Deliverable Report

Deliverable No:	2.1	**Delivery Month:**	18
Deliverable Title	SNP library and chip to genetically characterise meagre or to use in marker assisted breeding programs		
WP No:	2	**WP Lead beneficiary:**	P3. IRTA
WP Title:	Reproduction and Genetics – 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:	P1. HCMR		
Status:	Delivered	**Expected month:**	18
……..			

Lead Scientist preparing the Deliverable: Tsigenopoulos, C. (P1. HCMR)
Other Scientists participating: Duncan, N. (P3. IRTA), Estevez ,A. (P3. IRTA), Manousaki, T. (P1. HCMR)

Objective: The objective of this Deliverable was to identify SNPs in the genome of meagre using RNASeq which could be then used to genetically characterize individuals or to implement QTL analysis and marker assisted selective breeding programs. This led to a catalogue of polymorphic loci at the expressed part of the genome and sets the ground for understanding growth and other traits of interest in meagre.

Description: One of the principal bottlenecks to meagre production is variable growth rates, causing uncertainty in the prediction of total yield from each on-growing cycle. Fast and predictable growth is an important and highly desired trait, which affects the profitability of food animal production, since feed costs account for the largest proportion of production costs. SNPs explain the greatest part of the genetic differences between individuals and are suitable for genetic evaluation and strategies that employ molecular genetics for selective breeding. Therefore, this task aims at using SNPs to potentially identify markers and genes associated with genetic variation in growth through Next Generation Sequencing (NGS) of the whole transcriptome of 16 fish from different families and phenotypic size (of the same age) that will provide a data-set of thousands of markers. SNP and Short Tandem Repeats (STR or microsatellite) markers identified by RNA-Seq and will presumably be associated with growth traits in future studies are reported and catalogued.

Deliverable 2.1 SNP library and chip to genetically characterise meagre or to use in marker assisted breeding programs, provided an overall view of the transcriptome sequence of meagre and led to the discovery of SNP and STR markers that can be used in Marker-Assisted selection and breeding programs.

Material and Methods

Biological material

Sixteen meagre individuals were selected from four groups of fish (Lset1, Lset2, Mset1 and Mset2, see

Table 2.1.1). The groups were from two spawning dates that were one week apart; set 1 consisted of two families spawned on the 24/04/2014 and set2 of three families spawned on the 01/05/2014. The two sets were graded and fish for RNA extraction were selected from the largest (groups "L") and the medium grades (groups "M") in order to have the highest chances to sample from 4 families (1 & 4 plus 2 & 5, respectively; **Table 2.1.1**). Selection was based on the expected kinship of individuals targeting on fish that are theoretically not closely related. For this purpose, fish were selected from the four groups that contained five different meagre crosses (families) that resulted mostly from wild outbred parents (**Table 2.1.2**). Muscle and liver tissues were dissected and preserved in RNAlater (IRTA, Spain).

Table 2.1.1 Details for the *A. regius* fish used for RNA extraction and transcriptome sequencing from the six families initially formed and their stocking sets (rearing tanks) after size-grading for small (S), medium (M) and large (L) fish.

Rearing Tank	Number of fish	Proportion of each family at start				
		1	2	4	5	6
(L-set1)	12	38%		62%		
(M-set1)	46	61%		39%		
(S-set1)	224	51%	14%	33%		3%
(L-set2)	19		26%		74%	
(M-set2)	49		83%			17%
(S-set2)	66		83%			17%

Table 2.1.2 Characteristics of the six families formed in *A. regius*, spawning dates, ID number of breeder and origin (wild or cultured) of the breeders used.

Family	Tank (Spawning Date)	Female	Male
1	V8-1 (24/04/2014)	5-wild	19-wild
2	V8-1 (01/05/2014)	5-wild	20-wild
3	V8-2 (01/05/2014)	1-wild	19-wild
4	C2 (24/04/2014)	16-cultured	21-wild
5	C1 (01/05/2014)	2-wild	22-wild
6	V6 (01/05/2014)	13-cultured	17-wild

RNA extraction, Library preparation and Sequencing

Muscle and liver samples were further processed in HCMR, Greece. RNA extraction protocols have been completed for the liver and muscle tissues of 16 sampled individuals (four fish/family). For both tissue types, RNA was extracted after grinding the tissue with liquid nitrogen using pestle and mortar. In the case of liver tissues, total RNA was extracted with Qiagen's RNeasy Plus extraction kit, while muscle tissues were homogenized in TRIzol® reagent (Invitrogen, Carlsbad, CA, U.S.) and RNA was isolated according to the manufacturer's instructions. The quantity of the isolated total RNA was measured spectrophotometrically with NanoDrop® ND-1000 (Thermo Scientific), and quality was tested on an agarose gel (electrophoresis in 1.5% w/v).

