

## GENERATION AND ANALYSIS OF ESTs FROM THE GRASS CARP, *CTENOPHARYNGODON IDELLUS*

Bing Xu<sup>1</sup>, Shaolin Wang<sup>2</sup>, Yan Jiang<sup>1</sup>, Linlin Yang<sup>1</sup>,  
Ping Li<sup>2</sup>, Caixia Xie<sup>1</sup>, Jubin Xing<sup>1</sup>, Zhonghe Ke<sup>1</sup>, Jiale Li<sup>1</sup>,  
Junwei Gai<sup>1</sup>, Guimei Yang<sup>1</sup>, Baolong Bao<sup>1</sup>, and Zhanjiang Liu<sup>1,2</sup>

<sup>1</sup>*The Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Ministry of Education, Shanghai, China*

<sup>2</sup>*The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, Auburn University, Auburn, AL, USA*

*Grass carp, *Ctenopharyngodon idellus* (Valenciennes, 1844), is an economically important species widely cultured in the world, but its genome research resources are largely lacking. The objectives of this study were to construct normalized cDNA libraries for efficient EST analysis, to generate ESTs from these libraries, and to identify EST-related molecular markers such as microsatellites and single nucleotide polymorphisms (SNPs) for genetic analysis of this species. A total of 6,269 ESTs were generated representing 4,815 unique sequences, from which 105 putative microsatellites and 5,228 SNPs were identified. These genome resources provide the material basis for future genetic and functional analyses in this species.*

**Keywords:** EST; Genome; Grass carp; Marker; Microsatellite; SNP

Grass carp, *Ctenopharyngodon idella* (Valenciennes, 1844), native to Siberia and northern China,<sup>1</sup> is one of the most important aquaculture species in the world. However, diseases such as red spot disease have caused devastating losses to aquaculture in Asia.<sup>2,3</sup> In addition, habitat destruction has caused the decline of wild populations.<sup>4</sup> On the other hand, as an invasive species in North America, it has had a disastrous impact on native fish populations, threatening various native species in the Mississippi River and the Great Lakes.<sup>5–6</sup> In spite of the economic importance, genetic and genomic information of grass carp is largely lacking. In particular, expressed sequence tag (EST) resource was very small in spite of its utilities as

This work was supported by the key foundation program 06dj14003 from Shanghai Municipal Science Committee, China, and partially supported by National Natural Science Foundation of China 30771668, National Education Commission of China 06ZZ65, and the Key Discipline funding Y1101, Marine Biology funding J50701, and Hydrobiology funding S30701.

Address correspondence to Zhanjiang Liu or Baolong Bao, The Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Ministry of Education, Shanghai 201306, China. E-mails: liuzhan@auburn.edu or bilbao@shou.edu.cn

demonstrated in other aquaculture species.<sup>7–15</sup> The objectives of this project were to generate ESTs from grass carp and identify microsatellites and single nucleotide polymorphisms (SNPs) for future genetic analysis.

## MATERIAL AND METHODS

Total RNA was isolated from grass carp tissues of intestine, gill, and hepatopancreas<sup>16</sup> using the Trizol reagent (Invitrogen, Carlsbad, CA) following manufacturer's instructions. These tissues were used because of their importance in disease responses. Poly(A)<sup>+</sup> RNA was purified from total RNA using the Poly(A)<sup>+</sup> Pure kit (Ambion, Austin, TX) according to the manufacturer's instructions. The cDNA libraries were constructed using the pSPORT-1 Superscript Plasmid Cloning System (Invitrogen). Two micrograms of Poly(A)<sup>+</sup> RNA were used in each initial reaction. Procedures followed instructions provided by the manufacturer with the exception that ElectroMax DH12S cells (Invitrogen) were used for electroporation of the cDNA library.

Colony-lifting hybridization<sup>17</sup> was conducted using overgo probes to reduce the sequencing redundancy of a set of 200 genes determined to be highly expressed by preliminary sequencing of the libraries. After hybridization, negative colonies were picked for sequencing and manually arrayed into 96 well plates containing LB with antibiotics and 10% glycerin and stored at  $-80^{\circ}\text{C}$  until sequencing. Plasmid DNA was prepared using the AxyPrep Plasmid Miniprep Kit (Axygen, Union City, CA). One microliter of plasmid DNA (200 ng/ $\mu\text{l}$ ) was used in sequencing reactions. Chain termination sequencing was performed using BigDye Terminator Kit (Applied Biosystems, Foster City, CA).

