De novo assembly and characterization of the liver transcriptome of Mugil incilis (lisa) using next generation sequencing

Angela Bertel-Sevilla
University of Cartagena
Juan Alzate-Restrepo
University of Antioquia
Jesus Olivero-Verbel (joliverov@unicartagena.edu.co)
University of Cartagena

Research article

Keywords: Fish, Liver, Colombian Caribbean, signal transduction, immune response, Paired-end sequencing

Posted Date: May 14th, 2019

DOI: https://doi.org/10.21203/rs.2.9591/v1

Version of Record: A version of this preprint was published at Scientific Reports on August 18th, 2020.
See the published version at https://doi.org/10.1038/s41598-020-70902-5.

Abstract

Background Mugil incilis (lisa) is an important commercial fish species in many countries, living along the coasts of the western Atlantic Ocean. It has been used as a model organism for environmental monitoring and ecotoxicological investigations. Nevertheless, available genomic and transcriptomic information for this organism is extremely deficient. The aim of this study was to characterize M. incilis hepatic transcriptome using Illumina paired-end sequencing. Results A total of 32.082.124 RNA-Seq read pairs were generated utilizing the HiSeq platform and subsequently cleaned and assembled into 93.947 contigs (with N50 = 2.019 bp). The analysis of species distribution revealed that M. incilis contigs had the highest number of hits to Stegastes partitus (36.6%). Using a sequence similarity search against the public databases GO and KEGG, a total of 71.271 and 16.974 contigs were annotated, respectively. KEGG showed genes related to environmental information, metabolism and organismal system pathways were highly annotated. Complete or partial coding DNA sequences for several candidate genes associated with stress responses/detoxification of xenobiotics, as well as housekeeping genes, were employed to design primers that were successfully tested and validated by RT-qPCR. Conclusions This study presents the first transcriptome resources for Mugil incilis and provides basic information for the development of genomic tools, such as the identification of RNA markers, useful to analyze environmental impacts on this fish Caribbean species.

Background

The constant growth of the world population has brought a tremendous increase in the discharge of a wide diversity of pollutants to aquatic ecosystems, causing deleterious effects on organisms and fisheries [1-3]. Aquatic organisms, including fish, uptake and accumulate pollutants directly from water or sediments, and indirectly via the food chain [4]. Fish are often used as sentinels because they play a number of roles in the trophic web, bioaccumulate chemical substances, and respond to low concentrations of toxicants [5, 6].

Mugilidae (mullets) is a fish family widely distributed in tropical and subtropical waters around the globe, particularly in coastal and estuarine areas where they play an important ecological role, and provide biomass to support fisheries [7-9]. Over the past years, Mullets have been proposed as pollution bioindicators for environmental degradation [10]. Mugilids, in particular species of genus *Mugil*, such as *Mugil cephalus* and *Mugil incilis*, have been extensively employed on environmental monitoring programs, as well as in toxicological studies in coastal zones impacted by human activities [1, 10-16]. The use of these mugilids as sentinel species in these coastal systems arises from their wide geographical distribution; great abundance, salinity tolerance and bioaccumulation of land-based pollution, a feature largely enhanced by their consumption of benthic sediments along with their food [1, 17].

Mugil incilis, also known as mullet, lisa and lisa rayada, is one of the most abundant fish in the Caribbean. It is found in the western Atlantic Ocean, from Panama and Haiti to southeastern Brazil [18]. It is mainly found in brackish estuaries, but also in marine and hyper-saline waters [19]. Juvenile fish (<25 mm) are primarily planktonic or carnivorous feeders, whereas larger specimens switch their diet to detritus and benthic microalgae, ingesting large amounts of organic matter, sand or mud from the sediment [20]. *Mugil incilis* supports commercial and recreational fisheries, owing to its high protein and vitamin content, being a stable food source in many countries [15]. Its importance as food contrasts with its capacity to serve as intermediate host for several parasites [21]. In Colombia, *M. incilis* is one of the most widely distributed fish in the Caribbean coast [13, 22].

Along the Caribbean coast, Cartagena Bay is considered one of the ecosystems with high economic and environmental interest, as it hosts a great diversity of biological resources, but in the other hand, it receives many anthropic pressures, especially from industrial activities, such as oil refinery, pesticide packaging, metallurgical industry, and boatyards, including also naval and commercial shipping harbors. The direct and indirect discharges of urban and industrial wastes and runoff have led to the chemical contamination of the Bay [13, 23, 24]. Some studies with native species of this ecosystem, such as *M. incilis* (Lisa), have shown anthropic contamination by chemical substances, both organic and inorganic, and biological stressors, such as parasites. Heavy metals, such as mercury, have been detected in this species [13, 25], as well as polycyclic aromatic hydrocarbons (HAPs) and perfluorinated octyl sulphonates (PFOS) in bile [14, 23], and organochlorine pesticides in muscle (β-HCH, Aldrin, 4, 4'-DDE and endosulfan [16]. This species has also been found parasitized with nematodes of the Anisakidae family, in particular *Contracaecum* sp., as well as with Heterophydae trematodes, specifically *Ascocotyle (Phagicola) longa* in hepatic tissue [21, 26, 27].

As presented before, most studies on *M. incilis* from the Cartagena Bay environmental have included measurements of pollutant levels in different tissues. However, a better understanding of the impact of this pollution requires additional approaches to correlate the chemical or biological exposure with the molecular and physiological responses in these organisms, necessary to generate reliable information on their current health status and the effects of exposure to different pollutants [6, 28]. One of those approaches in the analysis of gene expression, a tool that provides multiple possibilities to evaluate molecular changes in exposed organisms [29, 30]. The understanding/comprehension of the biological response related to chronic exposition to environmental pollutants at the transcriptomic level is essential to safeguard the adaptive potential of populations under heavy anthropogenic pressure [31-33]. For example, *de novo* assembly of transcriptomes from important aquaculture species, such as *Epinephelus coioides*, *Larimichthys crocea*, *Scophthalmus maximus*, *Lateolabrax japonicas* and *Labeo rohita* (Hamilton) has revealed a huge number of molecular markers relevant in the immune response after exposition to pathogen microorganisms [34, 35]. Although experimentally controlled populations are ideal

for evaluating the transcriptional profile during exposure to specific environmental stressors [36], the study of toxic effects and transcriptional response in natural populations is essential for understanding the synergistic effects of multiple environmental stressors under field conditions [29, 30].

In fish, liver has been the focus of toxicological numerous studies as it is sensitive to pollutant exposure [1, 12, 37]. In this perspective, the assessment of hepatic defense and damage responses in a resident fish species becomes highly relevant in biomonitoring studies, especially considering the liver performs complex biological functions that are essential for survival, including nutrient synthesis, metabolite transformation and storage, and detoxification processes [38, 39]. However, the analysis of fish liver transcriptomes has been limited to only a few species and to our knowledge; no studies have been reported on the liver transcriptome of *M. incilis* or in other mugilid species. The purpose of this study was to characterize the hepatic transcriptome of *M. incilis* in order to facilitate future studies on gene expression related to the effects of environmental pollution in this Caribbean wild species.

Methods

Sample collection and handling

The study did not involve endangered or protected species and it was granted with a permission to access genetic resources from the Ministry of Environment, Housing and Territorial Development from Colombia (Contract No. 148; 09-11-17). One specimen of *M. incilis* (Lisa) was obtained with the help of local fishermen in Cartagena Bay (10°24'N 75°30'W). The specimen was stored in a plastic bag and transported on ice to the laboratory for immediate processing. Prior to RNA extraction, the liver sample was frozen in liquid nitrogen and grounded into a fine powder using a pre-cooled mortar and pestle.

RNA extraction, library construction and sequencing

RNA Extraction was performed according to the manufacturer's directions using Trizol reagent (Invitrogen, California, USA). RNA concentration and integrity of the sample was measured on an Agilent 2100 Bioanalyzer (Agilent Technologies) using an RNA Nano Bioanalysis chip. Mitochondrial RNA enrichment, library construction, and sequencing was conducted by Macrogen Inc. (Seoul, Korea) with Truseq mRNA kits and the Illumina HiSeq 2000 instrument, with 100 bp paired end (PE) reads.

