Iran. J. Ichthyol. (2025) 12(3): 233-247
P-ISSN: 2383-1561; E-ISSN: 2383-0964
© 2025 Iranian Society of Ichthyology

ORIGINAL ARTICLE

Transcriptome profiling of the liver among the spawning females of the Caspian Kutum (*Rutilus frisii*, Nordmann, 1840)

Saman FARASATI[1], Majidreza KHOSHKHOLGH[*1], Sajad NAZARI[2],

[1]Fisheries Department, Faculty of Natural Resources, University of Guilan, P.O. Box 1144, Sowmehsara, Iran.
[2]Shahid Motahary Cold-water Fishes Genetic and Breeding Research Center, Iranian Fisheries Sciences Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Yasouj, Iran.

Correspondence
majidreza@guilan.ac.ir

Article history:
Accepted 28 August 2025

Abstract
The Caspian Kutum (*Rutilus frisii*, Nordmann, 1840) is an economically significant species among bony fish stocks in the southern basin of Caspian Sea. Nevertheless, information for transcriptome, the physiological differences and the regulatory mechanism of controlling for two strains remain unclear. To address this issue, we performed *de novo* assembly of liver transcriptome from six female *R. frisii*, three from each strain, using RNA sequencing (RNA-seq). After quality control, the de novo transcriptome assembly carried out with Trinity software and then 110,381 transcripts were generated. A total of 12,569 unigenes were grouped into three major functional categories (biological process, cellular components and molecular functions) and 62 sub-categories. A total of 9,627 transcripts were assigned at least one GO term. According to the public protein databases, we observed 21 unigenes related to reproduction by comparing the assembled unigenes and then 12849 differentially expressed genes (DEGs) were identified. We compared the histology of ovaries between the two strains to determine the degree of sexual maturation. The results of the present study provide a basis for future studies on the genetic differences between these two strains of *R. frisii* and their impact on population structure.

Keywords: Caspian Kutum, Reproduction, RNA-seq, Transcriptome analysis

INTRODUCTION

The Caspian Kutum (*Rutilus frisii*) is a native semi-anadromous fish from the Caspian Sea and can reach up to 6kg of weight (Abdoli 1999). This species is an essential, economically unique fish belonging to the Leuciscid family, mainly distributed in the southern coastlines of the Caspian Sea. At the same time, it is a rarely seen fish in other regions of the Caspian Sea, such as the north Caspian and the Volga River (Valipour et al. 2009). *R. frisii* now comprises more than 70% of the total bony fish catch on the Iranian shores of the Caspian Sea and due to its good tasting meat and the culinary customs of the local people, has a great demand (Yousefian and Mosavi 2008; Eagderi et al. 2022).

Caspian Kutum live as adult fish in the sea and enter in flowing rivers to spawn when temperature and river flow are suitable for inducing migration (Holčik 1995). After spawning, adult fish return to the Caspian Sea and remain in deep water (Mashhor et al. 2011). Based on all available sources (Derzhavin 1934; Berg 1949; Belyaeva et al. 1989; Magomedov et al., 1989), this species has two strains, including the autumn strain (AS) and spring strain (SS). The AS stocks enter the Anzali lagoon from late September until middle December and spawn over the aquatic or flooded terrestrial vegetation, known as the phytophilous strain. While, the SS stocks undertaking migrations to main spawning rivers from the end March until the end of April. The SS seems to be lithophilous as its eggs are deposited on the stony bottom, and the spawning grounds are usually far from the river mouths (Shikhshabekov 1979). The SS stocks constitute more than 98 per cent of the Caspian Kutum population and only 2 per cent or less belong to the AS stocks (Valipour et al. 2009).

The AS stocks are highly endangered due to unstable condition of Anzali lagoon, overfishing, destruction of spawning grounds and, foremost lack of artificial reproduction. (Abdolhay et al. 2011; Fazli et al. 2012). The SS stocks are not only genetically at risk despite the artificial reproduction programs (Rahbar et

Journal homepage: ijichthyol.org/ DOI: https://doi.org/10.22034/iji.v12i3.1109

al. 2023) but are collapsing due to marine pollution. Hence, this valuable species is located in the "Conservation dependent organisms" list of the IUCN due to habitat limitation and decrease in population size (Esmaeili et al. 2015). Given the current situation of *R. frisii*, there are a range of studies that focused on biological aspects (Borujeni et al. 2015; Fazli & Daryanabard, 2020; Gheisvandi et al. 2015), reproduction (Heyrati et al. 2007) and trace elements determination (Sattari et al. 2020). However, research on population diversity, structure, genetic conservation and protection were limited (Kashiri et al. 2018). Limited genetic markers, including mtDNA (Kolangi Miandareh et al. 2015), DNA microsatellites (Rezaei et al. 2010; Abdolhay et al. 2012; Safari, 2016; Kashiri et al. 2018) and recently SNP based genome (Rahbar et al. 2023) has been used to study the population structure of *R. frisii*. However, most of these studies focused on comparison populations of SS stocks and research on differences between AS and SS are seldom. Nevertheless, despite representation of fisheries restocking programs (just for SS stocks), no scientific studies have focused on advancing the different genetic mechanisms between two strains. So, to better understand these apparatuses that lead us to reprogramming the breeding and restocking plans, the generation of genomic resources for this endangered non-model species is fundamental.

The use of next-generation sequencing (NGS) technology has recently become increasingly common in the field of aquaculture and fisheries (Kukurba & Montgomery 2015). This technology provides a general representation of almost all the transcripts (i.e., mRNAs) expressed in specific cells or organs at particular conditions and times that help obtain information about gene prediction, gene pathways and gene function (Zhang et al. 2018; Nazari & Pourkazemi 2021; Nazari et al. 2021). High-throughput RNA sequencing (RNA-Seq) is a potent way of whole transcriptome analysis that identifies and quantifies RNA transcripts (Wang et al. 2009). In absence of whole genome sequencing for many non-model species, transcriptome sequencing has become more attainable for a deep understanding of differences at the genomic level (Vasemägi & Primmer 2005). There are several reports on transcriptomic analysis in species closely related to Caspian Kutum using the NGS method, which have stimulated genetic and genomic studies of Cyprinid species (Fu & He 2012; Ji et al. 2012; Wang et al. 2012). Also, this high throughput technology has been repeatedly used for *de novo* assembly and annotation of transcriptomes in several non-model fish species (Calduch-Giner et al. 2013; Thanh et al. 2015 Pomianowski et al. 2021). Nevertheless, for asking this question, such as which certain genes are responsible for differentiation of *R. frisii* AS and SS stocks, a transcriptome based on genome makes it feasible and manageable to answer. In the current study, as shown in the Figure 1, we performed *de novo* comprehensive reference transcriptome profiles using paired-end RNA-seq approach from the hepatic tissue for two stocks of *R.frisii* and Comparative transcriptomic analyses of two stocks to identify the DEGs between them.

MATERIALS AND METHODS

Ethic statement: All experiments were conducted in accordance with the guidelines of the National Institute of Health Guide for the Care and Use of Laboratory Animals and approved by the Research Ethics Committee, University of Guilan, Rasht, Iran (approval ID: IR.GUILAN.REC.1401.066).

