Gene 502 (2012) 53–59

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

EST-based identification of immune-relevant genes from spleen of Indian catfish, *Clarias batrachus* (Linnaeus, 1758)

Akanksha Singh [a], Neeraj Sood [a], U.K. Chauhan [b], Vindhya Mohindra [a,*]

[a] National Bureau of Fish Genetic Resources, Canal Ring Road, PO Dilkusha, Lucknow-226002, U.P., India
[b] Awadesh Pratap Singh University, Rewa-486003, M.P., India

ARTICLE INFO

Article history:
Accepted 9 April 2012
Available online 21 April 2012

Keywords:
cDNA library
Clarias batrachus
Expressed sequence tags
Immune-relevant genes
EST-SSR, SNP

ABSTRACT

A normalized cDNA library from spleen of Indian catfish, *Clarias batrachus*, was constructed with a redundancy factor of 2.29. A total of 2045 clones from the library were single-pass sequenced, which generated 1937 high quality ESTs with an average read length of approximately 700 bp. Based on sequence similarities, 65 ESTs were found to be 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. The immune-relevant gene for CD141, thrombomodulin, has been identified in Teleosts for the first time. Six EST-SSRs and three SNPs were found associated with eight immune-relevant genes. These markers associated with important immune genes would be useful for the identification of trait associated alleles for marker-assisted selection. The identification of the putative immune-related genes provides a meaningful framework to understand the Indian catfish immune system and defense mechanisms.

1. Introduction

Clarias batrachus (*Clarias magur*), an Indian catfish species, is endemic to the Indian subcontinent and is distributed in Ganga and Brahmaputra river basins in northern and north-eastern India, Nepal, Bhutan and Bangladesh (Ng and Kottelat, 2008). It has been included in the threatened species list due to reduction in natural population size (Vishwanath, 2011). *C. batrachus* is high market value and a preferred and potential cultivable species (Hossain et al., 2006; Sahoo et al., 2004). The production of this species, however, is limited by its high disease susceptibility (Kanchanakhan, 2009). The disease in aquaculture is the end result of an interaction between three factors: host susceptibility, pathogen virulence and environmental stressors. Fish kept under intensive conditions are constantly exposed to a wide range of stressors and the fish attempt to adjust physiologically. However, any stressor that exceeds the ability of fish to adapt may be lethal or will facilitate the infection by opportunistic pathogens present in the water (Abiodun, 2009). This may be sufficient to trigger off a disease in the population. Under these circumstances, fish depend heavily on innate or non-specific immune responses (Camp et al., 2000). A successful, appropriate innate immune response can lead to rapid elimination of pathogens long before an antibody-based defense could potentially be mounted that may take several days to weeks (Baoprasertkul, 2006). Therefore, an important approach to disease prevention is to culture strains of fish with enhanced resistance to major diseases using molecular assisted selective breeding methods. However, genomic research of Indian catfish is still in its infancy, and genomic resources are largely unavailable.

An essential first step in the genomic characterization of a new species, is the generation of expressed sequence tag (EST) information. The EST method of analysis is widely employed as an effective means for gene discovery, gene expression profiling as well as identification of differentially expressed genes (Xu et al., 2010). This can form the basis for subsequent microarray design, SNP detection and the placement of novel markers on genetic linkage maps (Douglas et al., 2007). Putative immune-related genes including from spleen have been identified through EST approach in Asian seabass *Lates calcarifer* (Xia and Yue, 2010), miiuy croaker *Miichthys miiuy* spleen (Xu et al., 2010), Chinese amphioxus *Branchiostoma belcheri* (Liu et al., 2009), European seabass *Dicentrarchus labrax* (Sarropoulou et al., 2009), Atlantic cod (*Gadus morhua*) (Feng et al., 2009) and many other fish species. It has also identified polymorphic DNA markers, such as microsatellites (EST-SSR) and single nucleotide polymorphisms (SNP), in immune-relevant genes which are highly useful for genetic mapping and comparative genome analysis of fish (Vera et al., 2011).

In the present study, a normalized cDNA library was constructed from *C. batrachus* spleen, to identify the genes involved in immune response as a step towards understanding the molecular basis of host

Abbreviations: EST, expressed sequence tag; SSR, simple sequence repeats; SNP, single nucleotide polymorphism; cDNA, complementary DNA; PCR, polymerase chain reaction; GO, gene ontology; KO, KEGG orthology.

* Corresponding author. Tel.: +91 522 2442440; fax: +91 522 2442403.
E-mail address: vindhyamohindra@gmail.com (V. Mohindra).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2012.04.019

Table 1
Summary of analysis of expressed sequence tags in *Clarias batrachus*.

1. Total no. of clones sequenced	2045
2. No. of ESTs generated	1937
3. No. of contigs assembled	184
4. No. of sequences in contigs	423
5. No. of singletons	1514
6. No. of Unique ESTs[a] (3 + 5)	1698
7. Redundancy number	2.29
8. Total no. of immune relevant ESTs identified	65
9. ESTs with SSR and SNPs	7

[a] Number of putative unique transcripts equals the number of contigs plus the number of singletons.

defense and to provide resources for mapping and comparative genome analysis of *C. batrachus* immunity.

2. Materials and methods

2.1. Fish

Mature *C. batrachus* were obtained from commercial catches and acclimated for 2 weeks before sample collection. Fishes of mean weight 94.5 ± 20.8 g ($n = 12$,), adjudged by general appearance and level of activity were used in the experiment. The fish were killed by cervical incision. The spleen was aseptically removed from acclimated fishes and frozen in liquid nitrogen until RNA isolation.