Sequence assembly was conducted using Vector NTI suite 8.0 (Invitrogen). The consensus sequence of each contig and singletons were searched against NCBI nr database using BLAST2GO<sup>18,19</sup> with cutoff E-value  $1\text{e-}5$ . Gene ontology (GO) terms were assigned to the sequences with BLASTX hits according to the AmiGO database

**Table 1** Summary statistics of ESTs generated from grass carp (*Ctenopharyngodon idella*) and identification of putative microsatellites and SNPs

Total number of clones sequenced	7,342
Number of clean ESTs	6,269
Average length of clean ESTs (bp)	609
Number of contigs	867
Number of singletons	3,948
Number of unique sequences	4,815
Number of unique sequences with BLASTX hits	2,985 (62%)
Total SNPs detected within the 867 contigs	5,228
Transitions	2,243
Transversions	1,760
Indels	1,225
Putative SNPs within contigs with 2 sequences	2,969
Putative SNPs within contigs with 3 sequences	1,519
Putative SNPs within contigs with 4 sequences	698
Putative SNPs within contigs with 5–10 sequences	34
Putative SNPs within contigs with >10 sequences	8

**Table 2** Microsatellite-containing ESTs with sufficient flanking sequences for primers design

GenBank Accession no.	Clone no.	Putative identity	Length (bp)	Repeat motif and no.
GT226779	SFU-IT7-B067-F04	TGFBR1	755	(GT)9
GT226287	SFU-IT7-B027-E07	BCL7B	753	(AC)21
GT226294	SFU-GT7-A041-H10	CREM	785	(CA)9
GT226768	SFU-IT7-B063-B12	RRP7A	476	(AC)14
GT226307	SFU-LT7-C016-H08	CCL17	746	(AC)21, (AC)9
GT226308	SFU-IT7-B021-G01	CXCL12	400	(AC)8
GT226318	SFU-IT7-B069-F04	No name	821	(CA)9
GT226325	SFU-IT7-A002-G09	No name	761	(TCA)6
GT226328	SFU-IT7-B007-F08	CCNB3	513	(TC)16
GT226769	SFU-IT7-B069-B04	EPB41L3	822	(GAT)11
GT226780	SFU-IT7-B063-B04	GNA11	522	(CA)9
GT226770	SFU-IT7-B045-C09	No name	698	(TC)14
GT226360	SFU-IT7-B051-F09	HMCN1	730	(TAGA)6
GT226789	SFU-IT7-B043-F08	STEAP4	599	(TTC)7
GT226771	SFU-IT7-B003-C03	RABL4	742	(TG)14
GT226772	SFU-IT7-B006-C03	MBD5	719	(TG)11
GT226773	SFU-IT7-B081-F05	MTHFD1L	735	(GT)30
GT226781	SFU-GT7-A015-C02	No homologue	633	(GT)9
GT226486	SFU-GT7-A006-B12	No homologue	426	(GT)10
GT226507	SFU-GT7-A001-C08	No homologue	798	(AAAC)5
GT226508	SFU-GT7-A001-C10	No homologue	645	(TCTT)6
GT226518	SFU-GT7-A004-A10	No homologue	743	(TG)10
GT226783	SFU-GT7-A005-A10	No homologue	479	(CT)8
GT226521	SFU-GT7-A005-A05	No homologue	534	(TTCT)5
GT226522	SFU-GT7-A007-D08	No homologue	362	(AG)9
GT226532	SFU-IT7-B058-H03	No homologue	524	(TG)9
GT226534	SFU-IT7-B059-A10	No homologue	607	(TAGA)8
GT226542	SFU-IT7-B015-B06	No homologue	497	(TATC)6
GT226554	SFU-IT7-B017-B04	No homologue	412	(TG)8
GT226774	SFU-IT7-B017-C10	No homologue	392	(ATC)9
GT224958	SFU-IT7-B017-D07	No homologue	401	(TATT)5
GT226557	SFU-IT7-B019-G05	No homologue	374	(GAG)7
GT226563	SFU-IT7-B022-A03	No homologue	573	(ATTT)5
GT226572	SFU-IT7-B011-D01	No homologue	408	(CT)11
GT226784	SFU-IT7-B011-H05	No homologue	571	(CT)8
GT226785	SFU-IT7-B012-A05	No homologue	607	(TG)8
GT226574	SFU-IT7-B012-B11	No homologue	577	(GT)13
GT226577	SFU-IT7-B012-G12	No homologue	719	(AAAC)5
GT226786	SFU-IT7-B013-G07	No homologue	698	(AG)10
GT226582	SFU-IT7-B014-C01	No homologue	405	(GT)18
GT226586	SFU-IT7-B022-F08	No homologue	443	(AG)11
GT226587	SFU-IT7-B040-B09	No homologue	449	(GT)8
GT226590	SFU-IT7-B009-A02	No homologue	652	(TC)8
GT226595	SFU-IT7-B026-D05	No homologue	533	(CACT)6
GT226600	SFU-IT7-B038-A05	No homologue	742	(AT)9
GT226604	SFU-IT7-B055-A06	No homologue	616	(CT)13
GT226612	SFU-IT7-B025-H06	No homologue	773	(AAAG)5
GT226616	SFU-IT7-B049-C03	No homologue	782	(TAA)6
GT226630	SFU-IT7-B064-C06	No homologue	761	(AACA)6

