

# Characterization of Expressed Sequence Tags from Liver and Muscle Tissues of Walking Catfish, *Clarias macrocephalus*

Dutrudi Panprommin,<sup>a</sup> Supawadee Poompuang<sup>b\*</sup> and Prapansak Srisapome<sup>b</sup>

<sup>a</sup> Center for Agricultural Biotechnology, Kasetsart University, Nakhon Pathom 73140.

<sup>b</sup> Department of Aquaculture, Faculty of Fisheries, Kasetsart University, Bangkok 10900.

\* Corresponding author, E-mail: supawadee.p@ku.ac.th

Received 15 Mar 2007

Accepted 15 Aug 2007

**ABSTRACT:** Expressed sequence tags have been generated from cDNA libraries constructed from liver and muscle tissues of female walking catfish. Two thousand and twenty-nine randomly picked cDNA clones, 991 from the liver library and 1,038 from the muscle library were sequenced. A total of 1,334 EST clones showed significant sequence similarity to known genes in the databases, representing 303 genes from the liver library and 234 genes from the muscle library. Fifty-one full-length genes and 95 microsatellite repeat sequences were identified in both libraries. The majority of walking catfish EST sequences matched sequences identified from other Siluriformes, particularly channel catfish (513 EST clones, 38.5%) and zebrafish (284 EST clones, 21.3%). A number of identified genes appeared to be expressed in specific tissues. Vitellogenins were the most highly expressed genes in the liver of female walking catfish. Further, genes responsible for innate immune function of fish were found only in the liver. In contrast, genes encoding structural proteins were restricted to the muscle library. Analysis of the cDNA libraries indicates that EST approaches can provide effective ways to characterize expressed genes in walking catfish, and these libraries would be useful resources for identification of EST and microsatellite markers for mapping purposes.

**KEYWORDS:** *Clarias macrocephalus*, expressed sequence tags, liver, muscle.

## INTRODUCTION

Expressed sequence tags (ESTs) are partial sequences of mRNA molecules generated from complementary DNA library clones. The EST information is particularly useful for gene expression studies and genome mapping. As Type I markers or markers with known functions, ESTs are useful for comparative genome mapping between species.<sup>1,2</sup> Furthermore, ESTs can also provide candidate genes for mapping of quantitative trait loci.<sup>3</sup> From the aquaculture point of view, EST markers can be used in marker-assisted breeding programs for improvement of broodstock.

Large-scale efforts to characterize ESTs have focused on model fish species, particularly pufferfish<sup>4</sup> and zebrafish<sup>5,6</sup> for studies of vertebrate development. The EST approach also has been extensively used for analysis of expressed genes in various tissues from species of importance to world aquaculture, i.e., Atlantic salmon,<sup>7-9</sup> channel catfish,<sup>1,10</sup> rainbow trout,<sup>11</sup> and tilapia.<sup>12</sup>

Walking catfish (*Clarias macrocephalus*, order Siluriformes) is one important food fish species in the aquaculture industry in Thailand. The production of

this species, however, is limited by its slow growth rate and disease susceptibility. Selective breeding for growth traits has been practiced with slight improvement.<sup>13</sup> Molecular approaches can be combined with classical selection for genetic improvement of walking catfish. The application of genetic markers in the selection program for this fish requires information of its genome and a high-resolution linkage map, which is essential to mapping of quantitative trait loci.<sup>2</sup> Although the first generation of the walking catfish linkage map is available,<sup>14</sup> the resolution of the map is low. Moreover, a search of the GenBank database revealed only approximately 70 mRNA or gene sequences for *Clarias* spp (as of July, 2007).

To generate EST information for walking catfish tissues, we sequenced 2,029 clones from liver and muscle tissues of adult female walking catfish. The liver is the body's central metabolic factory, and is responsible for three major functions, metabolism, detoxification and synthesis of plasma protein. Therefore, a large number of gene transcripts would be expected in the liver library. In fish, muscle accounts for 45-50% of the body weight. Muscle growth and development directly affect growth of the animal. Because muscle is also an important trait in commercial fish farming, it is useful

to identify genes associated with muscle tissues for molecular breeding purposes. ESTs identified in walking catfish should be useful resources for genome studies in clariid catfishes and closely related species, e.g., pangasiidae. In this paper, we report the identification of ESTs in liver and muscle tissues of adult female walking catfish.