2.2. Construction of a normalized cDNA and generation of ESTs

From spleen of *C. batrachus* (ten individuals), a normalized cDNA library was constructed directionally in plasmid vector pDNR-LIB (Creator SMART cDNA Library Construction Kit, Clontech, Palo Alto, CA, USA). Briefly, total RNA was extracted from pooled spleen tissues (Trizol Reagent, Invitrogen, Carlsbad, CA) followed by mRNA isolation (Oligotex mRNA Mini Kits, Qiagen, Valencia, CA, USA). First strand cDNA was prepared using the CDS-3M adaptor (TRIMMER-DIRECT kit, Evrogen, Moscow, Russia). cDNA was amplified by LD-PCR, according to the Creator SMART cDNA method (Clontech) and normalized using the TRIMMER-DIRECT protocol (Evrogen). Products smaller than 500 bp were removed, after digestion with SfiI, using the Chroma Spin-400 column, as described in the Creator SMART protocol. The resulting cDNAs were directionally cloned into the SfiI sites of pDNR-LIB (Clontech) and transformed into ElectroMAX DH10B cells (Invitrogen) by electroporation (Gene Pulser Xcell, Bio-Rad, Hercules, CA). The clones were screened using colony PCR method and positive clones were sequenced from 5′ direction using primer pDNR.F2 (Douglas et al., 2007). The sequences obtained were cleaned using VecScreen and separate FASTA files were generated for each EST sequence.

Table 2
Most commonly represented KEGG classification of *Clarias batrachus* unique immune-related sequences.

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 578 are listed. The number (#) and percent (%) of ESTs in each category are shown.

2.3. Clustering and functional annotation of ESTs

ESTs generated from *C. batrachus* normalized spleen cDNA library were analyzed by cluster analysis using the CAP3 program (Huang and Madan, 1999). The linear assembly algorithm was used and the criteria for clustering were set at a minimum overlap of 30 bases (default is 20 bases). After the cluster analysis, each cluster was visually inspected to ensure fidelity of alignment to avoid pseudo-clusters caused by repetitive elements or long strings of microsatellite repeats. ESTs belonging to contigs and singletons were recorded. Redundancy number was calculated as the number of clones in the contigs divided by the number of contigs (Kocabas et al., 2002). To establish the identities of ESTs, BLAST searches were conducted using BLASTN and subsequently TBLASTX searches against the non-redundant (nr) database with a BLAST cut off of 1×10^{-10} and 1×10^{-5}, respectively, to confirm gene identities. All the BLAST results were visually inspected to ensure that the matches were not due to simple amino acid stretches or repeat regions. All sequences were aligned using clustalW (Kocabas et al., 2002) using default values for parameters, with pair-wise parameters set at gap opening penalty 10, gap extension penalty 0.1 and multiple alignment parameters set at gap opening penalty of 10, gap extension 0.2.

Gene ontology (GO) annotations for consensus and singleton sequences were assigned using the program BLAST2GO. Consensus and singleton sequences were submitted for GO annotation to the BLAST2GO program (Conesa et al., 2005). The BLAST2GO program uses BLAST to find homologous sequences for input sequences and extracts GO terms to each hit using existing annotations. These GO terms are assigned to the query sequence to give an assessment of the biological process, the molecular function and the cellular compartments

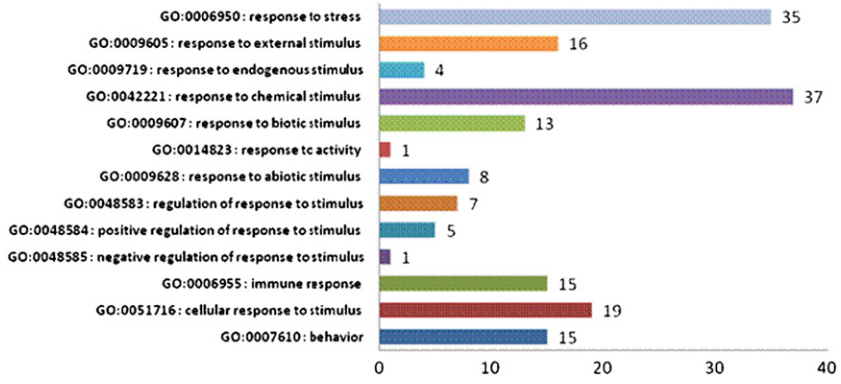

Fig. 1. ESTs represented by the response to stimulus category (GO: 0050896) in *Clarias batrachus*.

Table 3
List of immune-relevant expressed sequence tags identified in *Clarias batrachus*.

