



OPEN

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

Angela Bertel-Sevilla¹, Juan F. Alzate² & Jesus Olivero-Verbel¹✉

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. A total of 32,082,124 RNA-Seq read pairs were generated utilizing the HiSeq platform and subsequently cleaned and assembled into 93,912 contigs (N50 = 2,019 bp). The analysis of species distribution revealed that *M. incilis* contigs had the highest number of hits to *Stegastes partitus* (13.4%). Using a sequence similarity search against the public databases GO and KEGG, a total of 7,301 and 16,967 contigs were annotated, respectively. KEGG database 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. 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.

Industrialization and urbanization are the main causes of aquatic ecosystem degradation, discharging a wide variety of pollutants that promote deleterious effects on organisms, with typical responses such as inhibition of growth, alterations in sexual maturation (reproductive capacity) and immunodeficiency¹⁻³. The problem is magnified when aquatic organisms like fish, accumulate pollutants directly from water or sediment, and indirectly through the food chain⁴. Therefore, fish can serve as useful indicators of aquatic pollution as they play multiple roles across the trophic web, bioaccumulate chemical substances, and respond to low concentrations of toxicants^{5,6}.

Mugilidae (mullet) 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 fisheries⁷⁻⁹. Over the past years, Mulletts have been proposed as pollution bioindicators for environmental degradation¹⁰. 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 Brazil¹⁸. It can be found both in brackish estuaries and in marine and hyper-saline waters¹⁹. Juvenile fish (< 25 mm) are primarily planktonic or carnivorous feeders, whereas larger specimens switch their diet to detritus and benthic microalgae, ingesting large

¹Environmental and Computational Chemistry Group, School of Pharmaceutical Sciences, Zaragocilla Campus, University of Cartagena, 130015 Cartagena, Colombia. ²Centro Nacional de Secuenciación Genómica—CNSG, Sede de Investigación Universitaria-SIU, Universidad de Antioquia, Medellín, Colombia. ✉email: joliverov@unicartagena.edu.co

Item	Number
Total length of sequence	91,495,553 bp
Total number of sequences	93,912
Average contig length	974 bp
Largest contig	17,447 bp
Shortest contig	201 bp
N25 stats	25% of total sequence length is contained in the 4,329 sequences \geq 3,634 bp
N50 stats	50% of total sequence length is contained in the 12,855 sequences \geq 2,019 bp
N75 stats	75% of total sequence length is contained in the 30,528 sequences \geq 786 bp
Total GC count	43,537,515 bp
GC %	47.58%
Number of Ns	0
Ns %	0.00%

Table 1. Summary of the assembled transcriptome of *M. incilis*. Ns, ambiguous bases.

amounts of organic matter, sand or mud from the sediments²⁰. These fish are in high demand in commercial and recreational fishing due to their high protein and vitamin content, being an important source of food in many countries¹⁵. However, there are several studies that report this fish as an intermediate host of several parasites, which contrasts with its nutritional importance²¹. 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 chemical contamination of the bay has been the result of continuous, direct and indirect discharges of urban and industrial waste^{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 endosulfan¹⁶). This species has also been found parasitized with nematodes of the Anisakidae family, in particular *Contraecum* sp., as well as with Heterophyidae 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 is 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 pressure³¹⁻³³. For example, de novo assembly of transcriptomes from important aquaculture species, such as *Epinephelus coioides*, *Larimichthys crocea*, *Scophthalmus maximus*, *Lateolabrax japonicus* 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 stressors³⁶, 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 several toxicological studies and represents an essential organ for biomonitoring purposes^{1,12,37}, mainly because it is sensitive to pollutant exposure, and it allows the assessment of hepatic damage and responses. Furthermore, its role in detoxification, nutrient synthesis and the transformation and storage of metabolites, represents complex biological functions that are essential for the survival of the organism^{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.

Results

De novo assembly of the *Mugil incilis* hepatic transcriptome. Approximately 32,082,124 raw read 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 93,912 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 1). 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 40,296 contigs was considered from *Mugil incilis* (Table S1, Fig. S1).

The completeness of the *M. incilis* transcriptome assembly was quantified by comparing against a dataset of 3,640 actinopterygii single-copy orthologs (actinopterygii_odb10) using the BUSCO gene mapping method (BUSCO, RRID:SCR 015,008) for transcriptome assembly validation. This assessment reveals 61.0% completeness with the BUSCO Vertebrata gene set (48.9% complete single-copy BUSCOs, and 12.1% complete duplicated BUSCOs), 8.8% fragmented BUSCOs, and 30.2% of the 3,640 single-copy orthologs were classified as missing from our assembly (Table S2).

Similarity analysis of the *M. incilis* liver transcriptome. Species distribution analysis (Fig. S2) revealed that maximum percentage of *Mugil incilis* transcript contigs showed significant similarity with *Stegastes partitus* (13.4%), followed by *Amphiprion ocellaris* (7.9%), *Larimichthys crocea* (5.1%) and *Acanthochromis polyacanthus* (4.5%). In addition, 237 (0.6%) of the transcripts were annotated to other eukaryotes, i.e., mammals, teleost fishes, and few prokaryotes vertebrates. Notably, due to the scarcity of genome information, only 39 contigs (0.1%) were matched with the genes from mugilid species such as *Mugil curema*.