Sequence data processing and de novo assembly

Read quality was controlled using trimming threshold of Q30 bases with the program PRINSEQ [57]. After the quality filter, reads with less than 50 bases were rejected. The assembly step was carried out with the

program Trinity (http://trinityrnaseq.sourceforge.net/) employing default settings [58]. After the assembly, only the contigs longer than 200 bases were kept.

Functional contigs annotation and classification

Sequencing data were aligned using BLASTx alignment (E-value cut-off of 10^{-5}) with protein databases, including NCBI non-redundant (Nr) database and KEGG database. BLASTx comparisons and the software MEGAN were used to filter contigs that belong to *Mugil incilis*. Since *Mugil* sequences are scarce in the database, all sequences classified as Chordata were considered as of *Mugil* origin. Fish (Chordata) filtered contigs were loaded into the Blast2GO program (http://www. blast2go.org/) in order to obtain general gene ontology (GO) annotations. GO terms were assigned and annotated to the contigs/transcripts according to three main categories: cell component; molecular function; and biological process [59]. KEGG pathways were assigned to the transcripts using the online KEGG Automatic Annotation Sever (KAAS; http://www.genome.jp/kegg/kaas/) utilizing the bi-directional best-hit method. Several contigs/transcripts of the stress responses/detoxification and immune system were selected and annotated using the BLASTx results and the KEGG annotation. CDS prediction was carried out manually using Artemis visualizer (Sanger) [60].

Data deposition

The raw reads produced in this study were deposited in the NCBI database Sequence Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra) under the accession number SRP157095.

Real-time RT-PCR validation experiments

Validation of gene expression quantification in hepatic tissue of eight *M. incilis* specimens was conducted using real time quantitative PCR (RT-qPCR) for twenty genes associated with stress responses/detoxification (GAPDH, L13a, EF1, SOD1, CAT, HSP70, HIF1A, CYP1A1, CYP3A, RXRA X2, AHR, Bcl-X, BAX, CASP3, Gadd45A, TNF-α, IL6, MHC-II class alpha, PPARA and PPARG). Primers for each target gene were designed with “primer-BLAST” tool from Genbank (http://www.ncbi.nlm.nih.gov/tools/primer-blast) using the transcript sequences obtained from the transcriptome sequencing. The specificity of primers PCR was checked at the end of each amplification using a DNA melt curve analysis and by the observation of single amplification products of the expected size on agarose gels. Amplification efficiency was determined using a five-point 1:10 dilution series starting with cDNA corresponding to 100 ng of input total RNA [61]. Samples of cDNA from other fish species (*Prochilodus magdalenae* and *Hoplias malabaricus*) different from the mullets were included, as well as a sample of cDNA from *Mus*

musculus, in order to verify the specificity of the primers. The specific primers used for RT-qPCR are listed in Table 1.

The total RNA was isolated from liver tissue using RNeasy®Mini Kit (Qiagen, California, USA) as described by the manufacturer. The concentration of RNA was determined by spectrophotometry (A260) employing a NanoDrop 2000 Spectrophotometer (Thermo Scientific), and purity was assessed by the A260:A280 ratio (1.9–2.0). The integrity of RNA was checked by visual inspection of 28S and 18S ribosomal RNA on an agarose gel. Aliquots of RNA samples were stored at 80° C until analysis. For each sample, 1 µg of total RNA was reverse transcribed using QuantiTect® Reverse Transcription Kit (QiagenInc, Valencia, CA, USA). The resultant cDNA was employed as template in a 20 µL PCR reaction containing 10 pmol each of forward and reverse gene-specific primers. Real time-PCR was conducted utilizing a StepOnePlus®System (Applied Biosystems, Foster City, CA). The reactions were performed in MicroAmp optical 96-well reaction plates (Applied Biosystems) with SYBR® Green qPCR Master Mix (Thermo Scientific). Conditions for PCR were as follows: Initial denaturation and enzyme activation for 10 min at 95 °C, followed by 40 cycles of 95 °C for 15 s (denaturation) and 60 °C for 1 min (annealing/extension). A melt curve analysis was performed after the 40th cycle (10 s at 95°C, then 65–95°C in 0.5°C increments every 5 s). A negative control (without DNA template) was run in duplicate on each plate to verify the absence of cDNA contamination [62].

Results

De novo assembly of the Mugil incilis hepatic transcriptome

Approximately 32.082.124 raw reads pairs from the HiSeq platform were generated from the liver transcriptome sample from *M. incilis*. After cleaning with a Q30 base quality threshold, 31.061.840 read pairs remained. The clean read dataset was further assembled using Trinity software into 93.947 contigs larger than 200 bases that reached 92 Mb. Contig length ranged between 201 and 17.447 bases with an N50 value of 2.019 bases (Table 2). In order to identify contigs that belong solely to fish species, BLASTn and BLASTx comparison of the complete contig dataset was carried out and, using the software MEGAN, contigs were filtered according to its classification and only those belonging to Chordata were kept for further analysis. This filtered contig dataset that includes 13.474 contigs was considered from *Mugil incilis*.

Similarity analysis of the M. incilis liver transcriptome

Based on the BLASTx similarity analysis of the contigs, matched sequences suggested a variety of different species (Figure 1; Figure S1). Most contigs had greatest homology to genes from nineteen fishes. Among fish species with protein sequences in Genbank, *Mugil incilis* contigs had the highest

number of hits to *Stegastes partitus* (36.6%), followed by *Larimichthys crocea* (18.5%), *Oreochromis niloticus* (8.4%) and *Maylandia zebra* (3.4%). In addition, 378 (8.3%) of the transcripts were annotated to other eukaryotes, i.e., mammals, teleost fishes, and few prokaryotes vertebrates.

GO annotation of the lisa liver transcriptome

Gene Ontology (GO) is an international standardized gene functional classification system (Götz et al., 2008). Based on the Nr annotation from NCBI, 71.271 contigs were assigned GO terms. Using Blast2GO, 39.249 terms were assigned to different biological processes categories, 14.121 to cellular components, and 17.901 to molecular functions. These GO terms were summarized into 46 subcategories. Within the biological processes category, cellular (8.764; 22.3%), metabolic (8.456; 21.5%), single-organism (5.633; 14.4%), biological regulation (4.128; 10.5%), and response to stimulus (2.878; 7.3%) were the most dominant subcategories. In the cellular component category, the four most common categories were cell (4.865; 34.5%), membrane (3.298; 23.4%), organelle (3.215; 22.8%), and macromolecular complex (1.572; 11.1%). Regarding molecular functions, the majority of the contigs were assigned to binding (8.031; 44.9%), followed by catalytic activity (6.155; 34.4%), transporter activity (864; 4.8%), and molecular transduction activity (796; 4.4%) (Table S1). The top 10 groups in the three main categories are shown in Figure 2.

Functional classification by KEGG

To identify biological pathways in the liver of *M. incilis*, the contigs were mapped to the reference pathways recorded in the KEGG database. A total of 16.974 contigs were annotated in KEGG and located to 394 known KEGG pathways (Table S2). The top 20 pathways with the largest numbers of contigs are shown in Fig. 3. Among the 394 predicted KEGG pathways, the metabolic pathways represented the largest group, consisting of 812 contigs (4.78%), followed by cancer (297, 1.75%), biosynthesis of secondary metabolites (223, 1.31%), PI3K-Akt signaling (182, 1.07%) and human papillomavirus infection (182, 1.07%) (Figure 3, Table S2). These 394 pathways were subjected to six categories: human diseases (5.024, 29.6%), organismal systems (3.453, 20.3%), metabolism (3.249, 19.1%), environmental information processing (2.417, 14.2%), cellular processes (1.702, 10.0%), and genetic information processing (1.129, 6.7%) (Table S3).