## MATERIALS AND METHODS

### Construction of cDNA Libraries

Liver and skeletal tissues were collected from an adult female catfish (180 g). Total RNA was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The mRNA was isolated using the Quick Prep<sup>®</sup> Micro mRNA Purification Kit (Amersham Pharmacia Biotech, Uppsala, Sweden) according to the manufacturer's instructions. Approximately 5 mg each of liver and muscle mRNA was used for cDNA synthesis with the ZAP-cDNA synthesis kit (Stratagene, La Jolla, CA, USA). The synthesized cDNA was ligated into the Uni-ZAP XR insertion vector and subsequently packaged in Gigapack Gold III packaging extract (Stratagene). The libraries were amplified before *in vivo* excision. pBluescript phagemids were excised from the Uni-ZAP XR vector by co-infecting the amplified libraries and the ExAssist helper phage (Stratagene, USA) into *E. coli* XL-1 BLUE MRF<sup>+</sup> cells. The excised pBluescript phagemids were used to infect the *E. coli* SOLR cells and plated onto LB-ampicillin agar plates.

Random colonies were picked and grown overnight in LB-ampicillin broth. Colony PCR was performed using 0.5 ml of LB containing isolated colonies with the M13 forward and reverse primers to determine the insert size. The PCR profiles were: initial denaturation at 95°C for 5 min, and then 95°C for 30 s, 55°C for 30 s, and 72°C for 90 s for 30 cycles. PCR reaction products were separated by 1% agarose gel electrophoresis and were visualized by UV transillumination in the presence of ethidium bromide.

Plasmid DNA extraction was performed using QIAprep Miniprep (QIAGEN) according to the manufacturer's instructions. Purified DNA fragments with size greater than 300 bp were 5'-sequenced by the MacroGen Laboratory (Seoul, Korea) with an automated sequencer (ABI 3730 XL) and M13 reverse primer.

### Bioinformatics

The DNA sequences were checked for quality. Vector and polylinker sequences were removed with GENETYX version 7.0. The sequencing data was compared for sequence similarity with the available sequences in the GenBank data base (<http://www.ncbi.nlm.nih.gov>) using the BLASTN (for nucleotide similarity) and BLASTX

(for possible protein similarity) programs.<sup>15</sup> Matches were considered significant when *E* values were less than 10<sup>-4</sup> and there was a match of greater than 100 nucleotides for BLASTN and more than 10 amino acid residues matching for BLASTX. The EST sequences were classified into twelve functional categories<sup>16</sup> and were submitted to the GenBank dbEST database. All sequences were searched for microsatellite repeats with Sputnik 2 computer software (<http://espressoftware.com/pages/sputnik.jsp>).

## RESULTS

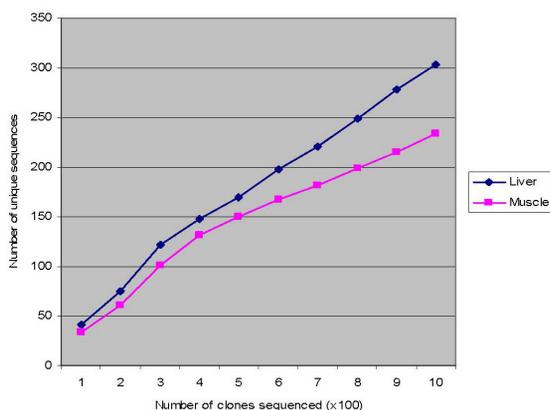
### Summary of EST Clones

A total of 2.6x10<sup>5</sup> pfu was obtained for the primary liver library and 1.6x10<sup>5</sup> pfu was obtained for the primary muscle library. The liver and muscle libraries contained 1,152 and 1,248 positive clones, respectively. After construction, the average size of the cDNA inserts was determined by PCR analysis from all clones. A total of 2,029 clones (991 from liver and 1,038 from muscle) with insert sizes between 300 to 3,000 bp were sequenced from the 5' region. Because the principal objective of this work was gene expression analysis, all of the EST clones were sequenced from the 5' end of the inserts. The 5' end of each clone is more likely to contain protein-coding sequence than the 3' ends, which often contain untranslated regions (UTRs). The average lengths of EST sequences were 560 and 545 bp for the liver and muscle libraries, respectively. All sequences have been deposited in the GenBank dbEST database under accession numbers EB 359601- EB 360662 and EG 631397-EG632363. A complete list of identified ESTs is available upon request.