S. no.	Gene identity	Accession number	Length (bp)	Similarity with organism	Expect
Humoral immunity					
1	BLNK; B-cell linker.	GW836286	485	*Ictalurus punctatus*	7E-47
2	NFKBIA; nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha.	GW492734	648	*Ictalurus punctatus*	5E-50
3	CD18/ITGB2; integrin beta 2	GW840364	533	*Ictalurus punctatus*	2.19E-49
4	Otu ubiquitin aldehyde binding 1	GW787381	1127	*Salmo salar*	1.06E-137
5	C-type lysozyme	GW836309	650	*Oreochromis aureus*	4.92E-48
Cellular immunity					
6	Alpha-2-macroglobulin	GW836104	1279	*Ctenopharyngodon idella*	1.79E-95
Complement and coagulation cascades					
7	Complement component 1, q subcomponent, B chain r	GW840387	390	*Ictalurus punctatus*	5E-42
8	C6; complement component 6	GW787388	613	*Danio rerio*	1.18E-05
9	C7; complement component 7	GW840704	571	*Danio rerio*	9.74E-14
10	CFB; complement factor b [EC:3.4.21.47]	GR955345 GW836313	1153 (AG)9	*Cyprinus carpio*	5.90E-41
Antigen processing and presenting					
11	Proteasome (prosome, macropain) subunit, beta type, 9a	GT271599	1070	*Oncorhynchus mykiss*	1.09E-80
12	Proteasome (prosome, macropain) subunit, beta type, 8	GW840607	829	*Danio rerio*	2.13E-133
13	Immunoglobulin light chain C region	GW836241	494	*Ictalurus punctatus*	9.36E-13
14	Immunoglobulin heavy chain mu	GT145380	574	*Ictalurus punctatus*	1.74E-26
15	MHC class ii beta chain	GW840560	761	*Leiocassis longirostris*	6.17E-89
16	Ig-binding receptor FcR	GW840553	644	*Ictalurus punctatus*	2e-27
Chemokines					
17	Cc chemokine scya102	GT145384	676	*Ictalurus punctatus*	2.18E-41
18	Cc chemokine scya104	GW8364176	529	*Ictalurus punctatus*	1.1E-31
19	Cc chemokine scya106	GW836321	1026	*Ictalurus punctatus*	8.64E-27
20	Cc chemokine scya108	GW840690	971	*Ictalurus punctatus*	4.77E-16
21	Interleukin 1 beta type a	GW840649	502	*Ictalurus punctatus*	8.88E-19
22	Interleukin-1 beta	GT271596	705	*Ictalurus punctatus*	1.61E-53
23	Interferon regulatory factor 1	GW672562	566	*Ictalurus punctatus*	1E-115
24	Platelet basic protein precursor/C-X-C motif chemokine 7	GT157715	513	*Ictalurus punctatus*	1.20E-13
CD molecules					
25	THBD, CD141; thrombomodulin	GW492696	671	*Gallus gallus* (chicken)	9.64E-18
26	P-selectin precursor (Granule membrane protein 140) (CD62P antigen)	GW707107	741	*Danio rerio*	3.14E-50
27	CD97 antigen-like	GW840705	742	*Salmo salar*	5.73E-69
28	Bone marrow stromal cell antigen 2	GW707092	979	*Danio rerio*	5e-36
29	Tumor necrosis factor receptor superfamily, member 9	GW787379	1121	*Danio rerio*	6.80E-44
30	CD9 antigen	GW787468	1059	*Danio rerio*	1.48E-37
Transcription factors					
31	Forkhead box K2	GW836231	817	*Xenopus laevis*	6e-42
32	V-rel reticuloendotheliosis viral oncogene homolog A	GW836160	386	*Danio rerio*	1.48E-34
Apoptosis					
33	NLRP1, CARD7; NACHT, LRR and PYD domains-containing protein 1	GT271615	685	*Danio rerio*	9.21E-19
34	PAK2; p21-activated kinase 2 [EC:2.7.11.1]	GW840560	708	*Danio rerio*	6.03E-32
35	PYCARD, ASC; apoptosis-associated speck-like protein containing a CARD	GT271615	765	*Danio rerio*	3.58E-20
Unclassified					
36	ACTB_G1; actin beta/gamma 1	GW836124	986	*Danio rerio*	8.89E-143
37	Dynamin GTPase [EC:3.6.5.5]	GW836317	735	*Danio rerio*	8e-36
38	DOCK2; dedicator of cytokinesis 2	GR955288	661	*Danio rerio*	4.25E-18
39	CDC42; cell division control protein 42	GW840592	814	*Salmo salar*	1.47E-78
40	IRAK4; interleukin-1 receptor-associated kinase 4 [EC:2.7.11.1]	GT145393 GW492690	732	*Oncorhynchus mykiss*	1.74E-29
41	MMP9; matrix metalloproteinase-9 (gelatinase B) [EC:3.4.24.35]	GW836477 GW840427 GW840408	588	*Ictalurus punctatus*	2.47E-18
42	ARPC3; actin related protein 2/3 complex, subunit 3	GW840665	821	*Esox lucius*	4.55E-96
43	EPOR; erythropoietin receptor	GW836104	964	*Danio rerio*	4.26E-26
44	HSP86; heat shock protein hsp 90-alpha	GW775018	743	*Danio rerio*	1.62E-30
45	Src-like-adapter 2	GW775005	824	*Danio rerio*	3.97E-15
46	Novel protein vertebrate valosin-containing protein	GT157698	612	*Danio rerio*	2.92E-09
47	Autoimmune regulator-like	GT271617	679	Danio rerio	3e-25
48	Manganese superoxide dismutase	GW787450	920	*Hemibarbus mylodon*	7.03E-107
49	Angiogenin 4	GW840656	720	*Danio rerio*	1.69E-21
50	WAS; Wiskott–Aldrich syndrome (eczema-thrombocytopenia) protein	GW397191 GW397105 GW672526	822	*Danio rerio*	8.86E-07

(*continued on next page*)

Table 3 (*continued*)

S. no.	Gene identity	Accession number	Length (bp)	Similarity with organism	Expect
Unclassified					
51	Mitochondrial ribosomal protein L33	GW836355	525	*Esox lucius*	3.46E-14
52	Collagenase 3 precursor	GT145389	727	*Danio rerio*	8.15E-56
53	Interferon-induced protein 44	GW836354	1007	*Salmo salar*	4.40E-44
Stress responsive genes					
54	Serpin peptidase clade a (alpha-1 antitrypsin) member 10	GW787336	1104	*Danio rerio*	3.72E-95
55	Cellular tumor antigen p53	GW787361	1183	*Danio rerio*	7.15E-23
56	Btb and cnc homology basic leucine zipper transcription factor 1	GW787371	908	*Danio rerio*	1.19E-50
57	Alsin	GW397085	568	*Danio rerio*	1.93E-14
58	Stress-associated endoplasmic reticulum protein 1	GW787373	926	*Salmo salar*	6e-26
59	DNA mismatch repair protein mlh3 isoform 2	GT145373	681	*Danio rerio*	1.87E-72
60	Tnf receptor-associated protein 1	GT271588	690	*Danio rerio*	1.77E-110
61	Sodium-dependent neutral amino acid transporter B(0)AT3	GT271601	533	*Danio rerio*	2e-08
62	Calcium and integrin-binding protein 1	GW836405	435	*Oncorhynchus mykiss*	3.08E-16
63	Eukaryotic translation initiation factor 2-alpha kinase 1	GW840329	915	*Danio rerio*	1.92E-72
64	Annexin A5	GW840658	638	*Ictalurus punctatus*	2.35E-15
65	Pre-mRNA-processing factor 19	GW836095 GW840493	812	*Ictalurus punctatus*	1.25E-29

