Indian J. Fish., 61(2) : 118-128, 2014

Note

Development of expressed sequence tags (ESTs) from the brain tissue of snowtrout *Schizothorax richardsonii* (Gray, 1832) (Family Cyprinidae) and its preliminary annotation

ASHOKTARU BARAT, CHIRAG GOEL, PRABHATI K. SAHOO AND A. R. RAO*

Directorate of Coldwater Fisheries Research, Bhimtal - 263 136, Nainital, Uttarakhand, India

**Indian Agricultural Statistics Research Institute, Pusa, New Delhi – 110 012, India*

e-mail: abart58@hotmail.com

ABSTRACT

The snowtrout (*Schizothorax richardsonii)* is an economically important fish in the north-east Himalayan region. However, genomic research on this species is still in its infancy, and genomic resources are largely unavailable. The objective of this study was to generate expressed sequence tags (ESTs) from a brain complementary DNA (cDNA) library and to identify the genes. A total of 1031 ESTs were sequenced, from which 73 contigs and 411 singletons were identified. BLAST homology analysis indicated that only 9.3% of these ESTs were homologues of known genes while the remaining 90.7% appeared to be novel sequences. Based on sequence similarities, 45 putative genes were identified that encodes stress proteins, enzymes and signal transduction regulators. Our study thus, provides a collection of novel transcripts and a partial annotation of genes that are expressed in brain tissue of the snowtrout species, *S. richardsonii*.

Keywords: cDNA Library, Expressed Sequence Tags (ESTs), *Schizothorax richardsonii,* Snowtrout

Indian snowtrout (*Schizothorax richardsonii*) is an important cold water fish, which inhabits fast flowing torrential snow fed streams and lakes with water temperature from 8 to 22 °C (Sharma, 1989; Pradhan, 1982) in the north-east Himalayan region. It is a short migratory fish and migrates upwards when water temperature rises and downwards when water temperature decreases. In India, this species is distributed in cold waters from Jammu and Kashmir (Sunder and Bhagat, 1979) to Assam and Eastern Himalayas through Bhutan and Sikkim at an altitude of 1180-3000 masl (Jhingran, 1982). Its size varies from 89 to 453 mm in total length and 120 to 450 g in weight. It is herbivore in nature, feeding mainly on algae, aquatic plants and detritus (Talwar and Jhingran, 1991). *S. richardsonii* is a commercially important fish for sport fishery and an excellent food fish. At present, this has been categorised as threatened in different water bodies (Camp, 1997), thus indicating a need for conservation and management of this species in its natural habitats.

S. richardsonii is a promising freshwater species for coldwater aquaculture in the Himalayan and sub-Himalayan regions, but little is known about the genetics or genomics of this fish, and lack of knowledge of their development and physiology have hindered their domestication. Several hatcheries are currently spawning partly domesticated wild caught broodstock. A number of factors such as egg quality, larval nutrition and weaning on to artificial diets, malpigmentation, growth optimisation, disease, and immune function still need to be understood to make snowtrout culture a viable industry.

Large scale cDNA sequencing projects have identified expressed sequence tags (ESTs) and provide an effective means of identifying expressed genes in organisms across all kingdoms. Ideally ESTs generated from a total cDNA library should represent all the expressed genes in the tissue from which the library was constructed (Reddy *et al.,* 2002). EST resource facilitates gene discovery and help in determination of how genes (proteins) evolve new functions and processes between species. It provides an opportunity for future microarray and QTL mapping. Study of *S. richardsonii* genome is important for achieving sustainable molecular breeding. To date only few novel microsatellite markers from a partial genomic library (A'Hara *et al.,* 2012) are available from this species

and so far no ESTs have been generated from any cDNA libraries in this species. Hence, a brain cDNA library was developed in this species to classify ESTs which represent a rich source of genes and will be useful for transcriptome analysis of *S. richardsonii* and related coldwater fish species.