Real-time qPCR assays

The *de novo* sequencing of the hepatic transcriptome of *M. incilis* generated sequence knowledge for a species with virtually no available genomic information. PCR primers of a set of known target genes related to responses to xenobiotic exposure were generated from the transcriptome sequences. Twenty-

one primers pairs were analyzed, including gene markers of heavy metal exposure, xenobiotic metabolism, nuclear receptor modulation, oxidative stress, DNA damage, inflammation and lipid metabolism. Accession numbers and primer sequences for the target genes are listed in Table 1.

The C_T (Threshold cycle) is the cycle number at which the fluorescence generated within a reaction crosses the threshold line. When C_T values were used to generate a log-linear regression plot, the standard curve for the target genes showed a strong relationship (R^2 ranging from 0.95 to 1; PCR efficiency ranging from 89 to 116%). A correlation coefficient greater than 0.99 shows good primer efficiency, and indicates a successful real-time PCR experiment. It is considered that a slope of the regression curve between −3.9 and −2.9 corresponds to PCR efficiencies ranging from 80% to 120% [40, 41]. For all the genes evaluated in this work we found efficiency percentages ranging from 90 to 110%. A melting curve analysis showed no indication of primer-dimerization for the tested candidate oligonucleotides. PCR efficiencies for each primer pair are shown in Table 3. Furthermore, the primers specificity for target genes was confirmed by gel electrophoresis of the PCR products. All the primer sets amplified one single band of the expected fragment, and most of them showed specificity for targets from *M. incilis*. Some of the primers, including housekeeping genes, amplified fragments from related fish species such as *P. magdalenae* and *H. malabaricus*. The electrophoresis of PCR products for the evaluated genes is presented in Figure 4.

Discussion

As high-throughput sequencing technology has developed rapidly, de novo transcriptome assembly has become an efficient research tool for differential transcriptome analysis in fish species [42]. 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). Despite the fact that Colombia is considered one of the most biodiverse countries in the world, there is little knowledge on fish species commonly found in its tropical ecosystems. The assembly of the transcriptome of *Mugil incilis* opens a large window of opportunities to develop toxicological studies with this organism, which can be found along the Caribbean.

Some reports on transcriptome data for Mugilids have been focused on population genetic structure [43] and innate defense of the host against pathogens [44, 45]. In this study the transcriptome of hepatic tissue from *M. incilis* was generated using Illumina HiSeq 2000 sequencing platform. A total of 93.947 contigs were assembled with an average length of 973 bp, a greater value than those obtained in previous studies using other *de novo* assembly methods for teleosts, such as viviparous blenny (*Zoarces viviparus*; 395 pb) [46], Guppy (*Poecilia reticulata*; 464 pb) [47], European eel (*Anguilla;* 531 pb) [48] and common carp (*Cyprinus carpio*; 888 pb) [49]. These numerous sequences can provide a sufficient

transcriptome sequence resource for discovering novel genes in *M. incilis*. These results showed that Trinity is a powerful and efficient tool for *de novo* assembly for organisms without reference genomes.

Homology search for the clustered sequences was carried out using BLASTX against the NCBI non-redundant (Nr) database. The distribution of top hit species suggested that most of the annotated sequences corresponded to known nucleotide sequences of a small group of fish species, *S. partitus*, *L. crocea* and *O. niloticus* (Figure 1). It was not surprising that *M. incilis* contigs shared the highest percentage with *S. partitus*, as both of them belong to the order Perciformes, although such finding may be due to the availability of the complete genome of *S. partitus,* providing sufficient gene sequences and annotations for comparison analyses. However, the transcriptomic similarities between *M. incilis* and *S. partitus* only reached 36.6%, suggesting these two fish species might be distantly related in subfamilies of Perciformes, Mugilidae and Pomacentridae for *M. incilis* and *S. partitus*, respectively.

GO Database annotation suggested genes associated with the cellular process (GO:0009987), metabolic process (GO:0008152) and single-organism process (GO:0044699) were significantly enriched in biological processes. For cellular components, the major category represented was cell (GO:0005623). For molecular function, binding (GO:0005488) was the most strongly represented GO term, followed by catalytic activity (GO:0003824), results that can be useful for gene functional studies. Furthermore, 16.974 contigs were annotated to 6 categories in KEGG, including cellular processes, environmental information processing, genetic information processing, metabolism, human diseases and organismal systems. Signal transduction had the most number of genes (2072) in environmental information. For metabolism, global and overview maps had the most number of genes (1473), whereas for organismal systems, the immune system comprises the most number of genes (1094) (Table S3). In the present study, PI3K-Akt, MAPK, Ras and Rap1 signaling pathways were found to be the most highly enriched. MAPK pathways are ubiquitous in all eukaryotic cells and transmit and process signals regulating a broad array of cell fate decisions, including proliferation, differentiation, motility, stress responses, and apoptosis [50]. It has been shown that it is involved in intracellular responses to diverse environmental stresses including cold, heat, reactive oxygen species, UV, desiccation and pathogen attack [51]. The PI3K-Akt signaling pathway is activated by many types of cellular stimuli or toxic insults, and regulates fundamental cellular functions such as transcription, translation, proliferation, growth, and survival. The transcription of genes related to these signaling pathways suggested *M. incilis* is a sensitive model for physiological studies.

The 1094 genes in the immune system were further grouped into 21 pathways (Figure S2). NOD–like receptor signaling pathway and chemokine signaling pathway were the top two pathways with the greatest number of expressed genes. Nucleotide-binding oligomerization domain (NOD)-like receptors

(NLRs) are a group of cytoplasmic pattern recognition receptors (PRRs), which are responsible for detecting various pathogens and generating innate immune responses. In fish, three distinct subfamilies of NLRs have been identified and characterized: NLR-A, NLR-B, and fish-specific NLR-C [52]. Expression of NLR-C subfamily (NLRCs) has been reported in many different teleost fish species such as zebrafish [53], Japanese flounder [54] and turbot [55]. These receptors are activated by a variety of bacterial pathogens or microbial ligands, including lipopolysaccharides (LPS), peptidoglycans (PGN) and polyinosinic-polycytidylic acid (Poly (I:C)), suggesting their participation in the fish innate immune response [52]. The other signaling pathway with greater annotation was related to Chemokines, small chemoattractant peptides that provide directional cues for the cell trafficking and thus are vital for protective host response. In response to a pathogenic exposure, chemokines not only promote leukocyte mobilization, but also regulate the immune responses and differentiation of the recruited cells to orchestrate the first steps of both innate and acquired immune responses [56]. Other important pathways include Fc gamma R-mediated phagocytosis, natural killer cell mediated cytotoxicity, leukocyte transendothelial migration, and T cell receptors, also critical in immune response. As *M. incilis* has been found parasitized with *Contracaecum* sp. and *Ascocotyle (Phagicola) longa* in the Colombian Caribbean [21, 26, 27], the characterization of these pathways will be an essential tool to study the interaction fish-parasites in this species, especially considering the high selectivity obtained for designed primers (Figure 4).

The transcriptome assembly of *M. incilis* generated in this work, as well as the PCR validation for several signaling pathways, will allow further molecular toxicology research on this species, as this fish has been used to monitor different chemical and biological pollutants in the Caribbean. The highly parasitized status of the fish, together with the different pollutant levels it may receive, creates a unique opportunity to study host-parasite interactions, and how environmental contaminants modulate this process.

Conclusions

This is the first report on the sequencing, de novo assembly, and functional annotation of the *Mugil incilis* transcriptome. The identification of the CDS sequences for several of these candidate genes will help in developing primers for qPCR, which will be used to asses gene expression levels in other Mugilids species.

