Mar. Biotechnol. 4, 526–536, 2002
DOI: 10.1007/s10126-002-0067-0

MARINE BIOTECHNOLOGY

Short Communication

Expression Profile of the Channel Catfish Spleen: Analysis of Genes Involved in Immune Functions

Arif M. Kocabas, Ping Li, Dongfeng Cao, Attila Karsi, Chongbo He, Andrea Patterson, Zhenlin Ju, Rex A. Dunham, and Zhanjiang Liu*

The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, Auburn University, Auburn, AL 36849, U.S.A.

Abstract: Both qualitative and quantitative patterns of tissue-specific gene expression can be determined using gene profiling. Expressed sequence tag (EST) analysis is an efficient approach not only for gene discovery and examining gene expression, but also for development of molecular resources useful for functional genomics. As part of an ongoing transcriptome analysis of channel catfish (*Ictalurus punctatus*), EST analysis was conducted for gene annotations and profiling using a complementary DNA library developed from messenger RNA of the spleen. A total of 1204 spleen cDNA clones were analyzed. Of the 1204 clones, 665 clones (55.2%) were identified as orthologs of known genes from other organisms by BLAST searches and 539 clones (44.8%) as unknown gene clones. In total 147 novel genes were identified, and annotations were made to 118 of them. In addition, 389 novel EST clusters were identified. Expression profile was analyzed in relation to metabolic functional groups. A total of 28 known genes were involved in immune functions, of which 10 were identified for the first time in channel catfish. Microsatellite-containing clones were also identified that may be potentially useful for genome mapping. This work contributed to the Catfish Gene Index, and toward a Unigene set useful for functional genomics research concerning spleen gene functions in relation to disease defenses.

Key words: expressed sequence tag, functional genomics, gene, spleen, expression, mapping.

INTRODUCTION

Transcriptome refers to the entire set of RNA transcribed from a genome. Compared with the relatively stable genome, the transcriptome is highly variable depending on differential gene expression in various tissues, developmental stages, or under variable environmental conditions. This variation is reflected not only by which genes are expressed, but also by levels of gene expression. Much of transcriptome analysis has examined expressed sequence tags (ESTs, Adams et al., 1991). When ESTs are available, robust analysis of the transcriptome can be conducted using complementary DNA microarrays (Schena et al., 1996).

EST analysis is an efficient method for gene discovery. Upon generation of short cDNA sequences, genes can be identified, annotated, and catalogued to establish orthologous relations with known genes from other organisms by comparison of similarities (Boguski and Schuler, 1995; Hishiki et al., 2000; Ju et al., 2000; Mekhedov et al., 2000; Cao et al., 2001); EST analysis is a convenient way to examine gene expression. Baseline expression levels of genes can be established by sequencing nonnormalized cDNA libraries from specific tissues. ESTs represent transcripts of

Received April 20, 2002; accepted August 12, 2002.

*Corresponding author: fax 334-844-9208; e-mail zliu@acesag.auburn.edu

genes; therefore, their sequencing can also provide information on alternative splicing and differential polyadenylation (Brett et al., 2000; Karsi et al. 2002b). ESTs are important resources for gene mapping. ESTs can be directly mapped by polymerase chain reaction (PCR) analysis for species in which radiation hybrid panels are available. For species without radiation hybrid panels, ESTs are a great resource for discovery of single nucleotide polymorphism (SNP) as well as polymorphic markers by microsatellite tagging (Liu et al., 1999b, 2001). Finally, ESTs are very useful in the development of microarrays for functional genomics, environmental genomics, and toxicogenomics.

Because many ESTs can be readily identified as orthologs of known genes by simple similarity searches, polymorphic ESTs provide a rich source for type I markers (O'Brien, 1991) including length polymorphisms (Liu et al., 1999a) and SNP markers (Wang et al., 1998; Emahazion et al., 1999). Type I genetic markers are markers associated with genes of known function, useful for comparative gene mapping (Womack and Kata, 1995).

Among molecular bioinformatic databases, dbEST is the fastest growing (Schuler, 1997; Quackenbush et al., 2000), with over 11 million sequence entries at the beginning of 2002. However, ESTs generated from teleosts account for only 2% of all ESTs in the dbEST. Almost 95% of the teleost ESTs were generated from zebrafish. Obviously, much effort needs to be directed toward EST development in aquaculture species. Channel catfish is the major aquaculture species in the United States, with production of about 300 million kg per year, 70% of U.S. aquaculture production (Waters, 2001). A catfish transcriptome project has been initiated, and initial EST analysis has been conducted using libraries made from brain, head kidney, and skin, from channel catfish (Ju et al., 2000; Cao et al., 2001; Karsi et al., 2002a).

