Mol Biol Rep (2013) 40:6965–6975
DOI 10.1007/s11033-013-2816-8

Transcriptome generation and analysis from spleen of Indian catfish, *Clarias batrachus* (Linnaeus, 1758) through normalized cDNA library

Akanksha Singh · A. S. Barman · Neeraj Sood · Vindhya Mohindra

Received: 8 November 2012 / Accepted: 16 October 2013 / Published online: 27 October 2013

Abstract Catfishes are commercially important fish for both the fisheries and aquaculture industry. *Clarias batrachus*, an Indian catfish species is economically important owing to its high demand. A normalized cDNA library was constructed from spleen of the Indian catfish to identify genes associated with immune function. One thousand nine hundred thirty seven ESTs were submitted to the GenBank with an average read length of approximately 700 bp. Clustering analysis of ESTs yielded 1,698 unique sequences, including 184 contigs and 1,514 singletons. Significant homology to known genes was found by homology searches against data in GenBank in 576 (34 %) ESTs, including similarity to functionally annotated unigenes for 158 ESTs. Additionally, 433 ESTs revealed similarity to unigenes and ESTs in the dbEST but the remaining 658 EST sequences (39 %) did not match any sequence in GenBank. Of a total of 1,698 ESTs generated, 65 ESTs were found to be associated with immune functions. Gene Ontology and KEGG pathway analyses of *C. batrachus* ESTs collectively revealed a preponderance of immune relevant pathways apart from the presence of pathways involved in protein processing, localization, folding and protein degradation. This study constitutes first EST analysis of lymphoid organ in aquaculturally important Indian catfish species and could pave the way for further research of immune-related genes and functional genomics in this catfish.

Keywords Indian catfish · *Clarias batrachus* · Expressed sequence tags · cDNA library · Normalisation · Functional annotation

A. Singh · A. S. Barman · N. Sood · V. Mohindra (✉)
National Bureau of Fish Genetic Resources (ICAR), Canal Ring Road, P. O. Dilkusha, Lucknow 226002, UP, India
e-mail: vindhyamohindra@gmail.com

Abbreviations

ESTs	Expressed sequence tags
ORF	Open reading frame
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
KAAS	KEGG automatic annotation server
FAO	Food and Agriculture Organization

Introduction

The demand for catfishes in the last decade has been steadily increasing throughout the world as represented by FAO estimates [1990–2012; 1]. *Clarias batrachus* belongs to the family Clariidae which represents one of the major economically important families of catfish [2]. The catfish is very popular owing to its favourable taste, medicinal and high market value [3, 4] as well as its suitability to culture in limited space, making it a preferred and potential cultivable species [5, 6]. However, the production of this species is limited by its high disease susceptibility. This species has been studied extensively by many workers in terms of physiology [7], biochemistry [8–10], toxicology [11, 12], host–parasite interaction [13, 14], pathology [15, 16], culture characters [17, 18] as well as its population genetics [19–21]. However, very few genomic or transcriptomic resources are available for *C. batrachus,* limited only to microsatellites [19, 21] and this has hindered the progress of immunological and developmental research. Thus, there is an urgent need to develop large number of molecular markers, the construction of framework genetic linkage maps, the identification of putative markers involved in performance traits and development of extensive genomic resources.

To overcome this obstacle, the transcriptome sequences of *C. batrachus* spleen tissue were generated. EST sequencing represents an attractive strategy because EST sequencing analyzes transcribed portions of the genome or the transcriptome, while avoiding non-coding and repetitive sequences that can make up much of the genome. In addition, EST sequencing is also an effective way for gene discovery, gene expression profiling and to develop 'functional' genetic markers that are highly useful for genetic mapping and comparative genome analysis [22]. Spleen, being a major lymphoid organ of the fish, protects the fish from blood borne pathogens [23], houses immune cells such as T and B cells that destroy the antigen or neutralizes it to prevent further manifestation during disease process [24]. The data obtained disclosed a broad catalogue of genes expressed in spleen tissue of *C. batrachus*, which are involved in several intracellular immune signalling pathways.

Materials and methods

Fish

Live specimens of *C. batrachus* (30–50 g) were procured from commercial catches from different collection sites. The fish were kept in FRP tanks of 100 l capacity, filled up to 25 l (water temperature 22 ± 3 °C). The fish were acclimated for 30 days before tissue collection. Fish were fed once daily in the evening, with processed goat liver or goat flesh and soybean powder. The feeding was stopped 48 h prior to tissue collection. Experimental fish were anesthetized on ice and sacrificed. Spleen was removed from the fish, pooled and stored in liquid nitrogen.

RNA extraction from spleen tissue

Total RNA was isolated using TRI Reagent (T9424; Sigma-Aldrich, St. Louis, USA) in 50 mg:1 ml ratio, according to manufacturer's recommendation. Poly A+ RNA was isolated from approximately 250 µg total RNA using an Oligo dT based matrix (70042; Qiagen Inc., Valencia, CA), eluted in 20 µl elution buffer and stored at −80 °C. The quality of mRNA was evaluated on 1 % denaturing agarose gel and quantitated using Picodrop Microliter UV/Vis Spectrophotometer (Picodrop Ltd, Saffron Walden, UK).