*(Continued)*

Table 2 Continued

GenBank Accession no.	Clone no.	Putative identity	Length (bp)	Repeat motif and no.
GT226641	SFU-IT7-B067-H12	No homologue	734	(GT)10
GT226644	SFU-IT7-B066-G03	No homologue	771	(TG)13
GT226650	SFU-IT7-B003-B11	No homologue	764	(TGTT)8
GT226653	SFU-IT7-B046-A03	No homologue	677	(GA)8
GT226657	SFU-IT7-B041-G12	No homologue	525	(AC)20
GT226661	SFU-IT7-B043-G08	No homologue	616	(CA)8
GT226663	SFU-IT7-B045-C06	No homologue	654	(GA)8
GT226677	SFU-IT7-B006-H11	No homologue	823	(TG)10
GT226683	SFU-IT7-B079-B03	No homologue	706	(TG)10
GT226684	SFU-IT7-B079-C02	No homologue	774	(CA)9
GT226692	SFU-IT7-B029-E03	No homologue	511	(CT)8
GT226697	SFU-IT7-B029-G08	No homologue	679	(AC)9
GT226709	SFU-IT7-B071-A04	No homologue	807	(TA)9
GT226713	SFU-IT7-B071-F02	No homologue	800	(GT)9
GT226716	SFU-IT7-B075-B01	No homologue	802	(CA)8
GT226721	SFU-IT7-B080-H12	No homologue	702	(TG)9
GT226731	SFU-IT7-B074-A10	No homologue	702	(CA)9, (GT)8
GT226733	SFU-IT7-B074-E11	No homologue	745	(AC)16
GT226748	SFU-IT7-B012-D08	No homologue	248	(TCA)8
GT226412	SFU-IT7-B057-F09	PIGM	233	(TG)11
GT226419	SFU-IT7-B052-F05	PRPF38A	241	(TTCG)5
GT226423	SFU-IT7-B070-E08	RFX1	825	(CA)8
GT226431	SFU-IT7-B080-H04	RYK	802	(CA)8
GT226788	SFU-IT7-B040-B08	PAF1	522	(GAA)7
GT226775	SFU-IT7-B077-G12	SUMO2	816	(TC)11
GT226439	SFU-IT7-B046-D08	SLC13A2	513	(GAG)6
GT226789	SFU-IT7-B043-F08	STEAP4	599	(CTT)7
GT226452	SFU-GT7-A019-F03	TFDP2	796	(TACA)5
GT226453	SFU-IT7-B042-A11	TMX2	774	(AG)10
GT226457	SFU-IT7-B069-H11	TRIM24	764	(CAG)6
GT226776	SFU-IT7-B070-D03	USP8	818	(TC)11
GT226777	SFU-IT7-B042-B03	UPF2	788	(CA)9, (AC)13
GT226468	SFU-GT7-A006-A06	ATP2B1	782	(TCTT)5
GT226278	SFU-GT7-A006-B11	ACTN3	737	(AG)8
GT226477	SFU-GT7-A008-H09	MKLN1	631	(TC)16
GT226790	SFU-GT7-A012-B01	DSP	583	(GGA)7
GT226778	SFU-GT7-A008-A06	No homologue	661	(CA)16, (AC)16
GT226792	SFU-GT7-A013-B12	No homologue	727	(TG)10, (GT)10