Abbreviations

RNA-seq, RNA sequencing

RT-qPCR, reverse transcription quantitative polymerase chain reaction

NGS, next-generation sequencing

CDS, coding sequences

SRA, Sequence Read Archive

bp, base pair(s)

cDNA, DNA complementary to RNA

RNA, Ribonucleic acid

GO, Gene Ontology

KEGG, Kyoto Encyclopedia of Genes and Genomes

CT, Threshold cycle

Nr, non-redundant

Ns, ambiguous bases

NCBI, National Center for Biotechnology Information

HAPs, Polycyclic Aromatic Hydrocarbons

PFOS, perfluorinated octyl Sulphonates

NLRs, Nucleotide-binding oligomerization domain (NOD)-like receptors

LPS, Lipopolysaccharides

PGN, Peptidoglycans

Poly (I:C), Polyinosinic-polycytidylic acid

SOD1, Superoxide dismutase 1

CAT, Catalase

HSP70, Proteína de shock térmico de 70 KDa

HIF1A, Factor Inducible por Hipoxia-1, Subunidad Alfa

CYP1A1, Cytochrome P4501A

CYP3A, Cytochrome P450, family 3, subfamily A

AHR, Aryl hydrocarbon receptor

Bcl-X, Apoptosis Regulator

BAX, Apoptosis Regulator BAX

CASP3, Caspase-3

Gadd45A, Growth arrest and DNA damage inducible alpha

TNF-α, Tumor necrosis factor alpha induced protein 3

IL6, Interleukin-6

MHC- class II alpha, Major Histocompatibility Complex, Class II, alpha chain

PPARA, Peroxisome proliferator-activated receptor alpha

PPARG, Peroxisome proliferator-activated receptor gamma

L13a, Ribosomal Protein L13a

EF1, Elongation Factor-1 alpha

GAPDH, Glyceraldehyde 3-phosphate dehydrogenase

Declarations

Ethics approval and consent to participate

Access to genetic material was granted by the Ministry of Environment, Housing and Territorial Development from Colombia.

Consent for publication

Not applicable.

Availability of data and material

Raw Illumina sequences were deposited in the National Center for Biotechnology Information (NCBI) and can be accessed in the Short Read Archive (SRA) database (http://trace.ncbi.nlm.nih.gov/Traces/sra/) under accession SRP157095.

Competing interests

The authors declare that they have no competing interest.

Funding

The fees for NGS sequencing, RT-PCR validation experiments and data processing were funded by the Vice-Presidency of Research (2017-2019), University of Cartagena, Colombia.

Authors' contributions

ABS, and JOV conceived and designed the experiment. ABS and JAR performed the RNA-Seq data analysis. ABS and JAR performed the pathways analysis. ABS developed the Real-time qPCR assays. ABS, JAR and JOV drafted the manuscript. All authors read and approved the final manuscript. Financial funds were obtained from Grants written by JOV.

Acknowledgments

The authors thank the National Program for Doctoral Formation (COLCIENCIAS, 727-2015); COLCIENCIAS (Bogotá, Colombia) and the University of Cartagena (Cartagena, Colombia).

References

1. Ameur WB, de Lapuente J, El Megdiche Y, Barhoumi B, Trabelsi S, Camps L, Serret J, Ramos-López D, Gonzalez-Linares J, Driss MR *et al*: **Oxidative stress, genotoxicity and histopathology biomarker responses in mullet (Mugil cephalus) and sea bass (Dicentrarchus labrax) liver from Bizerte Lagoon (Tunisia)**. *Mar Pollut Bull* 2012, **64**(2):241-251.
2. Brázová T, Torres J, Eira C, Hanzelová V, Miklisová D, Salamún P: **Perch and its parasites as heavy metal biomonitors in a freshwater environment: the case study of the Ružín water reservoir, Slovakia**. *Sensors (Basel)* 2012, **12**(3):3068-3081.
3. Goretti E, Pallottini M, Ricciarini MI, Selvaggi R, Cappelletti D: **Heavy metals bioaccumulation in selected tissues of red swamp crayfish: An easy tool for monitoring environmental contamination levels**. *Sci Total Environ* 2016, **559**:339-346.
4. Zhao S, Feng C, Quan W, Chen X, Niu J, Shen Z: **Role of living environments in the accumulation characteristics of heavy metals in fishes and crabs in the Yangtze River Estuary, China**. *Mar Pollut Bull* 2012, **64**(6):1163-1171.
5. Magnanou E, Klopp C, Noirot C, Besseau L, Falcón J: **Generation and characterization of the sea bass Dicentrarchus labrax brain and liver transcriptomes**. *Gene* 2014, **544**(1):56-66.
6. van der Oost R, Beyer J, Vermeulen NPE: **Fish bioaccumulation and biomarkers in environmental risk assessment: a review**. *Environ Toxicol Pharmacol* 2003, **13**(2):57-149.
7. **FAO. Fisheries and aquaculture software. FishStat Plus - Universal software for fishery statistical time series. In: FAO Fisheries and Aquaculture Department [online]. Rome. 2013. Accessed 4 May 2017.** [http://www.fao.org/fishery/statistics/software/fishstat/en]
8. Nash CE, Sehadeh ZH: **Review of Breeding and Propagation Techniques for Grey Mullet, Mugil cephalus L. ICLARM Studies and Reviews 3, 1- 87**, vol. 3; 1980.
9. Thomson JM: **The grey mullets. Oceanogr. Mar. Biol. Annu. Rev. 4, 301–355.**, vol. 4; 1966.

10. Waltham NJ, Teasdale PR, Connolly RM: **Use of flathead mullet (Mugil cephalus) in coastal biomonitor studies: Review and recommendations for future studies**. *Mar Pollut Bull* 2013, **69**(1):195-205.
11. Ameur WB, Trabelsi S, El Megdiche Y, Ben Hassine S, Barhoumi B, Hammami B, Eljarrat E, Barceló D, Driss MR: **Concentration of polychlorinated biphenyls and organochlorine pesticides in mullet (Mugil cephalus) and sea bass (Dicentrarchus labrax) from Bizerte Lagoon (Northern Tunisia)**. *Chemosphere* 2013, **90**(9):2372-2380.
12. Vasanthi LA, Revathi P, Mini J, Munuswamy N: **Integrated use of histological and ultrastructural biomarkers in Mugil cephalus for assessing heavy metal pollution in Ennore estuary, Chennai**. *Chemosphere* 2013, **91**(8):1156-1164.
13. Olivero-Verbel J, Caballero-Gallardo K, Torres-Fuentes N: **Assessment of mercury in muscle of fish from Cartagena Bay, a tropical estuary at the north of Colombia**. *Int J Environ Health Res* 2009, **19**(5):343-355.
14. Olivero-Verbel J, Tao L, Johnson-Restrepo B, Guette-Fernández J, Baldiris-Avila R, O'Byrne-Hoyos I, Kannan K: **Perfluorooctanesulfonate and related fluorochemicals in biological samples from the north coast of Colombia**. *Environ Pollut* 2006, **142**(2):367-372.
15. Crosetti D, Blaber SJM: **Biology, ecology and culture of grey mullet (Mugilidae). CRC Press, Taylor and Francis, Boca Raton** 2016.
16. Jaramillo-Colorado BE, Arroyo-Salgado B, Ruiz-Garcés LC: **Organochlorine pesticides and parasites in Mugil incilis collected in Cartagena Bay, Colombia**. *Environmental science and pollution research international* 2015, **22**(22):17475-17485.
17. Whitfield A: **Ecological Role of Mugilidae in the Coastal Zone. In: Crosetti, D. and Blaber, S.J.M. (Eds.), Biology, Ecology and Culture of Grey Mullet (Mugilidae). CRC Press, Boca Raton**. In.; 2016: 324-348.
18. Barletta M, Dantas D: **Biogeography and Distribution of Mugilidae in the Americas. In: Crosetti, D. and Blaber S.J.M. (Eds.), Biology, Ecology and Culture of Grey Mullet (Mugilidae). CRC Press, Boca Raton**. In.; 2016: 42-62.
19. Thomson JM: **The Mugilidae of the world**. *Mem Queensl Mus* 1997, **41**:457-562.
20. Cardona L: **Food and Feeding of Mugilidae. In: Crosetti, D. and Blaber, S.J.M. (Eds.), Biology, Ecology and Culture of Grey Mullet (Mugilidae). CRC Press, Boca Raton**. In.; 2016: 165-195.
21. Galván-Borja D, Olivero-Verbel J, Barrios-García L: **Occurrence of Ascocotyle (Phagicola) longa Ransom, 1920 (Digenea: Heterophyidae) in Mugil incilis from Cartagena Bay, Colombia**. *Veterinary Parasitology* 2010, **168**(1):31-35.
22. Arrieta L, de la Rosa Muñoz J: **ESTRUCTURA DE LA COMUNIDAD ÍCTICA DE LA CIÉNAGA DE MALLORQUÍN, CARIBE COLOMBIANO**. *Bol Invest Mar Cost* 2003, **32**:231-242.
23. Johnson-Restrepo B, Olivero-Verbel J, Lu S, Guette-Fernández J, Baldiris-Avila R, O'Byrne-Hoyos I, Aldous KM, Addink R, Kannan K: **Polycyclic aromatic hydrocarbons and their hydroxylated metabolites in fish bile and sediments from coastal waters of Colombia**. *Environ Pollut* 2008, **151**(3):452-459.