Normalized cDNA library construction

To achieve normalization of cDNA by DSN (Duplex Specific Nuclease) method SMART IV oligonucleotide primer and CDS-3M adaptor (included in the TRIMMER-DIRECT kit, NK002; Evrogen, Moscow, Russia), was used instead of the SMART CDSIII primer. First strand cDNA was prepared from 0.5 to 1.0 µg mRNA using SuperScript III reverse transcriptase (18080-093; Invitrogen; Breda, Netherlands) by CREATOR SMART cDNA method (634903; Clontech, Palo Alto, CA, USA). The ss cDNA obtained was amplified by LD-PCR, using the 5′ PCR primer and the LD-PCR products were purified using the QIAquick PCR Purification Kit (2810; Qiagen). Normalization of purified ds cDNA was carried out using DSN (Duplex Specific Nuclease) treatment; according to TRIMMER DIRECT (Evrogen) protocol. The normalized cDNA was digested with SfiI enzyme and the aliquots containing the desired size of ds cDNA were directionally cloned into the SfiI sites of vector pDNR-LIB and transformed into ELECTROMAX DH 10B (Invitrogen) *E. coli* cells by electroporation using BIORAD GENE PULSUR X cell system ((BioRad Laboratories, Hercules, CA). The transformation mixture containing 5 µl of ligation reaction and 25 µl electrocompetent cells was electroporated at 1.8 kV/cm, 200 Ω, and 25 µF in 1-mm pre-chilled electrocuvettes (Bio-Rad) using the preset protocol for *E. coli*. The normalised library of 10^6 cfu/ml was amplified according to CREATOR SMART protocol and stored at −80 °C for further screening of the library.

EST generation, filtering and assembly

The individual clones from the amplified cDNA library were screened using universal M13 forward and reverse primers, flanking the multiple cloning site of the vector pDNR-LIB, according to manufacturer's recommendation (Clontech). The culture aliquot was denatured at 95 °C for 5 min and 10 µl of PCR reaction was setup using 0.5 µl of denatured culture according to the following program: 94 °C for 5 min; 94 °C for 1 min, 56 °C for 1 min, 72 °C for 1 min for 35 cycles; and a final extension step of 72 °C for 10 min. Five µl of PCR reaction was electrophoresed on 1 % agarose gel against GeNei™ Low Range DNA ruler Plus (Genei, Bangalore, India) and the insert size of the clones was ascertained. Sequencing of 2,045 clones was performed from 5′ direction using pDNRF2 primer (GGCCGCATAACTTCGTATAGC) [25] to generate expressed sequence tags.

Vector sequences were removed using web based tool VecScreen (NCBI) and sequences were assembled into contigs using CAP3 program (http://deepc2.psi.iastate.edu/aat/cap/cap.html) [26]. The linear assembly algorithm was used and the criteria for clustering were set at a minimum overlap of 30 bases (default is 20 bases) and each cluster was visually inspected to ensure fidelity of alignment. Redundancy number was calculated [27] by dividing the total number of ESTs against the total number of contigs. Full length ESTs were identified using the Full-Lengther

[28] interface. Full-Lengther classifies ESTs as full-length, putative full-length or non full-length based on matches (similarities) found by executing BLAST against a protein database. The sequences were analysed using the e-value cut-off of $<10^{-6}$ and using TrEMBL when there were no hits found using Swiss-Prot.

Functional annotation of ESTs

To assign function to the transcripts, BLAST [27] searches were conducted for each EST using BLASTX, TBLASTX and subsequently BLASTN against the non-redundant (nr) and the dbEST database with a BLAST cut-off of 1×10^{-5} and 1×10^{-10}, respectively, to confirm the identity of genes. All the BLAST results were visually inspected to ensure that the alignment was of high quality and the matches were not due to simple amino acid stretches or repeat regions.

GO annotations were assigned using the program Blast2GO (version 2.4.8) (http://www.blast2go.de) [29]. The Blast2GO annotation procedure consists of three main steps: blast to find homologous sequences, mapping to collect GO terms associated to blast hits, and annotation to assign trustworthy information to query sequences. QBlast@NCBI was used to search molecular sequence in databases with the BLASTX algorithm, E-value $\leq 10^{-3}$ and an HSP length cut-off of 33 against the nr database. The unannotated sequences were further blasted using TBLASTX algorithm. Mapping was done for retrieving GO terms associated to the hits obtained after a blast search. Further the functionality of InterPro annotations in Blast2GO allowed the retrieval of domain/motif information in a sequence-wise manner against PFAM, SMART, GENE3d, PROSITE, PROFILE, PRODOM, SUPERFAMILY, PANTHER, PIR, PRINTS, and TIGRFAMs [30]. Corresponding GO terms were then transferred to the sequences and merged with already existent GO terms. Annotation step was done by assigning functional terms to query sequences from the pool of GO terms gathered in the mapping step according to the following parameters: a pre-E-value-Hit-Filter of 10^{-6}, a pro-Similarity-Hit-Filter of 15, an annotation cut-off of 55, and a GO weight of 5. In order to simplify the functional annotations to a set of broad terms, GO annotation results were summarized by mapping to the Generic GO-Slim terms. Once GO terms had been gathered, additional functionalities enabled processing and modification of annotation results. From these annotations, ESTs were categorized into biological process, molecular function and cellular component categories. GO graphs were generated using a sequence filter of 0, an alpha score of 0.6 and a 0 node score.

In addition to the GO classification, the ESTs were annotated according to the KEGG orthology (KO) by KAAS [31]. KEGG [32] is a database resource for understanding higher-order functions and utilities of the biological system, such as the cell or the organism, from genomic and molecular information. The sequences were analysed using the bi-directional best hit (BBH) method to obtain the KO terms for the query sequences with a blast threshold of 40. Once genes were assigned KO identifiers or K numbers by the ortholog annotation procedure, the collective body of K numbers was mapped to pathway maps and BRITE functional hierarchies to enable higher level functional interpretation of the transcriptome.

Results

Normalized spleen cDNA library and ESTs generated

Normalized cDNA library, constructed from pooled spleen tissue of adult *C. batrachus* in plasmid vector, had a titre of 10^6 cfu/ml with 95 % recombination efficiency. A total of 1,937 ESTs were generated (GenBank Accesssion No. GR955281-GR955351; GT145370-GT145406; GT157698-GT157747; GT271588-GT271629; GW397085-GW397200; GW492626-GW492739; GW672442-GW672586; GW706923-GW707142; GW774922-GW775089; GW787312-GW787471; GW836095-GW836479; GW840301-GW840715). The average read length of the ESTs was approximately 700 bp, with the majority of reads falling between 700 and 800 bp (Fig. 1). Clustering yielded 1,698 unique sequences, redundancy factor of 2.6 (Table 1), comprised of 184 contigs (423 ESTs, average read length 823 bp) and 1,514 singletons (average read length 681 bp). Only 23 % (43) contigs had 3 or more sequences (Fig. 2). The clones that were represented 4 times or more in library were eukaryotic translation initiation factor 4Aiii (9 times, GW836146; GW840422; GW836115; GW840435; GW840448; GW840509; GW840510; GW840684; GW840703), Gap junction cx32.2 protein (4 times, GW787428; GW836421; GT157702; GW840609) and tcb1 like transposase gene (4 times, GW787355;GW707061;GW397126; GW836246). Forty seven full length ESTs were identified based on BLAST results when the alignment started within 15 amino acids and contained an in-frame stop codon, while 17 were putative full length cDNA. The majority of ESTs showed complete ORFs and represented unannotated cDNAs.