GO annotation of the *M. incilis* liver transcriptome. GO terms were assigned to assemble contigs and provided defined ontologies to express gene product properties. Among the 40,296 contigs detected in the liver transcriptome, a total of 7,301 sequences had GO term assigned, which were subsequently categorized into 93 functional groups (Fig. S3). GO analysis assigned 4,695 (40%) contigs to biological process, 5,069 (44%) to cellular component, and 1,869 (16%) to molecular function. Briefly, for biological processes, genes involved in cellular process (GO:0009987) and multicellular organismal process (GO:0032501) were the most dominant categories; for molecular functions, binding (GO:0005488) were the most represented GO term, followed by catalytic activity (GO:0003824); while the principal categories observed in cellular component were cells (GO:0005623) and organelles (GO:0043226). We also identified genes involved in other major functions, such as developmental and metabolic processes, biological regulation, response to stimulus, and localization. The top 10 groups in the three main categories are shown in Fig. 1.

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 16,967 contigs were annotated in KEGG and located to 401 known KEGG pathways (Table S3). The top 20 pathways with the largest numbers of contigs are shown in Fig. 2. Among the 7 predicted KEGG pathways, the metabolic pathways represented the largest group, consisting of 898 contigs (5.29%), followed by cancer (290, 1.71%), biosynthesis of secondary metabolites (224, 1.32%), human papillomavirus infection (172, 1.01%) and PI3K-Akt signaling (170, 1.00%) (Fig. 2 and Table S3). These 401 pathways were subjected to six categories: human diseases (5,238, 30.9%), organismal systems (3,342, 19.7%), metabolism (3,269, 19.3%), environmental information processing (2,365, 13.9%), cellular processes (1,665, 9.8%), and genetic information processing (1,088, 6.4%) (Table S4).

Validation of primers for 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 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 2.

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 Fig. 3.

Discussion

The accelerated development of high-throughput sequencing technology has revolutionized our way of studying genetic diversity and gene expression. In this road, the RNA-Seq technology has become the method of choice for the study and characterization of dynamic transcriptomes, and the de novo transcriptome assembly provides a powerful research tool for differential transcriptome analysis in fish species⁴². In the past several years, transcriptomic studies in teleost fish have increased, leading to a tremendous pool of genetic knowledge (e.g. see database FISHIT [<https://www.fish-it.org/hcmr>] for 20 transcriptomes). However, although 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.

GO Functional classification

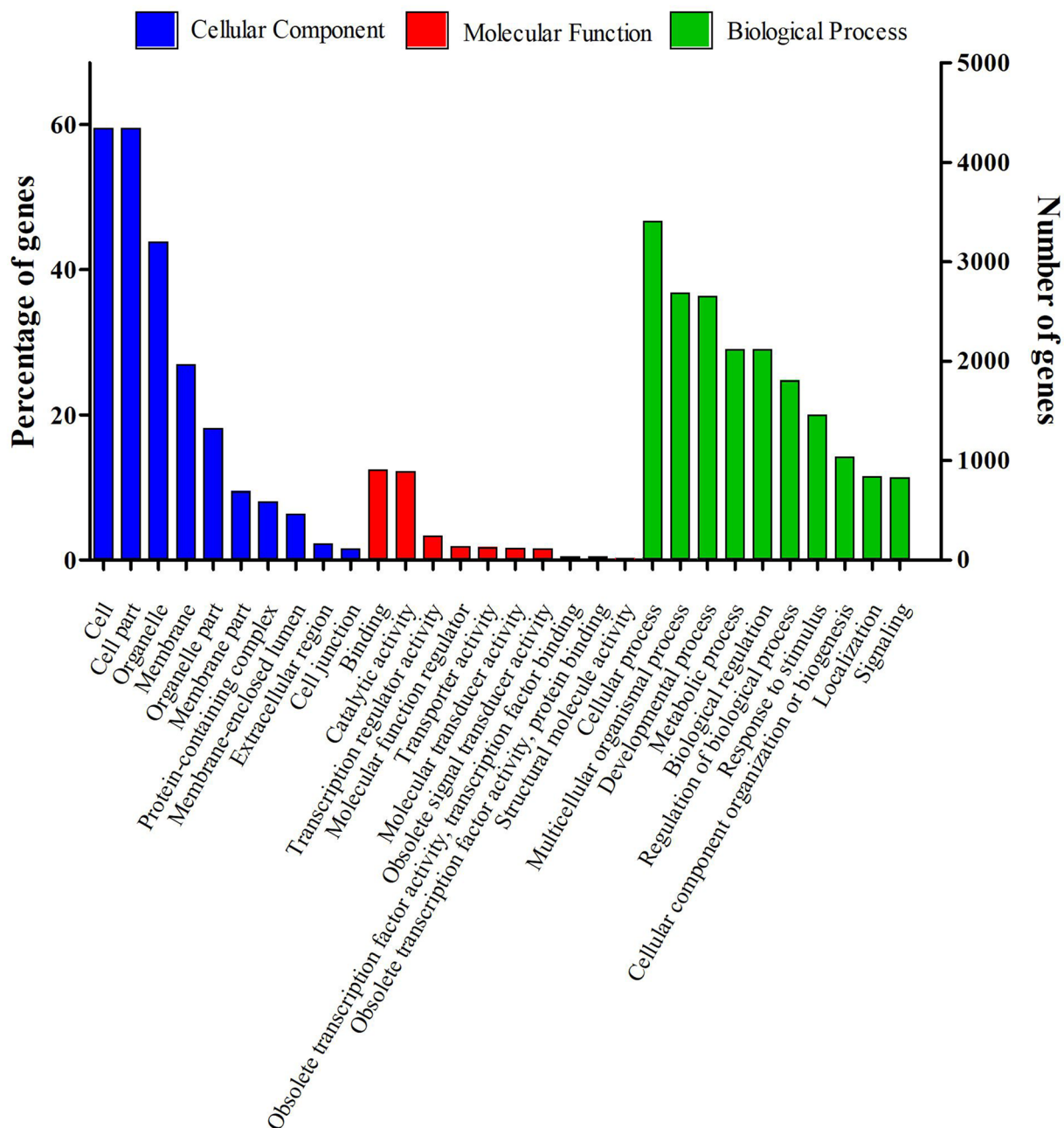


Figure 1. Function classifications of Gene Ontology (GO) terms of *M. incilis* transcriptome data. GO terms were annotated at level 2 of classification according to three main categories (Biological process, Cellular component and Molecular function). The *x*-axis indicates the subcategories, and the *y*-axis registers the percentages of genes (number of a particular gene divided by total gene number).