Fish sampling: Fish sampling was conducted in two periods: 1- In the middle of December for autumn strain (AS) and 2- In the middle of February for spring strain (SS). Three adults AS females were caught by dam traps from Anzali lagoon in the north part of Iran and three adult SS females were trapped by dams in the breeding station of Caspian Kutum in Havigh River. Mean body weights (BW) and lengths (L) of AS (1216.76±10.5g, 48.5±2.2cm) and SS (889.33±144.5g, 43.83±4.5cm) were recorded, respectively. The fishes were moved into breading facilities of university of Guilan-faculty of natural resources-Sowmehsara-Iran, and kept in 100L tanks for two days to stabilization. Then, each fish anesthetized separately in cloves powder (160ppm) to

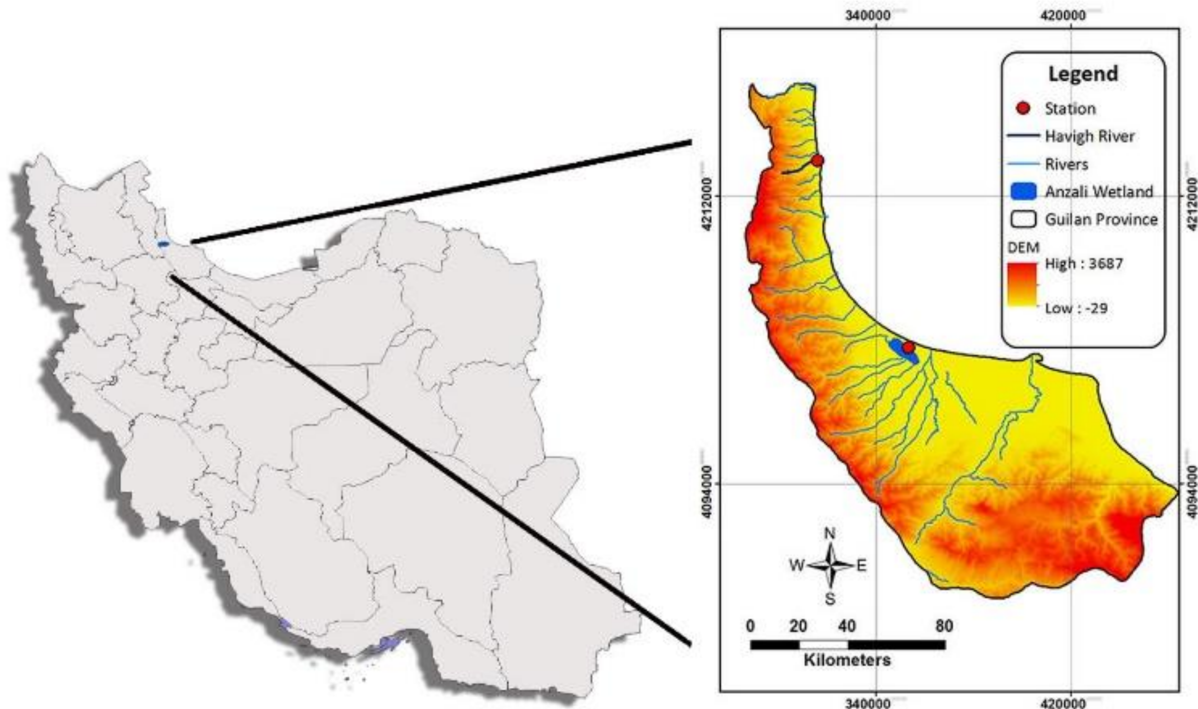

Fig.S1. Sampling stations for *Rutilus frisii*.

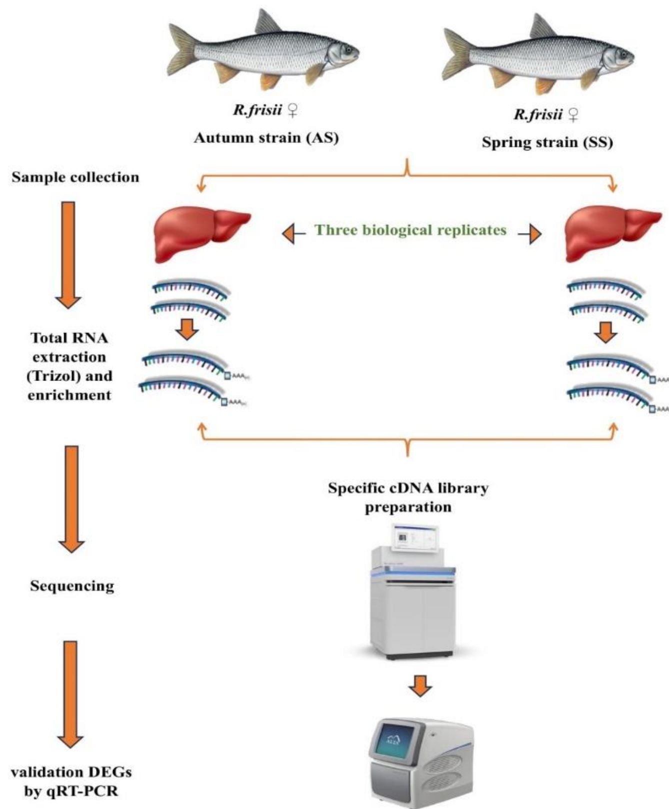

Fig.S2. The workflow of *de novo* transcriptome data analysis for *Rutilus frisii*.

minimize suffering. After complete anesthesia, the liver was collected from each fish. The tissues immediately preserved in RNA later, frozen in liquid nitrogen and, were stored at -80°C until RNA extraction and transcriptome analysis (Figs.1 and 2).

RESULTS

Histology: The ovaries of female, AS and SS fishes were obtained and one cm^2 of each ovary was sliced, fixed in Bouin's solution, dehydrated and embedded in paraplast. Hematoxylin-eosin method was used to staining the sections (5µm) were cut using a microtome (Leica, Wetzler, Germany). Photomicrographs were taken using an Olympus CX41 bright field light microscope (Olympus, Tokyo, Japan).

RNA extraction: Total RNA was isolated from liver samples from both strains of *R. frisii* using TRIzol reagent (Invitrogen; www.invitrogen.com). PrimerScript RT Reagent Kit with gDNA Eraser (TaKaRa; www.takarabio.com) was used to synthesize the first-strand complementary (c) DNA according to the manufacturer's instructions. The purified RNA was dissolved in RNase-free water, and genomic DNA contamination removed using DNase I (Invitrogen). RNA concentration and purity were measured using the NanoDrop spectrophotometer (ND- 1000, Nanodrop Technologies, and RNA integrity was checked using

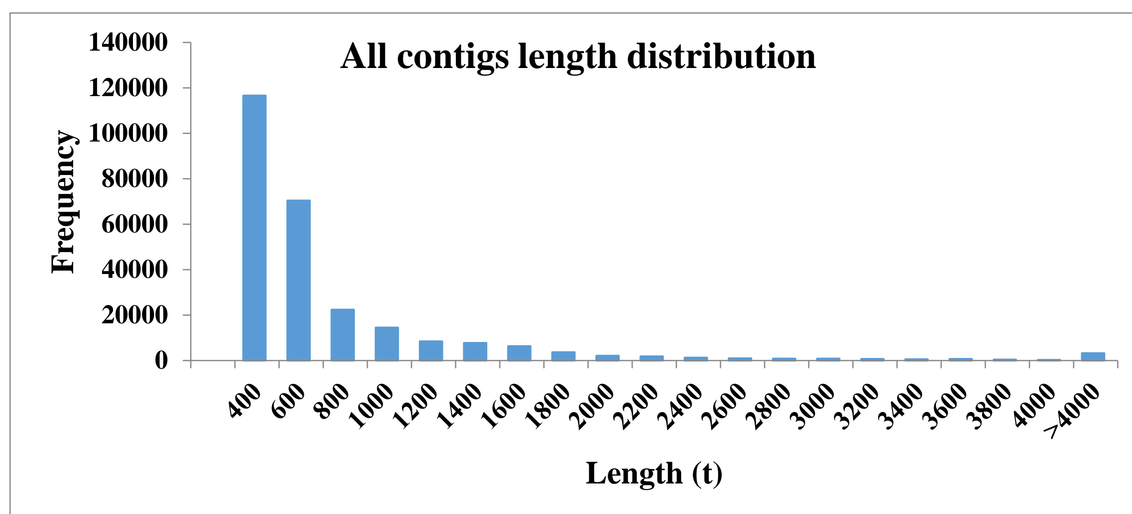

Fig.3. The distribution of all the *de novo* assembled transcripts with different size.