(<http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>).<sup>19</sup> Microsatellites were identified using Microsatfinder (<http://www.eusebius.mysteria.cz/microsatfinder/index.php>).<sup>20</sup> Putative SNPs were identified by using the AutoSNP program with default parameters.<sup>21</sup>

## RESULTS AND DISCUSSION

A total of 7,342 clones were sequenced, generating 6,269 clean ESTs with an average length of 609 bp. These ESTs have been deposited to GenBank with

**Table 3** Grass carp ESTs similar to gene potentially involved in immune and defense responses

GenBank Accession no.	Putative identity	Top hit species	E-value	Top hit ID
GT222377	cd83 antigen precursor	<i>Ginglymostoma cirratum</i>	1.09E-10	AAO62993.1
GT222410	chemokine (c-c motif) receptor 9	<i>Danio rerio</i>	1.33E-43	XP_001343888.1
GT222787	chemokine (C-C motif) receptor 6a	<i>Danio rerio</i>	1.24E-79	NP_001002125.1
GT223304	interleukin 6	<i>Danio rerio</i>	1.55E-18	XP_001919693.1
GT223345	neutrophil cytosolic factor 1	<i>Danio rerio</i>	4.89E-17	BAF73666.1
GT224155	wiskott-aldrich syndrome protein	<i>Danio rerio</i>	1.50E-62	CAQ15295.1
GT222206	apolipoprotein a-i	<i>Danio rerio</i>	5.8E-23	NP_571203.1
GT222365	ccat enhancer-binding protein beta	<i>Danio rerio</i>	1.15E-86	AAL54865.1
GT222411	chemokine (c-c motif) receptor-like 1	<i>Danio rerio</i>	2.27E-33	CAN88714.1
GT222413	c-x-c chemokine receptor 4	<i>Cyprinus carpio</i>	1.95E-54	BAA32797.1
GT222529	complement component 4b (childo blood group)	<i>Cyprinus carpio</i>	5.04E-35	BAB03284.1
GT222758	exosome component 9	<i>Danio rerio</i>	6.65E-115	NP_001006077.1
GT222793	fk506 binding protein 12-rapamycin associated protein 1	<i>Danio rerio</i>	6.02E-72	CAQ13429.1
GT222825	fyn binding protein	<i>Danio rerio</i>	1.03E-63	XP_001923593.1
GT222977	inhibitor of kappa light polypeptide enhancer in b-kinase complex-associated protein	<i>Danio rerio</i>	7.26E-13	XP_689534.2
GT223000	interleukin 2 gamma (severe combined immunodeficiency)	<i>Danio rerio</i>	6.89E-90	NP_001121743.1
GT223001	interleukin 6 signal transducer	<i>Danio rerio</i>	3.44E-43	BAH47260.1
GT223082	endoplasmic reticulum aminopeptidase 1	<i>Danio rerio</i>	8.45E-31	NP_955915.1
GT223175	Chitinase 1 (chitotriosidase)	<i>Monodelphis domestica</i>	6.53E-48	XP_001381999.1
GT223242	myelin basic protein	<i>Danio rerio</i>	9.54E-05	XP_001335551.1
GT223284	interleukin 4 receptor	<i>Danio rerio</i>	5.68E-13	XP_692125.3
GT223443	peroxisome proliferator activated receptor gamma	<i>Ctenopharyngodon idella</i>	8.52E-17	ACF70732.1
GT223453	glucose phosphate isomerase	<i>Ctenopharyngodon idella</i>	4.80E-19	AAS89352.1
GT223485	polymerase iii (dna directed) polypeptide 39 kda	<i>Danio rerio</i>	3.88E-22	AAT68074.1
GT223867	strawberry notch homolog 2	<i>Mus musculus</i>	3.78E-83	EDL31603.1
GT223905	tapasin	<i>Ctenopharyngodon idella</i>	1.75E-67	ABU86892.1
GT223970	transcription factor binding to IGHM enhancer 3a	<i>Danio rerio</i>	2E-10	CAQ14100.1
GT224041	tumor necrosis alpha-induced protein 1	<i>Danio rerio</i>	5.92E-07	NP_956093.1
GT224140	vitronectin	<i>Danio rerio</i>	3.10E-96	AAH55570.1
GT224232	cd164 sialomucin	<i>Danio rerio</i>	2.32E-36	NP_001038256.1