Some reports on transcriptome data for Mugilids have been focused on population genetic structure⁴³ 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 93,912 contigs were assembled with an average length of 974 bp, a greater value than those obtained in previous studies using other de novo assembly methods for teleosts, such as viviparous blenny (*Zoarcus viviparus*; 395 pb)⁴⁶, Guppy (*Poecilia reticulata*; 464 pb)⁴⁷, European eel (*Anguilla*; 531 pb)⁴⁸ and common carp (*Cyprinus carpio*; 888 pb)⁴⁹. All this transcriptome sequencing data may be sufficient and necessary to discover new genes and create new tools for the analysis of gene function and expression in *M. incilis*.

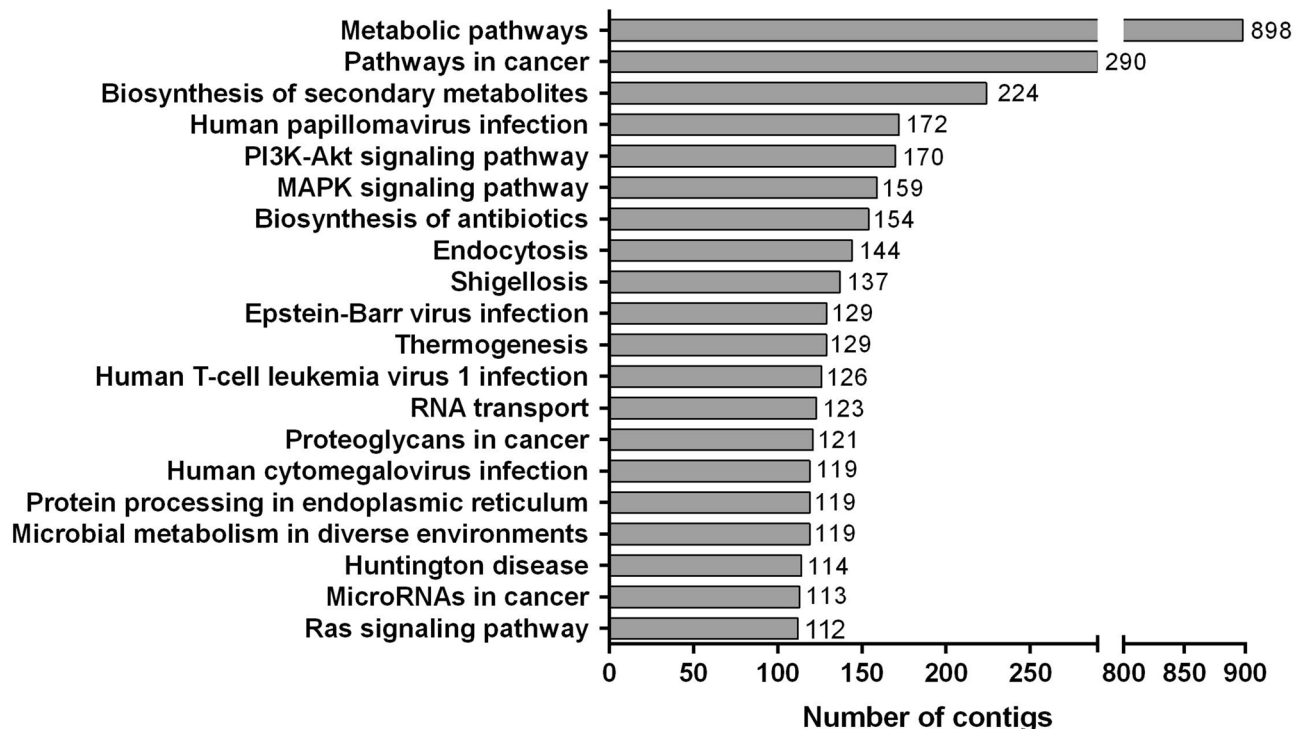


Figure 2. Top 20 pathways with the largest numbers of contigs based on KEGG annotation analysis.

To evaluate the quality and completeness of our transcriptome assembly we used the N50 reference-free metric and the reference-based BUSCO metric, which give us an idea of how good our assembly is in terms of continuity and coverage of the contigs respectively. The N50 measure showed that 50% of total sequence length is contained in sequences equal or greater than 2,019 bp, which is not a short length for the transcripts in our transcriptome assembly^{47,49,50}. On the other hand, the BUSCO analysis showed that most of the Actinopterygian single-copy orthologs (61%) were found completely in our assembly and 30.2% of the 3,640 orthologs were classified as missing from our assembly. This seems to be a relatively good level of completeness, when comparing with other assemblies reported for this class of organisms, especially if we consider that BUSCO recovery tends to be higher when multiple tissues and different stages are used to generate the assemblies, and our assembly was made from a single tissue of a single specimen^{50,51}. Furthermore, only 8.8% of the orthologs were recovered partially, indicating a low level of incomplete transcripts. We also ran a BUSCO analysis using the Vertebrata and Eukaryota datasets (see Table S2 in supplementary material), which showed a greater percentage of complete BUSCOs recovery (68% and 92.1% respectively) which may reflect evolutionary innovations in the *Mugil* genus.