Table 1. Number of raw reads, gene counts and specifically expressed genes of *Rutilus frisii* transcriptome sequencing samples.

Sample	Number of raw reads	Number of reads after trimmed	gene count	Specifically expressed genes
RFS1	67,554,345	63,155,214	67,415	215
RFS2	79,124,741	71,567,042	76,416	348
RFS3	65,746,442	61,354,963	101,733	196
RFA1	90,669,238	83,675,118	99,452	578
RFA2	72,285,476	68,719,062	115,371	214
RFA3	81,353,247	77,855,278	121,485	615
Total	456,777,489	426,326,677		

agarose gel electrophoresis and BioAnalyzer measurements. Only high-quality RNA samples (OD260/OD280 ranged 1.8-2.2, RIN≥8.0) were used to construct the sequencing library. The RNA of each strain's livers was pooled into three samples for later sequencing.

Illumina cDNA Library preparation and sequencing: The RNA purification and library construction procedures were performed using the TruSeq RNA sample prep kit (Illumina). Magnetic Oligo (dT) Beads (Invitrogen) for mRNA purification in each sample. Purified mRNA was randomly fragmented by the fragmentation buffer for synthesizing first-strand cDNA with random hexamer primers, followed by second-strand cDNA synthesis. Double-stranded cDNA was purified after agarose gel electrophoresis, further subjected to 3′ end single nucleotide A (adenine), and ligated to the sequencing linker. Suitable fragments were selected after purification for PCR amplification. The library was purified and quality assessed using an Agilent Bioanalyzer 2100 system. Finally, high-throughput sequencing was conducted on the Illumina sequencing platform (HiSeq™ 4000) (Illumina, San Diego, CA, USA) to generate 125bp paired-end reads.

Quality assessment and *de novo* transcriptome assembly: In the present study, the raw sequencing reads trimmed with FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to remove reads which contained adaptors, unknown sequences (N) over 5%, and those containing greater than 20% of low-quality nucleotides (Q-value lower than 10), (Yang et al. 2015). Quality control of sequenced reads was investigated based on each nucleotide by Fastx Toolkit (Q>20) and adapters, high poly-N and low-quality sequences were removed from the raw. The Trinity software package (v2.3.2)

Table 2. Statistics of *Rutilus frisii* transcriptome sequences assembly.

Summary statistics	
Not matched reads count	46,824,391
Number of contigs	138,462
Total reads in pairs	128,328,799
Broken paired reads	34,108,276
Mean length of reads in bp	1,108
Contigs measurements (including scaffolded regions)	
N75 of unigenes in bp	289
N50 of unigenes in bp	1,235
N25 of unigenes in bp	1,884
Minimum length in bp	201
Maximum length in bp	17,157
Mean length in bp	468
Total length	58,452,041
Contigs measurements (excluding scaffolded regions)	
N75 of unigenes in bp	319
N50 of unigenes in bp	867
N25 of unigenes in bp	1,905
Minimum length in bp	112
Maximum length in bp	15,305
Mean length in bp	515
Count	126,438
Total length	81,799,542

(http://trinityrnaseq.sf.net) was used to *de novo* assemble high-quality transcriptome reads using the paired-end method (Grabherr et al. 2010) with default parameters to get the unique transcript fragments (unigenes). Contigs were formed by combining specific reads with overlapping sequences, and then reads were transformed to a contigs map. Finally, contigs were clustered into transcripts and, unigenes were obtained by removing the redundancies in these transcripts.

In order to compare and analyze de novo and our reference-guided *de novo* assembly approaches (Grabherr et al. 2011), scaffolds numbers, longest scaffolds, the percentage of gaps in the assembly and scaffolds numbers larger than 1000bp, N50 of *all scaffolds* for quality metrics (the length of the shortest contigs for which larger and equal length contigs cover at least 50% of the assembly), in combination with other thresholds such as N25 and N75, were calculated.

Functional annotation of unigenes: Assembled unigenes were aligned against the NCBI non-redundant nucleotide database (NR) (http://www.ncbi.nih.gov), using BLASTn with an E-value cut-off of 10−5. Then, this data was explored in other public databases such as gene ontology (GO) using Blas2GO software, COG, KOG (Eukaryotic Ortholog Groups, http://www.ncbi.nlm.nih.gov/KOG/) employing BLASTX search with a cut-off E-value threshold lower than 10^{-5} (Altschul et al. 1990), and KEGG (Kyoto Encyclopedia of Genes and Genomes) database (http://www.genome.jp/kegg/pathway) using the BLASTX algorithm with E-values of less than 10^{-5} (Wixon & Kell 2000). Gene Ontology for Nr annotated unigenes was performed with Blast2GO package to estimate what the Go idiom is related to unigenes (Conesa et al. 2005), and subsequently, to classify function in all unigenes and dispensation of gene functions in *R. frisii* WEGO package (http://wego.genomics.org.cn/cgi-bin/wego) were employed (Ye et al. 2006).

Analysis of differentially expressed genes (DEGs): The sequence reads were *assembled* according to the *de novo* method and subsequently the reference transcriptome was created. The DESeq R package was used to identify differentially expressed genes (DEGs) in the liver tissues of each strain (P<0.05). Briefly,

cleaned reads were mapped back to the assembled contigs using the package Bowtie 2 (v. 2.2.5), and the number of *reads* count for genes were estimated from the results by RSEM software (v.1.2.12). Then, each gene was normalized into the Reads per Kilobase per Million mapped reads (RPKM) procedure to mapping the reads count (Roberts et al. 2011). In this paper, genes with a fold change≥2 and FDR≤0.05 (adjusted *P*≤0.05) were classified as significant DEGs.

RESULTS

In the present study, a total of 456,777,489 paired-end reads were created utilizing the HiSeq *technology sequencing* (Table 1, Fig. 3). After the ambiguous nucleotides (lower than 10%) and the low-quality reads (*quality*<20) were removed, over 426 million (93.34%) good quality reads were used for de novo transcriptome assembly. The de novo transcriptome *assembly* detected 138,462 contigs and the *length* of this contigs obtained varied from 201 to 17,157bp. The assembly had a N50 value of 1,235bp and the average length of the contigs was 468bp (Table 2).

Mappings of the reads to the reference genome: In this study the reads to several reference genomes were mapped and applied the resulting mapping data to improve the output of the de novo assembly in subsequent steps. For mapping onto the reference genome and after removing the contigs less than 200 bp, a total of 149,857 contigs were included into the assembly sequence which were longer than 220 bp. (The distribution of all the de novo assembled transcripts with different size was shown in Figure 3).

Six libraries from two strains of *R. frisii* were generated in which for the spring strain sample, the mean percentage of clean reads mapped and not mapped in pairs were 62.25% and 16.84%, respectively. The mean percentage of clean reads mapped in pairs for the autumn strain was 68.54%, while the percentage of clean reads not mapped was 19.35%, respectively. Moreover, a significant number of long sequences were generated in the assembly sequences. For example, a number of 58,325 transcripts were longer than 800 bp that means these transcripts accounted for 24.86% of whole transcripts.