To evaluate the redundancy of the identified clones, we determined the number of novel genes for every 100 clones sequenced in each library (Fig. 1). The number of new genes discovered in the first 100 randomly selected clones was 41 for the liver library to 34 for the muscle library. As the number of sequenced

**Table 1.** cDNA libraries from walking catfish liver and muscle tissues.

	Liver library	Muscle library	Total
Number of sequenced clones	991	1,038	2,029
Number of unmatched clones	345	350	695
Matched clones	646	688	1,334
<i>Clarias</i> spp.	4	3	7
Other Siluriformes	188	325	513
Zebrafish	148	146	294
Other teleost fish	226	131	357
Other Chordata	57	55	112
Other animals	23	28	51
Number of full-length genes	34	17	51



**Fig 1.** Number of unique sequences plotted against the total number of clones sequenced.

clones increased, the percent of new gene discovery gradually decreased for both libraries. For the first 500 clones, the liver library contained 170 novel genes, while the muscle library contained 150 new genes. However, estimates of the number of new genes discovered suggested that approximately 343 and 307 unique sequences would be identified when 1,000 clones had been sequenced in both libraries. The actual numbers of new genes were lower than the estimates, 303 and 243 from the liver and muscle libraries, respectively. The liver library appeared to be more diverse than the muscle library. The redundancy factor was calculated as the ratio between number of clones and number of genes for each functional category (Table 2). The redundancy factor is used to measure the frequency of repeated EST sequences for all categories of genes.<sup>17</sup>

Thirty-four and 17 full-length gene sequences were identified in the liver and muscle libraries (Tables 1 and

3). The libraries showed some contamination with sequences of ribosomal RNA genes and mitochondrial genes. In the liver library, 11 clones (1.1%) sequenced were ribosomal RNA genes and 52 clones (5.25%) were mitochondrial genes. The muscle library was less contaminated with 0.38% rRNA and 3.75% mitochondrial genes.

**Identification of cDNA Clones in the Liver Library**

A total of 991 clones were sequenced and 34.8% or 345 clones were unidentified or novel ESTs. Of 646 matched ESTs, a total of 303 encoding mRNA transcripts were identified (Tables 1 and 2). A homology search revealed that only four sequences had previously been described from clariid catfish (the cytochrome b gene in *C. fuscus*; two ribosomal RNA genes in *C. gariepinus* and *C. ngamensis*; and a *C. macrocephalus* microsatellite *Cma-42\** sequence). One hundred eighty-eight clones (29.1%) of the walking catfish ESTs showed significant sequence similarity to previously described genes from other fish of the order Siluriformes, particularly channel catfish. One hundred forty-eight clones (22.9%) matched genes from zebrafish. A large number of the walking catfish ESTs, 226 clones (34.98%), matched genes from other teleost fishes including carp and salmonids, while 57 clones (8.8%) matched genes from other Chordata.

The dominant functional class of messages from the liver library was that of signaling and communication, in which 204 clones (20.58%) represented only 37 genes with redundancy factor of 5.5. One hundred and eleven ESTs (11.2%) assigned to the category of defense and homeostasis encoded 47 different genes. Eighty-seven clones (8.8%) represented 52 different genes encoding ribosomal proteins. Fifty-nine ESTs (5.9%) encoded 28 metabolism-related proteins. Another 13 clones (1.11%),

**Table 2.** Summary of sequenced clones, number of genes, redundancy factor (RF) and functional classification of genes from the walking catfish EST libraries.

Classification	Liver			Muscle		
	Number of clones	Number of genes	RF	Number of clones	Number of genes	RF
Gene expression and protein synthesis	14	13	1.1	14	13	1.1
Internal/external structure and motility	13	6	2.2	338	51	6.6
Metabolism	59	28	2.1	43	28	1.5
Defense and homeostasis	111	47	2.4	127	17	7.5
Signaling and communication	204	37	5.5	7	6	1.2
Cell division/DNA synthesis	6	6	1.0	3	3	1.0
Ribosomal protein	87	52	1.7	61	41	1.5
Mitochondrial protein and rRNA	63	41	1.5	43	30	1.4
Transporter	12	10	1.2	9	8	1.1
Miscellaneous function	17	15	1.1	13	12	1.1
Unidentified-hypothetical protein	60	48	1.3	30	25	1.2
Subtotal	646	303	-	688	234	-
Unknown	345	-	-	350	-	-
Total	991	-	-	1,038	-	-

**Table 3.** List of 30 identified walking catfish EST clones containing full-length sequences from liver and muscle tissues.