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*, *A. ocellaris* and *L. crocea* (Fig. S2). 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 13.4%, suggesting these two fish species might be distantly related in subfamilies of Perciformes, Mugilidae and Pomacentridae for *M. incilis* and *S. partitus*, respectively.

The GO functional annotation indicated genes involved in cellular process and multicellular organismal were the most represented categories for biological processes. For cellular components, the major category represented was cell. For molecular function, binding was the most strongly represented GO term, followed by catalytic activity, results that can be useful for gene functional studies. Furthermore, 16,967 contigs were annotated to 6 categories in KEGG, including cellular processes, genetic information processing, environmental information processing, metabolism, human diseases and organismal systems. Signal transduction had the most number of genes (2,007) in environmental information. For metabolism, global and overview maps had the most number of genes (1,572), whereas for organismal systems, the immune system comprises the most number of genes (1,045) (Table S4). In the present study, PI3K-Akt, MAPK, Ras and Rap1 signaling pathways were found to be the most highly enriched. These pathways are ubiquitous and crosstalk in all eukaryotic cells. MAPK signaling, for example, regulates a wide variety of cellular processes, including proliferation, motility, differentiation, stress responses, and apoptosis⁵²; as well as cellular processes related to responses against diverse environmental stressors, such as heat, cold, UV radiation, reactive oxygen species, desiccation and pathogen attack⁵³. The transcription of genes related to these signaling pathways suggested *M. incilis* is a sensitive model for physiological studies.

The 1,045 genes in the immune system were further grouped into 21 pathways (Fig. S4). NOD-like receptor signaling pathway and chemokine signaling pathway were the top two pathways with the greatest number

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	CTGGATGTGACCCCTTGTC	GTTGGGATGGTGGTGTCTCT	86
Factor Inducible por Hipoxia-1, Subunidad Alfa	HIF1A	MH716027	TAGTTGAGAGCAGCCAACCC	CAGCTCTGCACACTTGTCTCT	80
Exposure to xenobiotics					
Cytochrome P4501A	CYP1A1	MH716028	TGTTCTTGCCATCCTCGTC	CAGTGGTCATCCCTCACTCG	142
Cytochrome P450, family 3, subfamily A	CYP3A	MK110662	GACTCTCCTGGTTGCGTTCA	CCCGTACTTCTTGTGGCACT	172
Retinoid X receptor alpha isoformX2	RXRA X2	MK110663	TTCAACCCAGACTCCAAAGG	CGGGATACTTGTGTTTGCAG	103
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	TTCACTTACCCTCATCCAG	TTCACCCCCACTCTGTTTTT	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	AGCTGCACTCGTTTCTTTTG	72
Lipid metabolism					
Peroxisome proliferator-activated receptor alpha	PPARA	MK183038	GAGGACTCGGTGTTGGACAG	CGTCGATGGACTGGGAGATG	91
Peroxisome proliferator-activated receptor gamma	PPARG	MH716033	CGGAGACAACATGCCTTTTCG	CCTGAAACTCCGTGTGCAGT	138
Housekeeping					
Ribosomal Protein L13a (RP-L13)	L13a	MH712490	TCCTGCGTAAGAGGATGAACAC	TGGTTTTATGGGGCAGCATG	108
Elongation Factor-1 alpha	EF1	MH712491	CGTGAAGCTGATCCACAGA	CCAGAGACCTCCTTGTACTCG	141
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH	MH716023	AAGTCGGTATCAATGGATTTGGC	TCCACCTTCTTGAGACGAA	73

Table 2. Primer sequences for real-time PCR.

of expressed genes. Nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) are a specialized group of cytoplasmic pattern recognition receptors (PRRs) that play an important role in the detection of various pathogens and regulation of innate immune responses. In fish, three distinct subfamilies of NLRs have been identified and characterized: NLR-A, NLR-B, and fish-specific NLR-C⁵⁴. Expression of NLR-C subfamily (NLRs) has been reported in many different teleost fish species such as zebrafish⁵⁵, Japanese flounder⁵⁶ and turbot⁵⁷. 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 response⁵⁴. The other signaling pathway with greater annotation was related to chemokines, a superfamily of small size chemoattractant peptides that provide directional cues for cell migration and positioning, fundamental for protective host response. The chemokine-mediated immune response to pathogens involves not only leukocyte migration, but also promotes differentiation of recruited cells to trigger the first steps of the innate and acquired immune response⁵⁸. 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 (Fig. 3).

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.

