* 2012) and digestion (CASTRO *et al.* 2013; PAPADAKIS *et al.* 2013). Although studies are accumulating for various fields of species biology, the genetic information and stock structure are only scarcely studied with the available information being limited to only 148 nucleotide and 71 protein entries in NCBI (as of 16 May 2017).

The paucity of available genetic resources is currently an impediment to any future effort for genetic improvement in the species. However, through next generation sequencing (NGS) technologies, and in particular RNA-Sequencing (RNA-Seq), one can collect sequence information for thousands of genes in a single experiment (WANG *et al.* 2009). Transcriptome characterization is one of the main applications of NGS as it lays the groundwork for future studies on physiology, genetics, immunology, etc., creates inventories and gives access to

thousand of single nucleotide polymorphisms (SNP) and short tandem repeats (STR) markers. Up to now, it has been widely used for numerous fish species leading to a tremendous pool of genetic knowledge (e.g. see database FISHIT [http://www.fish-it.org/hcmr/] for 20 transcriptomes). Especially for farmed species, RNA-Seq can be an invaluable source of genetic information that can facilitate research on reproduction and sex dimorphism (MANOUSAKI *et al.* 2014; PALSTRA *et al.* 2015), physiology (KAITETZIDOU *et al.* 2012; TELES *et al.* 2013; MININNI *et al.* 2014), growth (GARCIA DE LA SERRANA *et al.* 2015), metabolism (CEREZUELA *et al.* 2013; DE SANTIS *et al.* 2015; GLENCROSS *et al.* 2015), immunity and disease resistance (CALDUCH-GINER *et al.* 2012; SARROPOULOU *et al.* 2012; ALI *et al.* 2014; MARANCIK *et al.* 2015; VALENZUELA-MIRANDA *et al.* 2015) and genetic marker discovery (MANOUSAKI *et al.* 2014; YU *et al.* 2014).

The goal of this paper was two-fold. First, we sought to characterize the transcriptome of meagre and build a solid transcriptomic reference for the species. Then, we aimed at assessing the genetic polymorphism of the species by including a thorough SNP and STR discovery from multiple individuals of farmed meagre. The discovered markers will set the groundwork for future marker-assisted selection for the species.

Materials & Methods

Sample collection

Animal care was carried out according to the "Guidelines for the treatment of animals in behavioural research and teaching" (Animal Behaviour 2001). Fish were selected (aquaculture facilities, IRTA, Spain, 21 August 2014, Table 1) from five different meagre crosses (families) that resulted from a mix of cultured and wild outbred parents. Muscle and liver tissues were dissected and preserved in RNAlater® (Applied Biosystems, Foster City, CA, USA). Sixteen meagre individuals were randomly selected for RNA Sequencing analysis (Supplementary Table 1).

Table 1. MIxS information for transcriptome assembly of *Argyrosomus regius*.

Item	Description
Classification	Eukaryota; Animalia; Chordata; Vertebrata; Actinopterygii; Percomorphaceae; Sciaenidae; Argyrosomus regius
Investigation type	Eukaryote transcriptome
Project name	Meagre transcriptome
Environment	
Latitude, longitude	41.634502, 2.167185
Geographical location	IRTA, Spain
Collection date	21/8/2014
Biome	marine biome (ENVO_00000447)
Feature	fish farm (ENVO:00000294)
Material	sea water (ENVO:00002149)
Sequencing	
Sequencing method	Illumina HiSeq 2500 paired-end
Estimated size	100 Mb
Organ or tissue source	Liver, muscle tissue
Assembly	
Method	De novo assembly
Program	Trinity trinityrnaseq_r2013-02-25
Finishing strategy	High quality transcriptome assembly
Data accessibility	
Database name	NCBI
Project name	PRJNA397355, PRJNA399060
Sample name	SRR5903997, SRR5903998, SAMN07522546

RNA extraction, library preparation and sequencing

Muscle and liver tissues from the 16 individuals were collected in a sterile and RNase-free way. Following the manufacturer's recommendations, soaked tissues in RNAlater®, were stored at 4°C overnight and then were transferred to −80°C until further processing. For both tissue types the samples were grinded under liquid nitrogen using pestle and mortar. Liver is rich in RNA and thus a small amount of tissue was adequate to purify a high quality RNA using Qiagen's RNeasy Plus extraction kit (QIAGEN®). In contrary, because of the low cell density and the fibrous nature of muscle tissue, the yield of total RNA is low. In that case, a much larger proportion of tissue was grinded, focusing on pulverizing it into a fine powder while keeping it completely frozen. Complete homogenization achieved in TRIzol® reagent (Invitrogen, Carlsbad, CA, U.S.) using needle and syringe and high integrity total RNA was isolated according to the manufacturer's instructions.

The quantity of the isolated RNA was measured spectrophotometrically with NanoDrop® ND-1000 (Thermo Scientific), while its quality and integrity were tested on an agarose gel (electrophoresis in 1.5% w/v) and further on an Agilent Technologies 2100 Bioanalyzer (Agilent Technologies). All samples had an RNA Integrity Number (RIN) value higher than 8. Following extraction, RNA from different individuals was pooled in equal quantities for each of the two tissue types. Then, an RNASeq library was constructed for each tissue following standard Illumina TruSeq protocols. The two libraries were loaded into one lane of an Illumina HiSeq2500 instrument (2x100bp). Raw reads produced are available at NCBI SRA with the project ID PRJNA397355 (Table 1).