Walking catfish ESTs			Matching sequences			
Accession no.	Gene name	Category	Species	Accession no.	E-value	Identity (%)
<b>Liver</b>						
EB360377	nexin 12	2	<i>Ictalurus punctatus</i>	DQ086172	5E-153	93
EB360474	14 kDa apolipoprotein	3	<i>Carassius auratus gibelio</i>	AAW82445	2E-24	42
EB360248	endozepine	3	<i>Cyprinus carpio</i>	AAT00460	4E-34	80
EB360283	14 kDa apolipoprotein	3	<i>Danio rerio</i>	XP_698979	1E-31	50
EB360653	Metallothionein	4	<i>Ictalurus punctatus</i>	AF087935	1E-20	86
EB360634	fatty acid-binding protein, liver (L-FABP) (Liver basic FABP) (LB-FABP)	5	<i>Rhamdia sapo</i>	P80856	9E-56	85
EG631562	fatty acid binding protein 1b	5	<i>Danio rerio</i>	NP_001019822	2E-46	69
EG631653	zinc finger, CSL domain containing 2	6	<i>Danio rerio</i>	XP_698797	4E-26	91
EB360335	40S ribosomal protein S21	7	<i>Ictalurus punctatus</i>	AF402830	5E-116	91
EB360348	ribosomal protein L27	7	<i>Ictalurus punctatus</i>	AF401581	0	95
EB360387	ribosomal protein L37a	7	<i>Ictalurus punctatus</i>	AF401594	1E-179	95
EB360389	ribosomal protein L10a	7	<i>Ictalurus punctatus</i>	AF401564	0	93
EB360466	40S ribosomal protein S7	7	<i>Ictalurus punctatus</i>	AF402815	0	94
EB360477	40S ribosomal protein S16	7	<i>Ictalurus punctatus</i>	AF402825	0	96
EB360500	40S ribosomal protein S27-2	7	<i>Ictalurus punctatus</i>	AF402837	1E-154	90
EB360440	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1 (7.5kD, MWFE)	8	<i>Mus musculus</i>	NP_062316	1E-26	75
EB360356	hemoglobin beta chain	9	<i>Silurus asotus</i>	S83540	1E-47	85
EB360253	dexamethasone-induced protein	9	<i>Mus musculus</i>	NP_067403	4E-27	83
<b>Muscle</b>						
EB359774	U6 snRNA-associated Sm-like protein LSm6 (LOC573495)	1	<i>Danio rerio</i>	XM_689037	3E-48	86
EB359618	parvalbumin	2	<i>Ictalurus punctatus</i>	AF227795	0	94
EB359698	parvalbumin	2	<i>Danio rerio</i>	AAH76256	4E-36	85
EB359815	parvalbumin isoform 4a	2	<i>Danio rerio</i>	AAO33398	1E-35	71
EB359991	parvalbumin	2	<i>Cyprinus carpio</i>	AJ292211	1E-72	85
EB360100	troponin C	2	<i>Ictalurus punctatus</i>	AF227801	0	93
AF227801	myosin light chain 3	2	<i>Danio rerio</i>	AB042028	5E-97	88
EG631912	parvalbumin 7	2	<i>Danio rerio</i>	AAH76256	1E-35	85
EG632270	troponin I type 2 (skeletal, fast)	2	<i>Danio rerio</i>	AAH71462	5E-59	74
EB359739	40S ribosomal protein S1	7	<i>Ictalurus punctatus</i>	AF402820	5E-170	92
EB359905	ribosomal protein L37a	7	<i>Ictalurus punctatus</i>	AF401594	0	95
EB359989	ribosomal protein L37	7	<i>Ictalurus punctatus</i>	AF401593	5E-132	93

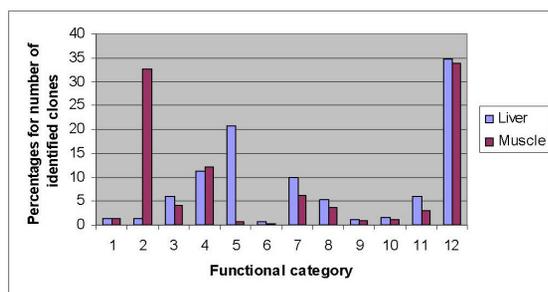
Functional categories include: (1) gene expression, regulation and protein synthesis, (2) internal/external structure and motility, (3) metabolism, (4) defense and homeostasis, (5) signaling and communication, (6) cell division/DNA synthesis, repair and replication, (7) ribosomal protein and rRNA, (8) mitochondrial protein, (9) transporter.

which encoded 6 different proteins, were categorized as structure and motility.