represented. Annotated accession numbers and GO numbers were derived with NCBI's QBLAST, with an expectation *E*-value $\leq 10^{-3}$ and an HSP length cut-off of 33. Contig sequences were then annotated 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. Graphs were generated using a sequence filter of 5, an alpha score of 0.6 and a 0 node score filter. From these annotations, ESTs belonging to biological process, molecular function and cellular components were identified using 2nd level GO terms.

In addition, the ESTs were annotated according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) by the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007). The query sequences are compared against the existing genes in KEGG using BLASTP for protein sequences and BLASTX and TBLASTN for nucleotide sequences. The sequences that were most similar to existing genes were then mapped onto the existing pathways. The sequences were analyzed 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 are assigned KO identifiers or K numbers by the ortholog annotation procedure were assigned to genes, the collective body of K numbers was mapped to BRITE functional hierarchies.

2.4. Identification of genes involved in immune process

To identify the genes and transcripts that play an important role in immune system process, the ESTs associated with GO terms under immune response (GO: 0006955) and immune system process (GO: 0006955) and with response to stress (GO: 0006950) were selected. Under KEGG BRITE, ESTs falling in the immune system category under organizational systems were also short listed. Additionally the annotated genes were searched through the IMMUNOME database (Rannikko et al., 2007) to identify the genes involved in immune process.

2.5. Identification of EST-SSRs and EST-SNPs

Identification of repeat motifs within the EST sequences identified under immune response was performed with the program Msatfinder http://www.genomics.ceh.ac.uk/msatfinder/ (Thurston and Field, 2005) with all the default values. The threshold limit for di-, tri-, tetra- and penta-nucleotide repeats was 6, 5, 5 and 5, respectively. For identification of SNPs, the CAP3 assembled contigs having at least three member ESTs were manually screened for the presence of SNPs, to rule out the putative SNPs, represent sequence errors and nature of polymorphism was identified.

3. Results and discussion

In the present study, the EST approach was used to identify the immune-related genes expressed in *C. batrachus* spleen, which is regarded as a major immune organ in fish. Similar to other vertebrates, the immune system of teleosts responds to the antigen in two ways i.e. non-specific and specific. The non-specific immune system provides an array of protective mechanisms that are inherently available and provide immediate protection against a wide variety of pathogens. The magnitude of this non-specific protection is usually consistent, regardless of the type of pathogen or the number of times the pathogen infects the animal. The

```
Equus caballus     CEYQCQPVGQ NDYRCICAEG FAPIPQDPDR CQMFCNQTAC PADCDPNNPS NCQCPEGYIL DDG----VIC TDINECDSGY  [ 80]
Rattus norvegicus  ........NS TH.N..... ...KLD.... .E....E.S. .......S.. F......F.. .E.----S.. ...D..SQ.E  [ 80]
Gallus gallus      .DHT.ED.PG G-.Q.S.YN. YVVN.KN.TE .LQK.EDGR. L.E..ASG-L Y.V....FL. .HLPNGVSL. V..D..E.NH  [ 80]
Clarias batrachus  ..H..KN.PG G-HM.F.NPK .R.SLKE.E. .ENY.DSDS. SRL.S----- T.E..K.F.K ..IR-----. A..D..K.HH  [ 80]

Equus caballus     CPG--ECRNL FGSYECICGP DSALAGQVAT DCDPKESKGD DEDSGSGEPP VSR-TPGTTA SPSPVAPLHS GVLIG-----  [160]
Rattus norvegicus  .LTN-..... P......... .T.....ISK ....IPVLE. S..G....H. S.NP.VVSST V.PSAR.M.. .....-----  [160]
Gallus gallus      .EH--N.T.T A...I.H.Q. G--YMPFDVN H.I.IS-QE. N..GY..D-- SGPPW.VPSH I.PKAEH..P .A.V.-----  [160]
Clarias batrachus  NNCEQK...M ..H.K.S.NE G--FMLVNKS K.L.LIKPAS VNYTLNHAAF ATPGGYIGLT LFIMLTICA. VG.LYYLRKR  [160]

Equus caballus     ---ISIASLS LVVALLALLC HLRKKQGAMP AELEYKYGAP -TKEVVLQHV GTERMPQKL [219]
Rattus norvegicus  ---....... ......... .......TAR ......CTSS -A........ R.D.TL..F [219]
Gallus gallus      ---.TMGV.C AALV...VGY ..AR.RCHP. SSM...C.S. RE..MG..Q. SASQKV--- [219]
Clarias batrachus  KSLNTFKQVK .IFVFMPSNN F.SDIVSFSI CLSI.VTFNT VLNFLH.T.I IH.NDGSLK [219]
```