Raw read quality control

Read quality was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and subjected to quality control following a pipeline including multiple steps and published elsewhere (ILIAS *et al.* 2015) Briefly, we first used Scythe - a bayesian adapter trimmer (version 0.994 BETA) (https://github.com/vsbuffalo/scythe), to identify adapter substrings in reads. Scythe

recognizes adapter sequences taking into account quality information especially at the 3' end where quality falls. Thus, this step was applied prior to any quality-based trimming (prior contamination rate set in 0.1 '-p 0.1'). Then, low quality (Phred quality threshold of 20 and minimum reads length of 45 nt) reads trimming was performed with Sickle (https://github.com/najoshi/sickle). Sickle scans the reads in sliding windows and based on the quality it determines whether a read requires trimming in the two ends or complete removal (parameters 'pe -g -t sanger -q 20 -l 45'). The surviving reads were used as input to Trimmomatic (BOLGER *et al.* 2014) to further remove 5' and 3' adaptor sequences and apply extra filtering steps (parameters 'PE -phred33 ILLUMINACLIP:adapter_file.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:25 MINLEN:45 CROP:99'). Finally, we used PrinSeq (SCHMIEDER AND EDWARDS 2011) to filter out low complexity sequences (threshold entropy value of 30) and perform poly A/T 5' tail (minimum of 5 A/T) trimming.

Transcriptome assembly and annotation

Following reads pre-processing, we pooled the filtered reads from both liver and muscle samples and built a transcriptome assembly using Trinity (GRABHERR *et al.* 2011) (trinityrnaseq_r2013-02-25; default kmer 25; minimum contig length of 200 nucleotides). This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFVG00000000, BioProject PRJNA399060 (Table 1).

To evaluate the completeness of the reconstructed assembly with used the software BUSCO v2 (SIMAO *et al.* 2015) through gVolante (https://gvolante.riken.jp/) selecting the Core Vertebrate Gene (CVG) set (HARA *et al.* 2015).

The assembled transcripts were annotated through a BLASTx similarity search against the SWISSPROT protein database (e-value threshold 10^{-5}; keeping the top twenty hits). To improve the speed of this long process, we implemented BLASTx in parallel using ParaNOBlast (https://github.com/jacqueslagnel/ParaNoBLast) described in (LAGNEL *et al.* 2009). Further, scan against protein domain signatures was done with InterProScan (JONES *et al.* 2014), which was run in parallel splitting the query in 100 subqueries and merging the

output with custom scripts. Blast and InterProScan results were input in Blast2GO V.2.8.0 (CONESA *et al.* 2005) where GO terms and Enzyme Commision (EC) numbers were retrieved and assigned to transcripts. Finally, sequences with EC numbers were further annotated with Kyoto Encyclopedia of Gene and Genome (KEGG) pathways using custom perl scripts.

Genetic marker discovery

To detect single nucleotide polymorphisms (SNPs) we used GATK pipeline (MCKENNA *et al.* 2010) according to the GATK best practices (DANECEK *et al.* 2011; VAN DER AUWERA *et al.* 2013). The implemented steps included mapping of the filtered reads to the assembled transcriptome using the highly accurate and fast aligner STAR (DOBIN *et al.* 2013), duplicate marking and sorting with Picard (https://github.com/broadinstitute/picard) and finally variant calling and filtering (options -window 35 -cluster 3 -filterName FS -filter "FS > 30.0" -filterName QD -filter "QD < 2.0"). The filtering options chosen filtered out SNPs that form clusters (more than 3 SNPs in a window of 35 bases), and variants with QualByDepth (QD) < 2.0 and FisherStrand (FS) > 30 accounting for variant quality and strand bias. Finally only SNPs with at least 15 reads coverage were kept.

Following the filtering steps conducted within GATK, variants without a “PASS” filter tag were excluded, ii. the variants that included insertions and deletions (indels), and SNPs with more than two alleles. Further, to avoid sampling the same SNP locus twice due to alternative splicing, we kept only those identified in the longest transcript of each locus. Finally, to check whether they belong to non-coding (3’UTR and 5’UTR) or coding regions (first, second or third codon positions), we excluded those that were identified in transcripts without ORF and analyzed the rest with a custom python script taking into account the SNP position in the longest predicted ORF and the ORF coordinates in each transcript.

To detect short tandem repeats (STRs) we scanned the longest transcript of each assembled locus using the software Phobos (http://www.ruhr-uni-bochum.de/ecoevo/cm/cm_phobos.htm). In particular, we detected non-exact STRs with 2–10 repeat unit length and a minimum length of 20 nucleotides. A custom Perl script was used to

parse the output. Once again, for markers included on transcripts with ORFs, STRs were categorized in coding, 3'UTR or 5'UTR according to the position in relation to the longest ORF within the longest transcripts per locus using python scripts.

Results & Discussion

Meagre transcriptome reconstruction and annotation

Illumina sequencing of the multi-individual liver and muscle libraries yielded in total 523,137,020 raw reads that were subjected to a series of quality control filters (Table 2). Following filtering, 341,439,304 paired reads were kept and used for assembly and downstream analyses.

Table 2. The raw read quality control process and the read survival following each filtering step.