Further analysis of the EST sequences revealed that genes associated with primary functions of liver were well represented in the walking catfish liver cDNA library (Fig. 2). One hundred and twelve of 303 (37%) identified genes encoded proteins associated with metabolism (apolipoprotein, gastric lipase and peroxiredoxin), defense and homeostasis (creatine kinase, fibrinogen, serine proteinase inhibitors, thrombin, heat shock proteins, transferrin, and complement C3), and signaling and communication (vitellogenin and fatty acid binding protein). Only ten transporter genes (0.3%) were identified followed by six genes (0.02%) responsible for structure and motility.

### Identification of cDNA Clones in Muscle Library

Of 1,038 clones sequenced, 688 matched 234 genes in the databases and 350 (33.7%) were unidentified (Tables 1 and 2). The majority of identified ESTs (325 clones, 47.23%) showed significant sequence similarities to genes previously identified in other fish of the order Siluriformes. One hundred and forty-six ESTs (21.22%) matched genes from zebrafish and 131 clones (19%) matched genes from other teleost fish. Fifty-five clones (8%) showed sequence similarities to those in other Chordata, while 28 (4.1%) clones matched genes from invertebrates. Only three EST clones matched sequences or genes of clariid catfish (an 18S rRNA gene from *C. camerunensis*, thyroid-stimulating hormone receptor from *C. gariepinus*, and mitochondrial control region from *C. fuscus*).



**Fig 2.** Distribution of ESTs by functional categories. Functional categories include: (1) gene expression, regulation and protein synthesis, (2) internal/external structure and motility, (3) metabolism, (4) defense and homeostasis, (5) signaling and communication, (6) cell division/DNA synthesis, repair and replication, (7) ribosomal protein and rRNA, (8) mitochondrial protein, (9) transporter, (10) miscellaneous function, (11) unidentified-hypothetical protein, and (12) unclassified.

ESTs represented internal/external structure and motility, were highly expressed in the muscle library, of which a total of 338 clones (32.56%) encoded 51 genes (redundancy factor of 6.6) followed by 127 clones (12.23%) represented 17 genes acting in defense and homeostasis. Forty-three clones (4.14%) corresponded to 28 different genes involved in metabolism. Sixty-one clones (5.8%) encoded 41 ribosomal proteins. Only seven clones in the muscle library represented six genes involved in signaling and communication.

Characterization of the ESTs revealed that genes associated with fish muscle growth and development, were relatively highly expressed in walking catfish muscle. Of 234 identified genes, 51 (21.8%) encoded proteins of structure and motility (actin, myosin, troponin, and parvalbumin), followed by 41 genes (17.5%) that encoded ribosomal proteins and 28 proteins (12%) involved in metabolism.

#### Identification of Microsatellites within EST Sequences

Of 2,029 sequenced clones, 89 (4.4%) contained 95 microsatellite repeat sequences. The majority of microsatellites were di- and tri- nucleotide perfect repeats (70.5%). Of the 95 microsatellite sequences identified, 15 were within the open reading frames (ORFs) and the remaining sequences were found in the 3' UTRs. Fifty-five microsatellite sequences were associated with known genes, including vitellogenin, heat shock protein, actin, myosin, and troponin (Table 4).

## DISCUSSION

A good EST resources should contain as many tissues at different developmental stages as possible.<sup>16</sup> It should be noted, however, that the walking catfish EST project

**Table 4.** Microsatellite sequences identified from walking catfish EST clones.

Accession no.	Gene name	Microsatellite sequence
<b>Liver library</b>		
EB360233	fibrinogen, B beta polypeptide (f <sub>g</sub> b)	(AC) <sub>18</sub>
EB360268	glucose transporter 2	(AAT) <sub>4</sub>
EG631840	vitellogenin	(AAT) <sub>4</sub>
EG631442	zinc finger protein 118	(AGC) <sub>4</sub>
EG631442	zinc finger protein 118	(AAT) <sub>5</sub>
EB360659	vitellogenin	(GCA) <sub>4</sub>
EB360661	heat shock 90kDa protein 1 beta isoform b	(AGG) <sub>4</sub>
<b>Muscle library</b>		
EG359605	actin	(TCC) <sub>4</sub>
EB359688	slow troponin T 2 (sTnT2)	(TC) <sub>6</sub>
EB359729	myosin light chain 1	(AC) <sub>11</sub>
EG632197	troponin T	(AG) <sub>4</sub>
EG632197	troponin T	(AAG) <sub>4</sub>
EB360094	troponin C, TnC	(AC) <sub>5</sub>
EG632266	fast skeletal myosin light chain 1a	(AC) <sub>11</sub>
EB360110	myosin heavy chain	(AAAC) <sub>3</sub>

was small scale compared to EST projects of other aquaculture species. Although different tissue expression patterns between sexes should be demonstrated in walking catfish, only adult female was chosen because it provided useful EST resource for expression analysis of liver cells. For example, liver genes are involved in vitellogenesis during the ovarian cycle of mature females, while in males, these genes are not expressed.