The spleen of fish is comparable to that of mammals and protects the fish from blood-borne pathogens (Ellis, 1992; Zapata et al., 1996). The spleen houses immune cells such as T and B cells that either destroy the antigen or neutralize it to prevent further manifestations during disease processes (Kaattari and Piganelli, 1996). Identification of functional genes in immune cells during disease processes may facilitate understanding of mechanisms involved in disease resistance and disease defenses. Thus identification of genes in the spleen and characterization of their expression profiles is one of the objectives of transcriptome analysis in catfish.

The objective of this study was to develop genomic resources from the channel catfish spleen including generation of EST sequences, establishment of orthologs, annotation of identified genes, determination of gene expression profiles, and identification of microsatellite-containing genes for potential polymorphic type I markers. Here we report analysis of 1204 ESTs from the channel catfish spleen cDNA library with emphasis on genes involved in immune functions.

Materials and Methods

Animals and Tissue Collection

All experimental channel catfish (*Ictalurus punctatus*) were raised in troughs placed inside the hatchery of the Auburn University Fish Genetics Facility until tissue collection under a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at Auburn University. Channel catfish fingerlings were kept under the same conditions for 4 weeks before experimentation. To create a resource cDNA library containing as many transcripts as possible including transcripts induced after disease infection, spleen tissues were collected from healthy fish as well as from infected fish. The channel catfish were challenged with *Edwardsiella ictaluri* using procedures adapted from Dunham et al. (1993). Fish were divided into 2 groups, the nonchallenged controls and the fish for challenge ($N = 240$). The fingerlings used for disease challenge were placed into a 150-L tank containing 1.1×10^6 *E. ictaluri* cells/ml for 1 hour. The challenged fish were then removed and stocked in a 1000-L tank. Spleen samples were collected from 10 fish daily and kept in liquid nitrogen until RNA isolation. Dead fish were removed, and their tissues were not collected.

RNA and messenger RNA Isolation

All spleen samples were mixed and ground with a mortar and pestle, and homogenized in RNA extraction buffer with a hand-held tissue tearer (model 985-370, Biospec Products Inc.). Total RNA was extracted following the guanidium thiocyanate method (Chomczynski and Sacchi, 1987). Poly(A)$^+$RNA was isolated from total RNA using the Poly(A)$^+$Pure kit (Ambion, Cat. No. 1915) according to the manufacturer's instructions, except that 2 rounds of purification were performed. The quality of RNA was checked by a Northern blot using a β-actin probe (Liu et al. 1989).

Complementary DNA Library Construction

A directional cDNA library was constructed using the pSPORT-1 Superscript Plasmid Cloning System from GIBCO/BRL. Two micrograms of Poly(A)$^+$ RNA was used for the initial reaction. The detailed protocols for construction of the cDNA library were provided by GIBCO/BRL, except that ElectroMax DH12S cells were used for electroporation of the cDNA library. This alteration gave effective electroporation and efficient production of single-stranded phagemids (GIBCO/BRL), a feature desirable for future development of normalized cDNA libraries. The quality of the spleen cDNA library was determined by number of primary recombinants and average insert size. The channel catfish spleen cDNA library had over 6×10^6 primary cDNA clones; its average insert size was estimated to be about 1000 bp by restriction digestion of 20 randomly picked clones. Before sequencing analysis, the primary cDNA library was amplified once (Sambrook et al., 1989). The pooled library was frozen in liquid nitrogen and stored at −80°C.

Plasmid Preparation and Sequencing Analysis

Random clones were picked from the spleen cDNA library and grown in 1.5 ml of LB medium overnight for plasmid preparation. Plasmid DNA was prepared by alkaline lysis method (Sambrook et al., 1989) using the Qiagen Spin Column Mini-plasmid kits. The prepared plasmid DNA (1 µl, about 200 to 500 ng) was used for all sequencing reactions. Nucleotide sequencing was conducted using the chain-termination method (Sanger et al., 1977). The single pass DNA sequencing reaction was accomplished in a thermocycler using cycleSeq-farOUT polymerase (Display Systems Biotech). An initial 2 minutes of extra denaturation at 94°C was always used. The PCR conditions were at 94°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute, for 30 cycles. All sequencing reactions were primed by using the M13 reverse sequencing primer to sequence the cDNAs from upstream (5′ of the transcripts) to reduce sequencing problems caused by poly(A)$^+$ at the 3′ end of cDNAs. The sequencing primer was labeled by fluorescence labels (Li-COR). All sequences were analyzed on automatic Li-COR DNA sequencers, Long ReadIR 4200 or Li-COR DNA Analyzer Gene ReadIR 4200.

Identification of Microsatellite-Containing cDNAs

After sequencing the EST clones, sequences were obtained by using the Li-COR Image Analysis software. Vector sequences were trimmed before analysis. The cDNAs containing simple sequence repeats (microsatellites) were identified by determining the minimal number of repeats in the microsatellite sequences: dinucleotide, 8 repeats; trinucleotide, 6 repeats, tetranucleotide, 5 repeats; pentanucleotide, 4 repeats. As the major objective of this research was to develop EST resources, polymorphic contents of the microsatellites within identified cDNAs were not determined.