Functional annotation and gene ontology terms

For 1,698 unique sequences generated, 576 (34 %) sequences had significant homology to known genes, which included 158 ESTs similar to functionally annotated unigenes (Table 2). Additionally, 433 (25.5 %) ESTs revealed similarity to unannotated unigenes and ESTs in the dbEST, but the remaining 658 sequences (39 %) did not match any sequence in any database (Fig. 3). In total, 27 % of the ESTs showed homology to proteins with

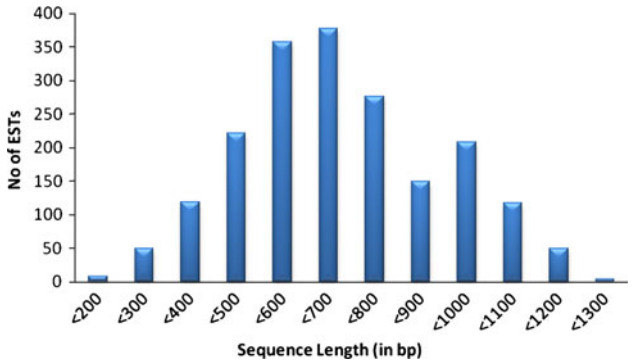


Fig. 1 Illustration of sequencing read lengths of *Clarias batrachus* spleen ESTs. Read lengths were grouped into 100 bp increments starting from < 200 to < 1,400 bp. The majority of ESTs were found to be 700–800 bp long

Table 1 Characteristics of ESTs generated from *Clarias batrachus* spleen normalized cDNA library

Total no. of ESTs generated	1,937
No. of singletons	1,514
No. of contigs (sequences in contigs)	184 (423)
Total no. of unique ESTs[a]	1,698
Redundancy number	2.29
Total no. of ESTs annotated with function	576

[a] Total no. of singletons + no. of contigs

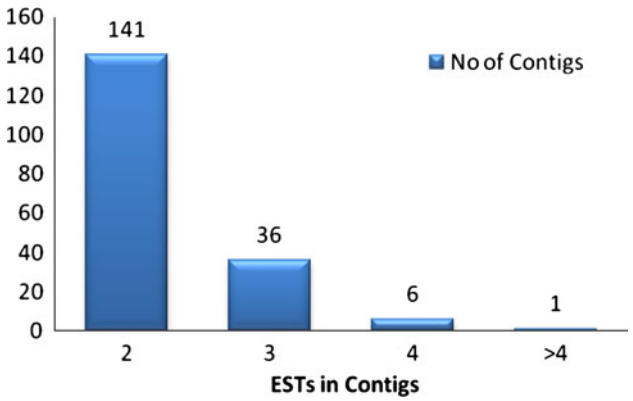


Fig. 2 Representation of number of *Clarias batrachus* spleen ESTs forming contigs. The majority of contigs contained only two ESTs and there was only one contigs containing more than 4 ESTs

known function, whereas 2 % of these homologies were to protein with no known function or hypothetical proteins. Additionally, functional information was derived for 9 % ESTs through the unigene database, the most common from channel catfish and blue catfish, followed by zebrafish unigenes. Eighty (5 %) EST showed similarity to unknown ESTs in the dbEST. Forty two (10 %) ESTs having similarity to genes and unigenes containing domain and 12 ESTs (2.8 %) with similarity to repeat containing proteins were detected.

Of the 576 functionally annotated and domain containing sequences, sequences with GO terms, corresponding to biological process fell into 58 categories, molecular function into 21 and cellular component into 38 categories (Fig. 4a–c). Interpro terms were assigned to 158 sequences having a total of 601 terms (Table 2).

Functional classification of *C. batrachus* spleen ESTs

KEGG categories were found for 185 sequences having 578 terms that fell into 149 pathways (Table 3).

Based on KEGG orthology (KO) terms and Enzyme Commission (EC) numbers

Classification based on EC and KO numbers yielded 165 sequences assigned with 84 EC and 311 KO mappings, respectively (Table 3). Only 65 unique sequences were assigned EC numbers and had 84 mappings to KEGG biochemical and metabolic processes (Table 3). The KEGG metabolic processes that were well represented were carbohydrate metabolism (9 enzymes), amino acid metabolism (8 enzymes), energy metabolism and metabolism of cofactors and vitamins (6 enzymes), lipid metabolism (5 enzymes) and nucleotide metabolism. Of these, 12 % of sequences belonged to the environmental information processing (EIP) category. The KEGG pathways well-represented under EIP and GIP (Genetic information processing) are given in Table 3. Immune (5 enzymes) and digestive systems (4 enzymes) were well represented under organisational systems.