The quality of cDNA libraries depends on the representation of gene diversity and the length of the inserts. A library should contain a minimum of 10<sup>5</sup> primary cDNA clones to provide a 99% probability that a rare cDNA is present in at least one clone<sup>7</sup>. Based on these measures, the liver and muscle libraries appeared to have a large diversity of genes and contained sufficient numbers of clones to cover low abundance transcripts. In addition, these libraries contained an average insert of about 750 bp, increasing the likelihood that each clone would contain coding region. Database searches of 2,029 clones, however, resulted in the putative identification of 538 walking catfish genes, in which 51 full-length genes were identified. The redundancy of clones in each category was due to the high expression of genes present in a particular tissue or at a limited period of time. Methods of normalization<sup>6</sup> and differential screening of cDNA libraries<sup>18</sup> have been developed to minimize redundancy of ESTs. Because the objective of this work was to obtain an overview of gene diversity and expression profiles in walking catfish tissues, we used non-normalized cDNA libraries to generate ESTs.

Of the 1,334 matched clones, nearly half of the EST sequences from both libraries (651 clones, 48.8%) showed significant sequence similarities to genes previously identified from other species of teleost fish, including zebrafish. The zebrafish ESTs represent large-scale EST projects and over 1,350,105 entries have been deposited in the public database (as of July, 2007). Approximately 38.5% of walking catfish ESTs (513 clones) showed similarities to sequences from other siluriformes. Among fishes of order Siluriformes, only the ESTs from channel catfish have been extensively identified with 44,767 entries (as of July, 2007). Further analysis showed that 112 clones (8.4%) matched genes previously identified from other Chordata including, humans, chicken, mouse, and pig. The remaining 51 clones (3.8%) showed sequence similarity to those of invertebrates, including insects. Despite the rapid progress in gene discovery in many species, a high percent of unidentified genes in the liver (345 clones, 34.81%) and muscle libraries (350 clones, 33.71%), representing putative novel genes, was found in walking catfish.

The liver and muscle libraries showed different expression patterns, indicating different functional activities of these tissues in walking catfish. A number of genes appeared to be highly expressed in a tissue-specific manner. Genes encoding proteins for signaling and communication (vitellogenins), defense (immune related genes) and homeostasis (metallothioneins) were expressed only in the liver. In contrast, a high proportion of genes responsible for internal/external structure and motility was found in the muscle library. The expression of genes involved in internal/external structure and motility, including actin, myosin light chain, tropomyosin, parvalbumin and troponin, were observed only in the muscle library. However, the common housekeeping genes encoding heat shock proteins, ribosomal proteins, translation factors and metabolic enzymes, which are essential to cell viability and expressed in all cells, were nearly equally expressed in both libraries.

Vitellogenin genes (Vtg) were highly expressed in the liver library of female walking catfish. Vitellogenin is an egg yolk precursor secreted in the liver of female fish in the reproductive stage. Expression of vitellogenins is under the control of estrogen and is tissue-dependent. During oocyte development, the deposition of vitellogenin results in the increase in oocyte size. Expression analysis of vitellogenin in females under normal conditions can be used to determine the stage of oocyte development and duration of ovarian cycle.<sup>19</sup> The high level of vitellogenin transcript presented in the liver library was not unexpected, because it was constructed from liver

tissue of adult female catfish. Vitellogenins are encoded by a multigene families, for example, 20 genes were identified in rainbow trout<sup>20</sup> and seven in zebrafish.<sup>21</sup> Five vitellogenin genes were identified in walking catfish, Vtg, Vtg1, Vtg2, Vtg3, and Vtg6. The vitellogenin genes identified in walking catfish share significant identity to those from common carp and zebrafish.

The defense genes found only in the liver of walking catfish were those responsible for acute phase response (APR) of fish, including transferrin, complement C3-H1, and pentraxin. APR is a physiological response of the body to injury, trauma and infection and hepatocytes are the major source of these acute phase proteins.<sup>22</sup> The plasma levels of these proteins can be used as a health parameter. Expression level of these immune genes in the liver of walking catfish was relatively low indicating that the fish was not under stress conditions.