Genes and pathways	Amplification efficiency (%)	R ²
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

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

Materials and methods

Sample collection and handling. 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 PRINSEQ⁵⁹. After the quality filter, reads with less than 50 bases were rejected. The assembly step was carried out with the program Trinity (<https://trinityrnaseq.sourceforge.net>) employing default settings⁶⁰. 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 eggNOG database (<https://eggnoG.embl.de>) in order to obtain general gene ontology (GO) annotations. The contigs/transcripts annotation was done assigning the GO terms according to three main categories: molecular function; cell component; and biological process⁶¹, and the gene ontology annotation results were plotted using the WEGO software at GO level 2 with default parameters (<https://wego.genomics.org.cn>). The assignment of KEGG pathways to the transcripts was done through the online KEGG Automatic Annotation Server (KAAS; <https://www.genome.jp/kegg/kaas>) using the bi-directional best-hit method. Several contigs/transcripts of the stress responses/detoxification and immune system

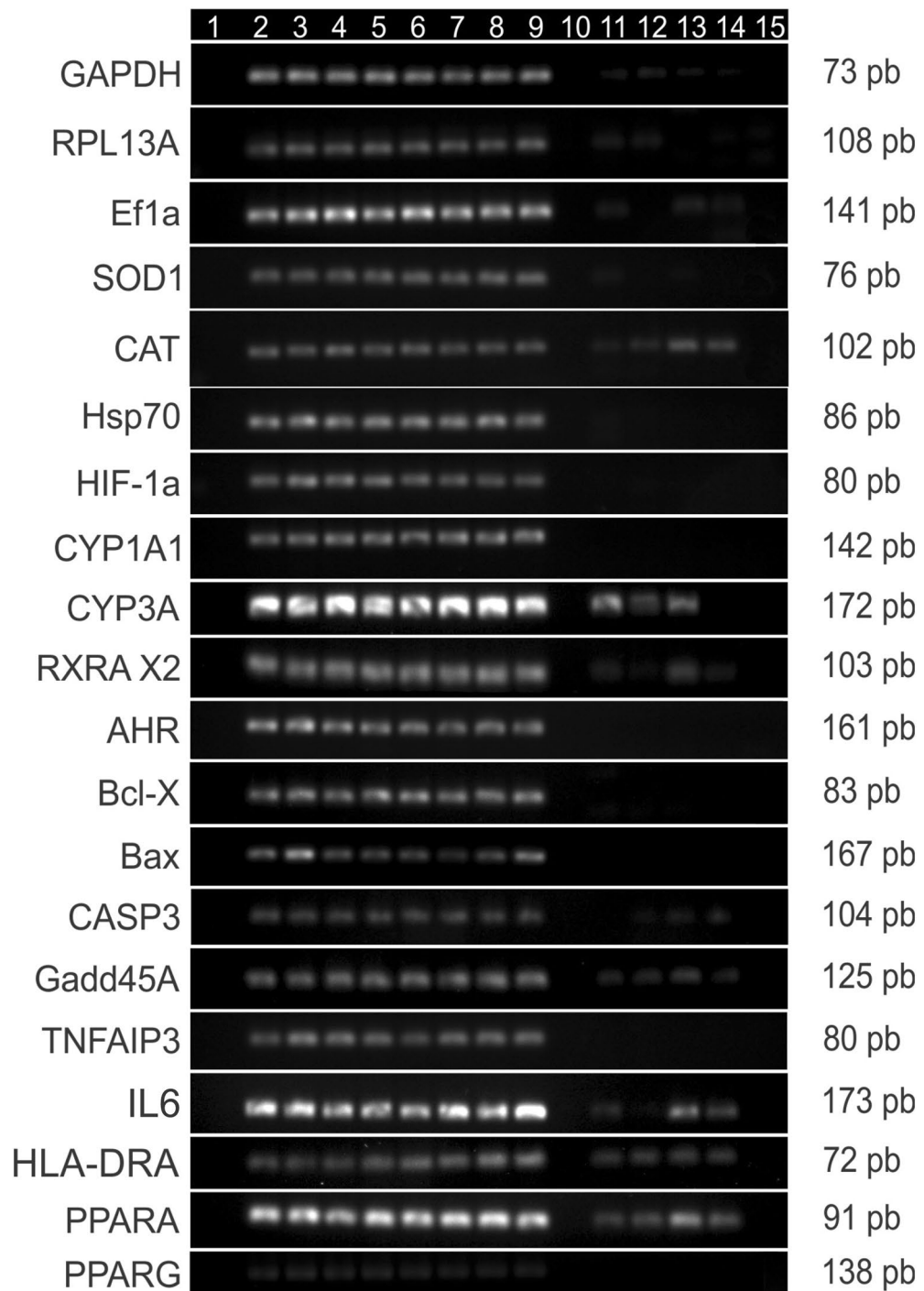


Figure 3. Gel electrophoresis of the RT-PCR products of genes associated with stress responses/xenobiotic detoxification. The lanes correspond to blank (1), cDNA from *M. incilis* (2–9), *P. magdalenae* (11–12), *H. malabaricus* (13–14) and *M. musculus* (15). All the gels were run under the same experimental conditions and are presented using cropped images. The entire gel photos of all genes evaluated can be found in Figure S5.

were selected and annotated using the BLASTx results and the KEGG annotation. CDS prediction was carried out manually using Artemis visualizer (Sanger)⁶².

Design of primers and validation for real-time qPCR assays. 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 (<https://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 RNA⁶³. 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 2.

The total RNA was isolated from liver tissue using RNeasy[®] Mini Kit (Qiagen, California, USA) as described by the manufacturer. The RNA was quantified by spectrophotometry (A260) using a NanoDrop 2000 Spectrophotometer (Thermo Scientific), and the A260:A280 ratio (1.9–2.0) was calculated to verify the RNA purity. Furthermore, the integrity of RNA was assessed by visual inspection of 28S and 18S ribosomal RNA after electrophoresis on an agarose gel. One microgram of total RNA was reverse transcribed, for each sample, using QuantiTect[®] Reverse Transcription Kit (Qiagen Inc, Valencia, CA, USA). The complementary DNA was used as template in a 20 µL PCR reaction containing 10 pmol each of forward and reverse gene-specific primers and 2 × SYBR[®] Green qPCR Master Mix (Thermo Scientific). Real time-PCR reactions were run in a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, CA), under the following conditions: 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 contamination⁶⁴.