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).

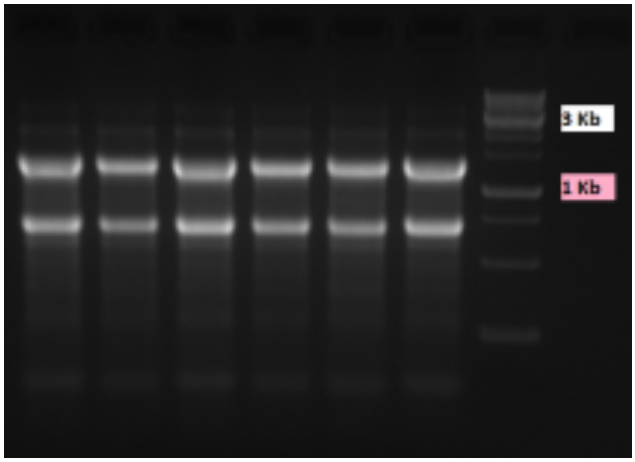

Figure 2.1.1 Total RNA extraction profile from meagre liver tissues. The size marker on the right side of the gel is the 1Kb DNA ladder RTU from Nippon Genetics GmbH.

Raw read pre-processing

Raw read quality was assessed with FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw data quality control took place in four steps. First, sequence contamination was removed from Illumina adapter sequences with Scythe (https://github.com/vsbuffalo/scythe). Scythe recognizes adapter sequences by taking into account quality information and thus increasing the efficiency of removing them especially at the 3' end of the reads where read quality drops. Second, Sickle (https://github.com/najoshi/sickle), a tool that uses sliding windows to identify reads with low-quality regions especially for 5' and 3' regions of the reads, was used to trim the low-quality ends of the reads. The next step was to use the general quality control software Trimmomatic (Bolger et al., 2014) that removes low quality reads applying various filters, including cutting adapter and other Illumina-specific sequences from the read, performing a sliding window trimming, cutting once the average quality within the window falls below a threshold and cutting bases off the start/end of a read if below a threshold quality. Finally, Prinseq (Schmieder and Edwards 2011) was used to remove any remaining sequences that are the result of adapter contamination. Each of the applied software tools applies alternative methodologies to remove errors. The combination of all four pieces of software has led to an efficient quality control.

Transcriptome assembly and annotation

To build the transcriptome assembly, quality-filtered reads were input in the software Trinity (version trinityrnaseq r20140717; Grabherr et al., 2011) and ran with default settings. Trinity is specialized in assembling Illumina reads and is considered one of the most efficient assembly software (Haas et al., 2013). The assembly process required ~ 100Gb of RAM memory and 10 CPUs and was run successfully at the computer cluster of IMBBC, HCMR.

To annotate the transcripts and get an idea of their potential role, a similarity search was conducted through *blastx* against Swiss-Prot protein database (*e*-value threshold 10^{-12}). *blastx* was run in parallel using NOblast (Lagnel et al., 2009) and the best hit was kept for each transcript. Blast output was summarized with custom shell commands/scripts.

Genetic marker discovery

The assembled sequences were scanned for STRs with Phobos (Mayer 2006-2010). Non-exact STRs were detected with a 2–10 repeat unit length and a minimum length of 20 nucleotides. A custom Perl script was used to parse the output.

SNP discovery took place using SAMTOOLS (Li et al., 2009), one of the most efficient SNP discovery tools for next generation data. First, quality-filtered reads from both muscle and liver samples were mapped onto the transcriptome assembly using the mapping software bowtie2 (Langmead and Salzberg 2012). Then, the alignment files (.sam file) were analyzed with samtools *mpileup* function. SNP calling was conducted with *bcftools call* command and then quality filtered with *bcftools filter* command keeping only SNPs with the above a strict quality threshold (Q>19). Finally, SNPs were further filtered based on the number of high quality read coverage (DP>9).