Live samples of snowtrout (*Schizothorax richardsonii*) were collected from a tributary of the river Teesta near Ranikhola, Sikkim, India during the month of December, 2010. Average size of the samples was 25 g. Brain tissues were collected in RNAlater solution (Ambion, USA) and stored at -80 ^{0}C until further use. RNA was isolated from 100-200 mg brain tissue. Tissues were first ground with mortar and pestle and the RNA was isolated using TRI-Reagent (Ambion, USA) protocol according to manufacturer's instruction. The quality of RNA was checked through agarose gel electrophoresis and absence of genomic DNA contamination was confirmed by PCR amplification of GAPDH gene (500 bp). First strand cDNA was synthesised by priming total RNA with the oligo (dT) primer using RevertAid™ First Strand cDNA Synthesis Kit (Fermentuas). Second strand cDNA synthesis was carried out following Sambrook *et al*. (2001). The final cDNA product was purified using PCR clean up kit (Fermentus). cDNA product was then ligated with *Hind*-III adapters (Fermentas). Then adapter ligated insert was ligated to pUC 18 cloning vector digested with *Hind*-III. The ligation mix was then transformed into highly efficient *E.coli* JM107 competent cells (Novagen, Merck) and grown on LB (Luria-Bertani) agar plates containing ampicillin, X- Gal and IPTG for selection of recombinant clones (Sambrook *et al.,* 2001).

cDNA library was screened using *Hind-III* digestion for average size of inserts before sequencing. The selected clones from replica plate were picked and grown in 2 ml liquid culture in LB medium overnight in culture tubes for plasmid preparation. Plasmid DNA were extracted using the plasmid isolation kit (Machery-Nagel, Germany) and stored at -20 °C. All positive clones were sequenced using vector specific primer (M13 reverse primer for pUC18) on ABI 3130 Genetic Analyser (Applied Biosystems, CA) with Big Dye Terminator version 3.1 cycle sequencing kit.

Assembly and annotation of EST sequences

The raw sequences were processed by PHRED program (Ewing *et al.,* 1998a, b) for base calling. Vector sequences were trimmed using CLC Genomics workbench (www.clcbio.com), while linker sequences were trimmed manually. High quality ESTs (>100 bp) were assembled and clustered into contiguous consensus sequences (contigs) and singletons (individual sequence reads) with Laser Gene SeqMan Pro Program (DNASTAR). The contigs were then checked manually to modify mosaic clones and mismatched sequences. Unigenes comprised of contigs and singletons were annotated by performing BLAST-X searches (Altschul *et al.,* 1990, 1997) in the NCBI non-redundant (nr) protein databases. Homology search for all the unigenes was conducted using BLAST-X in the NCBI nr database with a cutoff 'e value' of < 10^{-5} for the best hit. Annotations of transcripts subjected to GO classification were acquired by BLAST -X searching of the updated universal protein resource (UniProt) database using Blast 2 Go program (www.blast2go.org). GO terms were assigned to all annotated unigenes by performing UniProt 290 associations downloaded from the website of the Gene ontology consortium (Ashburner *et al.,* 2000). These terms were further classified and mapped to three GO categories. Further unigenes were screened for identification of SSR markers using Tandem Repeat Finder (Benson, 1999; http://tandem.bu.edu/trf/trf.html**)**. The EST resources were submitted to dbEST with the following accession numbers JK087360 to JK088390.

This is the first report to elucidate expressed genes in the brain tissue of *S. richardsonii*. EST clones from the library were randomly picked and sequenced from 5' end. 5' ESTs are more useful because most of the 5' sequences contain the coding sequences (Takasuga, 2001). A total of 1031 single colonies were randomly selected from the library for sequencing with M13 reverse primer resulting in 1023 EST sequences that were more than 100 bp, with average 442 bp, after the elimination of vector and linker sequences. Based on default criteria set in SeqMan Pro Program (DNASTAR), 484 of 1023 sequences were grouped in 73 contigs and 411 singletons (Table 1). Of the 484 unigenes obtained, 9.7% were formed by two EST sequences and 1.6% by more than ten ESTs (Fig. 1). A large fraction of the sequences ranged from 200 to 500 bp (Fig. 2). The average guanine-cytosine (GC) content of these ESTs was 40.8% with 81.32% of the sequences having GC content between 31% and 50% (Fig. 3).