Ethical statement. Protocols for the use of animals in this study were reviewed and approved by the Ethics Committee of the University of Cartagena (Act No. 123; 30-07-19). The field study did not involve endangered or protected species, and it was granted with permission to access genetic resources from the Ministry of Environment, Housing and Territorial Development from Colombia (Contract No. 148; 09-11-17). Conducted experiments were carried out in accordance with the International Guiding Principles for Biomedical Research Involving Animals (<https://www.cioms.ch/frame1985textsofguidelines.html>).

Conclusion

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 developing primers for qPCR, which would be used to assess gene expression levels in mugilid species.

Data availability

The raw reads produced in this study were deposited in the NCBI database Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/Traces/sra>) under the accession number SRP157095. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/ENA/GenBank under the accession GHNU00000000.

Received: 12 August 2019; Accepted: 24 July 2020

Published online: 18 August 2020

References

1. Ameur, W. B. *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.* **64**, 241–251. <https://doi.org/10.1016/j.marpolbul.2011.11.026> (2012).
2. Brázová, T. *et al.* Perch and its parasites as heavy metal biomonitors in a freshwater environment: the case study of the Ružín water reservoir, Slovakia. *Sensors (Basel)* **12**, 3068–3081. <https://doi.org/10.3390/s120303068> (2012).
3. Goretti, E., Pallottini, M., Ricciarini, M. I., 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.* **559**, 339–346. <https://doi.org/10.1016/j.scitotenv.2016.03.169> (2016).
4. Zhao, S. *et al.* Role of living environments in the accumulation characteristics of heavy metals in fishes and crabs in the Yangtze River Estuary, China. *Mar. Pollut. Bull.* **64**, 1163–1171. <https://doi.org/10.1016/j.marpolbul.2012.03.023> (2012).
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* **544**, 56–66. <https://doi.org/10.1016/j.gene.2014.04.032> (2014).
6. van der Oost, R., Beyer, J. & Vermeulen, N. P. E. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. *Environ. Toxicol. Pharmacol.* **13**, 57–149. [https://doi.org/10.1016/S1382-6689\(02\)00126-6](https://doi.org/10.1016/S1382-6689(02)00126-6) (2003).
7. FAO. 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. <https://www.fao.org/fishery/statistics/software/fishstat/en> (2013).
8. Nash, C. E. & Sehadah, Z. H. Review of breeding and propagation techniques for grey mullet, *Mugil cephalus* L.. *ICLARM Stud. Rev.* **3**, 1–87 (1980).
9. Thomson, J. M. The grey mullets. *Oceanogr. Mar. Biol. Annu. Rev.* **4**, 301–355 (1966).
10. Waltham, N. J., Teasdale, P. R. & Connolly, R. M. Use of flathead mullet (*Mugil cephalus*) in coastal biomonitor studies: review and recommendations for future studies. *Mar. Pollut. Bull.* **69**, 195–205. <https://doi.org/10.1016/j.marpolbul.2013.01.012> (2013).
11. Ameur, W. B. *et al.* Concentration of polychlorinated biphenyls and organochlorine pesticides in mullet (*Mugil cephalus*) and sea bass (*Dicentrarchus labrax*) from Bizerte Lagoon (Northern Tunisia). *Chemosphere* **90**, 2372–2380. <https://doi.org/10.1016/j.chemosphere.2012.10.028> (2013).
12. Vasanthi, L. A., 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* **91**, 1156–1164. <https://doi.org/10.1016/j.chemosphere.2013.01.021> (2013).