Fig. 2. Alignment of partial deduced amino acid sequence of *Clarias batrachus* thrombomodulin with *Equus caballus*, *Rattus norvegicus* and *Gallus gallus* (GenBank or EMBL accession numbers are XP_001915376.2, NP_113959.1 and XP_426101.1, respectively). Amino acid identity is represented by dots and gaps introduced into sequences to optimize alignments are represented by (−).

```
Esox lucius          MR-AVLVLLL VAVASAKVYD RCDLARRLKA AGMDGYYGNS LPNWVCLAKW ESSYNTQATN RNTDGSTDYG IFQINSRWWC  [ 80]
Scophthalmus rhombus ..-CL.F... ....G...FE ..E...L..S Y..NN.R.I. .AD....SQ. ......R... .......... ..........  [ 80]
Oreochromis aureus   ..SVFVF... IT......FE ...W..K... N.....R.V. .A.....T.H ..N...K... ........F. ..........  [ 80]
Clarias batrachus    .K-.LVFF.. L..V...Q.. ..E...AM.. Y.LA..R.I. .A.......H ..DF..R.L. ..S....... .....N....  [ 80]
Ictalurus punctatus  .K-.LVF... L..V...R.. ..E...AM.. N.L...H.I. .A.......H ..D...K.I. H......... .....N....  [ 80]

Esox lucius          DDGRTPRAKN GCGIRCSQLL TDDLTVAINC AKRVVRDPNG IGAWVAWRNR CKNRDLQSYV AGCGV [145]
Scophthalmus rhombus NN.Q..-TS. A...S..A.. ...VIA..A. .....V.... .......KSH .EG...SP.L ..... [145]
Oreochromis aureus   N.R.IN-SA. ..N.D..V.. A..V.S.T.. ...I..E-Q. .T........ .QGQ..RP.L S..RL [145]
Clarias batrachus    TN.QFR-SA. ..RLS..E.. ..NIAK.VE. ..TI..Q-Q. .T.......K .RG...S..T ..... [145]
Ictalurus punctatus  SN.SFR-SA. ..K.S.N... ..NIYQ.AQ. T.TI..Q-Q. .T......RY .RG..VG..T ..... [145]
```

Fig. 3. Alignment of partial deduced amino acid sequence of *Clarias batrachus* c-lysozyme with *Esox lucius*, *Scophthalmus rhombus*, *Oreochromis aureus* and *Ictalurus punctatus* (GenBank or EMBL accession numbers are ACO14488.1, BAF75844.1, ACF37258.1 and NP_001187718.1, respectively). Amino acid identity is represented by dots and gaps introduced into sequences to optimize alignments are represented by (−).

identification and characterization of genes involved in immune response have facilitated the study of expression of genes during the disease processes, which help in better understanding of the mechanisms underlying disease resistance/susceptibility. The altered expression of different cytokine and immune-related genes has been observed following infection (Liu et al., 2011; Mohanty and Sahoo, 2010; Saurabh et al., 2011), vaccination (Gioacchini et al., 2008; Jorgensen et al., 2008) and immuno-stimulation (Watanuki et al., 2006).

3.1. Normalized spleen cDNA library

Normalized cDNA library from spleen of control fish was constructed in plasmid vector having a titer of 1×10^6 cfu with 95% recombination efficiency. Out of 2045 clones sequenced, after trimming and vector removal, 1937 good sequences were obtained. Average read length of ESTs after trimming was approximately 700 bp. Clustering of the sequences using CAP3 assembler yielded 1698 unique sequences (redundancy factor of 2.29), demonstrating the excellent normalization achieved by this approach (Table 1). The unique ESTs consisted of 184 unique clusters and 1514 singletons. The normalization of the libraries presents an accurate analysis of transcript abundance although sequences that present multiple times in the normalized libraries presumably correspond to high abundance transcripts.

3.2. Annotation and functional classification

Of the 576 functionally annotated and domain name containing sequences, from multilevel analysis, sequences with GO terms corresponding to cellular component fell into 14 categories, molecular function into 17 categories and biological process into 27 categories. KEGG categories were found for 185 sequences having 578 terms that fell into 185 pathways. Interpro terms were assigned to 158 sequences having a total of 601 terms. Additionally 42 ESTs showing similarity to genes and uni-genes containing domain were found and 12 with similarity to repeat containing proteins were detected.

3.3. Immune-relevant genes

Genes in the GO subcategories “response to stimulus and immune system processes” are likely to be involved in the interaction of pathogen with its host (Fig. 1). A total of 65 genes were identified, which were mainly associated with response to stress (35), response to chemical stimulus (37), cellular response to stimulus (19), response to external stimulus (16), immune response (15) and regulation of response to stimulus (7). These immune-relevant pathways under which genes were identified were namely cytokine–cytokine receptor interaction (7), phagosome (7), chemokine signaling pathway (6), complement and coagulation cascades (5), Fc gamma R-mediated phagocytosis (4) and leukocyte trans-endothelial migration (4). The defense/immune-related genes that were found during this study in normalized *C. batrachus* spleen transcripts are given in Table 2. Cytokine related genes were predominant (22.8%) followed by CD molecules (12.2%), antigen processing and presenting related genes (10.5%) and equal proportion of humoral immunity and complement and coagulation cascades related genes (8.7%). The immune-related cDNAs were reported for the first time in *C. batrachus* and have important role in the immune response like B-cell linker, MHC light and heavy chains, complement components, cytokines, interleukins, CD molecules, interferon regulatory factor 1 and tumor necrosis factor (TNF) receptor among others (Table 3).