Filtering steps	Surviving reads Muscle	Surviving reads Liver
Raw	280,804,390	242,332,630
Scythe*	280,804,390	242,332,630
Sickle	250,526,202	216,487,756
Trimmomatic	209,252,073	181,232,317
PrinSeq**	182,802,502	158,636,802

*NOTE: Scythe does not eliminate sequences

**NOTE: Only paired reads surviving PrinSeq filtering step were used for assembly and downstream analyses

The high quality reads from both muscle and liver libraries were pooled and used for reconstructing the reference transcriptome of meagre. The strategy followed involved pooling all liver information from multiple individuals in a single library and the same for muscle. This design enables the identification of genetic markers across the individuals' transcriptome in a cost-effective manner, as only two libraries are constructed and sequenced deeply. However, it holds limitations, first in terms of identifying markers at the individual level, and second pooling individuals bears drawbacks especially regarding the unequal representation of each individual's alleles in the final read count leading to erroneous allele frequency estimations in population studies (SCHLÖTTERER *et al.* 2014), which were not the scope of this study.

Following assembly, the resulted transcriptome comprised of 95,945 transcripts belonging to 80,807 loci with N50 value of 2,183, average length of 1,059 nucleotides and 46.19% GC content. To evaluate to completeness of the assembly, we ran BUSCOv2 and gVolante to find that out of 233 queried genes, 208 (89.27) were complete, 15 were partial (summing up to 95.71% complete and partial genes) and only 10 genes (4.29%) were missing. The results revealed a satisfying assembled transcriptome covering the great majority of meagre genes. However, future sequencing of more tissue types would lead to a more complete transcriptome in the species.

To annotate the assembly, we conducted a blastx similarity search against the highly curated SWISSPROT database. The results revealed that 33,638 out of 95,945 transcripts were significantly homologous to a known SWISSPROT sequence. Finally, targeted BLASTN search of meagre transcripts against tilapia cDNA retrieved 15,589 unique tilapia genes as top hits, once again confirming the thorough representation of the expected geneset in the transcriptome.

Following similarity search through blast, GO mapping resulted in 31,986 annotated sequences. The most important GO terms in the 'biological process' ontology at the level 2 are shown in Figure 1. Search for InterPro domains resulted in 46,647 sequences annotated with protein domains and raised the number of GO annotated transcripts to 34,252. Then, EC

number mapping through GO terms resulted in 1,016 potential enzymes in 8,682 total transcripts with EC:6.3.2.19 ‘ubiquitin-protein ligase’ as the top enzyme group in meagre transcriptome, followed by EC:3.6.1.3 (adenosine triphosphatase) and EC:2.3.1.48 (histone acetyltransferase). Based on EC mapping, we identified the corresponding KEGG pathways to find that the EC-annotated 8,682 sequences are involved in 382 total KEGG pathways. The most highly represented pathway was MAPK signaling pathway, followed by Purine metabolism and PI3K-Akt signaling pathway. A summary of the annotation results is presented in Table 3 and detailed annotation is given in Supplementary Table 2.

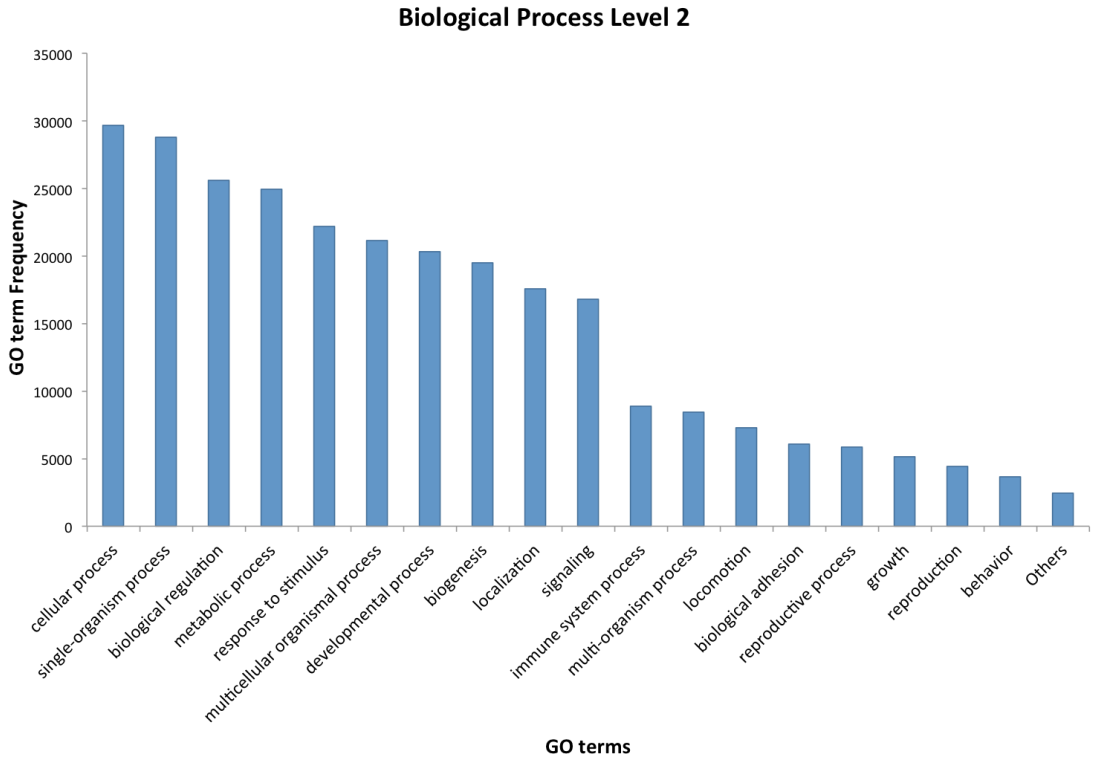

Figure 1. Gene Ontology functional characterization of meagre assembled transcriptome. Terms are shown for biological process level 2.