Table 1. General properties of the cDNA library

Description	Number
Total cDNAs sequenced	1031
Total analysed cDNAs	1023
Average EST size (bp after trimming)	442
Total number of Unigenes	484
Number of Contigs	73
Number of orphan sequences (Singletons)	411
Unigenes with known gene match, n (%)	45 (9.3%)
Unigenes with no match, n (%)	439 (90.7%)

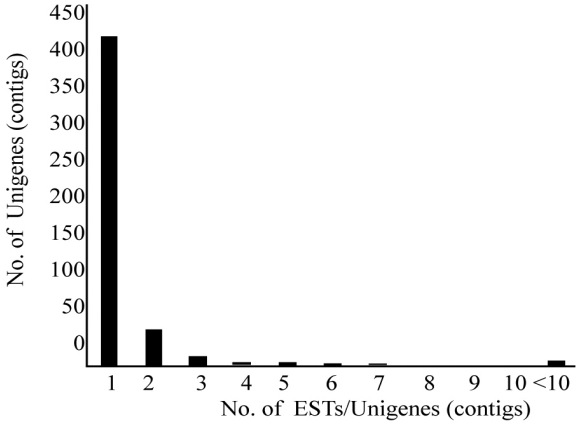

Fig. 1. Cluster analysis of 1023 ESTs from the cDNA library of *S. richardsonii*

Four hundred eighty four unigenes were searched for similarity using BLASTx and ranked by the expectation (e) values. "Little homology" indicates e> 1 X 10 $^{-10}$, and "no match" indicates the absence of homology to known sequences (Fig. 4). Fortyfive unigenes showed significant similarities to the known genes. ESTs were assigned putative functions according to the GO classification (Table 2). using Blast 2go. In this ontology, unigenes were classified into three functional categories, namely "Biological Process", "Cellular Component", and "Molecular Function". The snowtrout ESTs found highest similarity to the predicted protein sequences from various organisms *viz*. *Danio* sp. (36%), *Gallus* sp. (13%), *Homo sapiens* (7%), and *Medicago* sp. (7%) (Fig. 5).

However, no database matches were found for 439 (90.7%) sequences, suggesting of novel transcripts with unknown functions. It may be due to lack of genomic

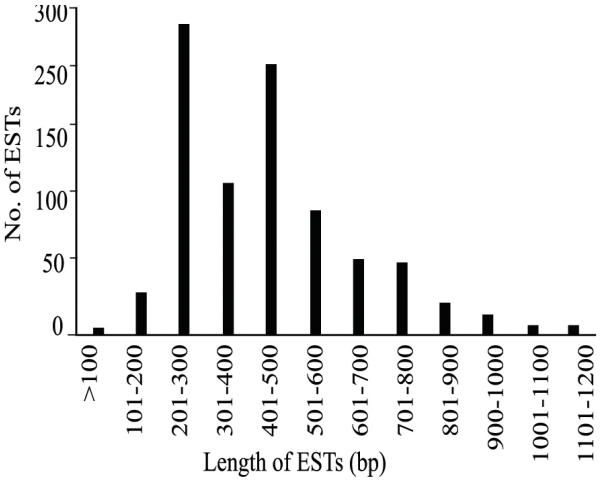

Fig. 2. Distribution of the readable lengths of the ESTs obtained from the cDNA library of *S. richardsonii*

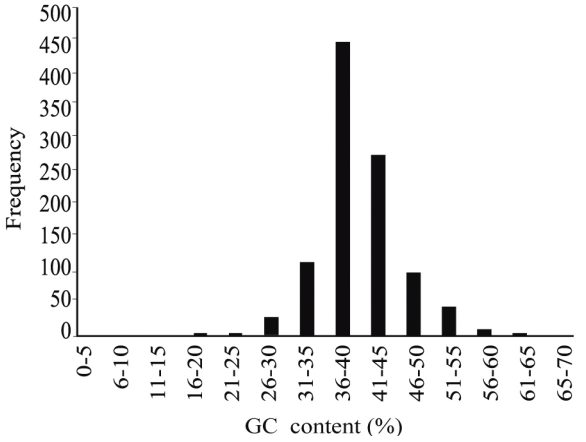

Fig. 3. The GC content of 1023 EST sequences from the cDNA library of *S. richardsonii*

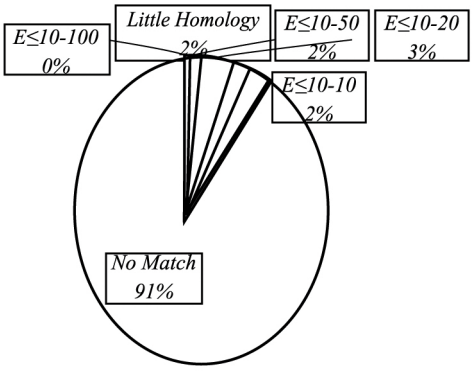

Fig. 4. Unigene distribution of the genome of *S.richardsonii.* Searched 484 unigenes in the public databases by BLASTx and ranked the annotation results by *e* values. *Little homology*- indicates e > 1 X 10^{-10}, *No match* - indicates the absence of homology to known sequences

resource of snowtrout and its related species. Besides, some of the 5' reads might have been limited to the 5' UTR region or poorly conserved end of the transcripts that could result in a high proportion of unmatched EST. The high volume of apparently novel sequences suggests that our data could be a useful resource for future genomic studies in other cold water fish species.