Similarity for one EST, accession no. GW492696, could not be found from the fish entries in databases. Through BlastX, it was found to have highest similarity with CD141, thrombomodulin gene from

```
Cyprinus carpio      MACHEYVHQL DLSEAFETDS AIYSDSADSD ELDCPDPQSM SCQCDMH-DI KLELSSHPHS MRQVVNIIIA VERLKHIKNM  [ 80]
Carassius auratus    .E.IG--.YI LH.D.LKK.- IT.P...GF. .P....LHG. M......E.. ...P.P.... .KK....... ........E.  [ 80]
Danio rerio          ...GQ.EVTI APKNLW...G .V....---. .M..S..LA. .YR....EG. R..MWTSQ.K .K.L..V... LN.M....PQ  [ 80]
Clarias batrachus    EEI.FFINLY P.YRHI.CG. GFD..DMGI. .M..S..LAL .GR..L.KGL RI.VTKE.L. ..SIA.VV.. LQ.....ERV  [ 80]
Ictalurus punctatus  ---MADKDL. M.ERY.DS.C GFD..DM.F. K...S..LA. .GR..L.KGL RI.VTKE.L. ..H.A.VV.. LQ...LTQ.I  [ 80]

Cyprinus carpio      SSGKFCDEEL LGFILENVIE ERLVKPLNET PI--YSKTSL TLQCTICDKY KKTMVQSNKL SDEPLHLKAV TLSAGAMQYK  [160]
Carassius auratus    ......EDA. .N........ ........VS QT--...S.R .......... ...L...... NN.D...... .....NI...  [160]
Danio rerio          .T-E.GEK.V .DMLMA...Q ..E.NVVDSV .S--.T..KN V.......Q. ..SL.R.G-- --GSP..Q.. ..R..SSDL.  [160]
Clarias batrachus    Q.TE.S.R.. FNIFM..... .TI.IN..CS ESKL..RQDK VV.......S ...L.H.--- ---FPY.L.F ..KG.NEVN.  [160]
Ictalurus punctatus  Q.TE.T.Q.. FNVFID.... .SM.IN.KC. ESKS..LQDK VVR......S .RAL..KE-- --KLPL.L.F ..KG.NKEN.  [160]

Cyprinus carpio      VQFSMSTFVS -SATQKE-AQ PVCLGISNSN LYLACTQLDG SSPVLILKEA SGSVNTIKAG DPN--DSLLF FRKETGTRYN  [240]
Carassius auratus    .R.....YL. -..P.NK-G. ....A..... ..I...ES.. ...I.....V ..PL....V. .Q.GY..... .......A..  [240]
Danio rerio          .R.....YA. P..PATS-.. .......K.. .....SPAE. .A.H.V...I ...LE..... ...GY.Q... ......SSI.  [240]
Clarias batrachus     --------- --------GK ---.T.KKKT .L.LTIMY-- ---.FNF.AV KIQS--KYII SHIRIRIVIT IVP-------  [240]
Ictalurus punctatus  AW.NL.AYTP PNC.ENKNG. ......VKT. .F.S..LEN- ET.F.G.E.V KDKERLKSIQ ENDGMERF.. ..NG..DSL.  [240]

Cyprinus carpio      TFESVKYPGW FISTAFDDWE KVEMNQMPTT RTTNFTLEDQ KRI-----  [288]
Carassius auratus    .......... Y......... R...I.V..D .........- --------  [288]
Danio rerio          ......C... .....YE.SQ M...DRKD.E .II..E.Q.K V..-----  [288]
#Clarias batrachus   ---------- LLVGSVFSHM PLHNLPFHF. YRLFCPYVY- --------  [288]
#Ictalurus punctatus .......... ..T.SK..DK P.QTCKQQSS HLQL...H.E TVVSQNEM  [288]
```

Fig. 4. Alignment of partial deduced amino acid sequence of *Clarias batrachus* interleukin 1, beta with *Cyprinus carpio*, *Carassius auratus*, *Danio rerio* and *Ictalurus punctatus* (GenBank or EMBL accession numbers are BAA24538.1, CAD12102.1, GT271596.1 and AAZ94731.1, respectively). Amino acid identity is represented by dots and gaps introduced into sequences to optimize alignments are represented by (−).

Table 4
EST-SSRs and SNPs identified in immune relevant genes in *Clarias batrachus*.

S. no.	Gene identity	Repeat region	SNP: position and nature	Accession number
1	Wiskott–Aldrich syndrome (eczema-thrombocytopenia) b	$(CA)_2GA(CA)_6$	419; C↔T	GW397191 GW397105 GW672526
2	Alsin, partial	$(GT)_6T(GT)_3G\ (GT)_4(CT)_4$	–	GW397085
3	P-selectin precursor (Granule membrane protein 140) (CD62P antigen)	$(TTCT)_3(TTTC)_{10}TA(TCTT)_9$	–	GW707107
4	Matrix metalloproteinase-9	–	409; C↔T	GW836477 GW840427 GW840408
5	Dedicator of cytokinesis	$(GT)_3AT(GT)_4AA(GA)_9$	–	GR955288
6	Ig-binding receptor FcR	$(AATA)_6$	–	GW840553
7	Forkhead box K2	–	660; G↔T	GW836231
8	Stress-associated endoplasmic reticulum protein 1	(ATC)6	–	GW787373

Gallus gallus (accession no. XP_426101.2), with maximum score and *E*-value of 85.5 and $1e^{-16}$, respectively. Maximum identity of *C. batrachus* thrombomodulin was found to be 28%, 31% and 32% with *Equus caballus*, *Rattus norvegicus* and *Gallus gallus*, respectively. Thrombomodulin, CD141, is an integral membrane protein expressed on the surface of endothelial cells and a key component of the anticoagulant protein C pathway (Sohn et al., 2005). In humans, thrombomodulin is encoded by the THBD gene and the protein consists of a single chain with 5 distinct domains (Wen et al., 1987). Alignment of the amino acids (Fig. 2) indicated that the EST putatively encoding for thrombomodulin in *C. batrachus* is incomplete and further investigations should be made to figure out the complete sequence of *C. batrachus* thrombomodulin.