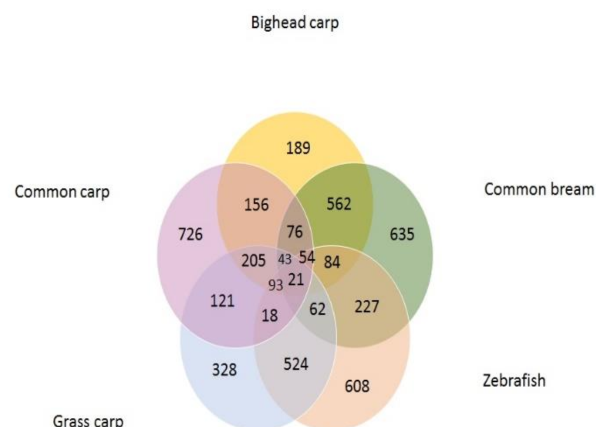

Fig.4. Venn diagram showing the number of overlapping proteins between the five cyprinid transcriptome assemblies. The diagram was built with a web-service (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Functional annotation: Of 138,462 assembled contig sequences, a total number of 11,343 contigs in which included 8.19% of the whole assembled contigs, showed significantly hit with those searching against the five reference protein sequence databases including common carp *Cyprinus carpio*, bighead carp *Hypophthalmichthys nobilis,* grass carp *Ctenopharyngodon idella*, zebrafish *Danio rerio*, and common bream *Abramis brama* using blasting program. According to the distribution of the similarities of the blasting results, a total of 5,364 unigenes shared among five cyprinid species (Fig. 4).

Specifically expressed genes and involved pathways: The gene ontology (GO) assignment procedures were used to categorize functional annotations of the assembled unigenes. All unigenes of *R. frisii* were classified into 62 sub-categories belonging to three major GO categories (biological process, cellular components and molecular functions) using the Blast2GO (Fig. 5). Among all the unigenes, 79.41% were assigned to cellular components, followed by biological processes 58.23% and 51.72% to molecular function. In the biological process category, cellular process, single organism cellular process and primary metabolic process were found to have the most common items. Within the cellular components category, the

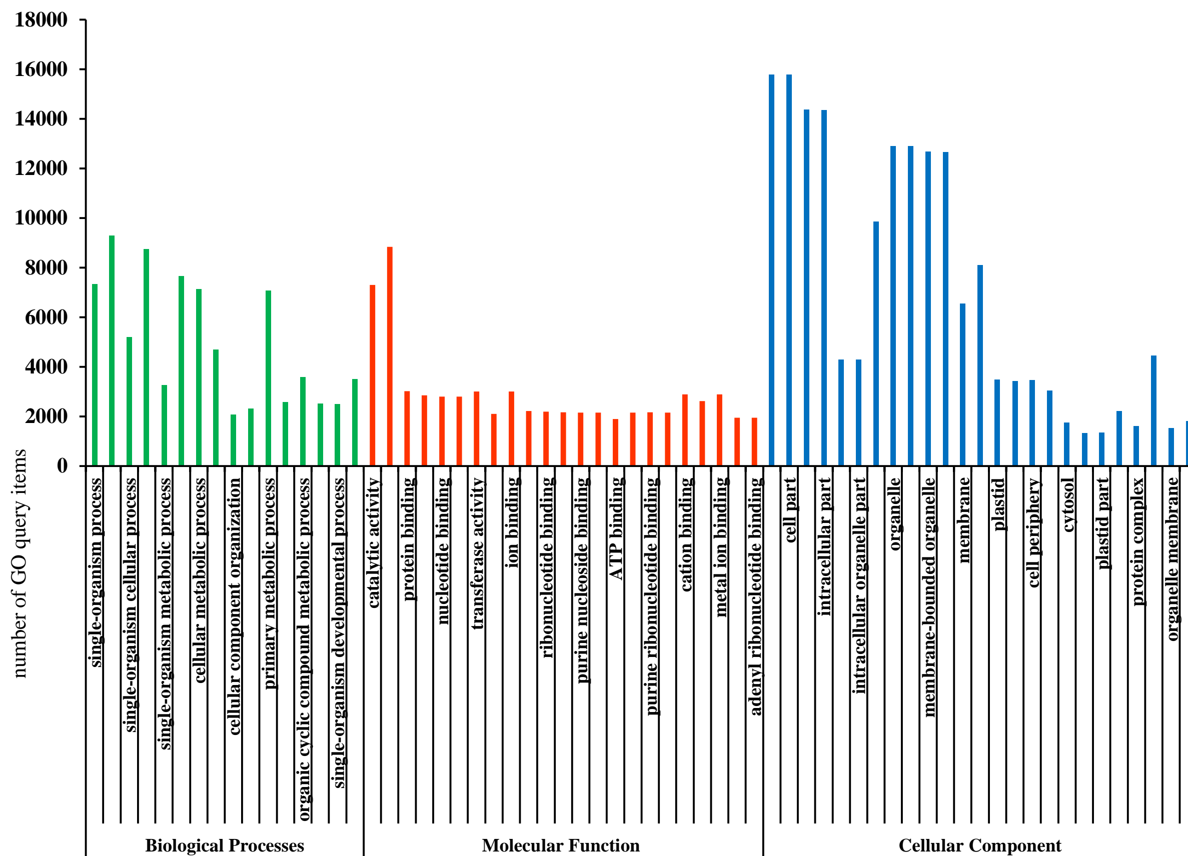

Fig.5. Distribution of all-unigenes and DEGs among GO terms in biological processes, cellular component, and molecular function. Unigenes in *Rutilus frisii* transcriptome are classified into biological processes, cellular component and molecular function.

predominant GO terms were grouped in cell part, intercellular part and membrane bounded organelle were found to have the most abundant terms. Within the molecular function, antioxidant activity and catalytic activity constituted the majority terms of the category. A number of the unigenes were grouped into the spawning migration related sub-categories such as, response to estrogen stimulus, reproduction and developmental process. These results indicate that such unigenes may be involved in development of female gonads in *R. frisii* and the DEGs with higher impacts on the spawning migration characteristics.

The contigs assembled of *R. frisii were* annotated based on previous known sequences in the *KEGG pathway analysis*. The results of the *KEGG pathway analysis* showed that 12099 unigenes were classified into different functional groups (Table 3). The greatest number of transcripts was for organismal system (31.03%, 3755), be contained in immune system (965), endocrine system (915), circulatory system (276), digestive system (402), excretory system (183), nervous system (542), sensory system (119) development (275) and environmental adaptation (98). Unigenes grouped in the *KEGG categories* into the metabolism (2844, 23.50%), in which the greatest number of transcripts involved in global and overview maps (1023), carbohydrate metabolism (285), lipid metabolism (221), nucleotide metabolism (156), amino acid metabolism (186), glycan biosynthesis and metabolism (274), metabolism of cofactors and vitamins (123).

The next common number of transcripts were for environmental information processing (2547 (21.05%)), cellular processes (1496 (12.36%)) and genetic information processing (1437 (11.87%)). Then, the unigenes were also *aligned* to the KOG database for predicting and classifying and functions of the unigenes. The possible functions of 8862 unigenes were clustered into 25 KOG classifications. The most abundant categories were general function prediction only, Signal transduction, posttranslational

Table 3. KEGG biomedical mapping for *Rutilus frisii*.