Several contigs matched with the mitochondrial genes. Other abundant genes include Dolichyl-diphospho-oligosaccharie protein glycosyltransferase, ORF2, reverse transcriptases and a number of hypothetical proteins.

ESTs are valuable resources for the development of useful genetic markers including microsatellite and single nucleotide polymorphism (SNP) markers. From the high

Table 2. ESTs that showed similarity to genes in the public databases along with Gene ontology (GO) classification of *S. richardsonii*.

Contig number	No. of ESTs	BLASTx hit	Accession	E-value	No. AA aligned	Gene ontology
9	2	Transmembrane protein 16C [*Danio rerio*]	AAH93348	1.01521E-19	41	F: ion channel activity; C: integral to membrane; C: membrane; P: ion transport; P: transport ; C: chloride channel complex; F: molecular function ;
17	2	Interferon-related developmental regulator 2 [*Danio rerio*]	CAP19399.1	2.39087E-26	60	F: binding ; C: cellular component ; F: molecular function
20	11	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3A [*Medicago truncatula*]	XP_003620144.1	2.29206E-68	127	P: protein amino acid glycosylation ; F: oligosaccharyl transferase activity ; F: transferase activity ; C: membrane ; F: dolichyl-diphosphooligosaccharide-protein
22	3	Zinc finger protein 831 [*Danio rerio*]	NP_001128615.1	4.81E-06	59	F: molecular function ; P: biological process ; C: cellular component
23	2	T-complex protein 1 subunit epsilon [*Danio rerio*]	NP_997778.1	1.61E-10	53	F: unfolded protein binding; P: cellular protein metabolic process; F: ATP binding; C: cytoplasm ; F: nucleotide binding ; P: protein folding; C: microtubule organizing center; C: nucleolus; F: protein binding; C: chaperonin-containing T-complex; P:'de novo' posttranslational protein folding; C: cytosol ; P: response to virus; C: cytoskeleton
39	3	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3A [*Medicago truncatula*]	XP_003620144.1	5.17E-67	132	P: protein amino acid glycosylation ; F: oligosaccharyl transferase activity; F: transferase activity; C: membrane; F: dolichyl-diphosphooligosaccharide-protein glycotransferase activity ; F: transferase activity, transferring glycosyl groups; C: endoplasmic reticulum; C: plasma membrane
51	17	Ferrichrome outer membrane transporter [*Escherichia coli* XH001]	EGV48412.1	5.86961E-52	103	F: iron ion binding; F: transporter activity; C: integral to membrane; C: membrane ; F: siderophore-iron transmembrane transporter activity; P: transport ; F: receptor activity; P: siderophore transport; C: cell outer membrane; C: plasma membrane