Results and Discussion

Raw data quality control

Illumina sequencing led to the production of 523,137,020 raw reads. Filtering steps reduced this to 341,439,304 (65%) high quality paired reads (182,802,502 for muscle and 158,636,802 for liver sample) (**Table 2.1.3**). Application of Sickle and Trimmomatic resulted in the greatest filtering. Prinseq removed relatively few more sequences, while Scythe trimmed adapter sequences from the reads' ends. Overall, read quality was significantly improved after the application of filtering criteria, especially at the 5' and 3' prime ends of the reads (**Fig. 2.1.2**).

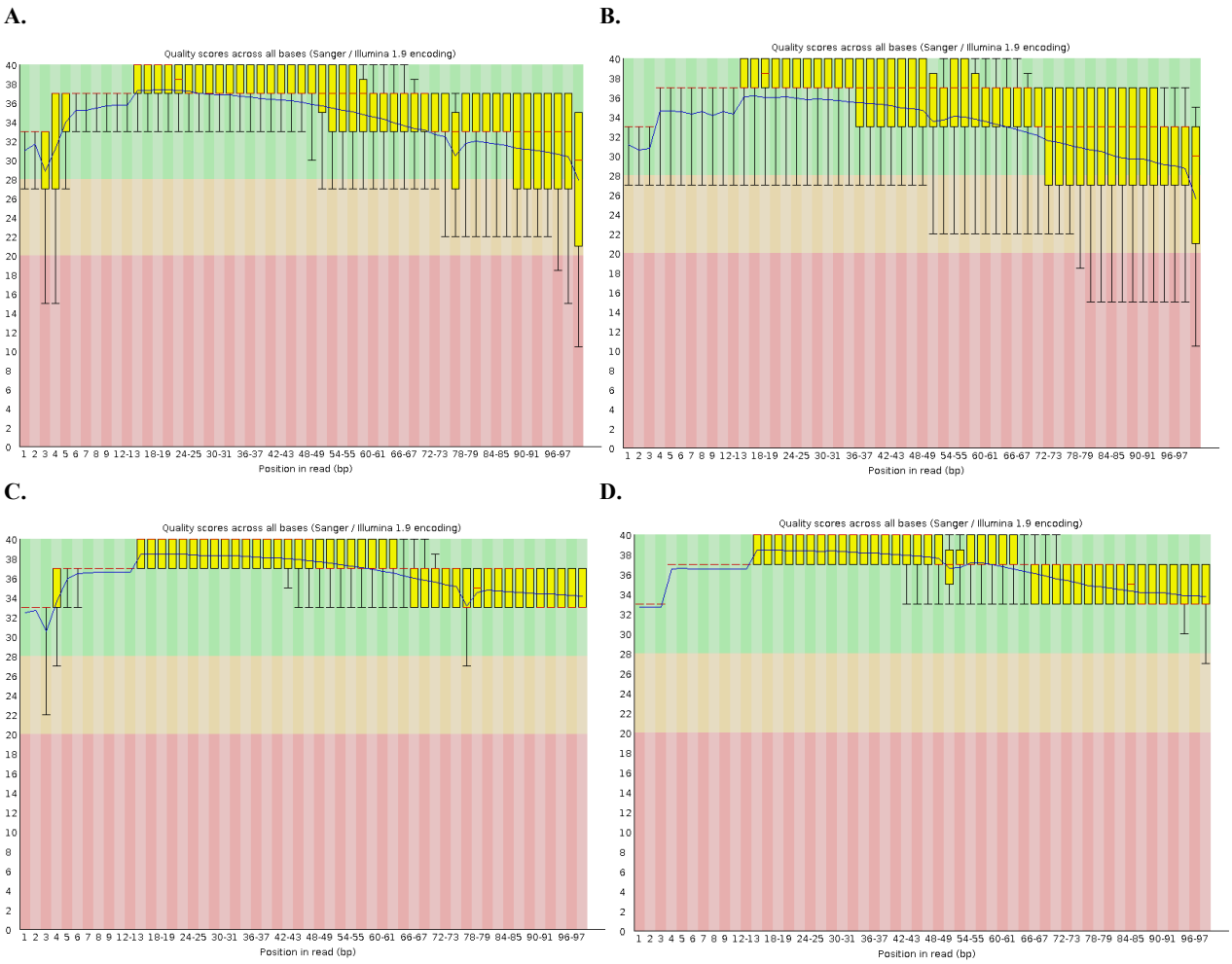

Figure 2.1.2 Quality filtering for the muscle sample. On top, the "per base" quality of read1 (A) and read2 (B) of the raw data is shown; on the bottom, the "per base" quality of read1 (C) and read2 (D) of the quality-filtered data are presented. Quality is measured in Illumina 1.9 Phred score. Plots were made with FASTQC toolkit.