Table 3. Annotation Summary

Annotation steps	Contigs	% Contigs
	95,945	100
BLAST	33,638	35.05
InterPro	46,647	48.61
With IPR number	27,907	29.09
With > =1GO	34,252	35.70
Blast2Go annotated	33,220	34.62
EC	8,440	8.80
KEGG 380 pathways	2,475	2.58

Meagre genetic markers

Following the transcriptome characterization of meagre, we aimed at scanning for both SNPs and STRs across meagre transcriptome.

Our SNP search revealed a total of 42,933 high quality markers located in 14,544 transcripts (Supplementary Table 3). A GO enrichment analysis (FDR 0.05) of the contigs containing SNPs compared to the assembly revealed a significant underrepresentation of genes related to the nervous system (Figure 2; Supplementary Table 4), which might reflect the evolutionary pressure for conservation in this group of genes.

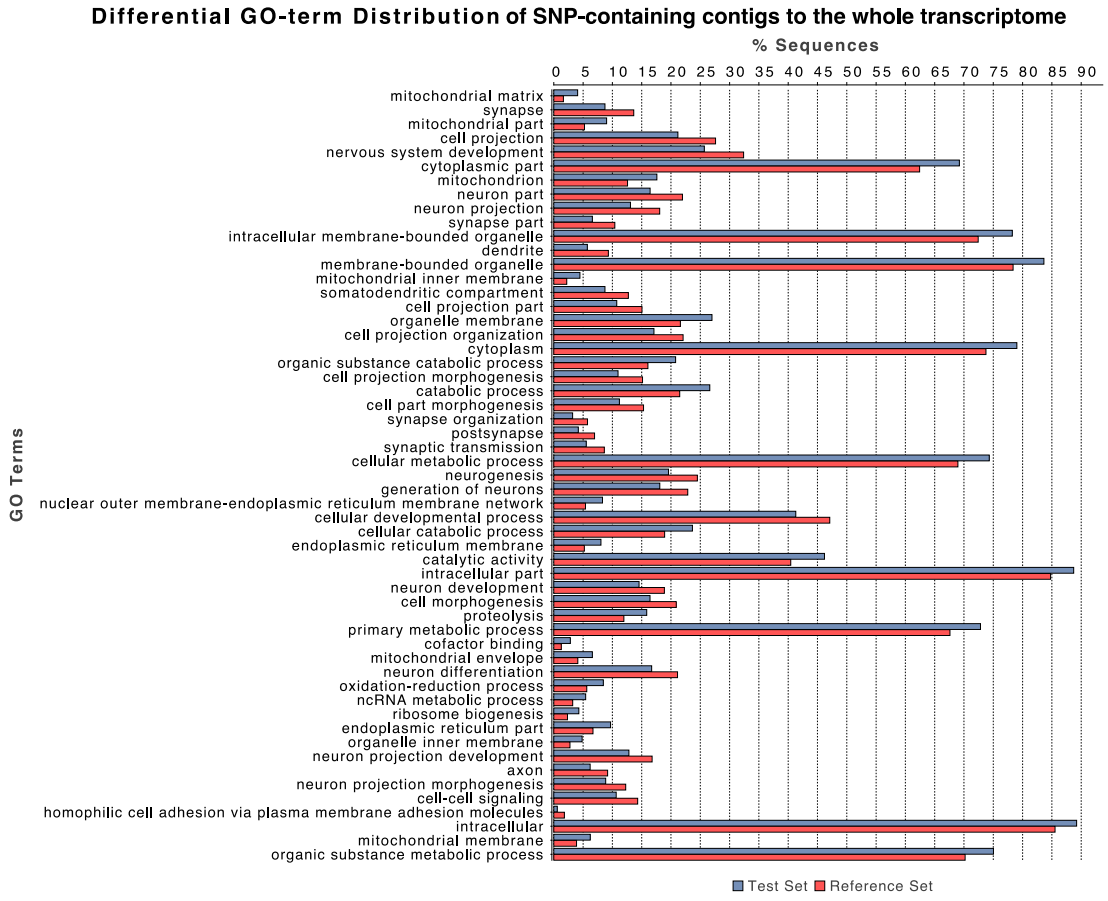

Figure 2. Top significantly over and under represented GO categories of SNP-containing contigs compared to the whole meagre transcriptome (only GO terms with FDR cut-off < 10^{-15} are shown).

Downstream analyses showed that most SNPs were located in the UTRs (15,149 SNPs) and at the third codon position (13,768 SNPs) in accordance to the sequence conservation pattern observed in coding sequences. SNPs that fall within the predicted open reading frame, result mostly to synonymous changes for the third codon position (731 non-synonymous and 13,037 synonymous SNPs), only in non-synonymous changes for the second codon position (2,937 SNPs) and mainly in non-synonymous changes for the first codon position (3,174 non-synonymous and 580 synonymous SNPs), as expected from the genetic code degeneracy (Figure 3).