Contig number	No. of ESTs	BLASTx hit	Accession	E-value	No. AA aligned	Gene ontology
61	2	TonB-dependent siderophore receptor precursor [*Escherichia coli* P12b]	AFG39026.1	1.42E-33	86	F: iron ion binding; F: transporter activity; C: integral to membrane; C: membrane; F: siderophore-iron transmembrane transporter activity; P: transport; F: receptor activity; P: siderophore transport; C: cell outer membrane; C: plasma membrane
70	5	Gag-like protein [*Danio rerio*]	BAC82616.1	3.88E-34	108	F: metal ion binding; F: nucleic acid binding; P: biological process; F: zinc ion binding; C: cellular component
71	2	Semaphorin 4b [*Danio rerio*]	ACI29017.1	8.33E-06	37	P: nervous system development; P: cell differentiation; F: receptor activity; P: multicellular organismal development; C: integral to membrane; C: membrane
72	2	Ras association domain-containing protein 6 [*Danio rerio*]	NP_001038804.1	1.18E-05	42	P: signal transduction; C: cellular component
76	1	Zinc finger protein 831 [*Danio rerio*]	NP_001128615.1	4.61E-06	59	F: molecular function; P: biological process; C: cellular component
83	1	L1 repeat, Tf subfamily, member 18 [*Mus musculus domesticus*]	AAC72805.1	6.38E-22	113	F: RNA binding; F: nucleotidyltransferase activity; F: transferase activity; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
89	1	Ferrichrome outer membrane transporter [*Escherichia coli* XH001]	EGV48412.1	4.28E-54	93	F: iron ion binding; F: transporter activity; C: integral to membrane; C: membrane; F: siderophore-iron transmembrane transporter activity; P: transport; F: receptor activity; P: siderophore transport; C: cell outer membrane; C: plasma membrane
91	1	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3A [*Medicago truncatula*]	XP_003620144.1	7.82E-73	125	P: protein amino acid glycosylation; F: oligosaccharyl transferase activity; F: transferase activity; C: membrane; F: dolichyl-diphosphooligosaccharide-protein glycotransferase activity; F: transferase activity, transferring glycosyl groups; C: endoplasmic reticulum; C: plasma membrane
92	1	Lipopolysaccharide biosynthesis protein [*Escherichia coli* P12b]	AFG40997.1	8.79E-45	90	C: membrane; P: polysaccharide biosynthetic process

Contig number	No. of ESTs	BLASTx hit	Accession	E-value	No. AA aligned	Gene ontology
98	1	Extracellular metalloprotease [*Simonsiella muelleri* ATCC 29453]	ZP_06754474.1	7.70E-09	68	F: phosphatase activity; P: dephosphorylation; F: hydrolase activity; P: protein amino acid dephosphorylation; F: protein tyrosine phosphatase activity; F: phosphoprotein phosphatase activity
102	1	CPSF (cleavage and polyadenylation specific factor), subunit A, putative [*Plasmodium falciparum* 3D7]	XP_001351247.1	4.95E-09	103	F: nucleic acid binding; C: nucleus
110	1	Peptidylprolyl isomerase domain and WD repeat-containing protein 1 [*Danio rerio*]	NP_001092228.1	1.06545E-29	64	F: peptidyl-prolyl cis-trans isomerase activity; P: protein folding; F: isomerase activity; C: cellular_component; C: spliceosomal complex; P: RNA splicing; P: mRNA processing; C: nucleus
112	1	Casein kinase 2 alpha 2a [*Danio rerio*]	NP_571315.1	3.24E-15	38	F: kinase activity; F:ATP binding; F: protein kinase activity; F: nucleotide binding; P: protein amino acid phosphorylation; F: protein serine/ threonine kinase activity
116	1	Seven transmembrane helix receptor [*Homo sapiens*]	BAC05810.1	8.06E-10	81	F: receptor activity; F: molecular_function; P: biological_process; C:cellular_ component
117	1	Calcium activated nucleotidase 1 [*Gallus gallus*]	NP_001026752.1	2.60E-16	58	C: Golgi cisterna membrane; C: integral to membrane; C: membrane; F: hydrolase activity; C: Golgi apparatus; F: nucleoside-diphosphatase activity; F: pyrophosphatase activity; C: endoplasmic reticulum membrane; F: signal transducer activity; P: positive regulation of I-kappaB kinase/NF-kappaB cascade; C: endoplasmic reticulum; F: calcium ion binding; P: ribonucleoside diphosphate catabolic process; F: guanosine-diphosphatase activity; C: membrane fraction; P: metabolic process; F: metal ion binding; F: uridine-diphosphatase activity
118	1	BTB (POZ) domain containing 8, isoform CRA_c [*Homo sapiens*]	EAW73104.1	2.23E-05	44	C: nucleus; C: nucleolus