24. Olivero-Verbel J, Johnson-Restrepo B, Baldiris-Avila R, Güette-Fernández J, Magallanes-Carreazo E, Vanegas-Ramírez L, Kunihiko N: **Human and crab exposure to mercury in the Caribbean coastal shoreline of Colombia: Impact from an abandoned chlor-alkali plant**. *Environ Int* 2008, **34**(4):476-482.
25. Alonso D, Pineda P, Olivero J, González H, Campos N: **Mercury levels in muscle of two fish species and sediments from the Cartagena Bay and the Ciénaga Grande de Santa Marta, Colombia**. *Environ Pollut* 2000, **109**(1):157-163.
26. Olivero Verbel J, Caballero-Gallardo K, Arroyo-Salgado B: **Nematode infection in fish from Cartagena Bay, North of Colombia**. *Veterinary Parasitology* 2011, **177**(1):119-126.
27. Olivero-Verbel J, Baldiris-Ávila R, Arroyo-Salgado B: **Nematode infection in Mugil incilis (lisa) from Cartagena bay and Totumo marsh, north of Colombia**. *Journal of Parasitology* 2005, **91**(5):1109-1112.
28. Lehtonen KK, Schiedek D: **Monitoring biological effects of pollution in the Baltic Sea: Neglected—but still wanted?** *Mar Pollut Bull* 2006, **53**(8):377-386.
29. Olsvik PA, Berg V, Lyche JL: **Transcriptional profiling in burbot (Lota lota) from Lake Mjøsa—A Norwegian Lake contaminated by several organic pollutants**. *Ecotoxicol Environ Saf* 2013, **92**:94-103.
30. Olsvik PA, Lindgren M, Maage A: **Mercury contamination in deep-water fish: Transcriptional responses in tusk (Brosme brosme) from a fjord gradient**. *Aquat Toxicol* 2013, **144-145**:172-185.
31. Olsvik PA, Amlund H, Torstensen BE: **Dietary lipids modulate methylmercury toxicity in Atlantic salmon**. *Food Chem Toxicol* 2011, **49**(12):3258-3271.
32. Uren Webster TM, Bury N, van Aerle R, Santos EM: **Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout**. *Environ Sci Technol* 2013, **47**(15):8869-8877.
33. Pierron F, Normandeau E, Defo M, Campbell P, Bernatchez L, Couture P: **Effects of chronic metal exposure on wild fish populations revealed by high-throughput cDNA sequencing**. *Ecotoxicology (London, England)* 2011, **20**(6):1388-1399.
34. Li G, Zhao Y, Liu Z, Gao C, Yan F, Liu B, Feng J: **De novo assembly and characterization of the spleen transcriptome of common carp (Cyprinus carpio) using Illumina paired-end sequencing**. *Fish Shellfish Immunol* 2015, **44**(2):420-429.
35. Zhu J, Li C, Ao Q, Tan Y, Luo Y, Guo Y, Lan G, Jiang H, Gan X: **Trancriptomic profiling revealed the signatures of acute immune response in tilapia (Oreochromis niloticus) following Streptococcus iniae challenge**. *Fish Shellfish Immunol* 2015, **46**(2):346-353.
36. Cossins A, Crawford D: **Fish as models for environmental genomics**. *Nat Rev Genet* 2005, **6**:324-333.
37. Gernhöfer M, Pawert M, Schramm M, Müller E, Triebskorn R: **Ultrastructural biomarkers as tool to characterize the health status of fish in contaminated streams**. *Journal of Aquatic Ecosystem Stress and Recovery* 2001, **8**:241-260.

38. Vega-Retter C, Rojas-Hernandez N, Vila I, Espejo R, Loyola DE, Copaja S, Briones M, Nolte AW, Véliz D: **Differential gene expression revealed with RNA-Seq and parallel genotype selection of the ornithine decarboxylase gene in fish inhabiting polluted areas**. *Sci Rep* 2018, **8**(1):4820-4820.
39. Oliveira M, Ahmad I, Maria V, Pacheco M, Santos M: **Antioxidant Responses Versus DNA Damage and Lipid Peroxidation in Golden Grey Mullet Liver: A Field Study at Ria de Aveiro (Portugal)**. *Arch Environ Contam Toxicol* 2010, **59**:454-463.
40. Rebrikov DV, Trofimov DY: **Real-time PCR: A review of approaches to data analysis**. *Applied Biochemistry and Microbiology* 2006, **42**(5):455-463.
41. Bustin S, Huggett J: **qPCR primer design revisited**. *Biomolecular Detection and Quantification* 2017, **14**:19-28.
42. Kumar G, Kocour M: **Applications of next-generation sequencing in fisheries research: A review**. *Fish Res* 2017, **186**:11-22.
43. Krück NC, Innes DI, Ovenden JR: **New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (Mugil cephalus)**. *Mol Ecol Resour* 2013, **13**(4):715-725.
44. Byadgi O, Chen Y-C, Barnes AC, Tsai M-A, Wang P-C, Chen S-C: **Transcriptome analysis of grey mullet (Mugil cephalus) after challenge with Lactococcus garvieae**. *Fish Shellfish Immunol* 2016, **58**:593-603.
45. Qi Z, Wu P, Zhang Q, Wei Y, Wang Z, Qiu M, Shao R, Li Y, Gao Q: **Transcriptome analysis of soiny mullet (Liza haematocheila) spleen in response to Streptococcus dysgalactiae**. *Fish Shellfish Immunol* 2016, **49**:194-204.
46. Kristiansson E, Asker N, Förlin L, Larsson DGJ: **Characterization of the Zoarces viviparus liver transcriptome using massively parallel pyrosequencing**. *BMC genomics* 2009, **10**:345-345.
47. Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA: **Sequencing and characterization of the guppy (Poecilia reticulata) transcriptome**. *BMC genomics* 2011, **12**:202-202.
48. Coppe A, Pujolar JM, Maes GE, Larsen PF, Hansen MM, Bernatchez L, Zane L, Bortoluzzi S: **Sequencing, de novo annotation and analysis of the first Anguilla anguilla transcriptome: EeelBase opens new perspectives for the study of the critically endangered European eel**. *BMC genomics* 2010, **11**:635-635.
49. Ji P, Liu G, Xu J, Wang X, Li J, Zhao Z, Zhang X, Zhang Y, Xu P, Sun X: **Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics**. *PLOS ONE* 2012, **7**(4):e35152-e35152.
50. Avruch J: **MAP kinase pathways: the first twenty years**. *Biochimica et biophysica acta* 2007, **1773**(8):1150-1160.
51. Sun Z, Tan X, Xu M, Liu Q, Ye H, Zou C, Ye C: **Liver transcriptome analysis and de novo annotation of the orange-spotted groupers (Epinephelus coioides) under cold stress**. *Comp Biochem Physiol Part D Genomics Proteomics* 2019, **29**:264-273.
52. Li T, Shan S, Wang L, Yang G, Zhu J: **Identification of a fish-specific NOD-like receptor subfamily C (NLRC) gene from common carp (Cyprinus carpio L.): Characterization, ontogeny and expression**

analysis in response to immune stimulation. *Fish Shellfish Immunol* 2018, **82**:371-377.