Transferrin, a plasma protein capable of binding iron was identified in the order siluriform for the first time. The deduced amino acid sequences of walking catfish transferrin showed highest similarity to those of red seabream, brown trout, European flounder and Japanese medaka. Complement component C3 is a gene involved in the complement system of innate immunity. The complement component C3 of walking catfish had significant similarity of sequence to that of common carp. In this study, four different proteins were identified including complement C3-H1, complement component 8, complement factor H precursor and complement control protein factor I-B. The complement components of walking catfish had significant similarity of sequence to those of common carp and zebrafish. Pentraxin is protein capable of activating complements, opsonizing bacteria, fungi, and parasites, and agglutinating particles. The pentraxin gene identified in walking catfish showed significant homology to Atlantic salmon.

The other gene found to be specific to liver tissues was metallothionein, a superfamily of ubiquitously expressed metal-binding proteins that can be induced by metal exposure, oxidative stress, and immune challenge. Metallothioneins play important roles in the homeostasis, detoxification and stress response of metals. A clone containing the complete coding sequence of metallothionein was identified in the present study. Further, we identified all three major families of heat shock proteins: Hsp90 (85–90 kDa), Hsp70 (68–73 kDa), and low molecular weight heat shock proteins Hsp 25 (16–47 kDa)<sup>23</sup> in the liver library. Heat shock proteins are highly conserved cellular proteins produced in response to a severe heat stress.

A large number of genes involved in muscle contraction and development were found only in the

muscle library of walking catfish. Among them, actin, myosin light chain, tropomyosin, parvalbumin and troponin were relatively highly expressed in the muscle of adult female catfish. Actin is one of the most highly conserved proteins in vertebrates. Myosin light chain and tropomyosin are genes encoding for myofibrillar proteins. Parvalbumins and troponins are calcium-binding proteins abundant in the white-muscle of fish, which act as soluble muscle-relaxing factors in active muscle contraction. Adult fish display different isotype characteristics of these muscle genes.<sup>24</sup> Most of muscle genes in this category showed sequence similarity to channel catfish, zebrafish, carp, and tuna.

Association of microsatellite sequences with ESTs has been reported in various species of fish, for example, channel catfish<sup>1</sup>, common carp<sup>25</sup> and Atlantic salmon.<sup>26</sup> EST sequences of fish tissues are found to contain microsatellite DNA, particularly in the non-coding sequence. Because mutation occurs more frequently in non-coding sequences of the genome than those in coding sequences, it is more likely that microsatellites found within non-coding regions of ESTs will be polymorphic. Isolation of microsatellite-associated ESTs can provide an effective way to develop polymorphic EST markers.<sup>1</sup> The microsatellites within the ESTs of walking catfish were abundant and most of them were found in the 3' UTRs. We identified 95 microsatellite repeat sequences, of which 55 were associated with known genes. Polymorphism of these microsatellite sequences has the potential to be utilized as markers for genetic linkage mapping. Additionally, microsatellites associated with specific genes of walking catfish are useful for quantitative trait (QTL) mapping, in which they can provide candidate genes for production traits (such as growth rate and disease resistance).

We have added 2,029 EST sequences to the public databases, increasing the number of clariid catfish gene sequences from approximately 70 to 592 genes. However, 695 of these walking catfish ESTs were unidentified. Although the characterization of walking catfish ESTs is based on only 537 identified genes, it provides baseline information on the number of genes involved in the biological functions of liver and muscle tissues and expression profiles of these genes. Moreover, these ESTs provide a useful resource for potential markers to increase the resolution of the linkage map and mapping of QTLs in walking catfish. Because a large number of walking catfish ESTs were unidentified, the classification of genes in the present study should be considered preliminary. Nevertheless, as more is known about functional roles of genes and proteins, it is expected that the number of walking catfish clones matching genes in databases will increase in the future.

## ACKNOWLEDGEMENTS

We thank Assoc. Prof. Dr. Nontawith Areechon and Prof. Dr. Anchalee Tassanakajon, for laboratory facilities. Comment from two anonymous reviewers is also appreciated. This work was supported by the Thailand Research Fund (TRF Grant # RSA4780004), Kasetsart University Development and Research Institute (KURDI), the Graduate School, and Center for Agricultural Biotechnology (CAB), Kasetsart University.