Gene Annotations

Vector sequences were removed before searching for homologies using the Basic Local Alignment Search Tool (BLAST) through the lnternet (National Center for Biotechnology Information) [NCBI], Bethesda, Md.). For establishing ortholog, procedures of Cao et al. (2001) were followed. Matches were considered significant only when the probability level (P) was less than 1×10^{-4} using BLASTN and BLASTX with all parameters at default except for a few cases in which more than one stretch of homologous sequence existed. Considering the biological meaning of the similarity searches, we specifically recorded queries that resulted in low scores but exhibited more than one stretch of sequence homologous to the target sequences. ESTs with significant similarities from BLASTN searches were considered an ortholog of known genes only if the significant similarities were not caused by simple sequences. ESTs failing to show significant similarities using BLASTN searches were further searched by BLASTX. Similarly, ESTs with significant similarities were evaluated for matches caused by simple amino acid matches or repeats. ESTs failing to match any known gene sequences were named unknown EST clones. Once orthologs were established, genes were annotated by examination of the gene identities of the homologous sequences from various organisms. Particular attention was paid to the Unigene annotations or gene names used in Linkage information of the human, rat, and zebrafish genes in the NCBI databases.

Results and Discussion

EST Identities and Gene Annotations

In total 1204 ESTs were analyzed from a cDNA library of the channel catfish spleen. Of the 1204 clones, 665 clones (55.2%) were identified as orthologs of known genes; 539 clones(44.8%) could not be identified by similarity com-

Table 1. Summary of BLAST searches of the 1204 EST Clones from Spleen of Channel Catfish (*Ictalurus punctatus*)

	BLASTN searches of GenBank NR database	BLASTX searches of GenBank NR database	BLASTN searches of dbEST database	Total
Orthologs of clones previously reported from channel catfish genes	152	17	–	169
Number of clones representing previously reported channel catfish genes	467	27	–	494
Average clones per gene	3.07	1.59	–	N/A
Orthologs established for the first time in channel catfish	52	65	–	117
Number of clones representing the novel channel catfish genes	52	87	–	139
Average clones per novel gene	1.00	1.34	–	N/A
Orthologs of ORFs previously reported from channel catfish	–	2		2
Number of clones representing the previously known channel catfish ORFs	–	2	–	2
Average clones per ORF		1.00	–	N/A
Orthologs of ORFs reported for the first time in channel catfish	8	22	–	30
Number of clones representing the novel channel catfish ORFs	8	22	–	30
Average clones per novel ORF	1.00	1.00	–	N/A
Number of unknown EST clusters previously reported from channel catfish	–	–	99	99
Number of clones representing the previously reported unknown catfish ESTs	–	–	132	132
Average clones per EST cluster	–	–		N/A
Number of unknown catfish EST clusters reported for the first time in the spleen	–	–	388	388
Number of clones representing the novel catfish EST	–	–	407	407
Average clones per EST cluster	–	–	1.05	N/A

[a]Analysis was conducted sequentially with the clones first being searched using BLASTN, followed by BLASTX for the clones whose identities could not be determined using BLASTN. The clones for which orthologs could not be established were searched against dbEST database to determine if they were novel catfish ESTs.

parison using BLAST searches and therefore were designated as unknown EST clones. The EST sequences have been deposited into dbEST database at NCBI with consecutive accession numbers from BM424289 to BM425492.

The 527 ESTs were identified to be the orthologs of known genes by BLASTN similarity searches including 8 clones homologous to open reading frames (ORFs) of unknown functions. A BLASTN search against the dbEST indicated that among the 212 genes represented by the 527 ESTs, 152 were previously identified from channel catfish, and 60 were identified from this project, including 8 ORFs (Table 1).

A BLASTN search is the first step for determination of gene identities because significant sequence conservation at the nucleotide level is relatively easy to interpret. However, caution must be exercised to avoid mistaken annotations because of significant sequence similarities beyond the coding regions. Short interspersed elements are associated with catfish transcripts, often located at the 3′-untranslated regions (UTRs, Kim et al., 2000). For example, IpSpn01066

exhibited significant similarities ($P = 3 \times 10^{-07}$) to gonadotropin β subunit 1 and 2 loci from common carp (accession numbers X59888 and X59889), but the homologous sequences lie outside the coding region at 3′ UTR. Further analysis indicated that the homologous sequences had no similarity to known genes, but shared similarity to several channel catfish ESTs including IpHdk02509 (accession number BE468332), IpCGBr1_15_G07_21 (BM494269), and IpLvr00905 (BM439008). These may represent some transcribed SINE-like sequences, but their identities need to be determined.