Contig number	No. of ESTs	BLASTx hit	Accession	E-value	No. AA aligned	Gene ontology
138	1	Phosphate transporter family protein [*Escherichia coli* DEC13E]	EHX71218.1	8.20E-39	89	P: phosphate transport; C: membrane; F: inorganic phosphate transmembrane transporter activity
158	1	Pleckstrin homology domain-containing family M member 1 [*Danio rerio*]	NP_001082872.1	4.48E-05	62	P: intracellular signaling pathway
161	1	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit C [*Danio rerio*]	NP_001018164.1	1.81E-27	44	F: molecular_function; P: biological_process; C: cellular_component; F: protein binding
162	1	Cytochrome oxidase subunit I [*Tor douronensis*]	ABF82296.1	1.51E-68	123	F: heme binding; F: iron ion binding; P: aerobic respiration; C: integral to membrane; C: membrane; C: mitochondrial inner membrane; F: electron carrier activity; C: respiratory chain; C: mitochondrion; P: transport; P: oxidation reduction; F: oxidoreductase activity; F: cytochrome-c oxidase activity; P: electron transport chain; F: metal ion binding
166	1	NEDD4 family-interacting protein 2 [*Danio rerio*]	NP_998512.1	7.97E-11	51	F: molecular_function; P: biological_process; C: cellular_component; C: integral to membrane; C: membrane; P: positive regulation of protein ubiquitination; P: negative regulation of transporter activity; C: endosome; C: mitochondrion; C: cytoplasm; P: negative regulation of protein transport; C: Golgi apparatus; C: perinuclear region of cytoplasm; P: negative regulation of gene expression; F: signal transducer activity; P: positive regulation of I-kappaB kinase/NF-kappaB cascade; C: endoplasmic reticulum; F: WW domain binding; C: Golgi membrane; C: endosome membrane; C: multivesicular body membrane; F: protein binding

Contig number	No. of ESTs	BLASTx hit	Accession	E-value	No. AA aligned	Gene ontology
185	1	Neurogranin (protein kinase C substrate, RC3) b [*Danio rerio*]	NP_001074129.1	6.39E-21	72	F: molecular_function; P: biological_process; C: cellular_component
217	1	Homeobox protein extradenticle [*Acromyrmex echinatior*]	EGI59311.1	1.23E-08	46	C: nucleus; F: transcription factor activity
275	1	Pol-like protein [*Biomphalaria glabrata*]	ABN58714.1	1.13E-20	171	F: RNA binding; F: nucleic acid binding; P: RNA-dependent DNA replication; F: ribonuclease H activity; F: RNA-directed DNA polymerase activity; F: zinc ion binding
278	1	Gag-like protein [*Danio rerio*]	BAC82614.1	2.34E-26	86	F: metal ion binding; F: nucleic acid binding; P: biological_process; F: zinc ion binding; C: cellular_component
313	1	Reverse transcriptase, partial [*Gallus gallus*]	AAA49027.1	7.08E-33	159	F: transferase activity; F: DNA-dependent protein kinase activity; C: nonhomologous end joining complex; F: DNA binding; C:nucleus; F: nucleotide binding; F: ATP binding; P: double-strand break repair *via* nonhomologous end joining; F:kinase activity; P: DNA repair; F: phosphotransferase activity, alcohol group as acceptor
315	1	DNA-dependent protein kinase catalytic subunit [*Gallus gallus*]	AAR28760.1	1.30E-21	127	F: transferase activity; F: DNA-dependent protein kinase activity; C: non-homologous end joining complex; F: DNA binding; C: nucleus; F: nucleotide binding; F: ATP binding; P: double-strand break repair *via* nonhomologous end joining; F: kinase activity; P: DNA repair; F: phosphotransferase activity, alcohol group as acceptor
336	1	POL-like, partial [*Gallus gallus*]	AAA49022.1	2.37E-08	40	F:transferase activity; F: DNA-dependent protein kinase activity; C: non-homologous end joining complex; F: DNA binding; C: nucleus; F: nucleotide binding; F: ATP binding; P: double-strand break repair *via* non-homologous end joining; F: kinase activity; P: DNA repair; F: phosphotransferase activity, alcohol group as acceptor
344	1	Adenylate cyclase [*Achromobacter xylosoxidans* A8]	YP_003978909.1	3.94E-05	66	F: peptidyl-prolyl cis-trans isomerase activity; P: protein folding; F: isomerase activity; C: outer membrane