## REFERENCES

1. Liu Z, Karsi A and Dunham, RA (1999) Development of polymorphic EST markers suitable for genetic linkage mapping of catfish. *Mar Biotechnol* **1**, 437-447.
2. Liu Z and Cordes J F (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1-37.
3. Grosse WM, Kappes S M and McGraw RA (2000) Linkage mapping and comparative analysis of bovine expressed sequence tags (ESTs). *Animal Genetics* **31**, 171-7.
4. Toramoto R, Ikeda D, Ochiai Y, Minoshima S, Shimizu N and Watabe S (2004) Multiple gene organization of pufferfish *Fugu rubripes* tropomyosin isoforms and tissue distribution of their transcripts. *Gene* **331**, 41-51.
5. Zeng S, and Gong Z (2002) Expressed sequence tag analysis of expression profiles of zebrafish testis and ovary. *Gene* **294**, 45-53.
6. Lo J, Lee S, Xu M, Liu F, Ruan H, Eun A, He Y, Ma W, Wang W, Wen Z, and Peng J (2003) 15,000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. *Genome Res* **13**, 455-66.
7. Davey GC, Caplice NC, Martin SA and Powell R (2001) A survey of genes in the Atlantic salmon (*Salmo salar*) as identified by expressed sequence tags. *Gene* **263**, 121-30.
8. Martin S A, Caplice NC, Davey GC and Powell R (2002) EST-based identification of gene expressed in the liver of adult Atlantic salmon (*Salmo salar*). *Biochem Biophys Res Commun* **293**, 578-85.
9. Rise ML, et al (2004) Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res* **4**, 478-90.
10. He C, Chen L, Simmons M, Li P, Kim S, and Liu ZJ (2003) Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Animal Genetics* **34**, 445-8.
11. Rexroad C, Lee Y, and Keele JW (2003) Sequence analysis of a rainbow trout cDNA library and creation of gene index. *Cytogenet Genome Res* **102**, 347-54.
12. Chu SL, Weng CF, Hsiao CD, Hwang PP, Chen YC, Ho JM, and Lee SJ (2006) Profile analysis of expressed sequence tags derived from the ovary of tilapia, *Oreochromis mossambicus*. *Aquaculture* **251**, 537-48.
13. Jarimopas P, Niyomkitsumlit A, Kumnanee A, and Wongchang S (1988) Preliminary study on mass selection of *Clarias macrocephalus* Guther for growth. *Technical Paper No. 88*. National Inland Fisheries Institute, Bangkok.
14. Poompuang S, and Na-Nakorn U (2004) A preliminary genetic map of walking catfish (*Clarias macrocephalus*). *Aquaculture* **232**, 195-203.

15. Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403-10.
16. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, et al (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651-6.
17. Ju Z, Karsi A, Kocabas A, Patterson A, Li P, Cao D, Dunham R and Liu Z (2000) Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. *Gene* **261**: 373-82.
18. Ponsuksili S, Wimmers K and Schellander K (2001) Application of differential display RT-PCR to identify porcine liver ESTs. *Gene* **280**, 75-85.
19. Linares-Casenave J, Kroll KJ, Van Eemennaam JP and Doroshov SI (2003) Effect of ovarian stage on plasma vitellogenin and calcium in cultured white sturgeon. *Aquaculture* **221**, 645-56.
20. Trichet V, Buisine N, Mouchel N, Moran P, Pendas AM, Le Pennec JP and Wolff J (2000) Genomic analysis of the vitellogenin locus in rainbow trout (*Oncorhynchus mykiss*) reveals a complex history of gene amplification and retroposon activity. *Mol Gen Genet* **263**, 828-37.
21. Wang H, Tan JTT, Emelyanov A, Korzh Vand Gong Z (2005) Hepatic and extrahepatic expression of vitellogenin genes in the zebrafish, *Danio rerio*. *Gene* **356**, 91-100
22. Bayne CJ and Gerwick L (2001) The acute phase response and innate immunity of fish. *Dev Com Immunol* **25**, 725-743.
23. Basu N, Todgham AE, Ackerman PA, Bibeau MR, Nkano K, Schulte PM and Iwama GK (2002) Heat shock protein genes and their functional significance in fish. *Gene* **295**, 173-83.
24. Ochiai Y, Huang MC, Fukushima H and Watabe S (2003) cDNA cloning and thermodynamic properties of tropomyosin from walleye Pollack *Theragra chalcogramma* fast skeletal muscle. *Fish Sci* **69**, 1033-41.
25. Yue G H, Ho MY, Orban L and Komen J (2004) Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. *Aquaculture* **234**, 85-98.
26. Vasemagi A, Nilsson J, and Primmer CR (2005) Seventy-five EST-linked Atlantic salmon (*Salmo salar* L.) microsatellite markers and their cross-amplification in five salmonid species. *Mol Ecol Notes* **5**, 282-8.