ELSEVIER

Gene 285 (2002) 157–168

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin

Attila Karsi, Dongfeng Cao, Ping Li, Andrea Patterson, Arif Kocabas, Jinian Feng, Zhenlin Ju, Kathryn D. Mickett, 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, USA

Received 11 November 2001; received in revised form 20 December 2001; accepted 15 January 2002
Received by J.A. Engler

Abstract

Previous molecular genetic studies on channel catfish (*Ictalurus punctatus*) have focused on limited number of genes and gene products. Recent advancement of molecular techniques made high throughput analysis of transcriptomes possible. As part of our transcriptome analysis of channel catfish, we have analyzed 1909 expressed sequence tags (ESTs) derived from a skin library. Of the 1909 ESTs analyzed, 1376 (72.1%) ESTs representing 496 unique genes had homologies with other organisms while 478 (25.0%) ESTs had no significant homologies and were designated as unknown. The remaining 55 (2.9%) EST clones were eliminated because of their low quality or short sequences. Of the 496 unique genes, 327 (65.9%) genes were singletons while 169 (34.1%) genes represented by two or more ESTs. A total of 1007 (52.8%) ESTs representing 235 unique genes matched previously reported channel catfish ESTs while 847 (44.4%) ESTs representing 261 unique genes were newly identified from this research. Functional categorization of the channel catfish genes indicated that the largest group was ribosomal proteins with 65 unique genes represented by 500 clones. The most abundantly expressed gene, the calcium binding protein ictacalcin, accounted for almost 5% of overall expression, indicating its important function in the skin. Sequence analysis of ESTs revealed the presence of 89 microsatellite-containing genes that may be valuable for future mapping studies.

Keywords: Fish; Skin; Expressed sequence tag; Marker; Simple sequence repeat

1. Introduction

Genomic approaches provide alternatives for addressing the mechanistic matters of gene expression. Expressed sequence tags (ESTs) are particularly useful for the development of cDNA microarrays that allow differentially expressed genes to be determined in a systematic way (Schena et al., 1996; Wang et al., 1999). ESTs are single pass sequences generated from random sequencing of cDNA clones (Adams et al., 1991). Large scale EST analysis is also an efficient way for identification of genes and for analysis of their expression by means of expression profiling (Franco et al., 1995; Azam et al., 1996; Lee et al., 2000). It offers a rapid and valuable first look at genes expressed in specific tissue types, under specific physiological conditions, or during specific developmental stages. ESTs have also been great resources for genomic mapping (Boguski and Schuler, 1995; Hudson et al., 1995; Schuler et al., 1996).

The potential use of ESTs for the discovery of new channel catfish genes has previously been shown (Ju et al., 2000; Cao et al., 2001). In addition, EST analysis offers opportunities for rapid identification and analysis of genes involved in specific biological pathways. For instance, using a transcriptomic approach, we previously characterized all the 47 ribosomal protein genes in the 60S ribosome (Patterson et al., 2002) and all the 32 ribosomal protein genes in the 40S ribosome (Karsi et al., 2002), which could otherwise have required many years of labor-intensive and expensive analyzes. Such systematic EST analyzes allowed for the

Abbreviations: aa, amino acids(s); bp, base pair(s); cDNA, complementary DNA; dNTP, deoxyribonucleotide triphosphate; ds, double stranded; EST, expressed sequence tag; kb, kilobase; LB, Luria-Bertani; NCBI, National Center for Biotechnology Information; PCR, polymerase chain reaction; ORF, open reading frame

* Corresponding author. Tel.: +1-334-844-4054; fax: +1-334-844-9208.
E-mail address: zliu@acesag.auburn.edu (Z. Liu).

0378-1119/02/$ - see front matter
PII: S0378-1119(02)00414-6

identification of alternatively spliced transcripts as well as alternatively polyadenylated transcripts, demonstrating the greater power and value of EST analysis than just for surveys of genes and their expression (Burke et al., 1998).