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.* **19**, 343–355. <https://doi.org/10.1080/09603120902749090> (2009).
14. Olivero-Verbel, J. *et al.* Perfluorooctanesulfonate and related fluorochemicals in biological samples from the north coast of Colombia. *Environ. Pollut.* **142**, 367–372. <https://doi.org/10.1016/j.envpol.2005.09.022> (2006).
15. Crosetti, D. & Blaber, S. J. M. *Biology, ecology and culture of grey mullet (Mugilidae)* (CRC Press, Taylor and Francis, Boca Raton, 2016).
16. Jaramillo-Colorado, B. E., Arroyo-Salgado, B. & Ruiz-Garcés, L. C. Organochlorine pesticides and parasites in *Mugil incilis* collected in Cartagena Bay, Colombia. *Environ. Sci. Pollut. Res. Int.* **22**, 17475–17485. <https://doi.org/10.1007/s11356-015-4986-5> (2015).
17. Whitfield, A. Ecological role of *Mugilidae* in the coastal zone. In *Biology, Ecology and Culture of Grey Mullet (Mugilidae)* (eds Crosetti, D. & Blaber, S. J. M.) 324–348 (CRC Press, Boca Raton, 2016).
18. Barletta, M. & Dantas, D. Biogeography and distribution of *Mugilidae* in the Americas. In *Biology, Ecology and Culture of Grey Mullet (Mugilidae)* (eds Crosetti, D. & Blaber, S. J. M.) 42–62 (CRC Press, Boca Raton, 2016).
19. Thomson, J. M. The *Mugilidae* of the world. *Mem. Queensl. Mus.* **41**, 457–562 (1997).
20. Cardona, L. Food and feeding of *Mugilidae*. In *Biology, Ecology and Culture of Grey Mullet (Mugilidae)* (eds Crosetti, D. & Blaber, S. J. M.) 165–195 (CRC Press, Boca Raton, 2016).
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. *Vet. Parasitol.* **168**, 31–35. <https://doi.org/10.1016/j.vetpar.2009.10.011> (2010).
22. Arrieta, L. & de la Rosa Muñoz, J. Estructura de la comunidad íctica de la Ciénaga de Mallorquín, Caribe Colombiano. *Bol. Invest. Mar. Cost.* **32**, 231–242 (2003).
23. Johnson-Restrepo, B. *et al.* Polycyclic aromatic hydrocarbons and their hydroxylated metabolites in fish bile and sediments from coastal waters of Colombia. *Environ. Pollut.* **151**, 452–459. <https://doi.org/10.1016/j.envpol.2007.04.011> (2008).
24. Olivero-Verbel, J. *et al.* Human and crab exposure to mercury in the Caribbean coastal shoreline of Colombia: impact from an abandoned chlor-alkali plant. *Environ. Int.* **34**, 476–482. <https://doi.org/10.1016/j.envint.2007.10.009> (2008).
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.* **109**, 157–163. [https://doi.org/10.1016/S0269-7491\(99\)00225-0](https://doi.org/10.1016/S0269-7491(99)00225-0) (2000).
26. Olivero Verbel, J., Caballero-Gallardo, K. & Arroyo-Salgado, B. Nematode infection in fish from Cartagena Bay, North of Colombia. *Vet. Parasitol.* **177**, 119–126. <https://doi.org/10.1016/j.vetpar.2010.11.016> (2011).
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. *J. Parasitol.* **91**, 1109–1112 (2005).
28. Lehtonen, K. K. & Schiedek, D. Monitoring biological effects of pollution in the Baltic Sea: neglected—but still wanted?. *Mar. Pollut. Bull.* **53**, 377–386. <https://doi.org/10.1016/j.marpolbul.2005.11.016> (2006).
29. Olsvik, P. A., Berg, V. & Lyche, J. L. Transcriptional profiling in burbot (*Lota lota*) from Lake Mjøsa—a Norwegian Lake contaminated by several organic pollutants. *Ecotoxicol. Environ. Saf.* **92**, 94–103. <https://doi.org/10.1016/j.ecoenv.2013.02.019> (2013).
30. Olsvik, P. A., Lindgren, M. & Maage, A. Mercury contamination in deep-water fish: transcriptional responses in tusk (*Brosme brosme*) from a fjord gradient. *Aquat. Toxicol.* **144–145**, 172–185. <https://doi.org/10.1016/j.aquatox.2013.10.002> (2013).
31. Olsvik, P. A., Amlund, H. & Torstensen, B. E. Dietary lipids modulate methylmercury toxicity in Atlantic salmon. *Food Chem. Toxicol.* **49**, 3258–3271. <https://doi.org/10.1016/j.fct.2011.09.025> (2011).
32. Uren Webster, T. M., Bury, N., van Aerle, R. & Santos, E. M. Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout. *Environ. Sci. Technol.* **47**, 8869–8877. <https://doi.org/10.1021/es401380p> (2013).
33. Pierron, F. *et al.* Effects of chronic metal exposure on wild fish populations revealed by high-throughput cDNA sequencing. *Ecotoxicology (London, England)* **20**, 1388–1399. <https://doi.org/10.1007/s10646-011-0696-z> (2011).
34. Li, G. *et al.* De novo assembly and characterization of the spleen transcriptome of common carp (*Cyprinus carpio*) using Illumina paired-end sequencing. *Fish Shellfish Immunol.* **44**, 420–429. <https://doi.org/10.1016/j.fsi.2015.03.014> (2015).
35. Zhu, J. *et al.* Transcriptomic profiling revealed the signatures of acute immune response in tilapia (*Oreochromis niloticus*) following *Streptococcus imiae* challenge. *Fish Shellfish Immunol.* **46**, 346–353. <https://doi.org/10.1016/j.fsi.2015.06.027> (2015).
36. Cossins, A. & Crawford, D. Fish as models for environmental genomics. *Nat. Rev. Genet.* **6**, 324–333. <https://doi.org/10.1038/nrg1590> (2005).
37. Gernhöfer, M., Pawert, M., Schramm, M., Müller, E. & Triebtskorn, R. Ultrastructural biomarkers as tool to characterize the health status of fish in contaminated streams. *J. Aquat. Ecosyst. Stress Recov.* **8**, 241–260. <https://doi.org/10.1023/A:1012958804442> (2001).
38. Vega-Retter, C. *et al.* Differential gene expression revealed with RNA-Seq and parallel genotype selection of the ornithine decarboxylase gene in fish inhabiting polluted areas. *Sci. Rep.* **8**, 4820–4820. <https://doi.org/10.1038/s41598-018-23182-z> (2018).
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.* **59**, 454–463. <https://doi.org/10.1007/s00244-010-9491-8> (2010).
40. Rebrikov, D. V. & Trofimov, D. Y. Real-time PCR: a review of approaches to data analysis. *Appl. Biochem. Microbiol.* **42**, 455–463. <https://doi.org/10.1134/s0003683806050024> (2006).
41. Bustin, S. & Huggett, J. qPCR primer design revisited. *Biomol. Detect. Quantif.* **14**, 19–28. <https://doi.org/10.1016/j.bdq.2017.11.001> (2017).
42. Kumar, G. & Kocour, M. Applications of next-generation sequencing in fisheries research: a review. *Fish Res.* **186**, 11–22. <https://doi.org/10.1016/j.fishres.2016.07.021> (2017).
43. Krück, N. C., Innes, D. I. & Ovenden, J. R. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). *Mol. Ecol. Resour.* **13**, 715–725. <https://doi.org/10.1111/1755-0998.12112> (2013).
44. Byadgi, O. *et al.* Transcriptome analysis of grey mullet (*Mugil cephalus*) after challenge with *Lactococcus garvieae*. *Fish Shellfish Immunol.* **58**, 593–603. <https://doi.org/10.1016/j.fsi.2016.10.006> (2016).
45. Qi, Z. *et al.* Transcriptome analysis of soiny mullet (*Liza haematocheila*) spleen in response to *Streptococcus dysgalactiae*. *Fish Shellfish Immunol.* **49**, 194–204. <https://doi.org/10.1016/j.fsi.2015.12.029> (2016).
46. Kristiansson, E., Asker, N., Förlin, L. & Larsson, D. G. J. Characterization of the *Zoarces viviparus* liver transcriptome using massively parallel pyrosequencing. *BMC Genom.* **10**, 345–345. <https://doi.org/10.1186/1471-2164-10-345> (2009).
47. Fraser, B. A., Weadick, C. J., Janowitz, I., Rodd, F. H. & Hughes, K. A. Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. *BMC Genom.* **12**, 202–202. <https://doi.org/10.1186/1471-2164-12-202> (2011).
48. Coppe, A. *et al.* Sequencing, de novo annotation and analysis of the first *Anguilla anguilla* transcriptome: EelBase opens new perspectives for the study of the critically endangered European eel. *BMC Genom.* **11**, 635–635. <https://doi.org/10.1186/1471-2164-11-635> (2010).
49. Ji, P. *et al.* Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. *PLoS ONE* **7**, e35152–e35152. <https://doi.org/10.1371/journal.pone.0035152> (2012).