Highly represented KEGG pathways

As a substitute method of categorizing clusters by GO, the non-enzyme based ESTs were assigned to metabolic pathways using the KEGG database (Table 4). 311 genes were assigned to 149 pathways under the KEGG pathway analysis and some had mappings in multiple pathways. The highest number (35 ESTs) of genes belonged to the immune system pathways and genetic information processing (19 ESTs). Among the individual KEGG pathways under the major categories, the most abundant pathways were ribosome (8) and protein processing in endoplasmic reticulum (8) falling under GIP, suggesting a high rate of protein metabolism in spleen. The individual KEGG pathways revealed a predominance of immune relevant pathways namely cytokine–cytokine receptor interaction (7), phagosome (7), chemokine signaling pathway (6), complement and coagulation cascades (5), Fc gamma

Table 2 Classification of unique EST sequences from spleen of *Clarias batrachus*

Classification	Classification method	Number of sequences
Unclassified	No BLAST hit	658
Unassigned protein	BLAST hit $>e^{-5}$ to unknown protein	31
Unknown EST	BLAST hit $>e^{-10}$ to unknown EST	80
Unassigned unigenes	BLAST hit $>e^{-10}$ to unknown EST	353
Functionally annotated unigenes	BLAST hit $>e^{-10}$	158
Functionally annotated protein	BLAST hit $>e^{-10}$ to known protein	418
Domain name-containing protein		42
Repeat containing protein		12
Informative terms		
Gene ontology (1,968)	295	
InterPro (158)	601	
KEGG (308)	165	
Total		1,698
Tandem repeats containing	221	
Containing more than one repeat	40	

The method of classifying functionally annotated proteins is indicated and the numbers of terms identified by GO, interpro and KEGG are given in brackets

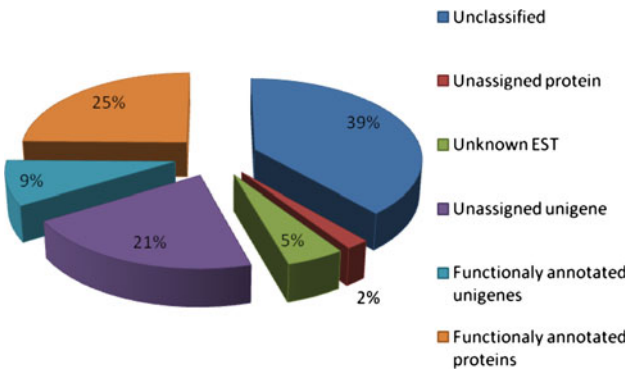


Fig. 3 Classification of unique EST sequences generated from *Clarias batrachus* spleen. 658 sequences (39 %) did not match any sequence in any database. In total, 27 % of the ESTs showed homology to proteins with known function including 2 % of the ESTs showing homology to hypothetical proteins

R-mediated phagocytosis (4) and leukocyte transendothelial migration (4).

Functional classification of ESTs based on interpro identifiers

The most informative domains were derived from Interpro (329), PANTHER (197), SUPERFAMILY (37), GENE3D (37) databases and almost negligible from PRODOM (1). The common Interpro protein domains identified in *C. batrachus* spleen ESTs are summarized in Table 5. Of the 329 Interpro protein domains, the most frequent interpro domain detected in *C. batrachus* transcripts were histone H5 (IPR005819), immunoglobulin-like fold (IPR013783) and immunoglobulin-like (IPR007110).

Discussion

The present study generated a broad catalogue of genes in spleen tissue by carrying out an EST analysis for the commercially important and aquaculture candidate catfish species *C. batrachus*. To take full advantage of the study and maximise the transcript coverage from the spleen, we used acclimated adult fishes as the RNA source for the study and DSN based normalization method [33, 34] was employed to increase the probability of rare transcript representation. Combined with the SMART approach, this method is ideally suited for rapid construction of enriched cDNA library and efficient identification of rare transcripts. High capacity libraries highly enriched in full-length inserts with low redundancy are useful tools in genetic analysis [35]. In the present study, a total of 1,698 unique sequences were generated from 1,937 expressed sequence tags with an average length of approximately 700 bp (500–800 bp) and the longer average contig length of 823 bp emphasizing the presence of longer transcripts in the spleen cDNA library using the SMART approach. The study revealed 1,514 singletons and 184 contigs from the normalized *C. batrachus* spleen cDNA library with a low redundancy factor (2.29). EST projects generate a large number of redundant sequences due to the random selection of cDNAs from tissue libraries [36], though interestingly the majority of contigs in this study were made up of only two ESTs sequences and only a 23 % of contigs contained 3 or more sequences. It suggested that the resultant library was efficiently normalized in reducing the representation of highly expressed genes and thus enriching the rarer transcripts.