Channel catfish (*Ictalurus punctatus*) is the most important aquaculture species in the United States (USDA, 2000). Diseases cause tremendous losses to the catfish industry each year. As the first layer of defense, skin tissue has been known to be involved in physical and immune defense reactions against invasive agents such as bacterial pathogens and parasites. Host responses against skin inhabiting parasites and infectious diseases have been commonly observed (Buchmann, 1999), but little is known concerning the mechanisms of how the skin tissue actually function in defense reactions against diseases in fish. As part of our long-term genome analysis of channel catfish, here we extend our transcriptome analysis of channel catfish by analyzing 1,909 ESTs from the skin tissue. This EST study established orthologs for 496 genes plus 478 unknown ESTs. Of the 496 orthologs, 261 (52.6%) were identified from channel catfish for the first time. Eighty-nine of the unique ESTs also contain microsatellites that may provide tools for comparative genomic studies as well as for linkage mapping.

2. Materials and methods

2.1. Tissue preparation and RNA isolation

All experimental fish were raised in troughs located in the hatchery of the Auburn University Fish Genetics Facility under the same conditions for 4 weeks prior to the initiation of the experiments. At the start of the experiment, MS222 at 300 ppm was used to euthanize the fish. Skin tissues were collected and cut into as small pieces as possible. Pooled skin tissues from 15 fish were rapidly frozen with liquid nitrogen and were ground with a mortar/pestle, and then homogenized with a hand-held tissue tearor in RNA extraction buffer following the guanidium thiocyanate method (Chomczynski and Sacchi, 1987). Poly(A)$^+$ RNA was purified from total cellular RNA using the Poly(A)$^+$ Pure kit (Ambion, Austin, TX) according to the manufacturer's instructions.

2.2. Construction of skin cDNA library

A directional cDNA library of the skin was constructed using the pSPORT-1 SuperScript Plasmid Cloning System (Life technologies, Bethesda, MD). Construction of the cDNA library followed the manufacturer's instructions and library was electroporated into ElectroMax DH12S cells. These cells are highly adapted to efficient electroporation and production of single-stranded phagemids (Life technologies), features advantageous to the development of normalized cDNA libraries. Over 7.5 million primary cDNA clones were obtained with an average insert size of 1.0 kb. The primary cDNA library was amplified once before colonies were picked for sequencing.

2.3. Plasmid preparation and sequencing

The plasmid cDNA library was plated to a density appropriate for picking individual colonies. Random clones were grown in 1.5-ml LB medium overnight in 12 × 75-mm culture tubes. Plasmid DNA was prepared by the alkaline lysis method (Sambrook et al., 1989) using the Qiagen Spin Column Mini-plasmid kits. Three microliters of plasmid DNA (about 0.5–1.0 μg) were used in sequencing reactions. Chain termination sequencing (Sanger et al., 1977) was performed using cycleSeq-farOUT™ polymerase (Display Systems Biotech, Vista, CA). The PCR profiles were: 95°C for 30 s, 55°C for 40 s, 72°C for 45 s for 30 cycles. An initial 2 min denaturation at 96°C and a 5 min extension at 72°C were always used. Sequences were analyzed on an automatic LI-COR DNA Sequencer Long ReadIR 4200 or LI-COR DNA Analyzer Gene ReadIR 4200.