Table 2.1.3 Illumina reads surviving in pairs (excluding orphans) after each quality-filtering step.

Filtering step	Software	Muscle tissue	Liver tissue	Total
1	Scythe	280,804,390	242,332,630	523,137,020
2	Sickle	250,526,202	216,487,756	467,013,958
3	Trimmomatic	183,041,946	158,882,592	341,924,538
4	Prinseq	182,802,502	158,636,802	341,439,304

Meagre Transcriptome

The transcriptome of meagre was reconstructed with the 341,439,304 paired filtered reads that passed through all quality control filters applied. The assembly consisted of 95,964 transcripts belonging to 80,824 loci or genes (**Table 2.1.4**). It has an average length of 1,058.83 bp, and N50 statistic (i.e. the length N for which half of all bases in the sequences are in a sequence of length L < N) of 2,183 bp. To understand the basic function of the assembled transcripts a *blastx* similarity search was made against Swiss-Prot. Out of 95,964 transcripts, approximately 37% (35,888) had a significant hit against Swiss-Prot (*e*-value 10^{-12}) and in particular against Nile tilapia (*Oreochromis niloticus*) proteins (**ANNEX 2.1,** at http://www.diversifyfish.eu/401/login.php?redirect=/deliverables.html. Part of the Annex is presented as **Annex 2.1a** at the end of the document). The produced assembly provides an excellent reference for future needs in terms of meagre coding sequences.

Table 2.1.4 Meagre transcriptome assembly statistics.

Assembly Statistic	Value
Number of Sequences	95,964
Total Length	101,609,879
Average Length	1,058.83
Median	472
Min	201
Max	25,456
% GC	46.19
N25	3,771
N50	2,183
bps at N50	50,805,786
N75	932
N90	365

*All statistics length is in bp.

Genetic markers discovery

Genetic marker discovery was two-fold. First, the transcripts were searched for non-exact short tandem repeats (STRs) of 2- up to 10-nucleotides at the assembled transcriptome of meagre. The search revealed 20,582 total STRs located in 16,517 transcripts belonging to 12,565 genes (summarized in **Table 2.1.5**). For detailed list and information for each locus see **ANNEX 2.1**

(http://www.diversifyfish.eu/401/login.php?redirect=/deliverables.html). Part of the Annex is presented as **Annex 2.1b** at the end of the document.

Table 2.1.5 Short Tandem Repeats (STRs or microsatellites) discovered in meagre transcriptome.

STR type	Number of STRs found
2-nucleotide	3,856
3-nucleotide	4,439
4-nucleotide	1,444
5-nucleotide	1,222
6-nucleotide	2,838
7-nucleotide	1,242
8-nucleotide	1,129
9-nucleotide	2,229
10-nucleotide	2,183
Total	20,582

The next step included a search for SNPs in the transcripts of meagre. A total of 133,613 SNPs were discovered. Quality filtering led to elimination of 46,259 SNPs and resulted to 87,354 high quality SNPs. Finally, further depth filtering resulted in 71,736 SNPs located in 20,309 transcripts belonging to 18,657 loci. For detailed list see **ANNEX 2.1** (http://www.diversifyfish.eu/401/login.php?redirect=/deliverables.html). Part of the Annex is presented as **Annex 2.1c** at the end of the document.

The SNP and STR loci discovered comprise a valuable source of genetic markers widely distributed in the transcriptome of meagre. To be able in the future to use markers not linked for genetic applications (e.g. SNP-chip construction, parentage analysis, QTL identification, etc.), indirect linkage information was extracted based on sequence similarity against Nile tilapia. The gene with the highest similar to a tilapia gene (top hit) was tracked to the genomic scaffold that this hit is located in tilapia genome (see ANNEX 2.1). The rationale behind indirect linkage information is that in cases where the top tilapia hits of two meagre genes are located on the same scaffold, it can be assumed that there is high chance that those two genes are linked in meagre genome too.