Contig number	No. of ESTs	BLASTx hit	Accession	E-value	No. AA aligned	Gene ontology
359	1	Reverse transcriptase, partial [*Gallus gallus*]	AAA49027.1	8.13E-19	129	F: transferase activity; F: DNA-dependent protein kinase activity; C: nonhomologous end joining complex; F: DNA binding; C: nucleus; F: nucleotide binding; F: ATP binding; P: double-strand break repair *via* non-homologous end joining; F: kinase activity; P: DNA repair; F:phosphotransferase activity, alcohol group as acceptor
362	1	Reverse transcriptase, partial [*Gallus gallus*]	AAA49027.1	3.89E-09	56	F: transferase activity; F: DNA-dependent protein kinase activity; C: non-homologous end joining complex; F: DNA binding; C: nucleus; F: nucleotide binding; F: ATP binding; P: double-strand break repair *via* non-homologous end joining; F: kinase activity; P: DNA repair; F: phosphotransferase activity, alcohol group as acceptor
379	1	Reverse transcriptase-like protein [*Homo sapiens*]	AAW63045.1	1.07E-10	107	P: inflammatory response; F: chemokine activity; C: extracellular region; P: signal transduction; P: chemotaxis; P: cell-cell signaling; F: heparin binding; P: immune response; C: extracellular space
400	1	Radical SAM domain protein [*Shigella sonnei* 53G	EFZ53827.1	5.28E-109	159	F: metal ion binding; F: catalytic activity; F: 4 iron, 4 sulfur cluster binding; F: iron-sulfur cluster binding; C: cellular_component
441	1	Carcinoembryonic antigen-related cell adhesion molecule 5 precursor [*Danio rerio*]	NP_001107266.1	1.38E-23	79	P: cell migration; P: integrin-mediated signaling pathway; C: integral to membrane; F: molecular_function; P: angiogenesis; C: extracellular region; C: membrane fraction; C: integral to plasma membrane; C: plasma membrane; P: homophilic cell adhesion; C: extracellular space; C: anchored to membrane; P: immune response
448	1	Ice nucleation protein [*Cricetulus griseus*]	EGW04298.1	4.43E-26	268	F: polysaccharide binding; F: scavenger receptor activity; C: extracellular region; P: immune response
463	1	UDP-N-acetylglucosamine diphosphorylase [*Propionibacterium acnes* HL096PA3]	EFT60788.1	1.49E-85	132	F: transferase activity; F: nucleotidyltransferase activity; F: UDP-N-acetylglucosamine diphosphorylase activity

Contig number	No. of ESTs	BLASTx hit	Accession	E-value	No. AA aligned	Gene ontology
475	1	Dual specificity protein phosphatase 10 [*Bos taurus*]	NP_001029897.2	1.94E-42	108	C: cytoplasm; F: MAP kinase tyrosine/ serine/threonine phosphatase activity; F: hydrolase activity; P: protein amino acid dephosphorylation; F: protein tyrosine phosphatase activity; C: nucleus; P: response to stress; P: JNK cascade; F: phosphoprotein phosphatase activity; C: Golgi apparatus; C: nucleolus
480	1	Centromere protein F-like, partial [*Saccoglossus kowalevskii*]	XP_002737030.1	1.67E-18	175	F: RNA binding; F: nucleic acid binding; P: RNA splicing; F: nucleotide binding; P: mRNA processing; C: nucleus

*Contigs with Acc. No. Contig 9 (JK087479, JK087593); 17 (JK087398, JK087673); 20 (JK087434 JK087460 JK087483 JK087547 JK087595 JK087665 JK087674 JK087766 JK087767 JK087780 JK088115); 22 (JK087501 JK088014 JK088015); 23 (JK087515 JK087857); 39 (JK087367 JK087761 JK087782); 51 (JK087397 JK087407 JK087417 JK087446 JK087528 JK087611 JK087656 JK087685 JK087689 JK087702 JK087707 JK087753 JK087802 JK087876 JK087884 JK087906 JK088116); 61 (JK087509 JK087764); 70 (JK088021 JK088025 JK088058 JK088072 JK088088); 71 (JK087521 JK087710); 72 (JK087624 JK087746); 76 (JK087830); 83 (JK087852); 89 (JK088386); 91 (JK087422); 92 (JK087368); 98 (JK087836); 102 (JK087645); 110 (JK087774); 112 (JK087554); 116 (JK088319); 117 (JK088381); 118 (JK088253); 138 (JK087544); 158 (JK087722); 161 (JK087382); 162 (JK087373); 166 (JK087612); 185 (JK087502); 217 (JK087690); 275 (JK087546); 278 (JK088055); 313 (JK088323); 315 (JK088378); 336 (JK088226); 344 (JK087529); 359 (JK088304); 362 (JK088280); 379 (JK087558); 400 (JK087808); 441 (JK088085); 448 (JK088217); 463 (JK087457); 475 (JK087523); 480 (JK088329).