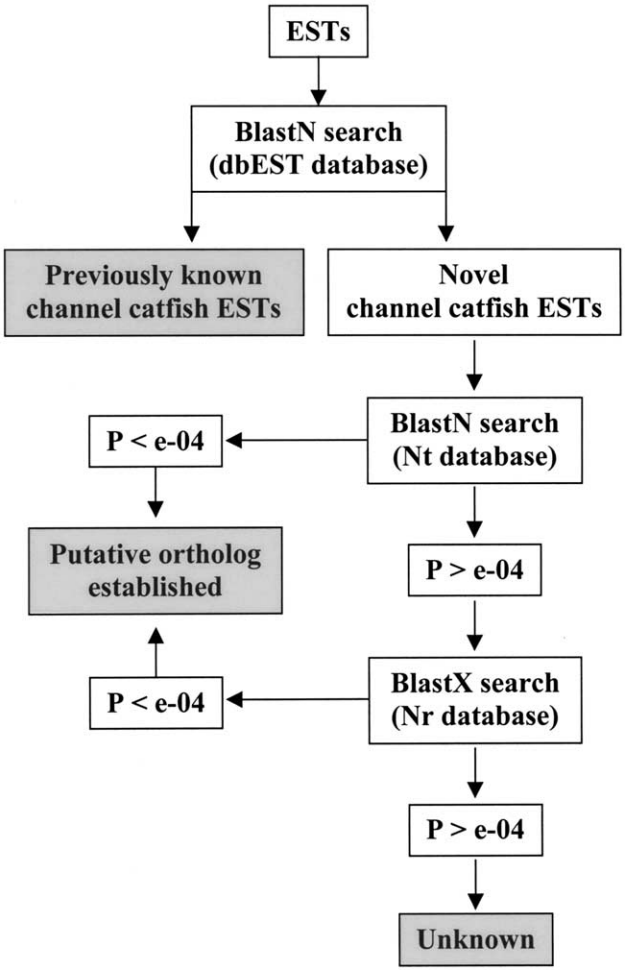

Fig. 1. Schematic presentation of sequence analysis and gene annotation using BLAST searches.

2.4. Bioinformatic analysis

BLAST searches (Altschul et al., 1990; Gish and States, 1993; Zhang and Madden, 1997) were conducted to determine gene identities. Procedures for establishing orthologs are shown in Fig. 1. A stepwise method was applied. First, BLASTN searches of dbEST database were conducted to determine which clones were homologous to the previously reported channel catfish ESTs. Second, non-redundant NT database was searched by BLASTN to establish orthologs relations. Finally, BLASTX searches were conducted on the NR non-redundant protein database for those ESTs failing to show significant similarities with BLASTN searches. Matches were considered significant only when the probability (P) was less than 1×10^{-4} using BLASTN and BLASTX with all parameters at the defaults. After the BLAST searches, a visual inspection was made to determine if the significant similarity was caused by simple sequences. ESTs with significant similarities in searches were considered orthologs of known genes only when the similarities were not caused by simple sequences. All ESTs that were not identified as orthologs of known genes were designated as unknown EST clones.

2.5. Identification of microsatellite containing cDNAs

During the compilation of EST sequences, genes that contained microsatellites were identified and their microsatellites were characterized in terms of complexity and repeat number. Clones containing microsatellites were identified by determination of a minimal number of repeats in the microsatellite sequences: dinucleotide, eight repeats; trinucleotide, six repeats; tetranucleotide, five repeats. Single nucleotide repeats were not included since they are not very useful for polymorphic markers. Some cDNA clones