The remaining 677 ESTs, whose identities were not revealed by BLASTN searches, were further examined for ammo acid similarities by BLASTX, resulting in identification of orthologs for an additional 138 clones. These 138 clones represented 106 unique genes. Among these 106 genes identified by BLASTX searches, 19 were previously identified from channel catfish and 87 were newly identified from this project, including the 22 ORFs of unknown function (Table 1).

In contrast to BLASTN searches, BLASTX searches depend on sequence similarities within the coding regions; therefore, problems in gene annotations caused by SINE and other simple sequences can be minimized. However, significant sequence similarities can be produced for many similar proteins, making annotation more complicated. In addition, the cutoff level of significant similarity in relation to biological importance is difficult to determine, and caution must be exercised when annotating genes based only on BLASTX searches. Although a significant P value of 1×10^{-5}, when coupled to examination of length homology, is often enough to determine identity using nucleotide similarities, a lower P value (often less than 1×10^{-9}) is needed to determine identity using amino acid similarities.

A total of 32 ESTs from the catfish spleen cDNA library was identified as orthologs of known ORFs from other organisms. These represent orthologs of genes of unknown functions. Of the 32 ORFs, 2 were previously identified from channel catfish and 30 were newly identified from this project (Table 1). Although the functions of these genes are currently unknown, their evolutionary conservation in model organisms such as human, rat, and *Caenorhabditis elegans* will facilitate functional annotations to these genes. As soon as the functions of these genes are determined from any of the model organisms, their functions can be inferred for channel catfish.

The proportion of known genes for the channel catfish spleen was comparable to that for the catfish brain, head kidney, and skin (Ju et al., 2000; Cao et al., 2001; Karsi et al., 2002a). The percentage of ESTs for which orthologs can be established is also a function of the EST length, which is important when attempting to compare data. Long ESTs are likely to produce sequences within the coding regions, while short ESTs often produce sequences in the UTRs. As sequence conservation is not prominent in UTRs, short sequences often lead to production of unknown ESTs. In addition, the establishment of orthologs depends on the contents of databases. Therefore, the ability to establish orthologs for aquaculture fish species should be greatly enhanced when the sequencing of the entire genome of zebrafish and *fugu* is completed at the conclusion of 2002.

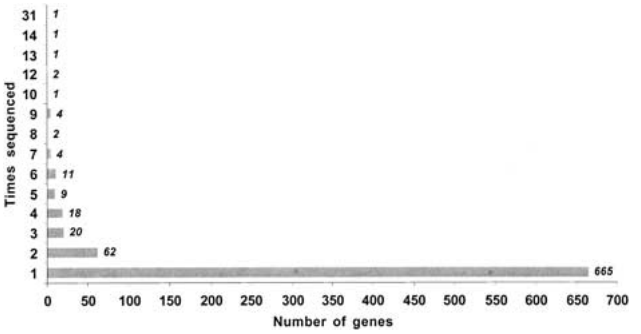

Figure 1. Expression profiles of the 1204 ESTs from the spleen of channel catfish.

Expression Profile in the Channel Catfish Spleen

Cluster analysis was conducted with the BLASTN searches against the dbEST database containing the spleen ESTs using all 1204 ESTs as queries to establish expression profiles of the genes in the spleen. Each search resulted in a list of homologous sequences. Identities of the homologous spleen ESTs were recorded, and they were regarded as belonging to the same EST cluster. In several cases, 2 subclusters were combined into a single EST cluster when the subclusters were homologous to different regions of the same gene, or when they shared similarities to a single EST, anchoring them into a single EST cluster.

A total of 801 EST clusters were identified from the 1204 EST clones. Possibly, some EST clusters will later be combined as more information becomes available—particular for the unknown EST clones. However, the actual number of genes represented by these EST clusters should be close to the number of clusters, especially when considering the long EST sequences produced in this project by sequencing using the Li-COR automated sequencers.

Table 2. Meta-analysis of Expression Profiles of the EST Clones for Which Orthologs Were Established from the Channel Catfish (*Ictalurus punctatus*) Spleen[a]

	No. of genes	No. of clones	(%)[a] Expression	Average clones/gene
Ribosomal protein genes	68	237	19.7	3.49
Immune-related genes	28	89	7.4	3.18
Structural genes	25	77	6.4	3.08
Oncoproteins and tumor-suppressor-related genes	5	10	0.8	2.00
Receptors and receptor-associated proteins	10	20	1.7	2.00
Ionic channels, metal metabolism, sorting proteins, and transporters	7	13	1.1	1.86
Hormones and regulatory proteins	18	32	2.7	1.79
Enzymes and enzyme activity regulators	61	84	7.0	1.38
Translational factors	3	4	0.3	1.33
Development and cell-cycle-related genes	15	19	1.6	1.27
Lectins, secreted proteins, and other genes	30	34	2.8	1.13
Transcriptional factors, DNA and RNA binding proteins	12	12	1.0	1.00
Chaperons and other induced genes	5	5	0.4	1.00
Open ORFS of unknown function	32	32	2.7	1.00

[a]Values are the percentage of the group of genes of the 1204 sequenced clones, rounded up or down as appropirate.

