Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Method paper

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

T. Manousaki[a], A. Tsakogiannis[a,b], J. Lagnel[c], D. Kyriakis[a,d], N. Duncan[e], A. Estevez[e], C.S. Tsigenopoulos[a,*]

[a] Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Greece
[b] Department of Biology, University of Crete, Heraklion, Greece
[c] INRA PACA, UR 1052 GAFL, 67 Allée des Chênes, CS 60 094, 84143 Montfavet Cedex, France
[d] School of Medicine, University of Crete, Heraklion, Greece
[e] IRTA, Sant Carles de la Rapita, Tarragona, Spain

ARTICLE INFO

Keywords:
Aquaculture
RNASeq
SNPs
STRs

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.

1. 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 (Estévez 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.,

* Corresponding author at: Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Greece.
E-mail address: tsigeno@hcmr.gr (C.S. Tsigenopoulos).

https://doi.org/10.1016/j.margen.2018.01.002
Received 7 August 2017; Received in revised form 8 January 2018; Accepted 8 January 2018

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 thousands 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.

2. Materials & methods

2.1. 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).

2.2. 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 (2 × 100 bp). Raw reads produced are available at NCBI SRA with the project ID PRJNA397355 (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

2.3. Raw read quality control

Read quality was assessed with FastQC (Andrews, 2010; 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 (Joshi and Fass, 2011; 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.

2.4. 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