Based on the frequency of SNPs found (1.22 SNPs per 1,000 bp), we describe a relatively low rate of polymorphism in the species probably reflecting the inbreeding that has taken place. Examples of other teleosts with reported higher SNP rate are turbot (one SNP per 302 bp; VERA et al. 2013), channel catfish (one SNP per 93 bp; LIU et al. 2016) and salmon with only slightly higher SNP rate (2.52 SNPs per 1,000 bp; TINE et al. 2014).

Our STR search revealed 20,581 markers ranging from 2-10 unit length transcriptome-wide (Supplementary Table 5). Following breaking down of the discovered STRs to those that fall within the coding regions (14,546 STRs) and those that do not, we show that 3-mer STRs (with unit length of 3, and then 6 and 9) fell into the first category (coding). The distribution of the 3-mer STRs are significantly higher than expected in the coding regions and significantly lower in the UTRs (chi-square p-value < 0.00) as expected due to the non-disturbance of the open reading frame by those repeats (Figure 4). The rest were mostly in the UTRs or in transcripts without ORFs.

The distribution of both marker types in the different regions of the transcriptome is consistent with the expected distribution of genetic variants as discovered in other studies as well (e.g. MANOUSAKI *et al.* 2014) and provide a high quality dataset for future genetic analysis in this new but important to aquaculture species.

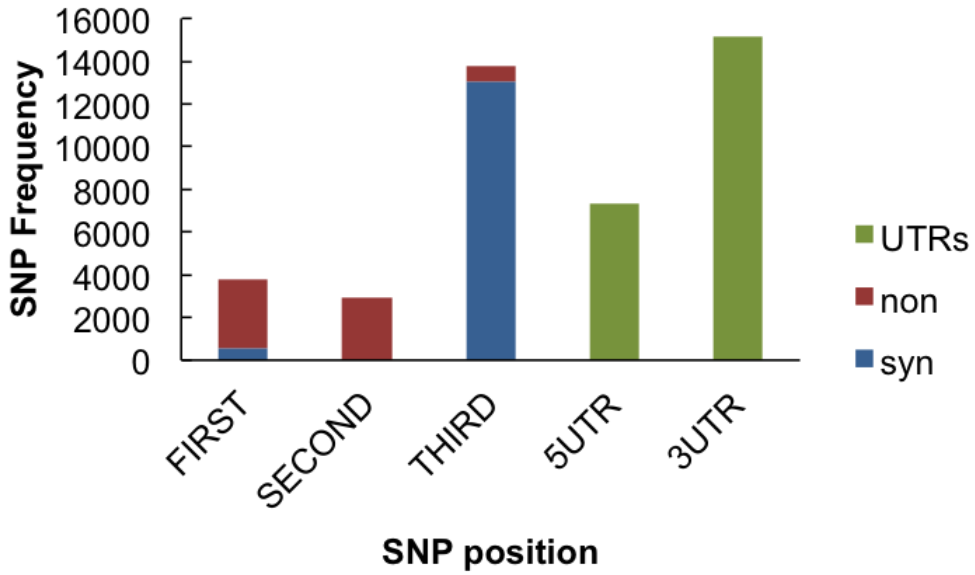

Figure 3. Distribution of SNPs along the coding and non-coding part of the transcripts. SNPs found within coding regions are separated to first, second and third position and are characterized as synonymous or nonsynonymous based on causing or not an amino acid shift in the protein sequence.

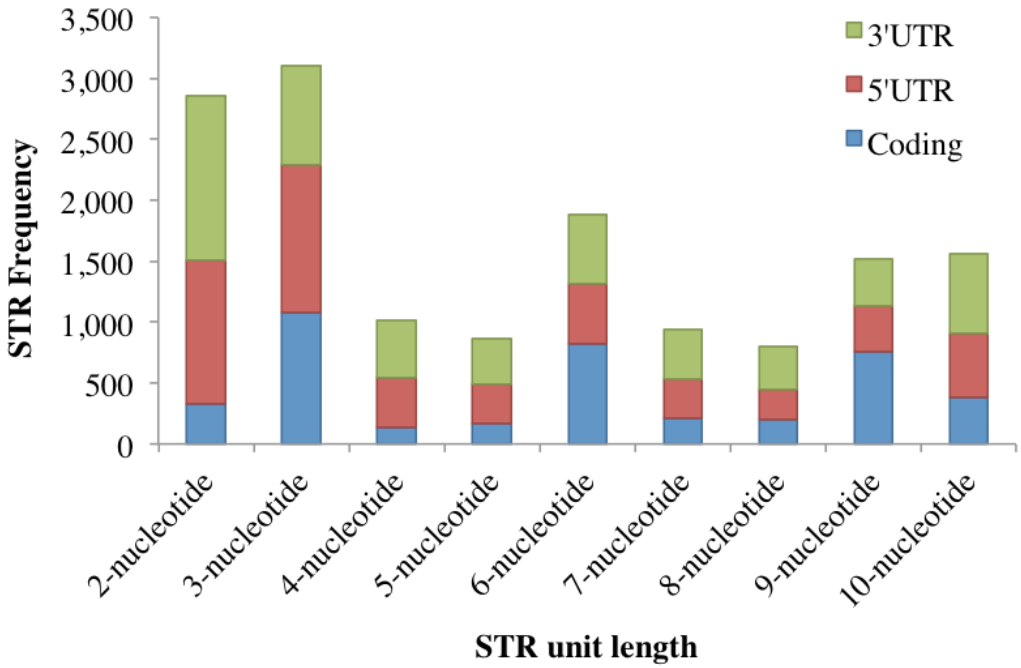

Figure 4. Distribution of STRs along the coding and non-coding part of the transcripts.