Under humoral immunity, lysozyme is an important defense molecule of leucocytic origin and is widely distributed in bacteriophages, microbes, plants, invertebrates and vertebrates (Jolles and Jolles, 1984). Lysozyme is found in a large variety of animal secretions such as mucus and saliva, in many tissues including blood and in the cell vacuoles of plants. Lysozymes are classified into five types; only the chicken-type (c-type) and goose-type (g-type) lysozymes have been reported in vertebrates including fish (Saurabh and Sahoo, 2008) The g-type lysozyme may be induced for the defense against bacterial infections, while c-type lysozyme might be the main molecule for the house-keeping defense under normal conditions (Ye et al., 2010). Only one EST putatively encoding for c-type was detected from our cDNA library. Alignment with reported sequences in other fishes indicated it to be with maximum identity of 65%, 64% and 64% with that of *Esox lucius*, *Scophthalmus rhombus* and *Oreochromis aureus*, respectively (Fig. 3).

Cytokines are low molecular weight, soluble proteins that are produced in response to an antigen (Ag) and function as chemical messengers for regulating the innate and adaptive immune systems. The cytokines are pleiotropic, redundant and multifunctional. Cytokines bind to specific receptors on the membrane of target cells, triggering signal-transduction pathways that ultimately alter gene expression in the target cells. Cytokines are subdivided into different families such as lymphokines, interleukins, growth factors, interferons and chemokines. Interleukin-1β (IL-1β) is a pro-inflammatory cytokine that is involved in the process of inflammation as well as the induction of other immunomodulatory cytokines in fish as well in mammals (Pressley et al., 2005). It is a key mediator in response to microbial invasion and tissue injury and can stimulate immune responses by activating lymphocytes or by inducing the release of other cytokines capable of triggering macrophages, NK cells and lymphocytes (Low et al., 2003). The increase in IL-1β gene expression may induce mucus secretion and activation of macrophages and up-regulate the expression of a number of NK-κB-dependent genes (Lindenstrøm et al., 2004). Comparison of the amino acid sequence of *C. batrachus* interleukin 1 beta type a (accession no. GW840649) from this study with that of other fishes showed 70%, 48%, 43% and 41% identity with *Ictalurus punctatus*, *Danio rerio*, *Carassius auratus* and *Cyprinus carpio* (Fig. 4). Phylogenetic analysis of the teleost IL-1β molecules groups them in a cluster separate from the molecules isolated from amphibians, birds, and mammals (Huising et al., 2004). One of the most striking differences in the amino acid sequences of fish IL-1 β, when compared with mammalian IL-1β, is the absence of a clear caspase 1 cut site (Zou et al., 1999).

3.4. Immune-relevant ESTs containing microsatellites and SNPs

Screening of immune-relevant EST sequences for microsatellites identified 6 microsatellites with ≥6 repeats and 3 with SNPs (Table 4). Out of these, 3, 1 and 2 ESTs had di-, tri- and tetra-repeats, respectively. Wiskott–Aldrich syndrome gene was found to contain both microsatellite and SNP. The identified microsatellites and SNPs in this study, when polymorphic, might prove to be useful markers in catfish genetic research, genomic mapping and for the identification of trait associated alleles for marker-assisted selection (Mohindra et al., 2011; Xia and Yue, 2010; Thompson et al., 1997).

In conclusion, the potential immune-related molecules identified in Indian catfish in the present study would form a basis for a better understanding of immunity and for developing effective strategies for immune protection against infections. The sequences information of markers will enable studies of the genetic variation, conservation genetics and molecular assisted selective breeding in the future.

Acknowledgments

Financial support received from the Department of Biotechnology, Government of India, for this study is thankfully acknowledged. Excellent technical assistance provided by Mr. R. S. Sah, Mr. Rajesh Kumar and Mr. Sree Ram is duly acknowledged.

References

Abiodun, A.A., 2009. Catfish Farming Business Ventures Handbook.
Baoprasertkul, P., 2006. Characterization of innate immune genes of catfish: Cxc chemokines and Toll-like receptors, Ph.D. Dissertation, Auburn University, Auburn, Alabama.
Camp, K.L., Wolters, W.R., Rice, C.D., 2000. Survivability and immune responses after challenge with *Edwardsiella ictaluri* in susceptible and resistant families of channel catfish, *Ictalurus punctatus*. Fish Shellfish Immunol. 10, 475–487.
Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
Douglas, S.E., Knickle, L.C., Kimball, J., Reith, M.E., 2007. Comprehensive EST analysis of Atlantic halibut (*Hippoglossus hippoglossus*), a commercially relevant aquaculture species. BMC Genomics 8, 144.
Feng, C.Y., et al., 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.
Gioacchini, G., Smith, P., Carnevali, O., 2008. Effects of Ergosan on the expression of cytokine genes in the liver of juvenile rainbow trout (*Oncorhynchus mykiss*) exposed to enteric red mouth vaccine. Vet. Immunol. Immunopathol. 123, 215–222.
Hossain, Q., Hossain, L.A., Parween, S., 2006. Artificial breeding of *Clarias batrachus*, Linnaeus 1758. Sci. World 4, 32–37.
Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.
Huising, M.O., Stet, R.J.M., Savelkoul, H.F.J., Verburg-van Kemenade, B.M., 2004. The molecular evolution of the interleukin-1 family of cytokines; IL-18 in teleost fish. Dev. Comp. Immunol. 28, 395–413.
Jolles, P., Jolles, J., 1984. What's new in lysozyme research? Always a model system, today as yesterday. Mol. Cell. Biochem. 63, 165–189.

Jorgensen, L.V.G., Nemli, E., Heinecke, R.D., Raida, M.K., Buchmann, K., 2008. Immune-relevant genes expressed in rainbow trout following immunisation with a live vaccine against *Ichthyophthirius multifiliis*. Dis. Aquat. Organ. 80, 189–197.