The expression profiles are shown in Figure 1 as determined by redundancy during sequencing. Overall, the majority of EST clusters were represented by singletons in the spleen. A total of 665 transcripts were sequenced only once (singletons), and 136 transcripts were sequenced more than once. Of the transcripts that were sequenced multiple times, 62 were sequenced twice; 20 were sequenced 3 times; 18 were sequenced 4 times; and 36 were sequenced 5 times or more (Figure 1).

Meta-analysis of the Expression Profile in the Channel Catfish Spleen

Expression of genes in the channel catfish spleen was analyzed according to their major metabolic functions (Table 2). Obviously, many genes have multiple functions, and their assignment into each group was based on their major functions (Ju et al., 2000). Because of the long list, gene names in each metabolic group were not presented, but will be available from our Web site (http://www.auburn.edu/genomics).

The largest group was the group involved in the translational machinery including ribosomal proteins and translational factors, followed by genes involved in immune functions, enzymes, and structural genes (Table 2). Immune-related genes accounted for 7.4% of gene expression in the spleen. The immune-related genes identified from the spleen are listed in Table 3. Many of the immune-related genes were expressed at high levels in the spleen including β_2-microblobulin (0.8%), monoclonal nonspecific suppressor factor β (0.8%), interleukin-8 (0.5%), interferon-induced transmembrane protein 2 (0.5%), immunoglobulin D heavy chain (0.5%), major histocompatibility complex (MHC) I α chain (0.3%), invariant chain-like protein (0.3%), MHC class II integral membrane protein α chain 3 (0.3%), and MHC II DAB gene (0.3%) (Table 3). Seven immune-related genes were identified from channel catfish for the first time: breakpoint cluster region protein 1, complement C4B, I κ B, interferon regulatory factor 4, interferon-regulated resistance GTP-binding protein Mx, interleukin-1 β, and B-cell receptor-associated protein 31. These and other immune-related EST clones should be useful for molecular characterization of these gene loci, and for functional analysis of these genes during disease defense reactions.

A comparison was made for expression of immune-related genes in the spleen (the present study), head kidney (Cao et al., 2001), and cultured leukocyte cell lines (available at http://morag.umsmed.edu/). A total of 6373 leukocyte ESTs were produced at the University of Mississippi Medical Center from 6 cDNA libraries made from an autonomous B cell line (JX11), an autonomous T cell line (JX12), mixed leukocytes (JX13), an autonomous micro-

Table 3. Comparison of Expression of the Immune-Related Genes of the Channel Catfish Spleen in Spleen, Head Kidney, and Leukocyte Cell Cultures

EST	Accession No.	Identity/ortholog	Frequency in spleen[a]	Frequency in head kidney	Frequency in leukocyte
IpSpn00671	BM424775	β_2-microglobulin	9	9.7	6.4
IpSpn01115	BM425098	Breakpoint cluster region protein 1 (barrier-to-autointegration factor)	1*	–	–
IpSpn01129	BM425108	CC chemokine-1	2	1.1	–
IpSpn01598	BM425331	Complement Clq A chain precursor	3	1.1	–
IpSpn01621	BM425349	Complement C4B	1*	–	–
IpSpn00865	BM424942	Complement subcomponent Clq chain B	3	1.1	–
IpSpn00167	BM424435	Cyclophilin D	4	3.8	1.1
IpSpn00180	BM424447	Embigin protein	1*	–	0.2
IpSpn00983	BM425006	Galectin like protein	3*	–	0.8
IpSpn00036	BM424321	IκB protein	1*	–	–
IpSpn00148	BM424419	MHC class II integral membrane protein α chain 3	4	1.6	0.8
IpSph00312	BM424566	*I. punctatus* clone PG7 sterile class G IgL chain	2	2.7	–
IpSpn01701	BM425401	*I. punctatus* MHC class I α chain	6	4.8	2.1
IpSpn00470	BM424704	MHC class II DAB gene	3	0.5	0.2
IpSpn00751	BM424845	Ig rearranged H-chain mRNA V-C-region	2	4.3	–
IpSpn00120	BM424393	Immunoglobulin D heavy chain	3	3.2	–
IpSpn00231	BM424491	Immunoglobulin light chain F class	3	7.5	–
IpSpn01512	BM425250	Immunoglobulin light chain L2	1	1.6	–
IpSpn00254	BM424512	Interferon regulatory factor 4	1*	–	–
IpSpn00363	BM424611	Interferon induced transmembrane protein 2 (1-8D)	6	0.5	–
IpSpn00717	BM424812	Interferon-regulated resistance GTP-binding protein Mx	1*	–	–
IpSpn01597	BM425330	Interleukin-1β	2*	–	–
IpSpn01565	BM425301	lnterleukin-8/γ interferon-induced monokine/Small inducible cytokine subfamily B (Cys-X-Cys), member 6 (granulocyte chemotactic protein 2)	6	2.2	–
IpSpn00384	BM424630	Invariant chain-like protein	4	0.5	0.2
IpSpn01500	BM425227	Monoclonal nonspecific suppressor factor β	9	3.8	1.3
IpSpn01188	BM425153	IgE (Fc-)-receptor γ-chain protein	1*	–	0.2
IpSpn00701	BM424800	B-cell receptor-associated protein 31	1*	–	–
IpSpn00812		Small inducible cytokine A3	2	–	5.3