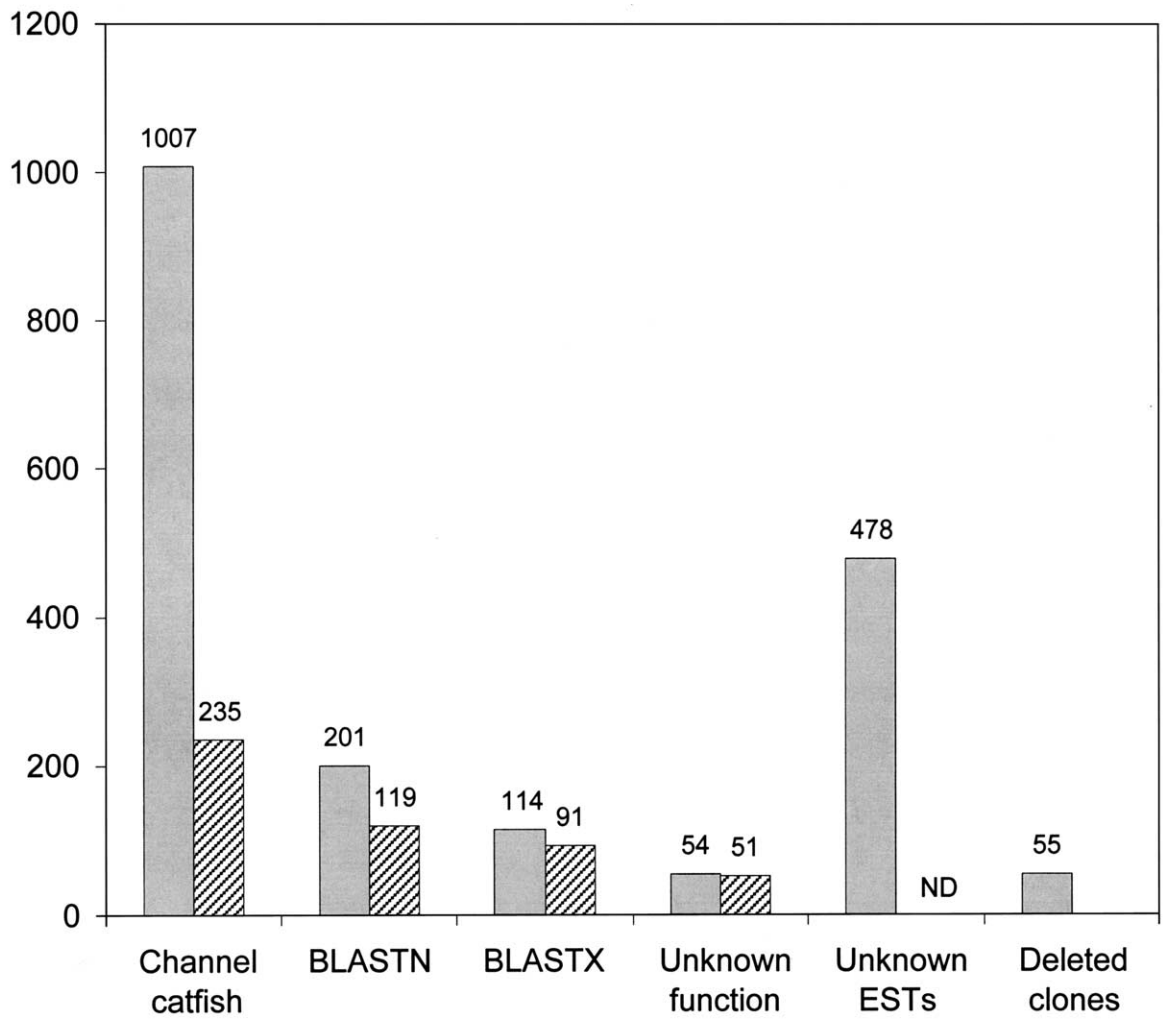

Fig. 2. Summary of the EST distribution in various groups and the number of genes they represent. Solid bars are number of ESTs and sketched bars are number of genes that they represent. Chanel catfish, ESTs that matched with previously identified channel catfish ESTs in the dbEST database; BLASTN, new ESTs identified by BLASTN searches; BLASTX, new ESTs identified by BLASTX searches; unknown function, ESTs similar to known sequences of unknown functions; and unknown ESTs, ESTs for which orthologs could not be established.