53. Laing KJ, Purcell MK, Winton JR, Hansen JD: **A genomic view of the NOD-like receptor family in teleost fish: identification of a novel NLR subfamily in zebrafish**. *BMC Evol Biol* 2008, **8**:42-42.
54. Unajak S, Santos MD, Hikima J-i, Jung T-S, Kondo H, Hirono I, Aoki T: **Molecular characterization, expression and functional analysis of a nuclear oligomerization domain proteins subfamily C (NLRC) in Japanese flounder (Paralichthys olivaceus)**. *Fish Shellfish Immunol* 2011, **31**(2):202-211.
55. Hou Z, Ye Z, Zhang D, Gao C, Su B, Song L, Tan F, Song H, Wang Y, Li C: **Characterization and expression profiling of NOD-like receptor C3 (NLRC3) in mucosal tissues of turbot (Scophthalmus maximus L.) following bacterial challenge**. *Fish Shellfish Immunol* 2017, **66**:231-239.
56. Bird S, Tafalla C: **Teleost Chemokines and Their Receptors**. *Biology* 2015, **4**(4):756-784.
57. Schmieder R, Edwards R: **Quality control and preprocessing of metagenomic datasets**. *Bioinformatics (Oxford, England)* 2011, **27**(6):863-864.
58. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nature Biotechnol* 2011, **29**(7):644-652.
59. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite**. *Nucleic Acids Res* 2008, **36**(10):3420-3435.
60. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation**. *Bioinformatics* 2000, **16**(10):944-945.
61. Hori TS, Gamperl AK, Booman M, Nash GW, Rise ML: **A moderate increase in ambient temperature modulates the Atlantic cod (Gadus morhua) spleen transcriptome response to intraperitoneal viral mimic injection**. *BMC genomics* 2012, **13**:431-431.
62. Guerrero-Castilla A, Olivero-Verbel J, Marrugo-Negrete J: **Heavy metals in wild house mice from coal-mining areas of Colombia and expression of genes related to oxidative stress, DNA damage and exposure to metals**. *Mutat Res Genet Toxicol Environ Mutagen* 2014, **762**:24-29.

Tables

Table 1. Primer sequences for Real-Time PCR.

Gene name	Gene abbreviation	Gene ID	Forward (5' to 3')	Reverse (5' to 3')	Amplicon size (pb)
Oxidative Stress					
Superoxide dismutase 1	SOD1	MH716024	TGGAGATAACACGAACGGGTG	ATTGGGGCCAGCGTGATTC	76
Catalase	CAT	MH716025	AAAGAGTGGTGCATGCCAAG	TGCCGACATGTTCAAACACC	102
Proteína de shock térmico de 70 KDa	HSP70	MH716026	CTGGATGTGACCCCCTTGTC	GTTGGGATGGTGGTGTTCCT	86
Factor Inducible por Hipoxia-1, Subunidad Alfa	HIF1A	MH716027	TAGTTGAGAGCAGCCAACCC	CAGCTCTGCACACTTGTCCT	80
Exposure to xenobiotics					
Cytochrome P4501A	CYP1A1	MH716028	TGTTCTTGGCCATCCTCGTC	CAGTGGTCATCCCTCACTCG	142
Cytochrome P450, family 3, subfamily A	CYP3A	MK110662	GACTCTCCTGGTTGCGTTCA	CCCGTACTTCTTGTGGCACT	172
Retinoid X receptor alpha isoformX2	RXRA X2	MK110663	TTCAACCCAGACTCCAAAGG	CGGGATACTTGTGTTTGCAG	103
The aryl hydrocarbon receptor	AHR	MH716029	CTCGGCTTTCACCAGACTGA	CAAGATGAGGTGGACTCGGC	161
Apoptosis/Necrosis					
Apoptosis Regulator	Bcl-X	MH716030	GAGTTTCAAGAAGTGGCTGCTG	ACAGTCGTTTCTGAGCGAAG	83
Apoptosis Regulator BAX	BAX	MH716031	AGGCGATCAAGGAAATGCCA	GTGCGAGCTTCTTGTGGTTG	167
Caspase-3	CASP3	MK183035	TCAGAGAAATGGCACAGACG	TCATTTGCTCGACTGACTGG	104
Growth arrest and DNA damage inducible alpha	Gadd45A	MK183036	TTCACTTCACCCTCATCCAG	TTCACCCCCACTCTGTTTTC	125
Tumor necrosis factor alpha induced protein 3	TNF-α	MH716032	ATGCTAGAGGGCTACTGCGA	ATAACCAGAGGTGGGGCGA	80
Immune Activation/inflammation					
Interleukin-6	IL6	MK183037	GAAAAACAAGAAGCGGGTCA	GCTTCTCCCTTCTGTCGATG	173
Major Histocompatibility Complex, Class II, alpha chain	MHC- class II alpha	MK183039	TTGGTCTGACTCTGGGTTTG	AGCTGCACTCGTTTCCTTTG	72
Lipid metabolism					
Peroxisome proliferator-activated receptor alpha	PPARA	MK183038	GAGGACTCGGTGTTGGACAG	CGTCGATGGACTGGGAGATG	91
Peroxisome proliferator-activated receptor gamma	PPARG	MH716033	CGGAGACAACATGCCTTTCG	CCTGAAACTCCGTGTGCAGT	138
Housekeeping					
Ribosomal Protein L13a (RP-L13)	L13a	MH712490	TCCTGCGTAAGAGGATGAACAC	TGGTTTTATGGGGCAGCATG	108
Elongation Factor-1 alpha	EF1	MH712491	CGTGAAGCTGATCCCACAGA	CCAGAGACCTCCTTGTACTCG	141
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH	MH716023	AAGTCGGTATCAATGGATTTGGC	TCCACCTTCTTGGAGACGAA	73

Table 2. Summary of the assembled transcriptome of *M. incilis*

Item	Number
Total length of sequence	91.486.749 bp
Total number of sequences	93.947
Average contig length	973 bp
Largest contig	17.447 bp
Shortest contig	201 bp
N25 stats	25% of total sequence length is contained in the 4.641 sequences >= 3.634 bp
N50 stats	50% of total sequence length is contained in the 13.471 sequences >= 2.019 bp
N75 stats	75% of total sequence length is contained in the 31.920 sequences >= 786 bp
Total GC count	43.533.422 bp
GC %	47.58 %
Number of Ns	0
Ns %	0.00 %

Ns, ambiguous bases

Table 3. PCR primer efficiencies used in Real-time qPCR assays.

Genes and pathways	Amplification Efficiency (%)	R²
Oxidative Stress		
SOD1	107.8	0.99
CAT	93.5	1.00
HSP70	112.8	0.98
HIF1A	100.3	1.00
Exposure to xenobiotics		
CYP1A1	99.7	1.00
CYP3A	90.1	1.00
RXRA X2	93.7	1.00
AHR	116.6	0.99
Apoptosis/Necrosis		
Bcl-X	100.9	1.00
BAX	116.1	1.00
CASP3	101.8	1.00
Gadd45A	106.0	0.95
TNF-α	113.3	0.98
Immune Activation/inflammation		
IL6	89.6	0.99
HLA-DRA	91.8	0.99
Lipid metabolism		
PPARA	94.2	1.00
PPARG	91.2	1.00
Housekeeping		
L13a	93.0	1.00
EF1	96.0	1.00
GAPDH	91.7	1.00