[a]Frequency in head kidney and in leukocyte cell cultures was normalized to have a total of 1204 each analyzed clones as in the spleen. Asterisk (*) indicates genes identified for the first time in channel catfish.

phage cell line (JX14), and nonautonomous T cell lines (JX15 and JX17) (Miller et al., 1994a, l994b; Wilson et al., 1997, 1998; Stuge et al., 2000; Zhou et al., 2001).

It should be noted that gene discovery through the EST approach depends on the sequencing depth into the libraries and the quality of the libraries. In this case the qualities of the spleen and head kidney libraries were similar, and they have been sequenced for 1204 and 2228 clones, respectively. Several genes exhibited differential rates of expression between the 2 tissues. Notably galectin-

Table 4. Highly Expressed Genes in the Spleen Tissue of Channel Catfish That Were Sequenced Three or More Times

Genes	Times sequenced	Number of genes
Hemoglobin β chain	31	1
α-Globulin	14	1
Ribosomal protein (RP) L32	13	1
RP S12, RP L41	12	2
RP L36	10	1
RP L38, β_2-microglobulin precursor, homologue of high-affinity IgE receptor γ, monoclonal non specific suppressor factor β_1 unknown gene	9	5
RP L22, RP L24	8	2
RP large P2, RP L11, RP L35, β-actin	7	4
RP L13a, RP L17, RP L30, RP L35a, RP S2, RP S15, Interferon-induced protein 2, MHC class 1 α chain, interleukin-8,1 unknown gene	6	11
RP S3, RP S11, RP S22, Clq related factor, cysteine-rich protein 1, ferritin heavy subunit, MHC class II β chain, Ubiquitin-like protein, 1 unknown gene	5	9
RP L7a, RP L9, RP L10a, RP L18, RP L34, RP L37a, RP L36a (RP L44), RP Sa (40 kDa RP protein), RP S14, RP S30, cytochrome *b*, cyclophilin D, invariant chain-like protein 2, 18S rRNA, cystatin E, trypsin IA, 2 unknown gene	4	18
RPL12, RP L14, RP S27, RP S29, cytochrome *c* oxidase subunit 1, cofilin 1, galectin-like protein, GTP-binding protein (rhoC), high-mobility group protein 17, immunoglobulin light chain F class, leukocyte DNA binding receptor, lysozyme precursor, NADH-ubiquinone oxidoreductase 15 kDa subunit, immunoglobulin D heavy chain, transposase, complement Clq A chain, 4 unknown genes	3	20
Total	415 clones	74 genes

like protein was sequenced 3 times from the spleen, but was not encountered from the head kidney (Cao et al., 2001). Other genes with an expression ratio greater than 5 included interferon-induced transmembrane protein 2 (12 times higher in the spleen), invariant chain-like protein (8 times higher in the spleen), and MHC II DAB gene (6 times higher in the spleen). It appeared that expression of immune-related genes was generally higher in the spleen than in the head kidney. Head kidney shares mixed immune and lymphohemopoietic functions with the spleen, but also has endocrine function. With the exceptions of β_2-microglobulin and the small inducible cytokine A3, most of the immune genes had a lower representation in the ESTs produced from the leukocyte cell lines than in the spleen or in the head kidney. This is the opposite of what one would expect because only a fraction of cells in the spleen and head kidney were leukocytes. However, the small inducible cytokine was highly expressed in the cultured leukocyte cell lines, and β_2-microglobulin had similar representations among the tissues and the cultured cell lines (Table 3). Expression in vivo may be quite different from that in vitro, complicating comparisons between the immune-related expression in live fish and that in cell lines.

The Most Abundantly Expressed Genes in the Channel Catfish Spleen

Highly expressed genes are repeatedly sequenced in EST projects using nonnormalized libraries. In the spleen many genes involved in the translational machinery were among those most frequently sequenced including 33 ribosomal proteins (Table 4). Hemoglobin β chain (2.6%) and α-globin (1.2%) were highly represented in the spleen cDNA library. In addition to the many structural genes, a few other genes, notably the genes involved in immune functions as discussed above, were also highly expressed in the spleen. The 74 most abundantly expressed genes accounted for 415 of the 1204 sequenced clones. This indicated that for most efficient gene discovery in an EST project, normalization alone may not be sufficient. Subtraction with selected clones may provide a more effective alternative.