KEGG categories represented		Number of unigene sequences	Number of mapped KO
Metabolism		2844	2496
	Global and overview maps	1023	1211
	Carbohydrate metabolism	285	285
	Energy metabolism	221	150
	Lipid metabolism	276	198
	Nucleotide metabolism	156	124
	Amino acid metabolism	186	134
	Metabolism of other amino acids	109	62
	Glycan biosynthesis and metabolism	274	142
	Metabolism of cofactors and vitamins	123	99
	Metabolism of terpenoids and polyketides	18	15
	Biosynthesis of other secondary metabolites	72	34
	Xenobiotics biodegradation and metabolism	101	42
Genetic information processing		1437	895
	Transcription	324	156
	Translation	484	298
	Folding, sorting and degradation	395	324
	Replication and repair	234	117
Environmental information processing		2547	1748
	Membrane transport	49	19
	Signal transduction	2012	1355
	Signaling moleculars and interaction	486	374
Cellular processes		1496	1117
	Transport and catabolism	514	326
	Cell motility	173	118
	Cell growth and death	286	298
	Cellular community	523	375
Organismal systems		3775	2606
	Immune system	965	721
	Endocrine system	915	586
	Circulatory system	276	238
	Digestive system	402	259
	Excretory system	183	98
	Nervous system	542	410
	Sensory system	119	59
	Development	275	177
	Environmental adaptation	98	58
Total		12099	8862

modification, protein turnover, chaperones while the nucleotide transport and metabolism, cell motility and nuclear structure were three smallest clusters (Fig. 6).

DISCUSSION

Using the NGS sequencing methods can give an overview at the whole transcriptome level of genomic activities, functional complexity of transcriptomes in many aquatic animals in different habitats including river, sea or physiological conditions (Martin & Wang 2011; Liu et al. 2012; Cui et al. 2020; Emam et al. 2022). Furthermore, identification of genes, gene associated markers, and regulatory non-coding RNAs; alternative splicing analysis can be understood using high-throughput sequencing (Zhang et al. 2017; Li et al. 2020). Among bony fish stocks, due to the importance of the Caspian Kutum *R. frisii*, in the southern basin of Caspian Sea, understanding of the knowledge of the molecular mechanisms of this species during seasonal spawning is particular. This information is useful for the future management strategies of broodstocks of *R. frisii*, as this species migrate to the river and Anzali

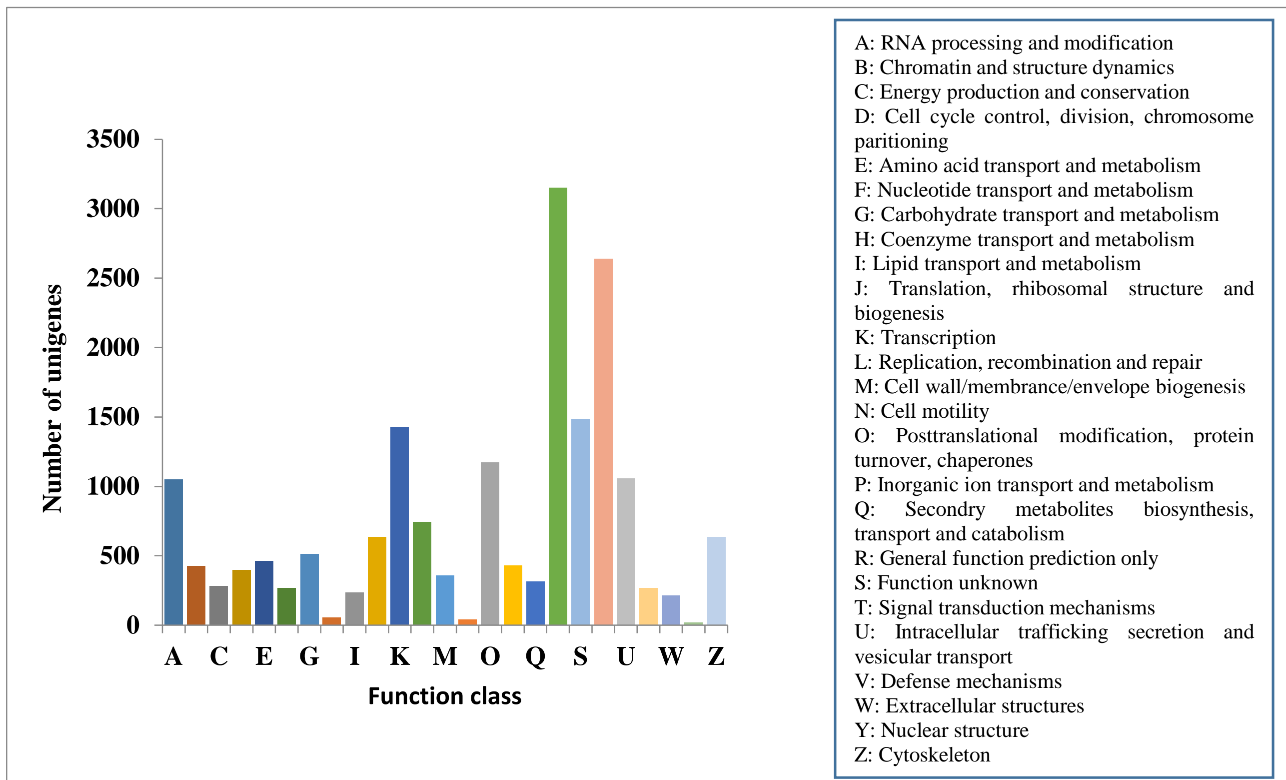

Fig.6. KOG classification of the assembled unigenes. The 25 group names (listed on the right) are indicated on the X-axis; the Y-axis shows the number of genes in the group out of annotated genes. Among them the general function prediction, signal transduction mechanisms and transcription were the most abundant terms.

lagoon in the southern parts of Caspian Sea to spawn in their life cycle. In fish like anadromous one, the liver plays vital roles as it is involved in the energy metabolism, including protein and lipids, to be processed and used for ovarian and follicular progress including formation yolk globules in the developing oocytes and to provide energy needed for keeping and, swimming and then during spawning in which metabolism becomes increasingly aerobic, and the capacity for fatty acid utilization enhances with some aspects phenotypic changes related with sexual maturation (Kiessling et al. 1995; Seli et al. 2014). To seek the physiological mechanisms underlying *R. frisii* spawning migration, liver tissues were applied to investigate the expression profiles of the DEGs between two strain samples using RNA-Seq technique. However, in comparison with other cyprinid species, little information is available for the transcriptome of different organs and tissues of *R. frisii.* The results of the present study can provide a reference genome for the Caspian Kutum transcriptome analyses and also more genomic data for the investigation of many putative genes that related to spawning migration and the controlling reproduction in the female *R. frisii.*

The transcriptome sequencing of *R. frisii* primarily was surveyed to detect a wide range of genes which could be related with spawning migration and reproduction. The results of annotations will acknowledge scientists to observe the correct gene annotation and also to detect target genes of interest more easily than would be obtained by degenerate PCR method, especially in cyprinid species, due to the paucity of functional information for anadromous cyprinid species genomes. To detect the spawning migration and reproduction-related genes, we searched our annotated transcriptome dataset using three principal strategies as described by Jung et al. (2013). In this manner, 25 reproduction–related unigenes were identified and afterwards classified into 24 categories. The results showed that most of the annotated sequences had high similarities of the mapped sequences with known sequences, indicated that the quality of the reference transcriptome of *R. frisii* had reliable assembly. The top-hit species distribution

analysis based on BLASTx results showed that the common bream provided most of the BLASTx hits for *R. frisii* annotated unigenes from E-value. Only, a low proportion of the unigenes had no BLASTx hits. This could be explained by some of the non-annotated transcripts were 5′ untranslated transcribed regions (UTR), 3′UTR regions of protein-coding and non-coding RNAs (lncRNAs), or short sequences containing no known protein domains (Gao et al. 2014).