Figures

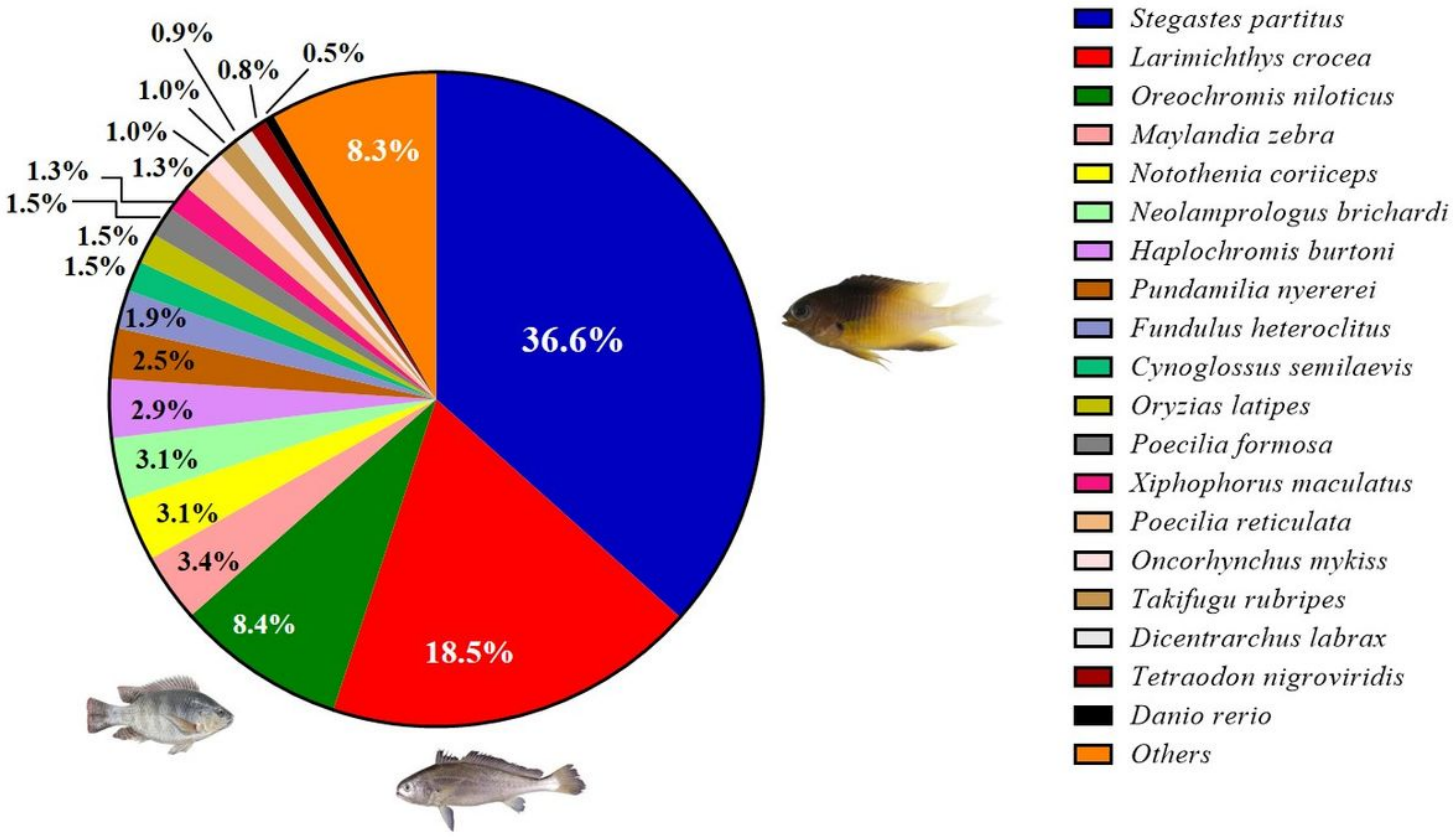

Figure 1

Similarity analysis based on the best hit.

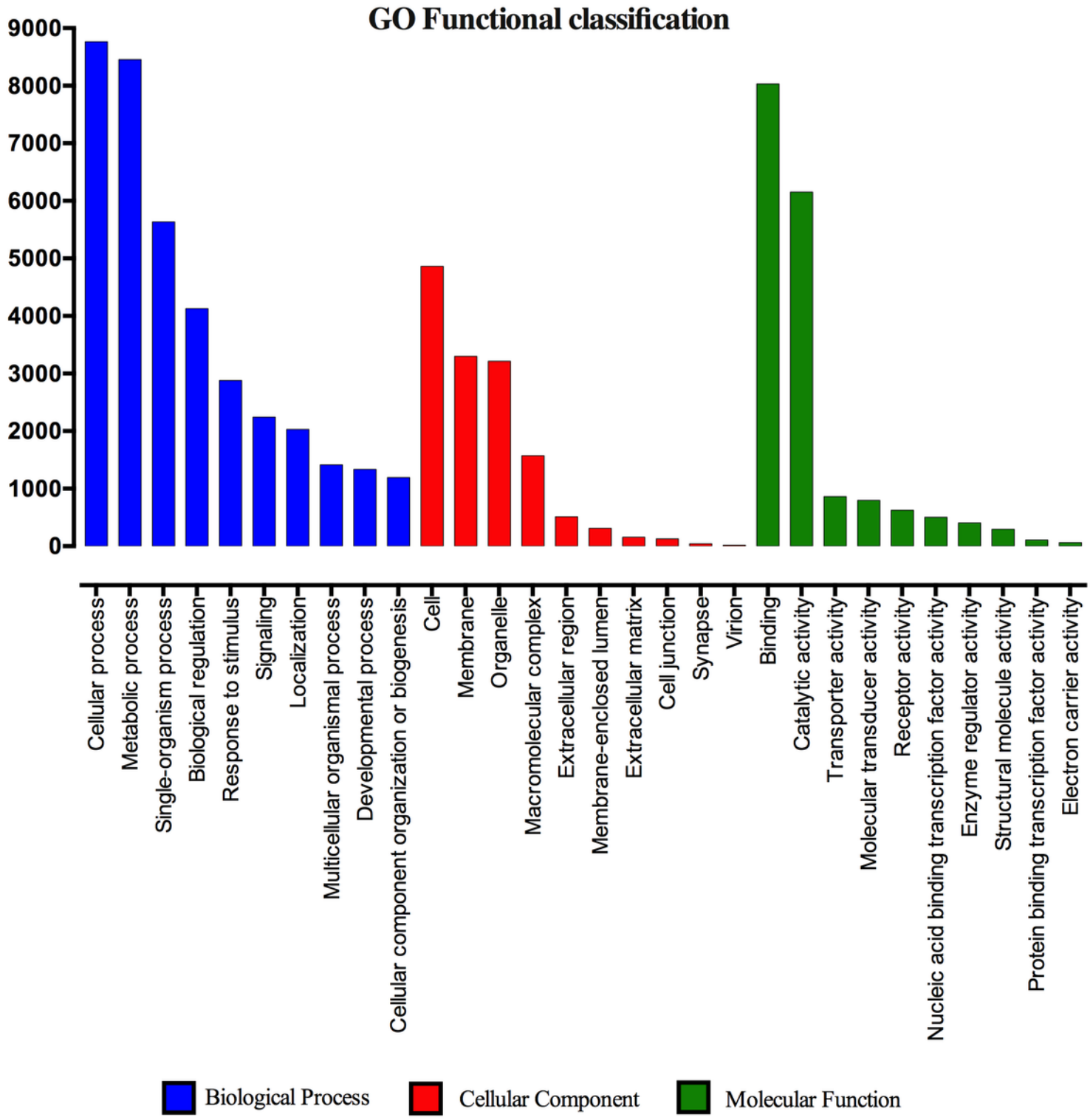

Figure 2

Gene ontology annotations based on Blast2GO analysis. GO terms were annotated at level 2 of classification according to three main categories (Biological process, Cellular component and Molecular function). The x-axis indicates the subcategories, and the y-axis registers the number of genes in the same category.