Meagre transcriptome gene content

Following the assembly annotation and genetic marker discovery, we sought to identify transcripts that might be involved in important biological functions. For example, growth is one of meagre's most important phenotypic traits for aquaculture. To that end, we extracted the sequences associated to growth by selecting transcripts with the search term 'growth' within GO annotation descriptions through Blast2GO. Our search revealed 7,121 sequences (Supplementary Table 6). The SNPs and STRs of those particular genes might serve as valuable resource for identifying variants linked to this critical trait of the species.

To further explore the gene content of meagre, we searched the assembled transcriptome for sequences that include representative terms in the GO annotations. More specifically, we found 3,144 sequences related to 'immune' functions, 15,274 genes involved in 'development', 4,300 genes involved in "stress", 1,012 genes involved in reproduction, 5,196 genes involved in metabolism and 2,168 genes involved in behavior (Supplementary Table 6).

Conclusions

Our study has built the first transcriptome assembly and at the same time the first next-generation based genomics resource for meagre. Following the annotated transcriptome, we launched a genetic marker discovery pipeline that led to the construction of a valuable dataset of SNPs and STRs transcriptome-wide coming out from a pool of 16 individuals. The low rate of polymorphism discovered imply that the transcriptome of meagre has been possibly shaped by inbreeding, a factor that raises even more the risk for further inbreeding through aquaculture. The provided assembly and genetic markers dataset will lay the groundwork for further studies of meagre biology and genetics and will set the basis for future applications of genetic breeding and marker-assisted selection for the species.

Acknowledgements

This study was funded by the European Union's Seventh Framework Programme for research, technological development and demonstration (KBBE-2013-07 single stage, GA 603121, DIVERSIFY). The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo (UiO) and supported by the 'Functional Genomics' and 'Infrastructure' programs of the Research Council of Norway and the South-Eastern Regional Health Authorities.

Supplementary Tables Legends

Supplementary Table 1. Summary including weight/length of sampled individuals.

Supplementary Table 2. Detailed annotation of meagre transcriptome including sequence description as defined by Blast2GO based on the annotation of the blast hits, the number of blast hits, the minimum e-value, the mean percentage of similarity, the number of GO terms, the assigned GO terms, EC numbers and InterProScan results.

Supplementary Table 3. The high quality SNP dataset discovered in meagre transcriptome. SNPs that remain after filtering and are located in the longest open reading frame of each gene are reported. For each SNP, provided information include: the respective contig, the open reading frame selected (ORF_Region), the starting point of the ORF (Start_ORF) and the position of the SNP (SNP_pos). Each SNP is characterized as coding or noncoding according to whether it falls inside or outside the coding regions and each noncoding SNP is annotated as upstream or downstream depending on whether it is found in the 5' or the 3' UTR of the gene. Further, SNPs that fall within the coding regions are broken down to those that fall within the first, second or third codon position and are also characterized as synonymous or nonsynonymous depending on whether the two alleles code for the same amino acid or not.

Supplementary Table 4. GO terms that are over- or under-represented in the SNP-containing genes compared to the whole assembly through a Fisher's exact test (FDR threshold 0.05). The test is run through Blast2GO for the three GO categories (P: Biological Process, C: Cellular Component, F: Molecular Function).

Supplementary Table 5. The high quality STR dataset discovered in meagre transcriptome. For each STR information regarding the respective contig (Seq ID), the unit length, the number of units in the reference (# of units), the start and stop position in the contig (start, stop), the total length (length), the length ignoring insertions/deletions (norm_length), number of mismatches (mis), number of insertions (ins), number of deletions (del), the unit motif (motif) and the total STR sequence (seq) are given.

Supplementary Table 6. The gene content of meagre transcriptome. The list and annotation of contigs that include "growth", "behaviour", "development", "reproduction", "metabolism",

"stress", "immune" within the GO terms description and may have a possible role in the respective functions.

References

Ali, A., C. Rexroad, G. Thorgaard, J. Yao and M. Salem, 2014 Characterization of the rainbow trout spleen transcriptome and identification of immune-related genes. Frontiers in Genetics 5.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bolger, A. M., M. Lohse and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30**:** 2114-2120.

Calduch-Giner, J. A., A. Sitja-Bobadilla, G. C. Davey, M. T. Cairns, S. Kaushik *et al.*, 2012 Dietary vegetable oils do not alter the intestine transcriptome of gilthead sea bream (*Sparus aurata*), but modulate the transcriptomic response to infection with Enteromyxum leei. BMC Genomics 13**:** 470.

Campoverde, C., and A. Estevez, 2017 The effect of live food enrichment with docosahexaenoic acid (22:6n-3) rich emulsions on growth, survival and fatty acid composition of meagre (*Argyrosomus regius*) larvae. Aquaculture 478**:** 16-24.

Cardeira, J., R. Vallés, G. Dionísio, A. Estévez, E. Gisbert *et al.*, 2012 Osteology of the axial and appendicular skeletons of the meagre *Argyrosomus regius* (Sciaenidae) and early skeletal development at two rearing facilities. Journal of Applied Ichthyology 28**:** 464-470.