(Continued)

Table 3 Continued

GenBank Accession no.	Putative identity	Top hit species	E-value	Top hit ID
GT224275	nuclear factor interleukin-3-regulated protein	<i>Salmo salar</i>	1.00E-11	ACN11325.1
GT226298	cathepsin s	<i>Danio rerio</i>	1.84E-121	AAH66625.1
GT226308	stromal cell-derived factor 1 precursor	<i>Danio rerio</i>	4.14E-09	NP_932334.1
GT226329	cytochrome b-alpha polypeptide	<i>Cyprinus carpio</i>	1.17E-76	BAF73665.1
GT226361	hepcidin	<i>Aristichthys nobilis</i>	8.32E-34	ACO51156.1
GT226374	invariant chain-like protein 1	<i>Danio rerio</i>	1.89E-08	AAD24542.1
GT226408	nucleoside phosphorylase	<i>Danio rerio</i>	1.21E-07	NP_998476.1
GT226423	regulatory factor 1 (influences hla class ii expression)	<i>Danio rerio</i>	1.24E-46	NP_001083022.1
GT226782	actin binding 1a	<i>Danio rerio</i>	1.61E-123	NP_957408.1
GT222377	cd83 antigen precursor	<i>Ginglymostoma cirratum</i>	1.09E-10	AAO62993.1
GT222410	chemokine (c-c motif) receptor 9	<i>Danio rerio</i>	1.33E-43	XP_001343888.1
GT222787	chemokine (C-C motif) receptor 6a	<i>Danio rerio</i>	1.24E-79	NP_001002125.1
GT223304	interleukin 6	<i>Danio rerio</i>	1.55E-18	XP_001919693.1
GT223345	neutrophil cytosolic factor 1	<i>Danio rerio</i>	4.89E-17	BAF73666.1
GT224155	wiskott-aldrich syndrome protein	<i>Danio rerio</i>	1.50E-62	CAQ15295.1
GT222136	adaptor-related protein complex beta 1 subunit	<i>Pan troglodytes</i>	5.89E-11	XP_001174130.1
GT222321	calcium binding atopy-related autoantigen 1	<i>Danio rerio</i>	1.44E-41	NP_001077302.1
GT222536	complement factor properdin	<i>Danio rerio</i>	1.57E-71	NP_001153598.1
GT222992	beta 1 (fibronectin beta antigen cd29 includes msk12)	<i>Danio rerio</i>	8.78E-49	AAI62706.1
GT223376	protein tyrosine receptor c	<i>Cyprinus carpio</i>	7.55E-18	BAA92179.1
GT223389	nuclear transcription x-box binding-like 1	<i>Danio rerio</i>	1.67E-41	XP_001346303.2
GT223415	Uromodulin precursor (Tamm-Horsfall urinary glycoprotein) (THP)	<i>Danio rerio</i>	8.23E-33	XP_001346047.2
GT223736	sh2 domain protein 1a	<i>Danio rerio</i>	1.84E-41	NP_001108163.1
GT223857	stabilin 1	<i>Danio rerio</i>	1.33E-137	XP_699221.3
GT223945	t-cell immune regulator 1	<i>Danio rerio</i>	1.75E-14	XP_001922910.1
GT223954	tnfaip3 interacting protein 1	<i>Danio rerio</i>	1.71E-31	AAH83289.1
GT223997	transmembrane 9 superfamily protein member 4	<i>Danio rerio</i>	1.42E-63	NP_956804.1
GT226289	bcl2 adenovirus e1b 19kda interacting protein 3	<i>Danio rerio</i>	3.55E-53	AAO85498.1