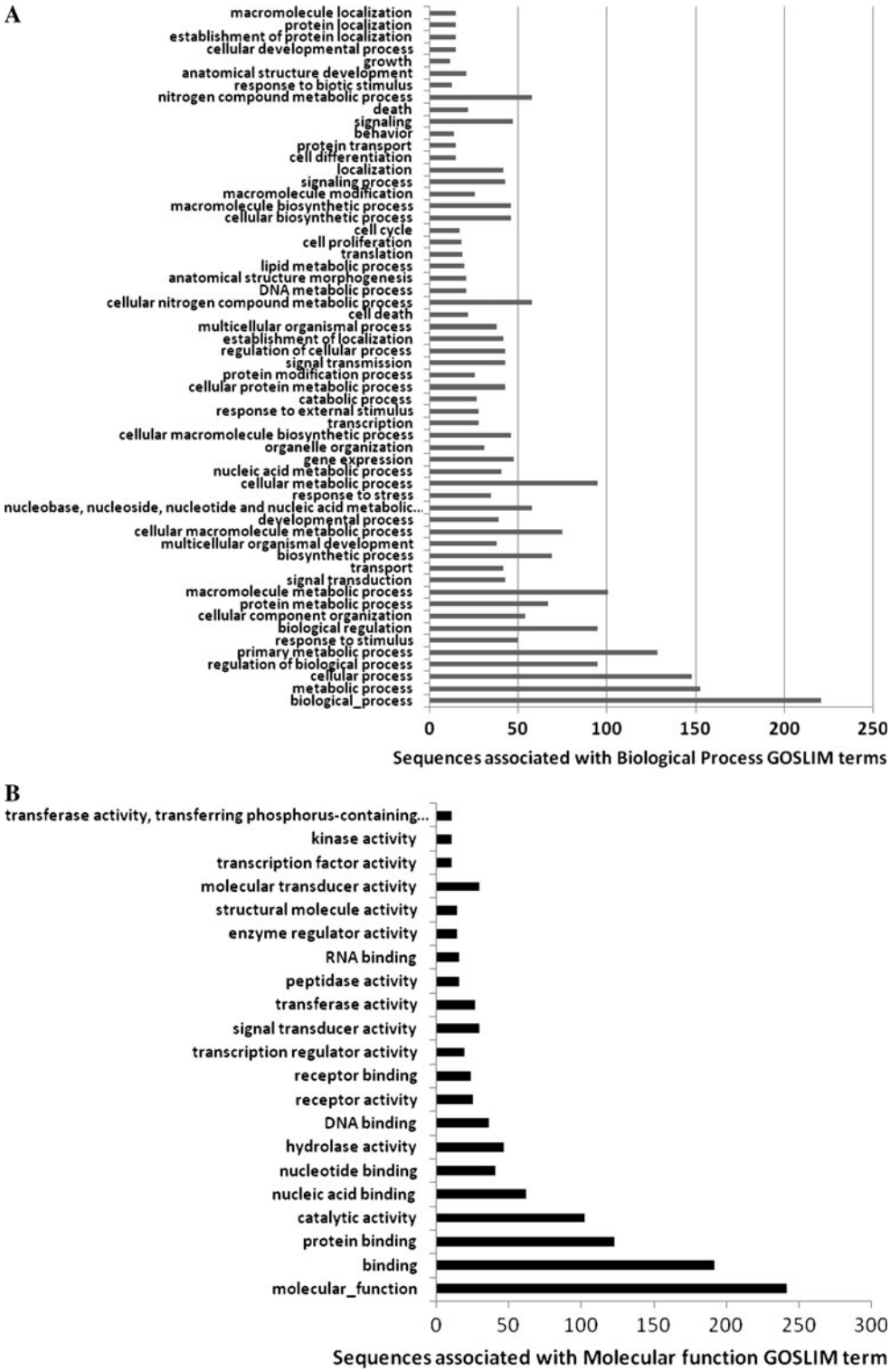
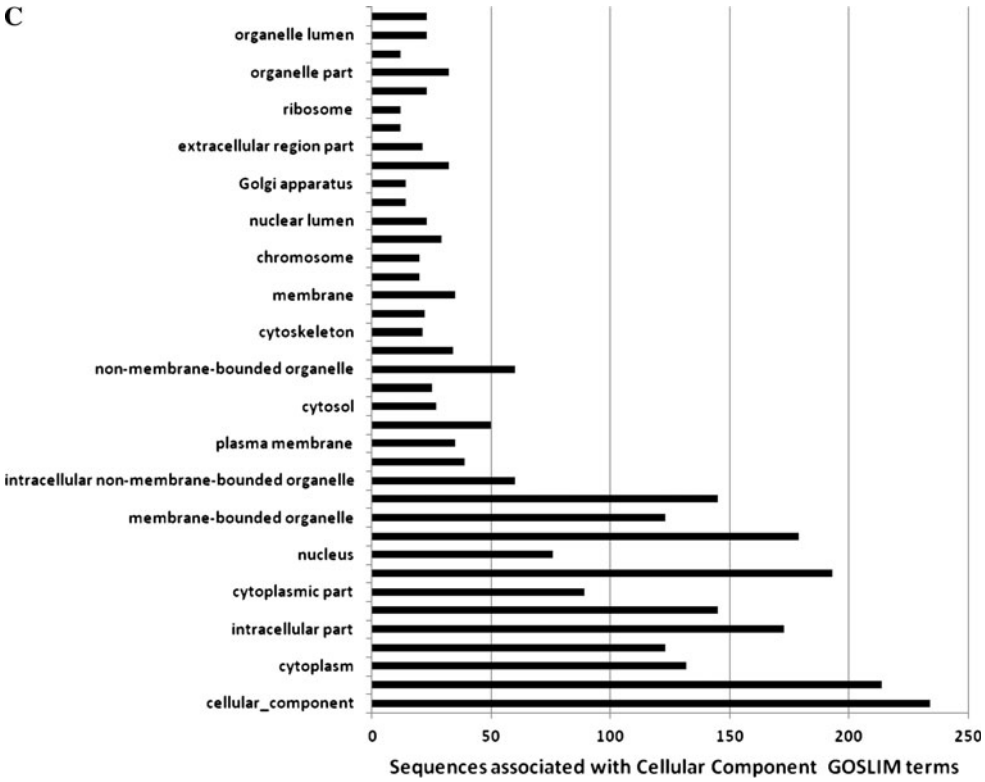


Fig. 4 **a** Classification of *Clarias batrachus* unique EST sequences according to Gene Ontology (GO) category: biological process. **b** Classification of *Clarias batrachus* unique EST sequences according to Gene Ontology (GO) category: molecular function. **c** Classification of *Clarias batrachus* unique EST sequences according to Gene Ontology (GO) category: cellular component

In contrast to the notion that the normalization of the libraries contradicts the purpose of an accurate analysis of transcript abundance in a tissue, although sequences present multiple times in the normalized libraries presumably correspond to high abundance transcripts [25], three highly abundant transcripts were found, one of them was tcb1 like transposase gene, whose multiple copies have been reported in several fish species from different orders [37–40]. A wide distribution of Tc1 transposons is believed to account for the limited requirements in the host cellular factors. Tc1-like transposons are co-regulated with a group of genes that are implicated in the defense response, signal

Fig. 4 continued



transduction and regulation of transcription and Tc1-like fragments reside in a number of immune and stress-related genes [41]. Up to 4.2 % of the channel catfish genome has been shown to be composed of Tc1-like transposon-related sequences [42] and thus supports their abundance in *C. batrachus* genome as well. The other two genes with high abundance were eukaryotic translation initiation factor 4Aiii and Gap junction cx32.2 protein (GJB). The eIF4AIII protein, one of the core components of EJC (exon junction complex) is essential for nonsense-mediated decay, EJC providing a link between pre-mRNA splicing and its downstream events [43]. Teleost 32.2 connexion gene encoded protein has been shown to be required for the formation of heterocellular gap junctions. Gap junctions are specialized regions of the plasma membrane of animal cells characterized by intercellular, cytoplasmic channels of communication formed by two connexions (hemichannels), one contributed by each of the adjacent cells that allow direct cell-to-cell passage of small molecule. Connexions have been implicated in physiological functions including paracrine intercellular signaling and in induction of cell death under pathological conditions [44, 45].