Kanchanakhan, S., 2009. Epizootic ulcerative syndrome. OIE Manual of Diagnostic Tests for Aquatic Animals, pp. 188–199.

Kocabas, A.M., et al., 2002. Expression profile of the channel catfish spleen: analysis of genes involved in immune functions. Mar. Biotechnol. 4, 526–536.

Lindenstrøm, T., Secombes, C.J., Buchmann, K., 2004. Expression of immune response genes in rainbow trout skin induced by *Gyrodactylus derjavini* infections. Vet. Immunol. Immunopathol. 97, 137–148.

Liu, Z., Li, L., Li, H., Zhang, S., Ji, G., Sun, Y., 2009. EST analysis of the immune-relevant genes in Chinese amphioxus challenged with lipopolysaccharide. Fish Shellfish Immunol. 26, 843–849.

Liu, H.P., Chen, R.Y., Zhang, Q.X., Peng, H., Wang, K.J., 2011. Differential gene expression profile from haematopoietic tissue stem cells of red claw crayfish, *Cherax quadricarinatus*, in response to WSSV infection. Dev. Comp. Immunol. 35, 716–724.

Low, C., Wadsworth, S., Burrells, C., Secombes, C.J., 2003. Expression of immune genes in turbot (*Scophthalmus maximus*) fed a nucleotide-supplemented diet. Aquaculture 221, 23–40.

Mohanty, B.R., Sahoo, P.K., 2010. Immune responses and expression profiles of some immune-related genes in Indian major carp, *Labeo rohita* to *Edwardsiella tarda* infection. Fish Shellfish Immunol. 28, 613–621.

Mohindra, V., Singh, A., Barman, A.S., Tripathi, R.K., Sood, N., Lal, K.K., 2011. Development of EST derived SSRs and SNPs as a genomic resource in Indian catfish, *Clarias batrachus*. Mol. Biol. Rep., http://dx.doi.org/10.1007/s11033-011-1404-z

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185.

Ng, H.H., Kottelat, M., 2008. The identity of *Clarias batrachus* (Linnaeus, 1758), with the designation of a neotype (Teleostei: Clariidae). Zool. J. Linn. Soc. 153, 725–732.

Pressley, M.E., Phelan, P.E., Witten, P.E., Mellon, M.T., Kim, C.H., 2005. Pathogenesis and inflammatory response to *Edwardsiella tarda* infection in the zebrafish. Dev. Comp. Immunol. 29, 501–513.

Rannikko, K., Ortutay, C., Vihinen, M., 2007. Immunity genes and their orthologs: a multi-species database. Int. Immunol. 19, 1361–1370.

Sahoo, S.K., Giri, S.S., Sahoo, A.K., 2004. Effect of stocking size of *Clarias batrachus* fry on growth and survival during fingerling hatchery production. Asian Fish. Sci. 17, 229–233.

Sarropoulou, E., et al., 2009. Profiling of infection specific mRNA transcripts of the European seabass *Dicentrarchus labrax*. BMC Genomics 10, 157.

Saurabh, S., Sahoo, P.K., 2008. Lysozyme: an important defence molecule of fish innate immune system. Aquacult. Res. 39, 223–239.

Saurabh, S., Mohanty, B.R., Sahoo, P.K., 2011. Expression of immune-related genes in rohu *Labeo rohita* (Hamilton) by experimental freshwater lice *Argulus siamensis* (Wilson) infection. Vet. Parasitol. 175, 119–128.

Sohn, R.H., et al., 2005. Regulation of endothelial thrombomodulin expression by inflammatory cytokines is mediated by activation of nuclear factor-kappa B. Blood 105, 3910–3917.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The clustal_x Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Thurston, M.I., Field, D., 2005. Msatfinder: Detection and Characterisation of Microsatellites. CEH Oxford, Mansfield Road, Oxford OX1 3SR. Distributed by the authors at http://www.genomics.ceh.ac.uk/msatfinder/.

Vera, M., et al., 2011. Validation of single nucleotide polymorphism (SNP) markers from an immune Expressed Sequence Tag (EST) turbot, *Scophthalmus maximus*, database. Aquaculture 313, 31–41.

Vishwanath, W., 2011. *Clarias magur*. IUCN 2011. IUCN Red List of Threatened Species. Version 2011.1. <www.iucnredlist.org>. Downloaded on 27 July 2011.

Watanuki, H., Ota, K., Tassakka, A., Kato, T., Sakai, M., 2006. Immunostimulant effects of dietary *Spirulina platensis* on carp, *Cyprinus carpio*. Aquaculture 258, 157–163.

Wen, D.Z., Dittman, W.A., Ye, R.D., Deaven, L.L., Majerus, P.W., Sadler, J.E., 1987. Human thrombomodulin: complete cDNA sequence and chromosome localization of the gene. Biochemistry-US 26, 4350–4357.

Xia, J.H., Yue, G.H., 2010. Identification and analysis of immune-related transcriptome in Asian seabass *Lates calcarifer*. BMC Genomics 11, 356.

Xu, T.J., Meng, F.X., Sun, Y.N., Shi, G., Wang, R.X., 2010. Identification of immune genes of the miiuy croaker (*Miichthys miiuy*) by sequencing and bioinformatic analysis of ESTs. Fish Shellfish Immunol. 29, 1099–1105.

Ye, X., Zhang, L., Tian, Y., Tan, A., Bai, J., Li, S., 2010. Identification and expression analysis of the g-type and c-type lysozymes in grass carp *Ctenopharyngodon idellus*. Dev. Comp. Immunol. 34, 501–509.

Zou, J., Grabowski, P.S., Cunningham, C., Secombes, C.J., 1999. Molecular cloning of interleukin 1 beta from rainbow trout *Oncorhynchus mykiss* reveals no evidence of an ICE cut site. Cytokine 11, 552–560.