accession numbers of GT222067-GT226792, and GW811622-GW813164. A total of 4,815 unique sequences were identified including 867 contigs and 3,948 singletons (Table 1).

From the 867 contigs, 5,228 putative SNPs were identified that included 2,243 transitions, 1,760 transversions, and 1,225 indels (Table 1). These SNPs represented a rate of 7.7 SNP per kilo base pairs. The majority of the putative SNPs were identified from contigs with only 2 sequences (56.1%) or 3 sequences (28.7%). In order to

make these useful as molecular markers, additional validation is required because of the low sequence redundancy. A total of 51 SNPs were identified with the presence of minor allele frequency at least twice from contigs containing four or more sequences. These SNPs should have a greater success rate during validation.<sup>22</sup>

A total of 111 microsatellites were identified from 105 unique sequences, with 88 have sufficient flanking sequences for primer design, and are therefore potentially useful for genetic linkage mapping (Table 2). Thirty-five microsatellites-containing ESTs have significant hits to the nr database and, therefore, represent known genes. These microsatellites need to be tested for their polymorphism and practical utilities.<sup>23</sup>

Of the 4,815 unique sequences, 2,985 had significant hits ( $E < -5$ ) with BLASTX searches. Of the genes with putative identities, 52 unique genes were related to immune or defense responses (Table 3). Of these, only three genes, peroxisome proliferator activated receptor gamma, glucose phosphate isomerase, and tapasin, have been identified previously in grass carp. These immune and defense response genes are of particular interest, and their further study is warranted because of their involvement in immunity functions. Farmed grass carp are more susceptible to various diseases, including hemorrhagic disease, bacterial septicemia, bacterial enteritis, and bacterial gill-rot disease.<sup>2,3,24,25</sup> The immune-related gene, tapasin, was shown to be responsive to infections in parasitic copepods.<sup>26</sup>

A total of 2,197 unique sequences were assigned at least one Gene Ontology (GO) category term using BLAST2GO. Cellular Process (78.6%) was the most dominant 2nd level term of Biological Process out of the 1,169 unique sequences. Binding (68.6%) was the most dominant out of 1,349 ESTs with Molecular Function category at 2nd level. Cell Part (96.5%) was the most dominant out of 996 ESTs in the Cellular Component GO category (data not shown).

In summary, a relatively large number of ESTs were generated from this work, from which type I molecular markers can be developed.<sup>15</sup> As genes collectively account for phenotypes,<sup>27</sup> gene associated markers are preferable. In addition to functional utilities, sequence-tagged type I markers are also more useful for comparative genome studies including synteny discovery between fish genomes.<sup>28</sup> The microsatellites and SNPs identified from the ESTs, therefore, should be useful genome resources for genetic analysis of grass carp.

## REFERENCES

1. Cudmore B, Mandrak NE. Biological synopsis of grass carp (*Ctenopharyngodon idella*). *Can Manus Rep Fish Aquatic Sci.* 2705 2004;V + 44p.
2. Zhong J, Wang Y, Zhu Z. Introduction of the human lactoferrin gene into grass carp (*Ctenopharyngodon idellus*) to increase resistance against GCH virus. *Aquaculture.* 2002;214:93–101.
3. Zhang Q, Ruan H, Li Z, Zhang J, Gui J. Detection of grass carp hemorrhage virus (GCHV) from Vietnam and comparison with GCHV strain from China. *High Technol Lett.* 2003;9:7–13.
4. Ding W, Cao L, Cao Z. Isolation, Cloning and sequencing for SRAP markers related to germ degeneration in *Ctenopharyngodon idellus*. *J Guangdong Ocean Univ.* 2007;27: 13–17.