From a total of 576 annotated genes with functional information, the highest informative category was the molecular function. A remarkably high proportion of annotated sequences was categorized as binding and catalytic activity followed by molecular transducer and signal transducer in the *C. batrachus* ESTs. Most of the binding functions were at the intracellular level rather than external and included nucleotide and nucleic acid binding, protein binding and ion binding. Catalytic activities included hydrolase, transferase, peptidase and oxidoreductase activities. Each one of these subcategories represents catalytic activities that are important for a cell to survive. From results of these, 12 % of sequences belonged to the EIP category, majority included signal transduction and signalling molecules and interaction, indicating higher activities of stress and chaperone related genes in spleen.

The category of biological process includes genes encoding for proteins associated with cellular process, metabolic process and biological regulation, which were the three largest annotated subcategories in this study. The second largest group was genes encoding products related to regulation of biological process, response to stimulus and biosynthetic process. The remaining genes encoded products were involved in many other diverse biological processes. With respect to cellular components, we observed that a large proportion of sequences in this study were classified into cell, followed by organelle, macromolecular complex, extracellular region and membrane-enclosed lumen.

Among the 149 pathways, six immune related pathways, cytokine–cytokine receptor interaction (7; map no. 04060), phagosome (7; map no. 04145), chemokine signalling pathway (6; map no. 04062), complement and coagulation cascades (5; 04610), Fc gamma R-mediated

Table 3 KEGG biochemical mappings for ESTs from *Clarias batrachus* spleen

KEGG categories represented	Unique sequences (number of enzymes)	Percentage[a]
Metabolism	44 (43)	19
Carbohydrate metabolism	9 (9)	
Amino acid metabolism	8 (8)	
Energy metabolism	6 (6)	
Metabolism of cofactors and vitamins	6 (6)	
Lipid metabolism	5 (5)	
Nucleotide metabolism	4 (4)	
Metabolism of other amino acids	2 (2)	
Glycan biosynthesis and metabolism	2 (1)	
Xenobiotics biodegradation and metabolism	2 (2)	
Genetic information processing	46 (15)	20
Folding, sorting and degradation	19 (8)	
Translation	10 (2)	
Replication and repair	9 (3)	
Transcription	8 (2)	
Environmental information processing	30 (7)	12
Signal transduction	15 (5)	
Signalling molecules and interaction	14 (2)	
Membrane transport	1	
Cellular processes	43 (8)	19
Transport and catabolism	15 (4)	
Cell growth and death	14 (2)	
Cell communication	8 (1)	
Cell motility	6 (1)	
Organisational system	63 (11)	28
Immune system	35 (5)	
Digestive system	10 (4)	
Endocrine system	6	
Excretory system	4	
Nervous system	3 (1)	
Development	2 (1)	
Sensory system	1	
Circulatory system	1	
Environmental adaptation	1	

[a] Percentage based on total 165 unique transcripts in 224 mappings with significant similarity to sequences in database. Mappings to human diseases are not shown

phagocytosis (4; map no. 04666) and leukocyte transendothelial migration (4; map no. 04670) were well represented in this study. These pathways represent the components directly linked to the innate antibacterial immune response, chemotactic signalling, antigen processing and presentation complement pathway and migration of leukocytes [46] emphasizing the role of spleen as an immune responsive lymphoid tissue similar to the observations in many teleost fish [47–49]. In addition, 4 ESTs were harboured in "MAPK signaling pathway" (map no. 04010), an important signal transduction pathway involved in immune response [50]. Future research is needed to fully understand the roles of these transcripts in *C. batrachus* as these pathways are constructed mainly based on model animals.

In addition to immune relevant pathways, the pathways involved in protein processing, localization, folding and protein degradation were highly represented in *C. batrachus* spleen transcripts. These pathways are also shown to be differentially expressed during response to pathogen [51]. The pathways were protein processing in endoplasmic reticulum (map no. 04141), ribosome (map no. 03010) and ubiquitin-mediated proteolysis (map no. 04120) represented by 8, 8 and 5 transcripts, respectively. The high percentage of transcripts in the enzymes and transcription/translation factors categories may possibly reflect the

Table 4 Most commonly represented KEGG classifications of unique EST sequences from *Clarias batrachus* spleen

Category	Number	Percentage
Ribosome	8	2.5
Protein processing in endoplasmic reticulum	8	2.5
Cytokine–cytokine receptor interaction	7	2.2
Phagosome	7	2.2
Spliceosome	7	2.2
Regulation of actin cytoskeleton	6	1.9
Chemokine signaling pathway	6	1.9
Pathways in cancer	5	1.6
Complement and coagulation cascades	5	1.6
Ubiquitin mediated proteolysis	5	1.6
Fc gamma R-mediated phagocytosis	4	1.2
Leukocyte transendothelial migration	4	1.2
Bacterial invasion of epithelial cells	4	1.2
MAPK signaling pathway	4	1.2

Categories with more than three sequences out of a total of 311 ESTs listed in 149 pathways. The number (#) and percent (%) of ESTs in each category is shown

Table 5 Interpro protein domains found in unique EST sequences from *Clarias batrachus* spleen

Domain name	No of ESTs	Frequency	Description
IPR005819	7	0.021276596	Histone H5
IPR013783	4	0.012158055	Immunoglobulin-like fold
IPR007110	3	0.009118541	Immunoglobulin-like
IPR011046	3	0.009118541	WD40 repeat-like-containing domain
IPR013032	3	0.009118541	EGF-like region, conserved site
IPR013753	3	0.009118541	Ras
IPR016040	3	0.009118541	NAD(P)-binding domain
IPR008967	3	0.009118541	p53-like transcription factor, DNA binding

complexity of the immune system and demonstrated significant involvement of other proteins and intermediate molecules required to instigate an immune response.