Table 5. Novel Microsatellite-Containing cDNAs from the Spleen ESTs

Clone name	Accession number	Gene identity	Simple sequence repeat	Microsatellite location
IpSpn00061	BM423338	Unknown	(AC)20	3′ UTR
IpSpn00099	BM424374	Unknown	(CA)10	N/D
IpSpn00129	BM424402	Unknown	(TA)26	3′ end
IpSpn00183	BM424450	Unknown	(GT)16	N/D
IpSpn00300	BM424555	Unknown	(CA)11, (CA)14	3′ end
IpSpn00332	BM424585	Ribonucleoprotein L	(TGA)5	3′ UTR
IpSpn00370	BM424618	Unknown	(ATT)24	3′ UTR
IpSpn00378	BM424625	Unknown	(AT)8	N/D
IpSpn00404	BM424648	Unknown	(GT)15	N/D
IpSpn00468	BM424702	Histone H3	(TG)10, (GT)20	3′ UTR
IpSpn01750	BM425443	Unknown	(AT)16	N/D

Microsatellite-Containing Spleen Genes

Of the 1204 ESTs, 28 unique ESTs contained simple sequence repeats. Among the 28 microsatellite-containing ESTs, 23 clones included dinucleotide repeats, 4 clones included trinucleotide repeats, and 1 clone included pentanucleotide repeats. Seventeen of the 28 clones were previously reported, and 11 were newly identified from this project (Table 5). These microsatellites may potentially be polymorphic and useful for genetic mapping of the catfish genome, but their potential polymorphisms were not determined.

Conclusions

A total of 1204 ESTs were analyzed from the spleen of channel catfish. This work allowed establishment of orthologs for 318 genes. A total of 147 genes (including the 29 novel genes orthologous to genes of known functions) were identified from channel catfish for the first time, including 7 genes involved in immune functions. Annotations were made for 118 newly identified channel catfish genes, contributing to the characterization of the catfish transcriptome and establishing a Unigene set for channel catfish. These ESTs should serve as a resource toward the Gene Index of Catfish (TIGR, available at http://www.tigr.org). Along with other ongoing efforts in EST analysis in channel catfish, these ESTs should be useful for functional genomics research concerning genes specifically expressed after disease infection. As catfish are widely distributed in various watersheds, their cDNA microarrays may become extremely useful for environmental and toxicogenomic studies. The microsatellites that were identified may be useful for genetic linkage mapping. As some of the microsatellites are tagged to genes, development and mapping of type I markers will further facilitate comparative genomics concerning genome organization and evolution.

Acknowledgments

This project was supported by a grant from the U.S. Department of Agriculture NRI Animal Genome Basic Genome Reagents and Tools Program (USDA/NRICGP 2003-35205-12827). We appreciate the support of Auburn University Department of Fisheries and Allied Aquacultures, College of Agriculture, and the Vice President for Research for their matching funds to USDA NRI Equipment Grants (98-35208-6540 and 99-35208-8512).

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., and Venter, J.C. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651–1656.

Boguski, M.S., and Schuler, G.D. (1995). ESTablishing a human transcript map. *Nat Genet* 10:369–371.

Brett, D., Hanke, J., Lehmann, G., Haase, S., Delbruck, S., Krueger, S., Reich, J., and Bork, P. (2000). EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett* 474:83–86.

Cao, D., Kocabas, A., Ju, Z., Karsi, A., Patterson, A., and Liu, Z. (2001). Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles of the head kidney. *Animal Genet* 32:169–188.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analyt Biochem* 162:156–159.

Dunham, R.A., Brady, Y., and Vinitnantharat, S. (1993). Response to challenge with *Edwardsiella ictaluri* by channel catfish, *Ictalurus punctatus,* selected for resistance to *E. ictaluri. J Appl Aquacul* 3:211–222.

Ellis, A.E. (1992). Fish immunology. In *Encyclopedia of Immunology.* Vol. III., Roitt, I.M., Delves, P.J. (eds.). San Diego: Academic Press, 536–539.

Emahazion, T., Jobs, M., Howell, W.M., Siegfried, M., Wyoni, P., Prince, J.A., and Brooks, A.J. (1999). Identification of 167 polymorphisms in 88 genes from candidate neurodegeneration pathways. *Gene* 238:315–324.

Hishiki, T., Kowamoto, S., Morishita, S., and Okubo, K. (2000). Body Map: a human and mouse gene expression database. *Nucl Acids Res* 28:136–138.

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* 261:373–382.

Kaattari, S.L., and Piganelli, J.D. (1996). The specific immune response: Humoral Defense. In *The Fish Immune System; Organism, Pathogen and environment*, Iwama G., and Nakanishi, T. (eds.). San Diego: Academic Press, 207–254.