5. Stott B, Robson T. Efficiency of grass carp (*Ctenopharyngodon idella* Val.) in controlling submerged water weeds. *Nature*. 1970;226:870.
6. Raibley PT, Blodgett D, Sparks RE. Evidence of grass carp (*Ctenopharyngodon idella*) reproduction in the Illinois and upper Mississippi Rivers. *J Freshwater Ecol*. 1995;10: 65–74.
7. Clark MS, Edwards YJ, Peterson D, et al. Fugu ESTs: New resources for transcription analysis and genome annotation. *Genome Res*. 2003;13:2747–2753.
8. Rise ML, von Schalburg KR, Brown GD, et al. Development and application of a salmonid EST database and cDNA microarray: data mining and inter-specific hybridization characteristics. *Genome Res*. 2004;14:478–490.
9. Rexroad CE 3rd, Lee Y, Keele JW, et al. Sequence analysis of a rainbow trout cDNA library and creation of a gene index. *Cytogenet. Genome Res*. 2003;102:347–354.
10. Cao D, Kocabas A, Ju Z, et al. Transcriptome of channel catfish (*Ictalurus punctatus*): Initial analysis of genes and expression profiles of the head kidney. *Anim Genet*. 2001;32: 169–188.
11. Ju Z, Karsi A, Kocabas A, et al. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. *Gene*. 2000;261:373–382.
12. Kocabas AM, Kucuktas H, Dunham RA, Liu Z. Molecular characterization and differential expression of the myostatin gene in channel catfish (*Ictalurus punctatus*). *Biochim Biophys Acta*. 2002;1575:99–107.
13. Karsi A, Cao D, Li P, et al. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin. *Gene*. 2002;285:157–168.
14. Li P, Peatman E, Wang S, et al. Towards the ictalurid catfish transcriptome: Generation and analysis of 31,215 catfish ESTs. *BMC Genomics*. 2007;8:177.
15. Wang S, Peatman E, Abernathy J, et al. Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biol*. 2010;11(1):R8.
16. Li JL, Zhu ZY, Wang GL, Bai ZY, Yue GH. Isolation and characterization of 17 polymorphic microsatellites in grass carp. *Mol Ecol Notes*. 2007;7:1114–1116.
17. Bao B, Peatman E, Xu P, et al. The catfish liver-expressed antimicrobial peptide 2 (LEAP-2) gene is expressed in a wide range of tissues and developmentally regulated. *Mol Immunol*. 2006;43:367–377.
18. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21:3674–3676.
19. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: Tool for the unification of biology. *Nat Genet*. 2000;25:25–29.
20. Quilang J, Wang S, Li P, et al. Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers. *BMC Genomics*. 2007;8:157.
21. Barker G, Batley J, H OS, Edwards KJ, Edwards D. Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics* (Oxford, England) 2003;19:421–422.
22. Wang S, Sha Z, Sonstegard TS, et al. Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics*. 2008;9:450.
23. Serapion J, Kucuktas H, Feng J, Liu ZJ. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar Biotechnol*. 2004;6:364–377.



24. Wang WS, Wang DH. Enhancement of the resistance of tilapia and grass carp to experimental *Aeromonas hydrophila* and *Edwardsiella tarda* infections by several polysaccharides. *Comp Immunol Microbiol Infect Dis.* 1997;20:261–270.
25. Xu D, Wei H, Zhang W. The therapeutic effect of onion on the pathogenic bacteria of gill rot disease in grass carp. *Ag Sci Technol.* 2008;9:117–120.
26. Chang MX, Nie P, Liu GY, Song Y, Gao Q. Identification of immune genes in grass carp *Ctenopharyngodon idella* in response to infection of the parasitic copepod. *Sinergasilus major.* *Parasitol Res.* 2005;96:224–229.
27. Eujayl I, Sorrells ME, Wolters P, Powell W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor Appl Genet.* 2004;104:399–407.
28. Shimizu N, Sasaki T, Asakawa S, et al. Comparative genomics of medaka and fugu. *Comp Biochem Physiol D.* 2006;1:6–12.