Castro, C., A. Pérez-Jiménez, F. Coutinho, P. Pousão-Ferreira, T. M. Brandão *et al.*, 2013 Digestive enzymes of meagre (*Argyrosomus regius*) and white seabream (*Diplodus sargus*). Effects of dietary brewer's spent yeast supplementation. Aquaculture 416-417**:** 322-327.

Cerezuela, R., J. Meseguer and M. A. Esteban, 2013 Effects of dietary inulin, Bacillus subtilis and microalgae on intestinal gene expression in gilthead seabream (*Sparus aurata L.*). Fish Shellfish Immunol 34**:** 843-848.

Chao, L. N. (1986). Sciaenidae. In Fishes of the Eastern Atlantic and Mediterranean Poissons de l'Atlantique du nord-est et de la Méditerranée (Whitehead, P. J. P., Bauchot, M.-L., Hureau, J. C. & Tortonese, E., eds), pp. 865–874. Paris: Unesco.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon *et al.*, 2005 Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27.

De Santis, C., J. F. Taylor, L. Martinez-Rubio, S. Boltana and D. R. Tocher, 2015 Influence of development and dietary phospholipid content and composition on intestinal transcriptome of Atlantic salmon (*Salmo salar*). PLoS One 10**:** e0140964.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29**:** 15-21.

Duncan, N., A. Estévez, J. Porta, I. Carazo, F. Norambuena *et al.*, 2012 Reproductive development, GnRHa-induced spawning and egg quality of wild meagre (*Argyrosomus regius*) acclimatised to captivity. Fish Physiology and Biochemistry 38**:** 1273-1286.

El Kertaoui, N., C. M. Hernández-Cruz, D. Montero, M. J. Caballero, R. Saleh *et al.*, 2017 The importance of dietary HUFA for meagre larvae (*Argyrosomus regius*; Asso, 1801) and its relation with antioxidant vitamins E and C. Aquaculture Research 48**:** 419-433.

Estévez, A., Treviño, L., Gisbert, E. 2007. La densidad larvaria inicial afecta al crecimiento pero no a la supervivencia de las larvas de corvina (*Argyrosomus regius*) en cultivo. Paper presented at the XI Spanish Aquaculture Congress, Vigo, Spain, 24-28 September 2007.

FAO, 2015 http://www.fao.org/home/en/

Fernández, J., M. Á. Toro, A. K. Sonesson and B. Villanueva, 2014 Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. Frontiers in Genetics 5.

Fernandez-Palacios, H., Schuchardt, D., Roo, J., Izquierdo, M., Hernandez-Cruz, C.M., Duncan, N. 2014. Dose-dependent effect of a single GnRHa injection on the spawning of meagre ("*Argyrosomus regius*") broodstock reared in captivity. Spanish journal of agricultural research, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 12(4): 1038-1048.

Garcia de la Serrana, D., R. H. Devlin and I. A. Johnston, 2015 RNAseq analysis of fast skeletal muscle in restriction-fed transgenic coho salmon (*Oncorhynchus kisutch*): an experimental model uncoupling the growth hormone and nutritional signals regulating growth. BMC Genomics 16**:** 564.

Glencross, B. D., C. De Santis, B. Bicskei, J. B. Taggart, J. E. Bron *et al.*, 2015 A comparative analysis of the response of the hepatic transcriptome to dietary docosahexaenoic acid in Atlantic salmon (*Salmo salar*) post-smolts. BMC Genomics 16**:** 684.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29**:** 644-652.

Guidelines for the treatment of animals in behavioural research and teaching. Anim Behav. 2001, 61 (1): 271-275.

Hara, Y., K. Tatsumi, M. Yoshida, E. Kajikawa, H. Kiyonari *et al.*, 2015 Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation. BMC Genomics 16**:** 977.

Ilias, A., J. Lagnel, D. E. Kapantaidaki, E. Roditakis, C. S. Tsigenopoulos *et al.*, 2015 Transcription analysis of neonicotinoid resistance in Mediterranean (MED) populations of *B. tabaci* reveal novel cytochrome P450s, but no nAChR mutations associated with the phenotype. BMC Genomics 16**:** 939.

Joshi NA, Fass JN. 2011. https://github.com/najoshi/sickle.

Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li *et al.*, 2014 InterProScan 5: genome-scale protein function classification. Bioinformatics 30**:** 1236-1240.

Kaitetzidou, E., D. Crespo, Y. Vraskou, E. Antonopoulou and J. V. Planas, 2012 Transcriptomic response of skeletal muscle to lipopolysaccharide in the gilthead seabream (*Sparus aurata*). Mar Biotechnol (NY) 14**:** 605-619.

Lagnel, J., C. S. Tsigenopoulos and I. Iliopoulos, 2009 NOBLAST and JAMBLAST: new options for BLAST and a java application manager for BLAST results. Bioinformatics 25.

Liu, Z., S. Liu, J. Yao, L. Bao, J. Zhang et al., 2016 The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. Nat Commun 7: 11757.

Manousaki, T., A. Tsakogiannis, J. Lagnel, E. Sarropoulou, J. Z. Xiang *et al.*, 2014 The sex-specific transcriptome of the hermaphrodite sparid sharpsnout seabream (*Diplodus puntazzo*). BMC Genomics 15**:** 1-16.

Marancik, D., G. Gao, B. Paneru, H. Ma, A. G. Hernandez *et al.*, 2015 Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with *Flavobacterium psychrophilum*. Frontiers in Genetics 5.