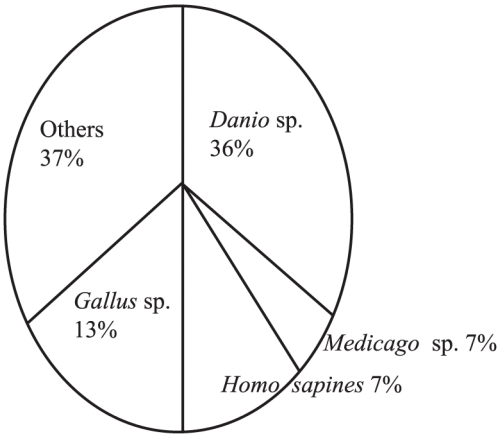

Fig. 5. The proportion of putative transcripts with BLASTx matches in the non-redundant database classified by the organism of the "best hit" protein sequence

quality unigenes, 26 unique microsatellites containing ESTs were identified. Thus, the EST approach offers an efficient opportunity for the development of microsatellite markers. Moreover, many novel cDNA sequences were generated during this study. The database generated in this study will support discovery of essential genes that play key role in physiological and biochemical pathway relevant to defense, growth mechanisms and temperature adaptation of this species leading to improved management of cultured stocks.

Acknowledgements

The authors are highly grateful to the National Agricultural Innovative Project (ICAR) for financial support to conduct this work. They also acknowledge Dr. P. C. Mahanta, Director, DCFR, Bhimtal for his continuous guidance and providing infrastructure facilities to carry out the work. The authors are also thankful to Dr. Mathew L. Rise, Associate Professor and Mr. Charles Y. Feng, Post Graduate Student, Ocean Sciences Laboratory, Memorial University of Newfoundland, Canada, for their generous help in transcriptome analysis of this cDNA library.

References

A' Hara, S.W., Amouroux, P., Argo, E. E., Avand-Faghih, A. and Barat, A. 2012. Molecular ecology resources primer development consortium, Permanent genetic resources added to Molecular ecology resources database 1, August 2011–30 September 2011. *Mol. Ecol. Resour.,* 12:185-189.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. *J.* Mol. *Biol.,* 215**:** 403 –410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.,* 25: 3389– 3402.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S. and Eppig, J. T. 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.,* 25: 25– 29.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.,* 27 (2): 573-580.

CAMP 1997. *Report of the workshop on "Conservation Assessment and Management Plan (CAMP) for freshwater fishes of India".* Zoo Outreach Organization and NBFGR, Lucknow, p. 1-158.

Ewing, B., Hillier, L. D., Wendl, M. C. and Green, P. 1998a. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.,* 8: 175 –185.

Ewing, B., Hillier, L. D., Wendl, M. C. and Green, P. 1998b. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.,* 8: 186 –194.

Jhingran, V. G. 1982. *Fish and fisheries of India*. Hindustan Publishing Corporation, India.

Pradhan, B. R. 1982. Preliminary studies of SyarpuDaha-A mid-hill lake in Rukum District, Nepal. *Proc. Nat. Sc. Tech. Cong. Kath*., p. 261-264.

Reddy, A. R., Ramakrishna, W., Chandra Sekhar, A., Ithal, N., Babu, P. R., Bonaldo, M. F., Soares, M. B. and Bennetzen, J. L. 2002. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). *Genome,* 45: 204-211.

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. *Molecular cloning: a laboratory manual,* 2nd edn. Cold Spring Harbor Laboratory Press, New York, USA.

Sharma, B. P. 1989. Status of *Schizothorax* spp. in India-Chinese Sub-continent. *FAO. Fish Tech Rep*., 405: 90-94.

Sunder, S. and Bhagat, M. J. 1979. A note on the food of *Schizothorax plagiostomus* (McClelland) in the Chenab drainage of Jammu Province during 1973-74. *J. Inland Fish. Soc. India*, 11 (1): 117-118.

Takasuga, A. 2001. Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs. *Nucleic Acids Res.,* 29: E108.

Talwar, P. K. and Jhingran, A. 1991. *Inland fishes of India and adjacent countries*, vol.1-2. Oxford and IBH publishing Co. Pvt. Ltd, 1158 pp.

Date of Receipt : 01-12-2012

Date of Acceptance : 22.11.2013