During the spawning migration, the reproductive stage is the most sensitive and vulnerable stage in fish that is considerably affected by external environmental factors. Variations in the flow, salinity, water temperature and turbidity affect the three main stages of reproduction, namely induction (the beginning of oogenesis), vitellogenesis (yolk formation) and maturation (including ovulation and oviposition) in the fish. The reproduction process, generally including gonad differentiation, development, maturation and gametogenesis, is more complex in aquatic species than in other vertebrate species and in addition to genetic, environmental factors has a profound effect on reproduction in aquatic species. Previous studies have indicated that information of the mechanisms related to reproductive regulation could vigorously enhance the breeding efficiency of aquaculture species (Yang et al. 2016; Yang et al. 2018; Tian et al. 2019). In comparison with other commercial aquaculture species, the genetic knowledge available for the Caspian Kutum is limited. Notwithstanding, reproduction-related genes and molecular aspects of reproductive activity have been hardly ever described in *R. frisii* and a basic knowledge of their expression profiles is still lacking. Prior to clarifying the specific functions of these genes, it is necessary to enrich to genomic or transcriptome information.

The results of the GO assignment results revealed that most annotated unigenes were classified into 62 sub-categories under three main ontologies (biological process, cellular components and molecular functions), which was in congruent with previous work. In this study, the GO functional annotation results indicated that various transcripts in our transcriptome data for the Caspian Kutum were categorized and could be useful for describing the actions of a gene product at the molecular level and therefore predict the possible physiological roles of all unigenes. The GO analysis was mostly in consistent with other analysis of DEG(s). Genes involved in some important biological processes such as positive regulation of gene expression, reproduction, response to estrogen stimulus and signalling were also determined. Moreover, KEGG pathways, which is an alternative method to classify gene interaction and biological functions with the emphasis on biochemical pathways, were divided into several categories (32 subcategories), organismal, immune and endocrine system, suggesting the significance of these systems in reproductive process and function in the Caspian Kutum. Obviously, the genes in such functions may be more conserved across different species by their binary presence/absence and are thus straightforward to annotate in the database. Functional classification provided according to the KEGG pathways showed that many significant metabolic pathways within the Caspian Kutum are still unrevealed and need to be more investigated.

Our findings reveal that *R. frisii* makes an enormous investment in gene transcription, translation and folding, sorting and degradation, as well as membrane transport, signal transduction and signalling molecular and interaction. Most of the top hit pathways were involved in organismal systems, metabolism and environmental information processing, whereas the other pathways were related to those involved in genetic information processing and cellular process. Briefly, functional analysis indicated that RNA-seq-based de novo transcriptome sequencing and gene expression profiling for *R. frisii* will accelerate more studies associated with basic physiological and metabolic functions and molecular genetics of *R. frisii* or related species.

In aquatic species especially invertebrates, it is widely accepted that gonadotropin-releasing hormone 1 (*gnrh1*) (Guan et al. 2014), D(2) dopamine receptor (*drd 2*) (Cournil et al. 1995; Zhu et al. 2005), GtH II beta subunit (*gth 2*) and putative gonadotropin-releasing hormone II receptor (*gnrhr2*) (Kawada et al. 2013; Plachetzki et al. 2016) play pivotal roles in growth,

molting and reproductive process via the function as neuropeptides that directly regulate target tissues (Suwansa-ard et al. 2016). It is reported that the expression of GnRH receptors (GnRHRs) as an important neuropeptides (NPs), in various tissues also supports non-hypothalamic functions of invertebrate GnRHs, as these species lack orthologs of gonadotropin hormones and pituitary glands (Millar 2005). Moreover, it has been shown that these NPs regulate exceptionally complicated signaling pathways involving ligand-receptor selectivity, coupling with multiple G-protein subtypes, and receptor hetero-dimerization (Millar et al. 2008). Although the androgenic hormones are commonly present in many aquatic species (Ngernsoungnern et al. 2009), the pathway through which it mediates spawning migration and reproductive process, growth and gonadal development is poorly understood in cyprinids. For the past several years, little improvement had occurred in the description of other gene variation in aquatic invertebrates. Recent genomic studies have revealed the existence and evolution of cytochrome P450 superfamily across other aquatic species in general. The superfamily cytochrome P450's including cytochrome P450 family 19 sub-family A polypeptide 1b (*cyp 19 a 1 b*) detected for *R. frisii*, are involved in the metabolism of endogenous molecules (including steroids, fatty acids and ecdysteroids) and xenobiotic compounds (e.g. hydrocarbons, pesticides, drugs) (Dauphin-Villemant et al. 1999; Snyder 2000; Wang & Guo 2007; Jiang et al. 2014; Khoshkholgh & Nazari 2019).

CONCLUSION

We collected wild female samples of *R. frisii*, an economically important cyprinid fish in Caspian Sea, and conducted RNA-Seq experiment using liver tissues to obtain the transcriptome profile using Illumina™ platform HiSeq technology. We further identified a number of well-known transcription factors that are possibly involved in spawning migration and reproduction. This study will enrich genetic information for *R. frisii* and provide a basis for further research into the regulatory mechanisms controlling spawning migration and reproduction at the molecular level in *R. frisii*.

ACKNOWLEDGEMENTS

We would like to the university of Guilan and Iran national science foundation (INSF) for supporting the research and financing. We are also appreciating our colleagues in the college of fisheries science.

FUNDING

Funding was provided by Iran National Science Foundation (INSF) research fellowship to M. Khoshkholgh (INSF-4013197).

REFERENCES

Abdolhay, H.; Daud, S.K.; Rezvani, S.; Pourkazemi, M.; Siraj, S.S.; Laloei, F.; Javanmard, A. & Hassanzadeh Saber, M. 2012. Population genetic structure of Mahi Sefid (*Rutilus frisii* kutum) in the South Caspian Sea: Implications for fishery management. Iranian Journal of Animal Biosystematics 8(1): 15-26.

Abdolhay, H.A.; Daud, S.K.; Rezvani Ghilkolahi, S.; Pourkazemi, M.; Siraj, S.S. & Abdul Satar, M.K. 2011. Fingerling production and stock enhancement of Mahisefid (*Rutilus frisii* kutum) lessons for others in the south of Caspian Sea. Reviews in Fish Biology and Fisheries 21(2): 247-257.

An, H.S. & Lee, J.W. 2012. Development of microsatellite markers for the Korean mussel, *mytilus coruscus* (Mytilidae) using next-generation sequencing. International Journal of Molecular Sciences 13(8): 10583-10593.

Belyaeva, V.N.; Kazancheev, E.N. & Raspopov, V.M. 1989. The Caspian Sea: Ichthyofauna and commercial resources. House Nauka, Moscow.

Berg, L.S. 1949. Freshwater fishes of the USSR and adjacent countries. Program for scientific translations, Jerusalem 2: 496.

Borujeni, P.A.; Steinshamn, S.I.; Mortazavi, S.A. & Salehi, H. 2015. Enhance and advance: The benefits of recruitment enhancement in the case of the Iranian Kutum Fishery. Marine Policy 61: 23-32.

Calduch-Giner, J.A.; Bermejo-Nogales, A.; Benedito-Palos, L.; Estensoro, I.; Ballester-Lozano, G.; Sitjà-Bobadilla, A. & Pérez-Sánchez, J. 2013. Deep sequencing for *de novo* construction of a marine fish (*Sparus aurata*) transcriptome database with a large

coverage of protein-coding transcripts. BMC Genomics 14(1): 1-11.

Conesa, A., Götz, S.; García-Gómez, J.M.; Terol, J., Talón, M., & Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18): 3674-3676.