The Interpro domain also pointed towards the preponderance of immune related transcripts in the *C. batrachus* spleen cDNA library with immunoglobulin-like fold and immunoglobulin-like protein domains being highly represented. Domains with an Ig-like fold can be found in many, diverse proteins in addition to immunoglobulin molecules. Ig-like domains are involved in a variety of functions, including cell–cell recognition, cell-surface receptors, muscle structure and the immune system [52]. Expression of a number of genes associated with immune functions, which were mainly associated with response to stress, response to chemical stimulus, cellular response to stimulus, response to external stimulus, immune response and regulation of response to stimulus have been reported in spleen [53].

About 39 % of unique ESTs could not be annotated using the public databases [54]. These may correspond to 3′ or 5′ untranslated regions, non-coding RNAs, or short sequences not containing known protein domains. Considering that, in general, the longer the sequence the higher the chance of annotation and number of GO terms recovered, a large number of these unannotated unique ESTs (i.e., those longer than 500 bp) may correspond to novel or undescribed genes.

The EST resources generated in the present study will lay a foundation for functional genomics research on *C. batrachus,* concerning aquaculture associated traits and their underlying mechanisms. Molecular markers associated with immune-relevant genes would be useful for the identification of trait associated alleles for marker-assisted selection. The identification of the putative immune-related genes would provide a meaningful framework to understand the Indian catfish immune system and defense mechanisms.

Acknowledgments This work was carried out under the project funded by Department of Biotechnology (DBT), Government of India and financial support provided by DBT is thankfully acknowledged.

References

1. FAO FishStatJ (2011) FishStatJ. Software for fishery statistical time series. http://www.fao.org/fishery/statistics/software/fishstatj/en
2. Sudarto H (2002) Systematic revision and phylogenetic relationships among populations of Clariid species in Southeast Asia (Thesis). University of Indonesia, Depok, West Java, Indonesia
3. Sahoo S, Giri S, Sahu A (2004) Effect of stocking density on growth and survival of *Clarias batrachus* (Linn.) larvae and fry during hatchery rearing. J Appl Ichthyol 20:302–305
4. Debnath S (2011) *Clarias batrachus*, the medicinal fish: an excellent candidate for aquaculture and employment generation. In: International Conference on Asia Aagriculture and Animal Singapore, vol 13, pp 32–37
5. Chondar SL (1999) Biology of finfish and shellfish. SCSC Publishers (India), Howrah, West Bengal, India, pp 422–433
6. FAO © (2005–2012) National Aquaculture Sector Overview. India. National Aquaculture Sector Overview Fact Sheets. Text by Ayyappan, S. In: *FAO Fisheries and Aquaculture Department* [online]. Rome. Updated 1 June 2005
7. Srivastava S, Choudhary SK (2010) Effect of artificial photoperiod on the blood cell indices of the catfish, *Clarias batrachus*. J Stress Physiol Biochem 6:22–32
8. Rao A, Mukhopadhyay P, Rao G, Sahoo S (2011) Effects of some non-conventional animal proteins on growth and certain tissue biochemical response in *Clarias batrachus* fingerlings. Indian J Fish 44:57–62
9. Saha N, Jyrwa LM, Das M, Biswas K (2011) Influence of increased environmental water salinity on gluconeogenesis in the