The use of individuals from multiple families increases the chance of including in our data multiple polymorphisms including those involved in growth. The implemented experimental design allowed the identification of both SNP and STR markers in unprecedented magnitude. It is anticipated that future analyses based on those markers will lead to a better understanding of meagre genetics. Finally, in combination with the QTL mapping approach, which is in progress and expected to be delivered by Month 36 and will greatly expand the SNP catalogue for meagre.

Conclusions

In this deliverable, the muscle and liver transcriptome of meager was sequenced and characterized. The outcome from meagre transcriptome sequencing is two-fold. First, it provides information on meagre gene content and sequence -on a global scale- allowing the further study of any gene family or genetic pathway of the species. Second, it allowed a transcriptome-wide scan for genetic marker discovery. A thorough marker discovery pipeline was implemented that led to thousands of SNP and STR markers that can be useful in future marker-assisted selection or other analyses. Our current work, in combination with the forth-coming

SNP information that will be produced from the planned linkage analysis using full- and half-sib families in the next months (Deliverable D2.4 Identification of genetic markers related to growth for use in marker assisted breeding programs for meagre, Month 36), will establish a new standard in meagre genetics setting the groundwork for deeper studies on growth and other traits of the species.

Deviations: There were no major deviations from the general outline in the DOW. According to the changes proposed for D2.4 an SNP- chip will not be implemented to genetically characterize meagre or to use in marker assisted breeding programs; instead, the ddRAD approach will be followed first to identify genetic markers related to growth for use in marker assisted breeding programs and then to genetically characterize fast and slow growing meagre.

References

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. May 15;29(7):644-52.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. Aug;8(8):1494-512.

Lagnel, J., Tsigenopoulos, C.S., Iliopoulos I. (2009) NOBLAST and JAMBLAST: new options for BLAST and a java application manager for BLAST results**.** Bioinformatics, **25**(6)**:**824-826.

Langmead, B., Salzberg, S. (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods, 9, 357-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.

Mayer, C., (2006-2010). Phobos 3.3.11. http://www.rub.de/spezzoo/cm/cm_phobos.htm.

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


Co-funded by the Seventh Framework Programme of the European Union


Annex 2.1a Example of the results from the *blastx* similarity search against Nile tilapia (*Oreochromis niloticus*) proteins in Swiss-Prot (*e*-value 10^{-12}) for the assembled transcripts in meagre. **