Cui, W.; Ma A.; Huang Z.; Wang XA.; Sun Z.; Liu Z.; Zhang W.; Yang J.; Zhang J.; Qu J. 2020. Transcriptomic analysis reveals putative osmoregulation mechanisms in the kidney of euryhaline turbot Scophthalmus maximus responded to hypo-saline seawater. Journal of Oceanology and Limnology 38(2): 467-479

Dauphin-Villemant, C. D.; Böcking, M.; Tom, M. Maïbèche, R.; Lafont, 1999. Cloning of a novel cytochrome P450 (CYP4C15) differentially expressed in the steroidogenic glands of an arthropod, Biochemical and Biophysics Res. Communications. 264(1999): 413-418

Derzhavin, A.N. 1934. Freshwater fish of the southern coast of the Caspian region. Tr Azerb Otd Zakavkaz Fil Sektor Zoology 7(3): 91-126.

Eagderi, S.; Mouludi-Saleh, A.; Esmaeli, H. .; Sayyadzadeh, G. & Nasri, M. 2022. Freshwater lamprey and fishes of Iran; a revised and updated annotated checklist-2022. Turkish Journal of Zoology 46(6): 500-522.

Emam, M.; Caballero-Solares, A.; Xue X.; Umasuthan, N.; Milligan, B.; Taylor, RG.; Balder R. & Rise, ML. 2022. Gill and liver transcript expression changes associated with gill damage in Atlantic Salmon (*Salmo salar*). Frontiers in Immunology 13(13): 806484

Fazli, H.; Daryanabard, G.; Salmanmahiny, A.; Abdolmaleki, S.; Bandani, G. & Bandpei, M.A.A. 2012. Fingerling release program, biomass trend and evolution of the condition factor of Caspian Kutum during the 1991-2011 period/Programme d'alevinage et evolution de la biomasse et du coefficient de condition du kutum de la mer Caspienne au cours de la periode 1991-2011. Cybium, International Journal of Ichthyology 36(4): 545-551.

Fu, B. & He, S. 2012. Transcriptome analysis of silver carp (*Hypophthalmichthys molitrix*) by paired-end RNA sequencing. DNA Research 19(2): 131-142.

Gheisvandi, N.; Hajimoradloo, A.; Ghorbani, R. & Hoseinifar, S.H. 2015. The effects of gradual or abrupt changes in salinity on digestive enzymes activity of Caspian kutum, *Rutilus kutum* (Kamensky, 1901) larvae. Journal of Applied Ichthyology 31(6): 1107-1112.

Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R. .; Zeng, Q. and Chen, Z. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7): 644-652.

Guan, Z.B.; Shui Y.; Liao X.R.; Xu Z.H. & Zhou X. 2014. Primary structure of a novel gonadotropin-releasing hormone (GnRH) in the ovary of red swamp crayfish *Procambarus clarkii*. Aquaculture 418-419: 67-71.

Holčik, J. 1995. New data on the ecology of kutum, *Rutilus frisii* (Nordmann, 1840) from the Caspian Sea. Ecology of Freshwater Fish 4(4):175-179.

Jiang, Q.; Min Y.; Yang, H.; Wan, W. & Zhang, X. 2019. De novo transcriptome analysis of eyestalk reveals ovarian maturation related genes in *Macrobrachium rosenbergii*. Aquaculture 505(): 280-288.

Kashiri, H.; Shabani, A.; Gorgin, S.; Rezaii, M. & Jabale, A.R. 2018. Microsatellite DNA Marker Analysis of Genetic Variation in Wild and Hatchery Populations of Caspian Kutum (*Rutilus kutum*). Turkish Journal of Fisheries and Aquatic Sciences 18(9): 1101-1111.

Kiessling, A.; Larsson, L.; Kiessling, K.H.; Lutes, P.B.; Storebakken, T. & Hung, S.S. 1995. Spawning induces a shift in energy metabolism from glucose to lipid in rainbow trout white muscle. Fish Physiology and Biochemistry 14(6): 439-448

Kolangi Miandareh, H.; Shabany, A. & Hojati, M., 2015. Investigating genetic diversity of *Rutilus kutum* (Kamenskii, 1901) in some rivers in southern of the Caspian Sea using Cytochrome b gene (mtDNA-Cytb) sequences. Journal of Applied Ichthyology Research 3(2): 1–12.

Kukurba, K.R. & Montgomery, S.B. 2015. RNA sequencing and analysis. Cold Spring Harbor Protocols 2015(11): 084970.

Li, J.; Xue L.; Cao, M.; Zhang, Y.; Wang, Y.; Xu, S.; Zheng, B. & Lou, Z. 2020. Gill transcriptomes reveal expression changes of genes related with immune and ion transport under salinity stress in silvery pomfret (*Pampus argenteus*). Fish Physiology and Biochemistry 46(4): 1255-1277.

Liu, S.; Zhang, Y.; Zhou, Z.; Waldbieser, G.; Sun, F.; Lu, J.; Zhang, J.; Jiang, Y.; Zhang, H.; Wang, X.; Kucuktas, H.; Peatman E. & Liu, Z. 2012. Efficient

assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote. BMC Genomics 5 (13), 595.

Livak, K.J. & Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods 25(4): 402-408.

Magalhães, M.G.P.; Moreira, D.A.; Furtado, C. & Parente, T.E. 2019. Development of SNVs and indels markers mined out of the first multi-organ transcriptomes from *Hypancistrus zebra* (Loricariidae), an endangered Amazonian catfish. Genomics 112(1), 971-980.

Magomedov, G. M.; Aliev, D.A. & Proskurina, V. O. 1987. Reproduction of Kutum off Dagestan coast of the Caspian Sea. Rybnoye Khoziaistvo 1(5): 37-39.

Martin, J.A. & Wang, Z. 2011. Next-generation transcriptome assembly. Nature Reviews Genetics 12(10): 671-682.

Mashhor, M.A.M.; Abdolmaleki, S.H.; Najafpour, S.H.; Bani, A.; Pourgholam, R.; Fazli, H.; Nasrolahzadeh, H. & Janbaz, A.A. 2011. The environmental effect on spawning time, length at maturity and fecundity of Kutum (*Rutilus frisii* kutum, Kamensky, 1901) in southern part of Caspian Sea, Iran. Iranian (Iranica) Journal of Energy & Environment 2(4): 374-381.

Millar, R.P.; Pawson, A.J.; Morgan, K.; Rissman, E.F.; Lu, Z.L. 2008. Diversity of actions of GnRHs mediated by ligand-induced selective signaling. Frontiers in Neuroendocrinology 29(1): 17-35.

Nazari, S. & Pourkazmi, M. 2021. Isolation and characterization of SNP markers of rainbow trout (*Onchorhynchus mykiss* Walbaum, 1792) from transcriptomic sequences. Molecular Biology Reports 48(1): 989-995.

Nazari, S.; Pourkazemi, M.; Paknejad, H.; Kazemi, E.; Ghaderi, M. & Eslamloo, K. 2021. Transcriptome profiling of farmed rainbow trout (*Oncorhynchus mykiss*) liver from different sources of dietary zinc. Aquaculture 543(15): 737017.

Ngernsoungnern, P.; Ngernsoungnern, A.; Sobhon, P. & Sretarugsa, P. 2009. Gonadotropin-releasing hormone (GnRH) and a GnRH analog induce ovarian maturation in the giant freshwater prawn, *Macrobrachium rosenbergii*. Invertebrate Reproductive Development 53(3): 125-135

Plachetzki, D.C.; Tsai, PS.; Kavanaugh, S.I. & Sower, S.A. 2016. Ancient origins of metazoan gonadotropin-releasing hormone and their receptors revealed by phylogenomic analyses. General and Comparative Endocrinology 234(1): 9-10

Pomianowski, K.; Burzyński, A. & Kulczykowska, E. 2021. A *de novo* Transcriptome Assembly of the European Flounder (Platichthys flesus): The Preselection of Transcripts Encoding Active Forms of Enzymes. Frontiers in Marine Science, *8*, 618779.