50. Magalhães, M. G. P., Moreira, D. A., Furtado, C. & Parente, T. E. Development of SNVs and indels markers mined out of the first multi-organ transcriptomes from *Hypancistrus zebra* (Loricariidae), an endangered Amazonian catfish. *Genomics* **112**, 971–980. <https://doi.org/10.1016/j.ygeno.2019.06.013> (2020).
51. Carruthers, M. *et al.* De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genom.* **19**, 32. <https://doi.org/10.1186/s12864-017-4379-x> (2018).
52. Avruch, J. MAP kinase pathways: the first twenty years. *Biochim. Biophys. Acta* **1773**, 1150–1160. <https://doi.org/10.1016/j.bbamc.2006.11.006> (2007).
53. Sun, Z. *et al.* Liver transcriptome analysis and de novo annotation of the orange-spotted groupers (*Epinephelus coioides*) under cold stress. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **29**, 264–273. <https://doi.org/10.1016/j.cbd.2018.12.008> (2019).
54. 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.* **82**, 371–377. <https://doi.org/10.1016/j.fsi.2018.08.045> (2018).
55. Laing, K. J., Purcell, M. K., Winton, J. R. & Hansen, J. D. A genomic view of the NOD-like receptor family in teleost fish: identification of a novel NLR subfamily in zebrafish. *BMC Evol. Biol.* **8**, 42–42. <https://doi.org/10.1186/1471-2148-8-42> (2008).
56. Unajak, S. *et al.* Molecular characterization, expression and functional analysis of a nuclear oligomerization domain proteins subfamily C (NLRC) in Japanese flounder (*Paralichthys olivaceus*). *Fish Shellfish Immunol.* **31**, 202–211. <https://doi.org/10.1016/j.fsi.2011.05.007> (2011).
57. Hou, Z. *et al.* Characterization and expression profiling of NOD-like receptor C3 (NLRC3) in mucosal tissues of turbot (*Scophthalmus maximus* L.) following bacterial challenge. *Fish Shellfish Immunol.* **66**, 231–239. <https://doi.org/10.1016/j.fsi.2017.05.014> (2017).
58. Bird, S. & Tafalla, C. Teleost chemokines and their receptors. *Biology* **4**, 756–784. <https://doi.org/10.3390/biology4040756> (2015).
59. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)* **27**, 863–864. <https://doi.org/10.1093/bioinformatics/btr026> (2011).
60. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. <https://doi.org/10.1038/nbt.1883> (2011).
61. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435. <https://doi.org/10.1093/nar/gkn176> (2008).
62. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945. <https://doi.org/10.1093/bioinformatics/16.10.944> (2000).
63. Hori, T. S., Gamperl, A. K., Booman, M., Nash, G. W. & Rise, M. L. A moderate increase in ambient temperature modulates the Atlantic cod (*Gadus morhua*) spleen transcriptome response to intraperitoneal viral mimic injection. *BMC Genom.* **13**, 431–431. <https://doi.org/10.1186/1471-2164-13-431> (2012).
64. 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* **762**, 24–29. <https://doi.org/10.1016/j.mrgentox.2013.12.005> (2014).

Acknowledgements

The authors thank the National Program for Doctoral Formation (COLCIENCIAS, 727-2015); COLCIENCIAS (Bogotá, Colombia), the Ministry of Environment, Housing and Territorial Development from Colombia (Contract No. 148; 09-11-17), and the University of Cartagena (Support to Research Groups, 2018–2020), Cartagena, Colombia.

Author contributions

A.B.S., and J.O.V. conceived and designed the experiment. A.B.S. and J.F.A. performed the RNA-Seq data analysis. A.B.S. and J.F.A. performed the pathways analysis. A.B.S. developed the Real-time qPCR assays. A.B.S., J.F.A. and J.O.V. drafted the manuscript. All authors read and approved the final manuscript. Financial funds were obtained from Grants written by J.O.V.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-70902-5>.

Correspondence and requests for materials should be addressed to J.O.-V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020