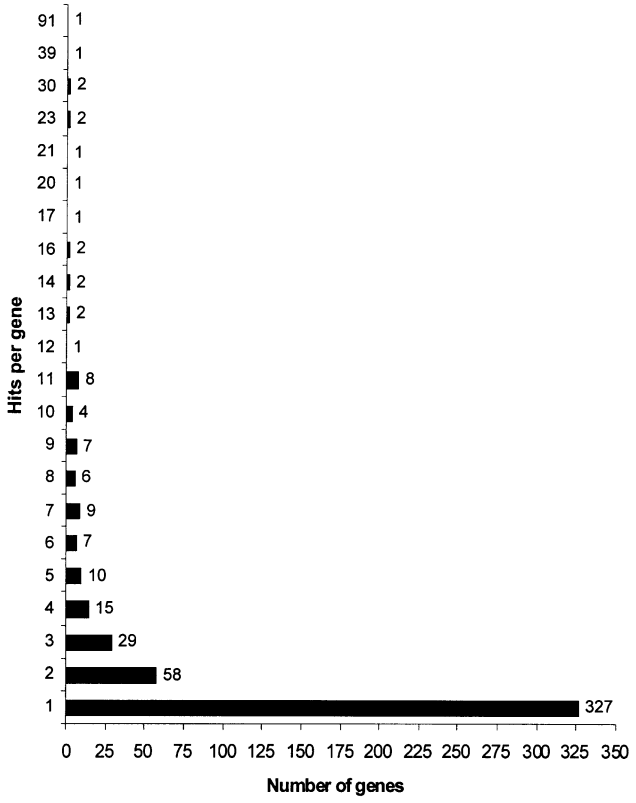

Fig. 3. Expression profiles and sequencing redundancy in the analysis of ESTs from the channel catfish skin tissue.

contain more than one type of repeat, in which case these clones were categorized according to the longest repeats.

3. Results and discussion

3.1. Newly identified channel catfish ESTs

Single pass sequencing of 1909 cDNA clones was conducted using a skin cDNA library to produce expressed sequence tags (ESTs). Our strategies for the analysis of the ESTs are outlined in Fig. 1. Gene annotation procedures were completed in three steps. First, homology searches of the sequenced ESTs have been locally done by BLASTN software on the dbEST database obtained from the National Center for Biotechnology Information (NCBI at http://www.ncbi.nlm.nih.gov). This analysis resulted in the identification of clones homologous to the previously submitted channel catfish ESTs. Those that had not been identified previously were searched using BLASTN on the non-redundant NT database. This homology comparison allowed us to establish new channel catfish orthologs from other organisms. Finally, remaining unknown ESTs were searched by BLASTX for amino acid similarity comparisons. Among the 1909 skin EST clones, 1007 (52.8%) were identified as homologous to the previously reported channel catfish ESTs, 847 (44.4%) clones were channel catfish ESTs sequenced for the first time in this research, and 55 sequenced clones produced either short or ambiguous sequences and were eliminated from further analysis. Among the 847 EST clones, 201 clones representing 119 unique genes were identified as orthologs of known genes from other organisms by BLASTN searches; 114 clones representing 91 unique genes were identified as orthologs of known genes using BLASTX searches; In addition, orthologs were established for 54 clones representing 51 genes of known sequences with unknown functions through both the BLASTN and BLASTX searches; 478 EST clones remains unknown in terms of their gene identity (Fig. 2).

Table 1
The ten most highly expressed channel catfish genes in the skin tissue

Putative identification	Frequency (%)
Ictacalcin	4.8
Ribosomal protein S02	2.0
Ribosomal protein L11	1.6
Ribosomal protein L41	1.6
Creatine kinase	1.2
Keratin type I	1.2
Ribosomal protein S20	1.1
Ribosomal protein S09	1.0
Dopamine receptor	0.9
Ribosomal protein L35	0.8

Table 2
Expression profiles of previously known channel catfish ESTs in the skin tissue

Category	# of genes	# of clones	Average redundancy
Ca^{2+}-binding proteins	3	106	35.3
Ribosomal protein genes	65	500	7.7
Cytoskeleton	5	38	7.6
Kinases/phosphatase	4	30	7.5
Receptors	5	32	6.4
Mitochondrial genes	16	74	4.6
Other enzymes	9	21	2.3
Immune-related genes	11	22	2.0
All other genes	36	75	2.0
Proteases	4	7	1.8
Translational factors	6	9	1.5
Transcriptional factors, zinc finger proteins	5	7	1.4
Unknown	64	82	1.3
Known ESTs of unknown functions	3	4	1.3

3.2. Expression profile in the channel catfish skin

Expression profiles of the known genes identified from the channel catfish skin is shown in Fig. 3. Among 496 identified distinct known genes, 327 known genes (65.8%) were sequenced only once; 111 genes (22.6%) were sequenced 2–5 times; 58 genes (11.7%) were sequenced

Table 3
New genes of channel catfish identified by BLASTN similarity comparison and their expressions