Rahbar, M.; Khoshkholgh, M. & Nazari, S. 2023. Population structure of Caspian Kutum (*Rutilus frisii*, Nordmann, 1840) in the southern coast of Caspian Sea using genome-wide single nucleotide polymorphism markers. Fisheries Research 257(1): 106499.

Rezaei, M.; Shabani, A.; Shabanpour, B. & Kashiri, H. 2010. Microsatellite diversity and population genetic structure of *Rutilus frisii* kutum in Mazandaran coasts. Iranian Journal of Biology 25(4): 548–558

Safari, R. 2016. Population structure of *Rutilus frisii kutum* in Iranian coastline of the Caspian Sea using microsatellite markers. Environmental Recourses Research 4(1): 65-74.

Sattari, M.; Imanpour Namin, J.; Bibak, M.; Forouhar Vajargah, M.; Bakhshalizadeh, S. & Faggio, C. 2020. Determination of trace element accumulation in gonads of *Rutilus kutum* (Kamensky, 1901) from the south Caspian Sea trace element contaminations in gonads. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences. 90(4): 777–784.

Seli, E.; Babayev, E.; Collins, S.C.; Nemeth, G. & Horvath, T.L. 2014. Minireview: Metabolism of female reproduction: Regulatory mechanisms and clinical implications. Molecular Endocrinology, 28(6): 790-804.

Shikhshabekov, M. M. 1979. The reproductive biology of kutum, *Rutilus frisii* kutum, the Asp, *Aspius aspius*, Vimba, *Vimba vimba persa*, and the Rudd, *Scardinus erithrophethalmus*, in the water of Daghestan. Ichthyology 19: 98-105.

Snyder, M.J. 2000. Cytochrome P450 enzymes in aquatic invertebrates: recent advances and future directions. Aquatic Toxicology 48(2): 529–547.

Suwansa-ard S.; Zhao M.; Thongbuakaew T.; Chansela P.; Ventura T. & Cummins, S.F. 2016. Gonadotropin-releasing hormone and adipokinetic hormone/corazonin-related peptide in the female prawn. General Comparative Endocrinology 236(30): 70-82.

Thanh, N.M.; Jung, H.; Lyons, R.E.; Njaci, I.; Yoon, B.H.; Chand, V.; Tuan, N.V.; Thu, V.T.M. & Mather, P.

2015. Optimizing *de novo* transcriptome assembly and extending genomic resources for striped catfish (*Pangasianodon hypophthalmus*). Marine Genomics 23(2): 87-97.

Valipour, A.R. & Khanipour, A.A. 2009. Kutum: jewel of the Caspian Sea. Iranian Fisheries Research Organization. 145P.

Vasemägi, A. & Primmer, C.R. 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Molecular Ecology 14(12): 3623-3642.

Wang, H.; Tang, L.; Wei, H.; Lu, J.; Mu, C. & Wang, C. 2018. Transcriptomic analysis of adaptive mechanisms in response to sudden salinity drop in the mud crab, *Scylla paramamosain*. BMC Genomics 19(9): 421.

Wang, R.; Sun, L.; Bao, L.; Zhang, J.; Jiang, Y.; Yao, J.; Song, L.; Feng, J.; Liu, S. & Liu, Z. 2013. Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. BMC Genomics 14(1): 1-18.

Yang, M.; Zhu, L.; Pan, C.; Xu, L.; Liu, Y.; Ke, W. & Yang, P. 2015. Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical lotus (*Nelumbo nucifera*). Scientific Reports 5(1), 13059.

Ye, J.; Fang, L.; Zheng, H.; Zhang, Y.; Chen, J.; Zhang, Z.; Wang, J.; Li, S.; Li, R.; Bolund, L. & Wang, J. 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Research 34 (suppl_2): W293-W297.

Yousefian, M. & Mosavi, H. 2008. Spawning of South Caspian kutum *Rutilus frisii* kutum in most migratory river of South Caspian Sea. Asian Journal of Animal and Veterinary Advances 3(6): 437-442.

Zhang, H.; He, L. & Cai, L. 2018. Transcriptome sequencing: RNA-seq. Computational Systems Biology: Methods and Protocols 15-27.

Zhang, P.F.; Zhou, X.R.; Pang, B.P.; Tan, Y.; Chang, J. & Gao, L.J. 2016. High-throughput discovery of microsatellite markers in *Galeruca daurica* (Coleoptera: Chrysomelidae) from a transcriptome database. Jornal of Appllied Entomology 53(1): 1058–1064.

Zhang, X.; Wen H.; Wang, H.; Ren Y.; Zhao, J. & Li, Y. 2017. RNA-Seq analysis of salinity stress-responsive transcriptome in the liver of spotted sea bass (*Lateolabrax maculatus*). PLoS ONE 12(3): e0173238.

Iran. J. Ichthyol. (2025) 12(3): 233-247
P-ISSN: 2383-1561; E-ISSN: 2383-0964
http://www.ijichthyol.org
© 2025 Iranian Society of Ichthyology
doi: 10.22034/iji.v12i3.1109

مقاله کامل

بررسی ترانسکریپتوم کبد در ماهی سفید مولد ماده دریای خزر (*Rutilus frisii*, Nordmann, 1840)

سامان فراستی[۱]، مجیدرضا خوش خلق*[۱]، سجاد نظری[۲]

[۱]گروه شیلات، دانشکده منابع طبیعی، دانشگاه گیلان. صومعه‌سرا، ایران.

[۲]مرکز تحقیقات ژنتیک و اصلاح نژاد ماهیان سردآبی شهید مطهری، موسسه تحقیقات علوم شیلاتی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، یاسوج، ایران.

چکیده: ماهی سفید خزر (*Rutilus frisii*, Nordmann, 1840) یک گونه اقتصادی مهم در میان ذخایر ماهیان استخوانی در حوضه جنوبی دریای خزر است. با این وجود، اطلاعات مربوط به رونوشت، تفاوت‌های فیزیولوژیکی و مکانیسم تنظیمی کنترل دو سویه هنوز مشخص نیست. برای پرداختن به این موضوع، ما مونتاژ de novo رونوشت کبد از شش ماهی سفید ماده، سه عدد از هر سویه، را با استفاده از توالی‌یابی RNA (RNA-seq) انجام دادیم. پس از کنترل کیفیت، مونتاژ de novo رونوشت با نرم‌افزار Trinity انجام و سپس ۱۱۰۳۸۱ رونوشت تولید شد. در مجموع ۱۲۵۶۹ تک‌ژن در سه دسته اصلی عملکردی (فرآیند بیولوژیکی، اجزای سلولی و عملکردهای مولکولی) و ۶۲ زیرگروه گروه‌بندی شدند. در مجموع ۹۶۲۷ رونوشت حداقل یک عبارت GO به آنها اختصاص داده شد. طبق پایگاه‌های داده پروتئین، با مقایسه تک‌ژن‌های مونتاژ شده، ۲۱ تک‌ژن مرتبط با تولیدمثل مشاهده و سپس ۱۲۸۴۹ ژن با بیان افتراقی (DEGs) شناسایی شدند. بافت‌شناسی تخمدان‌ها بین دو سویه مقایسه شدند تا درجه بلوغ جنسی تعیین شوند. نتایج مطالعه حاضر مبنایی برای مطالعات آینده در مورد تفاوت‌های ژنتیکی بین این دو سویه از *R. frisii* و تأثیر آنها بر ساختار جمعیت فراهم می‌نماید.

کلمات‌کلیدی: ماهی سفید، تولیدمثل، توالی‌یابی RNA، ترانسکریپتوم