Q Name	Best Hit	Associated gene name	Description
c100_g2_i1	ENSONIG00000019171	dph3	diphthamide biosynthesis 3 [Source:ZFIN;Acc:ZDB-GENE-050522-545]
c10004_g1_i1	ENSONIG00000019577	si:dkey-94l16.4	si:dkey-94l16.4 [Source:ZFIN;Acc:ZDB-GENE-121214-291]
c10012_g1_i1	ENSONIG00000004579	dbt	dihydrolipoamide branched chain transacylase E2 [Source:ZFIN;Acc:ZDB-GENE-05032
c10025_g1_i1	ENSONIG00000006407	0	Uncharacterized protein {ECO:0000313|Ensembl:ENSONIP00000008078} [Source:Un
c10027_g1_i1	ENSONIG00000019313	MYOG	myogenin [Source:RefSeq peptide;Acc:NP_001266455]
c10027_g2_i1	ENSONIG00000019313	MYOG	myogenin [Source:RefSeq peptide;Acc:NP_001266455]
c10028_g1_i1	ENSONIG00000011699	apoba (2 of 2)	apolipoprotein Ba [Source:ZFIN;Acc:ZDB-GENE-070702-4]
c1004_g1_i1	ENSONIG00000014380	CD200 (2 of 2)	CD200 molecule [Source:HGNC Symbol;Acc:HGNC:7203]
c10040_g1_i1	ENSONIG00000004509	faima	Fas apoptotic inhibitory molecule a [Source:ZFIN;Acc:ZDB-GENE-040718-323]
c10041_g1_i1	ENSONIG00000006282	cyp7a1a	cytochrome P450, family 7, subfamily A, polypeptide 1a [Source:ZFIN;Acc:ZDB-GENE-
c10042_g1_i1	ENSONIG00000005895	C3orf58	chromosome 3 open reading frame 58 [Source:HGNC Symbol;Acc:HGNC:28490]
c10047_g1_i1	ENSONIG00000001892	haus5	HAUS augmin-like complex, subunit 5 [Source:ZFIN;Acc:ZDB-GENE-041114-150]
c10056_g1_i1	ENSONIG00000011745	MAPKBP1 (2 of 2)	mitogen-activated protein kinase binding protein 1 [Source:HGNC Symbol;Acc:HGNC
c10057_g1_i1	ENSONIG00000012495	armc9	armadillo repeat containing 9 [Source:ZFIN;Acc:ZDB-GENE-100922-67]
c10066_g1_i1	ENSONIG00000011504	ppiab	peptidylprolyl isomerase Ab (cyclophilin A) [Source:ZFIN;Acc:ZDB-GENE-030131-7459
c10071_g1_i1	ENSONIG00000008937	POLQ	polymerase (DNA directed), theta [Source:HGNC Symbol;Acc:HGNC:9186]
c10078_g1_i1	ENSONIG00000001401	myo7aa	myosin VIIAa [Source:ZFIN;Acc:ZDB-GENE-020709-1]
c10079_g1_i1	ENSONIG00000001550	rhbg	Rhesus blood group, B glycoprotein [Source:ZFIN;Acc:ZDB-GENE-030131-9542]
c10079_g2_i1	ENSONIG00000001550	rhbg	Rhesus blood group, B glycoprotein [Source:ZFIN;Acc:ZDB-GENE-030131-9542]
c10082_g1_i1	ENSONIG00000014910	cyb5d1	cytochrome b5 domain containing 1 [Source:ZFIN;Acc:ZDB-GENE-050417-173]
c10083_g1_i1	ENSONIG00000001322	psmg3	proteasome (prosome, macropain) assembly chaperone 3 [Source:ZFIN;Acc:ZDB-GEN
c10093_g1_i1	ENSONIG00000004599	rp2	retinitis pigmentosa 2 (X-linked recessive) [Source:ZFIN;Acc:ZDB-GENE-040426-2795]
c10093_g1_i2	ENSONIG00000004599	rp2	retinitis pigmentosa 2 (X-linked recessive) [Source:ZFIN;Acc:ZDB-GENE-040426-2795]
c10100_g1_i1	ENSONIG00000005827	arl14ep	ADP-ribosylation factor-like 14 effector protein [Source:ZFIN;Acc:ZDB-GENE-060825-2
c10107_g1_i1	ENSONIG00000015386	tert	telomerase reverse transcriptase [Source:ZFIN;Acc:ZDB-GENE-080405-1]
c1011_g1_i1	ENSONIG00000010896	pbx3a	pre-B-cell leukemia transcription factor 3a [Source:ZFIN;Acc:ZDB-GENE-031218-1]
c10118_g1_i1	ENSONIG00000005878	zic4	zic family member 4 [Source:ZFIN;Acc:ZDB-GENE-040622-4]

** *Only the first page is presented. The full list is available in an excel file "Annex 2.1"*

Annex 2.1b Example of the short tandem repeats (STRs) of 2- up to 10-nucleotides at the assembled transcriptome of meagre. **