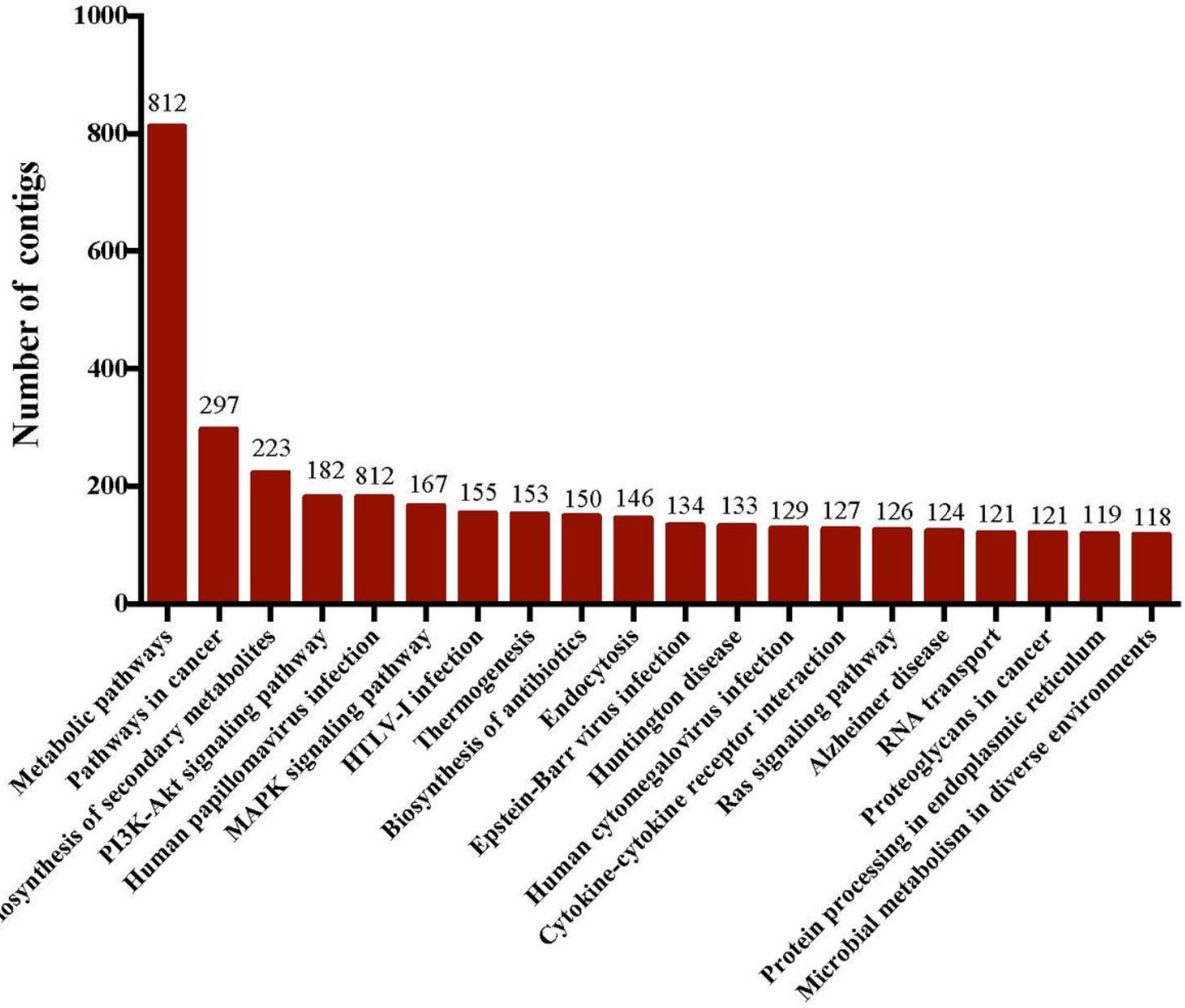

Figure 3

Top 20 pathways with the largest numbers of contigs based on KEGG annotation analysis.

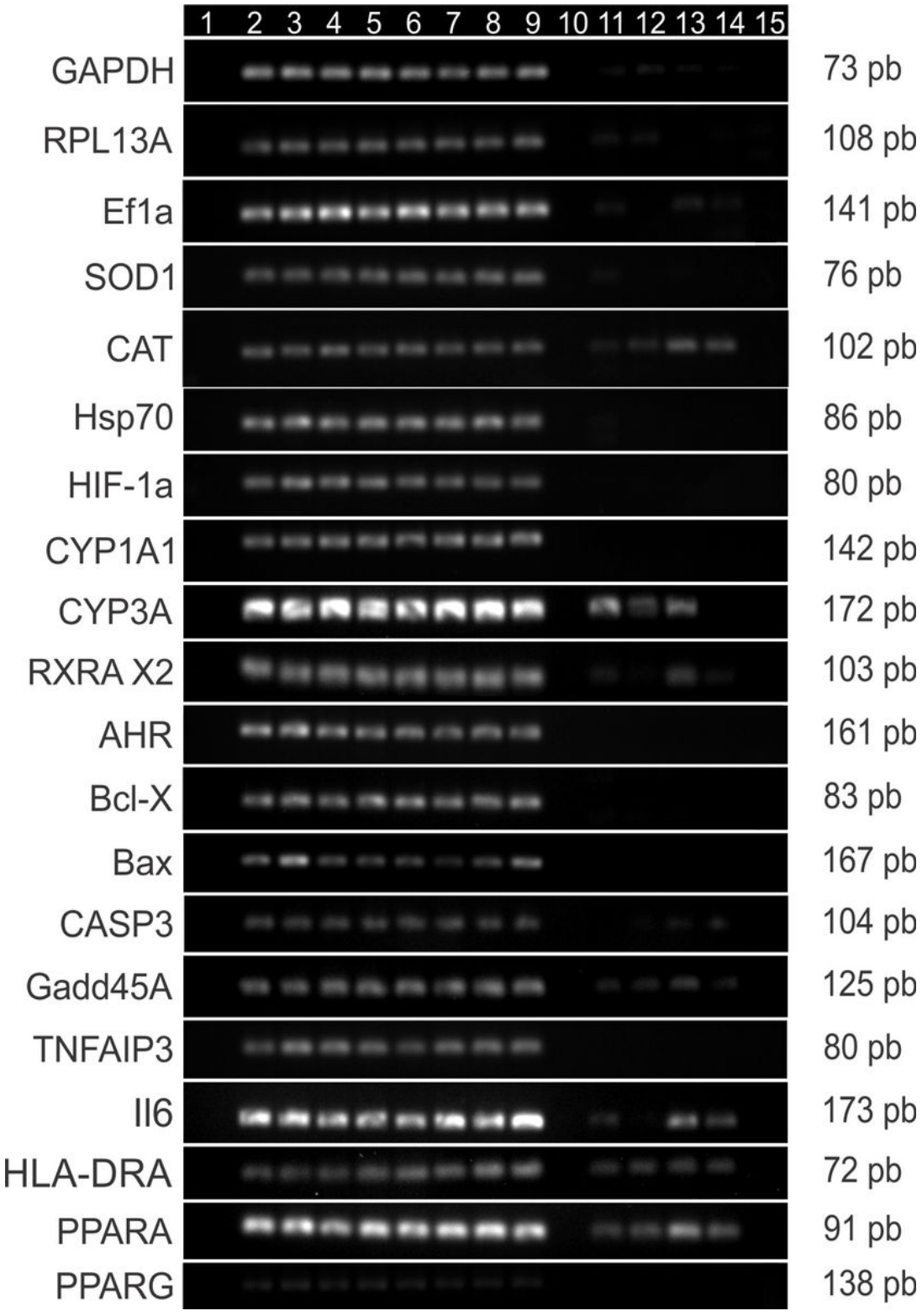

Figure 4

Gel electrophoresis of the RT-PCR products of genes associated with stress responses/xenobiotic detoxification. The lanes correspond to blank (1), cDNA from M. incilis (2-9), P. magdalenae (11-12), H. malabaricus (13-14) and M. musculus (15).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.docx
- supplement2.docx
- supplement3.docx
- supplement4.jpg
- supplement5.jpg
- supplement6.png