air-breathing walking catfish, *Clarias batrachus*. Fish Physiol Biochem 37:681–692
10. Chowdhary S, Srivastava P, Mishra S, Yadav A, Dayal R, Lakra WS (2012) Synergistic effects of dietary glucosamine and plant/animal proteins on the growth performance of Asian catfish (*Clarias batrachus*) juveniles. Online J Anim Feed Res 2:50–57
11. Moitra S, Bhattacharjee R, Sen N (2012) Histopathological Changes in the gills of air breathing teleost *Clarias batrachus* Linn. Exposed to endosulfan. Asian J Exp Sci 26:23–26
12. Joshi PS (2012) Studies on the effects of zinc sulphate toxicity on the detoxifying organs of fresh water fish *Clarias batrachus* (Linn.). Rev Res J 1:1–4
13. Hiware C, Pawar R (2011) Two new species of the genus *Lytocestus* (Caryophyllidea-Lytocestidae) from freshwater catfish, *Clarias batrachus* Linnaeus (1758). Recent Res Sci Technol 3(12):25–28
14. Ash A, Scholz T, Oros M, Kar PK (2011) Tapeworms (Cestoda: Caryophyllidea), parasites of *Clarias batrachus* (Pisces: Siluriformes) in the Indomalayan Region. J Parasitol 97:435–459
15. Balasundaram C, Harikrishnan R (2009) P26: immunomodulatory effect of *Andrographis paniculata* on Asian catfish *Clarias batrachus* against *Aeromonas hydrophila*. Exp Toxicol Pathol 61:293–294
16. Verma V, Prasad Y, Singh BR (2011) Effect of pH and salinity on pathogenicity of *Flavobacterium columnare* and *Myxobacterium* sp. in Indian cat fish, *Clarias batrachus* (Linn.) and *Heteropneustes fossilis* (Bloch.). J Environ Biol 32:573
17. Hossain Q, Hossain MA, Parween S (2006) Artificial breeding and nursery practices of *Clarias batrachus* (Linnaeus, 1758). Sci World 4:32
18. Sahoo S, Giri S, Maharathi C, Sahu A (2011) Effect of salinity on survival, feed intake and growth of *Clarias batrachus* (Linn.) fingerlings. Indian J Fish 50:119–123
19. Yue G, Kovacs B, Orban L (2003) Microsatellites from *Clarias batrachus* and their polymorphism in seven additional catfish species. Mol Ecol Notes 3:465–468
20. Islam MN, Islam MS, Alam MS (2007) Genetic structure of different populations of walking catfish (*Clarias batrachus* L.) in Bangladesh. Biochem Genet 45:647–662
21. Mohindra V, Singh A, Barman A, Tripathi R, Sood N, Lal KK (2012) Development of EST derived SSRs and SNPs as a genomic resource in Indian catfish, *Clarias batrachus*. Mol Biol Rep 39:5921–5931
22. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol 17:1636–1647
23. Ag Z, Chiba A, Varas A (1996) Cells and tissues of the immune system of fish. Fish Physiol 15:1
24. Kaattar SL, Piganelli JD (1996) The specific immune system: humoral defense. Fish Physiol 15:207
25. Douglas SE, Knickle LC, Kimball J, Reith ME (2007) Comprehensive EST analysis of Atlantic halibut (*Hippoglossus hippoglossus*), a commercially relevant aquaculture species. BMC Genomics 8:144
26. Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877
27. Kocabas AM, Li P, Cao D, Karsi A, He C, Patterson A, Ju Z, Dunham RA, Liu Z (2002) Expression profile of the channel catfish spleen: analysis of genes involved in immune functions. Mar Biotechnol 4:526–536
28. Lara A, Pérez-Trabado G, Villalobos D, Díaz-Moreno S, Canton F, Claros M (2007) A Web tool to discover full-length sequences-full-lengther. In: Corchado E, Corchado JM, Abraham A (eds) Innovations in Hybrid Intelligent Systems, Springer, Berlin, pp 361–368
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
30. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676
31. Labarga A, Valentin F, Anderson M, Lopez R (2007) Web services at the European bioinformatics institute. Nucleic Acids Res 35:W6–W11
32. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185
33. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res 32:e37
34. Xia J, Radford C, Guo X, Magor KE (2007) Immune gene discovery by expressed sequence tag analysis of spleen in the duck (*Anas platyrhynchos*). Dev Comp Immunol 31:272–285
35. Marques MC, Alonso-Cantabrana H, Forment J, Arribas R, Alamar S, Conejero V, Perez-Amador MA (2009) A new set of ESTs and cDNA clones from full-length and normalized libraries for gene discovery and functional characterization in citrus. BMC Genomics 10:428
36. Pardo BG, Fernández C, Millán A, Bouza C, Vázquez-López A, Vera M, Alvarez-Dios JA, Calaza M, Gómez-Tato A, Vázquez M (2008) Expressed sequence tags (ESTs) from immune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens. BMC Veterinary Res 4:37
37. Goodier JL, Davidson WS (1994) Tc1 transposon-like sequences are widely distributed in salmonids. J Mol Biol 241:26–34
38. Radice AD, Bugaj B, Fitch DHA, Emmons SW (1994) Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. Mol Gen Genet MGG 244:606–612
39. Reed KM (1999) Tc 1-like transposable elements in the genome of lake trout (*Salvelinus namaycush*). Mar Biotechnol 1:60–67
40. Leaver MJ (2001) A family of Tc1-like transposons from the genomes of fishes and frogs: evidence for horizontal transmission. Gene 271:203–214
41. Krasnov A, Koskinen H, Afanasyev S, Mölsä H (2005) Transcribed Tc1-like transposons in salmonid fish. BMC Genomics 6:107
42. Nandi S, Peatman E, Xu P, Wang S, Li P, Liu Z (2007) Repeat structure of the catfish genome: a genomic and transcriptomic assessment of Tc1-like transposon elements in channel catfish (*Ictalurus punctatus*). Genetica 131:81–90
43. Kanno S, Saffitz JE (2001) The role of myocardial gap junctions in electrical conduction and arrhythmogenesis. Cardiovasc Pathol 10:169–178
44. Bruzzone R, White TW, Yoshizaki G, Patiño R, Paul DL (1995) Intercellular channels in teleosts: functional characterization of two connexins from Atlantic croaker. FEBS Lett 358:301–304
45. Sáez JC, Berthoud VM, Branes MC, Martinez AD, Beyer EC (2003) Plasma membrane channels formed by connexins: their regulation and functions. Physiol Rev 83:1359–1400
46. Secombes C, Hardie L, Daniels G (1996) Cytokines in fish: an update. Fish Shellfish Immunol 6:291–304
47. Rise ML, Hall J, Rise M, Hori T, Kurt Gamperl A, Kimball J, Hubert S, Bowman S, Johnson SC (2008) Functional genomic analysis of the response of Atlantic cod (*Gadus morhua*) spleen to the viral mimic polyriboinosinic polyribocytidylic acid (pIC). Dev Comp Immunol 32:916–931
48. Feng CY, Johnson SC, Hori TS, Rise M, Hall JR, Gamperl AK, Hubert S, Kimball J, Bowman S, Rise ML (2009) Identification and analysis of differentially expressed genes in immune tissues of Atlantic cod stimulated with formalin-killed, atypical *Aeromonas salmonicida*. Physiol Genomics 37:149–163

49. Liu G, Zheng W, Chen X (2007) Molecular cloning of proteasome activator PA28-beta subunit of large yellow croaker (*Pseudoscianacrocea*) and its coordinated up-regulation with MHC class I α-chain and beta 2-microglobulin in poly I: C-treated fish. Mol Immunol 44:1190–1197
50. Xia JH, Yue GH (2010) Identification and analysis of immune-related transcriptome in Asian seabass *Lates calcarifer*. BMC Genomics 11:356
51. Wei W, Xu H, Wang Q, Zhang X, Chang K, Wu C, Zhang Y (2009) Identification of differentially expressed genes in large yellow croaker (*Pseudosciaena crocea*) induced by attenuated live *Vibrio anguillarum*. Aquaculture 291:124–129
52. Teichmann SA, Chothia C (2000) Immunoglobulin superfamily proteins in *Caenorhabditis elegans*. J Mol Biol 296:1367–1383
53. Singh A, Sood N, Chauhan U, Mohindra V (2012) EST-based identification of immune-relevant genes from spleen of Indian catfish, *Clarias batrachus* (Linnaeus, 1758). Gene 502:53–59
54. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36:3420–3435