Transcript	unit length	# of units	start	stop	length	norm_length	score	%perfect	mismatch	insertion	deletion	motif
c100_g1_i1	3	7	1	23	21	21	20	100	0	0	0	TCA
c100_g2_i1	3	7	1	22	21	21	19	100	0	0	0	GAT
c10005_g1_i1	10	2	477	496	20	20	10	100	0	0	0	ATAACAGACT
c10007_g1_i1	3	7	1	22	21	21	19	100	0	0	0	TAT
c10011_g1_i1	2	10	672	691	20	20	11	95	1	0	0	TG
c10012_g1_i1	6	3	288	307	18	18	14	100	0	0	0	GGACCA
c10017_g1_i1	6	3	27	47	17	18	9	95,455	0	0	1	TACAAT
c1002_g1_i1	3	7	2	24	21	21	6	91,304	2	0	0	AAT
c10032_g1_i2	10	2	149	169	20	20	11	100	0	0	0	GTACACTGTA
c10065_g1_i1	5	5	384	408	25	25	13	96	1	0	0	GCTAA
c10068_g2_i1	6	3	184	202	17	18	7	95	0	0	1	TTTTAC
c10090_g1_i1	6	3	52	71	18	18	14	100	0	0	0	ATGAAG
c10100_g1_i1	6	3	1270	1290	19	18	8	95	0	1	0	CTCCAC
c10100_g1_i1	4	5	1965	1987	21	20	12	95,455	0	1	0	AAAT
c10118_g1_i1	9	2	567	587	18	18	12	100	0	0	0	CATCACCAT
c10129_g1_i1	3	7	207	227	20	21	12	95,455	0	0	1	TCT
c10145_g1_i1	7	3	548	568	20	21	8	95,455	0	0	1	TATATAT
c10146_g1_i1	3	6	1	20	18	18	17	100	0	0	0	GGA
c10152_g1_i1	2	10	427	447	20	20	12	95,238	1	0	0	TG
c10158_g1_i1	3	6	1	20	18	18	17	100	0	0	0	TCC
c10158_g1_i1	9	3	231	258	28	27	12	96,296	0	1	0	AAAACAAAA
c10158_g2_i1	5	4	8	28	19	20	10	95,455	0	0	1	AACAA
c10192_g1_i1	6	6	119	154	36	36	23	97,222	1	0	0	GGTTCT
c10194_g1_i1	10	2	26	45	20	20	10	100	0	0	0	GTCTCTCTCT
c10197_g1_i1	7	3	95	119	21	21	11	96	1	0	0	TGACCTC
c10205_g1_i1	4	5	1	23	20	20	19	100	0	0	0	TATG
c10205_g2_i1	5	4	1	20	20	20	15	100	0	0	0	ATTTT

** *Only the first page is presented. The full list is available in an excel file "Annex 2.1"*

Annex 2.1c Example of the single nucleotide polymorphisms (SNPs) identified at the assembled transcriptome of meagre.***

Transcript	position	Ref allele	Alternative allele	Quality	Depth	Hit in tilapia	Scaffold in tilapia
c100_g1_i1	1	T	G	83	10	#N/A	#N/A
c100_g1_i1	2	C	G	26,0947	17	#N/A	#N/A
c10004_g1_i1	553	A	G	115	12	ENSONIG00000019577	GL831136.1
c10004_g1_i1	547	C	T	109	13	ENSONIG00000019577	GL831136.1
c10007_g1_i1	1045	T	C	65	32	#N/A	#N/A
c10007_g1_i1	394	T	G	42,4248	33	#N/A	#N/A
c10007_g1_i1	872	G	C	41,0108	45	#N/A	#N/A
c10007_g1_i1	910	A	G	33,8222	49	#N/A	#N/A
c10007_g1_i1	66	T	C	53	556	#N/A	#N/A
c10018_g1_i1	476	G	A	49,4513	12	#N/A	#N/A
c10025_g1_i1	285	C	T	34,271	10	ENSONIG00000006407	GL831138.1
c10025_g1_i1	482	T	C	111	13	ENSONIG00000006407	GL831138.1
c1004_g1_i1	474	C	T	34,5193	13	ENSONIG00000014380	GL831258.1
c10042_g1_i1	50	A	T	82	11	ENSONIG00000005895	GL831138.1
c10042_g1_i1	367	G	A	48,371	16	ENSONIG00000005895	GL831138.1
c10044_g1_i1	121	A	T	51	10	#N/A	#N/A
c10046_g1_i1	167	T	C	61	12	#N/A	#N/A
c10048_g1_i1	485	G	A	40,5792	14	#N/A	#N/A
c10054_g1_i1	174	C	T	28,1449	16	#N/A	#N/A
c10056_g1_i1	1338	G	A	35,7649	15	ENSONIG00000011745	GL831368.1
c10057_g1_i1	154	A	G	34,919	14	ENSONIG00000012495	GL831204.1
c10083_g1_i1	663	C	T	413	82	ENSONIG00000001322	GL831149.1
c10098_g1_i1	1216	G	A	55	34	#N/A	#N/A
c10100_g1_i1	503	G	A	34,6693	16	ENSONIG00000005827	GL831227.1
c10100_g1_i1	971	T	C	180	23	ENSONIG00000005827	GL831227.1
c10100_g1_i1	993	T	A	45,688	29	ENSONIG00000005827	GL831227.1
c10100_g1_i1	992	T	A	196	30	ENSONIG00000005827	GL831227.1

** *Only the first page is presented. The full list is available in an excel file "Annex 2.1*