Clone #	Accession number	Putative identification	Homologous sequence	Probability	Frequency
IpSkn01020	BM027865	14-3-3 protein, signal transduction	NM_006761.1	5e-031	1
IpSkn00915	BM027858	16.5 kDa secretory protein	AF291663.1	3e-012	1
IpSkn01609	BM027904	26S proteasome, subunit p112	AJ006340.1	1e-004	1
IpSkn00675	BM027845	Acetyl-coA dehydrogenase, medium chain	NM_007382.1	5e-019	2
IpSkn00263	BM027829	Actin 2, actin-like protein	D12816.1	1e-012	1
IpSkn01850	BM027915	Adaptor-related protein complex AP-3, sigma 2	NM_009682.1	9e-038	2
IpSkn01162	BM027873	Amyloid precursor protein homolog HSD-2	NM_016160.1	5e-041	1
IpSkn01706	BM027909	Annexin II type 1	M60768.1	7e-018	2
IpSkn01651	BM027907	Annexin max3	Y11254.1	4e-016	6
IpSkn01578	BM027902	Anterior gradient 2 homolog	NM_011783.1	3e-009	1
IpSkn01367	BM027892	Aquaporin 3	NM_004925.2	2e-011	2
IpSkn00546	BM027839	Archain 1	NM_001655.1	5e-009	1
IpSkn00797	BM027852	Aspartyl-tRNA synthetase	NM_001349.1	2e-015	2
IpSkn00776	BM027851	B-cell receptor-associated protein 37	NM_007531.1	3e-038	1
IpSkn01704	BM027908	Beta-*N*-acetylhexosaminidase beta subunit	AF014805.1	1e-013	1
IpSkn00203	BM027827	Calcium and integrin-binding protein	AF136585.1	2e-004	1
IpSkn00854	BM027854	Calpactin I heavy chain	NM_019905.1	2e-008	1
IpSkn02511	BM027931	CDP-diacylglycerol synthase 2	AF069532.1	2e-020	1
IpSkn02173	BM027926	Cellular factor (p15)	X79805.1	4e-012	1
IpSkn01003	BM027863	Cellular retinoic acid-binding protein	M35523.1	6e-019	2
IpSkn01612	BM027905	Centrin	U37538.1	6e-036	1
IpSkn01920	BM027918	Chloride channel 7	NM_011930.1	3e-034	1
IpSkn02052	BM027921	Choline/ethanolaminephosphotransferase	NM_006090.1	3e-022	1
IpSkn02379	BM027927	CMP-*N*-acetylneuraminic acid synthase	NM_018686.1	8e-007	1
IpSkn00084	BM027819	Collagen, alpha 1 type I	AB008373.1	2e-022	1
IpSkn01263	BM027880	Collagen, alpha 1 type II	NM_001844.1	4e-006	1
IpSkn00110	BM027820	Collagen, alpha 1 type V/XI	AB045975.1	7e-030	1
IpSkn00374	BM027835	Collagen, alpha 2 type I	AB008372.1	2e-090	3
IpSkn00532	BM027838	Connectin	D83008.1	1e-009	1
IpSkn02119	BM027924	Cullin 1	NM_012042.1	1e-050	1
IpSkn02057	BM027922	Cyclooxygenase-2	U97696.1	2e-020	1
IpSkn01279	BM027882	Cyclophilin B	AF071225.1	7e-069	2
IpSkn01482	BM027895	Dedd1	AF232226.1	4e-007	1
IpSkn02130	BM027925	EIG-1	NM_020031.1	2e-008	1
IpSkn01054	BM027867	Electron transfer flavoprotein alpha	NM_000126.1	9e-023	1
IpSkn00171	BM027823	Elongation factor 1 beta	AF001098.1	2e-017	1
IpSkn00005	BM027815	Elongation factor 1 gamma	S69726.1	7e-018	1
IpSkn00315	BM027832	F-box leucine-rich repeat protein 13	AF176354.1	3e-013	1
IpSkn00200	BM027826	FK506, rapamycin-binding protein	D82876.1	9e-007	1
IpSkn00881	BM027855	gC1qBP, a glycoprotein that binds to C1q	NM_007573.1	3e-004	1
IpSkn00998	BM027862	GDP-mannose 4, 6-dehydratase	NM_001500.1	8e-036	1
IpSkn01291	BM027884	Glucose transporter 1A	AF247728.1	1e-025	1
IpSkn00736	BM027848	Glyceraldehyde-3-phosphate dehydrogenase	AB029337.1	e-139	3
IpSkn01360	BM027890	Gu protein, a new subgroup of RNA helicases	U41387.1	1e-018	1
IpSkn00994	BM027861	Heparan glucosaminyl *N*-deacetylase	AF042084.1	1e-007	1
IpSkn01450	BM027893	Keratin type I	L09743.1	1e-083	23
IpSkn02068	BM027923	Keratin type II	AF134850.1	0.0	8
IpSkn00616	BM027842	K-ras	U53782.1	e-180	2
IpSkn02494	BM027930	Large multifunctional protease 7	AF032390.1	5e-077	1
IpSkn00592	BM027840	Leukemia-associated gene	NM_019641.1	1e-018	1
IpSkn01363	BM027891	LIM-domain protein CRP1	Z28333.1	2e-029	1
IpSkn01170	BM027874	MCT-1, a novel oncogene in T-cell malignancy	NM_014060.1	4e-032	1
IpSkn01356	BM027889	Myosin heavy chain IIB	AJ278733.1	2e-010	1
IpSkn01809	BM027914	Myosin light chain 1	L38596.1	8e-039	2
IpSkn00024	BM027816	Myosin light chain 2	X07314.1	2e-037	4
IpSkn01341	BM027887	Myosin light chain 3	D85141.1	4e-098	4
IpSkn00526	BM027837	NADH dehydrogenase Fe-S protein 4	NM_010887.1	1e-018	1
IpSkn00720	BM027847	NADH ubiquinone oxidoreductase	J02877.1	1e-008	1
IpSkn01307	BM027885	N-terminal acetyltransferase complex ard1	NM_016100.1	2e-004	1
IpSkn01452	BM027894	Nucleolar protein KKE/D repeat	NP_006383.1	5e-052	2