Mayer, Christoph, Phobos 3.3.11, 2006-2010 Available online at: http://www.ruhr-uni-bochum.de/ecoevo/cm/cm_phobos.htm

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20**:** 1297-1303.

Mininni, A. N., M. Milan, S. Ferraresso, T. Petochi, P. Di Marco *et al.*, 2014 Liver transcriptome analysis in gilthead sea bream upon exposure to low temperature. BMC Genomics 15**:** 765.

Monfort M.C. 2010 Present market situation and prospects of meagre (*Argyrosomus regius*), as an emerging species in Mediterranean aquaculture. General Fisheries Commission for the Mediterranean: Studies and Reviews, No. 89, Food and Agriculture Organization of the United Nations, Rome.

Mylonas, C. C., E. Fatira, P. Karkut, M. Papadaki, I. Sigelaki *et al.*, 2015 Reproduction of hatchery-produced meagre *Argyrosomus regius* in captivity III. Comparison between GnRHa implants and injections on spawning kinetics and egg/larval performance parameters. Aquaculture 448**:** 44-53.

Mylonas, C. C., N. Mitrizakis, C. A. Castaldo, C. P. Cerviño, M. Papadaki *et al.*, 2013a Reproduction of hatchery-produced meagre *Argyrosomus regius* in captivity II. Hormonal induction of spawning and monitoring of spawning kinetics, egg production and egg quality. Aquaculture 414**:** 318-327.

Mylonas, C. C., N. Mitrizakis, M. Papadaki and I. Sigelaki, 2013b Reproduction of hatchery-produced meagre *Argyrosomus regius* in captivity I. Description of the annual reproductive cycle. Aquaculture 414**:** 309-317.

Nishimura & Kuraku, 2016 gVolante. Available online at: https://gvolante.riken.jp/

Palstra, A. P., K. Fukaya, H. Chiba, R. P. Dirks, J. V. Planas *et al.*, 2015 The olfactory transcriptome and progression of sexual maturation in homing Chum salmon *Oncorhynchus keta*. PLoS One 10**:** e0137404.

Papadakis, I. E., M. Kentouri, P. Divanach and C. C. Mylonas, 2013 Ontogeny of the digestive system of meagre *Argyrosomus regius* reared in a mesocosm, and quantitative changes of lipids in the liver from hatching to juvenile. Aquaculture 388-391**:** 76-88.

Roo, J., C. M. Hernández-Cruz, C. Borrero, D. Schuchardt and H. Fernández-Palacios, 2010 Effect of larval density and feeding sequence on meagre (*Argyrosomus regius;* Asso, 1801) larval rearing. Aquaculture 302**:** 82-88.

Sarropoulou, E., J. Galindo-Villegas, A. Garcia-Alcazar, P. Kasapidis and V. Mulero, 2012 Characterization of European sea bass transcripts by RNA SEQ after oral vaccine against V. anguillarum. Mar Biotechnol (NY) 14**:** 634-642.

Schlotterer, C., R. Tobler, R. Kofler and V. Nolte, 2014 Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nat Rev Genet 15: 749-763.

Schmieder, R., and R. Edwards, 2011 Quality control and preprocessing of metagenomic datasets. Bioinformatics 27**:** 863-864.

Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31**:** 3210-3212.

Soares, F., L. Ribeiro, M. Gamboa, S. Duarte, A. C. Mendes *et al.*, 2015 Comparative analysis on natural spawning of F1 meagre, *Argyrosomus regius*, with wild broodstock spawns in Portugal. Fish Physiology and Biochemistry 41**:** 1509-1514.

Teles, M., S. Boltana, F. Reyes-Lopez, M. A. Santos, S. Mackenzie *et al.*, 2013 Effects of chronic cortisol administration on global expression of GR and the liver transcriptome in *Sparus aurata*. Mar Biotechnol (NY) 15**:** 104-114.

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.

514 Valenzuela-Miranda, D., S. Boltana, M. E. Cabrejos, J. M. Yanez and C. Gallardo-Escarate,
515 2015 High-throughput transcriptome analysis of ISAV-infected Atlantic salmon
516 *Salmo salar* unravels divergent immune responses associated to head-kidney, liver
517 and gills tissues. Fish Shellfish Immunol 45**:** 367-377.
518 Vallés, R., and A. Estévez, 2013 Light conditions for larval rearing of meagre (*Argyrosomus*
519 *regius*). Aquaculture 376**:** 15-19.
520 Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel *et al.*, 2013 From
521 FastQ data to high confidence variant calls: the Genome Analysis Toolkit best
522 practices pipeline. Curr Protoc Bioinformatics 43**:** 11 10 11-33.
523 Vera, M., J. A. Alvarez-Dios, C. Fernandez, C. Bouza, R. Vilas et al., 2013 Development and
524 validation of single nucleotide polymorphisms (SNPs) markers from two
525 transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput
526 genotyping. Int J Mol Sci 14: 5694-5711.
527 Wang, Z., M. Gerstein and M. Snyder, 2009 RNA-Seq: a revolutionary tool for
528 transcriptomics. Nat Rev Genet 10.
529 Yu, Y., J. Wei, X. Zhang, J. Liu, C. Liu *et al.*, 2014 SNP Discovery in the Transcriptome of
530 White Pacific Shrimp *Litopenaeus vannamei* by Next Generation Sequencing. PLOS
531 ONE 9**:** e87218.
532