Karsi, A., Cao, D., Li, P., Patterson, A., Kocabas, A., Feng, J., Ju, Z., Mickett, K., and Liu, Z.J. (2002a). Transcriptome analysis of channel catfish (*Ictalurus punctatus*): Initial analysis of gene expression and microsatellite-containing cDNAs in the skin. *Gene* 285:157–168.

Karsi, A., Patterson, A., Feng, J., and Liu, Z.J. (2002b). Translational machinery of channel catfish: I. A transcriptomic approach to the analysis of 32 40S ribosomal protein genes and their expression. *Gene* 291:177–186.

Kim, S., Karsi, A., Dunham, R.A., and Liu Z. (2000). The skeletal muscle alpha-actin gene of channel catfish (*Ictalurus punctatus*) and its association with piscine specific SINE elements. *Gene* 252:173–181.

Liu, Z., Karsi, A., and Dunham, R.A. (1999a). Development of Polymorphic EST Markers Suitable for Genetic Linkage Mapping of Catfish. *Mar Biotechnol* 1:437–447.

Liu, Z.J., Li, P., Kocabas, A., Ju, Z., Karsi, A., Cao, D., and Patterson, A. (2001). Microsatellite-containing genes from the channel catfish brain: evidence of trinucleotide repeat expansion in the coding region of nucleotide excision repair gene *RAD23B. Biochem Biophys Res Comm* 289:317–324.

Liu, Z.J., Tan, G., Li, P., and Dunham, R. (1999b). Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus. Biochem Biophys Res Comm* 259:190–194.

Liu, Z.J., Zhu, Z., Roberg, K., Faras, A., Guise, K., Kapuscinski, A., and Hackett P. (1989). The β-actin gene of carp (*Ctenopharyngodon idella*). *Nucl Acids Res* 17:5850.

Mekhedov, S., de Ilarduya, O.M., and Ohlrogge, J. (2000). Toward a functional catalog of the plant genome: A survey of genes for lipid biosynthesis. *Plant Physiol* 122:389–401.

Miller, N.W., Chinchar, V.G., and Clem, L.W. (1994a). Development of leukocyte cell lines from the channel catfish (*Ictalurus punctatus*). *J Tissue Cult Meth* 16:117–123.

Miller, N.W., Rycyzyn, M.A., Wilson, M.R., Warr, G.W., Naftel J.P., and Clem, L.W. (1994b). Development and characterization of channel catfish long term B cell lines. *J Immunol* 152:2180–2189.

O'Brien, S.J. (1991). Molecular genome mapping: lessons and prospects. *Curr Opin Genet Dev* 1:105–111.

Quackenbush, J., Liang, F., Holt, I., Pertea, G., and Upton, J. (2000). The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Res* 28:141– 145.

Sambrook, J., Frisch, E.F., and Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual,* Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 93:10614–10619.

Schuler, G.D. (1997). Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J Mol Med* 75:694–698.

Stuge, T.B., Wilson, M.R., Zhou, H., Barker, K.S., Bengtèn, E., Chinchar, V.G., Miller, N.W., Clem, L.W. (2000). Development and analysis of various clonal alloantigen-dependent cytotoxic cell lines from channel catfish. *J Immunol* 164:2971–2977.

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, J., Lander, E.S., et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077–1082.

Waters, P.J. (2001). Response and correlated response to selection of channel catfish, *Ictalurus punctatus* for resistance to *Edwardsiella ictaluri, Flexibacter columnaris,* low levels of dissolved oxygen and toxic levels of nitrite in MS thesis. Auburn University, AL, USA.

Wilson, M.R., Bengtèn, E., Miller, N.W., Clem, L.W., DuPasquier, L., and Warr, G.W. (1997). A novel chimeric immunoglobulin heavy chain from a teleost fish shares similarities to IgD. *Proc Nat Acad Sci USA* 94:4593–4597.

Wilson, M.R., Zhou, H., Bengtèn, E., Clem, L.W., Stuge, T.B., Warr, G.W., and Miller, N.W. (1998). T-cell receptors in channel catfish: structure and expression of TCR alpha and beta genes. *Mol Immunol* 35:545–557.

Womack, J.E., and Kata, S.R. (1995). Bovine genome mapping: evolutionary inference and the power of comparative genomics. *Curr Opin Genet Dev* 5:725–733.

Zapata, A.G., Chiba, A., and Varas, A. (1996). Cells and tissues of the immune system of fish. In *The Fish Immune System; Organism, Pathogen and environment*, Iwama, G., and Nakanishi, T. (eds). San Diego, Academic Press, 1–62.

Zhou, H., Stuge, T.B., Miller, N.W., Bengtèn, E., Naftel, J.P., Bernanke, J.M., Chinchar, V.G., Clem, L.W., and Wilson, M. (2001). Heterogeneity of channel catfish CTL with respect to target recognition and cytotoxic mechanisms employed. *J Immunol* 167:1325–1332.