(continued overleaf)

Table 3 (*continued*)

Clone #	Accession number	Putative identification	Homologous sequence	Probability	Frequency
IpSkn00050	BM027818	NUDT4 gene	AF191653.1	4e-024	1
IpSkn01145	BM027871	Oligophrenin-1	AJ248245.1	3e-010	2
IpSkn01004	BM027864	Oracle 2 protein	AF228058.1	2e-006	1
IpSkn01505	BM027897	Osteonectin precursor	AF077327.1	8e-084	1
IpSkn00964	BM027860	p21-activated protein kinase I	U46915.1	3e-076	1
IpSkn02481	BM027929	p53 regulated PA26 nuclear protein	NM_014454.1	4e-005	1
IpSkn01243	BM027879	Parvalbumin	AF180888.1	9e-064	7
IpSkn01498	BM027896	Pescadillo gene	U77627.1	9e-028	1
IpSkn01989	BM027919	Plexin 6	NM_019587.1	1e-034	1
IpSkn00193	BM027824	Processing peptidase beta subunit	L12965.1	6e-048	1
IpSkn01041	BM027866	Proline oxidase 2	NM_016335.1	2e-011	2
IpSkn01771	BM027912	Propionyl-CoA carboxylase alpha subunit	AF080073.1	6e-015	1
IpSkn02047	BM027920	Prostaglandin E2 receptor EP4 subtype	AF177934.1	2e-020	1
IpSkn01775	BM027913	Prostaglandin endoperoxide synthase-2	AF158373.1	7e-018	1
IpSkn01096	BM027870	Proteasome subunit beta 7	AF155581.1	2e-043	1
IpSkn00132	BM027821	Protein kinase C beta	X04795.1	5e-017	1
IpSkn01351	BM027888	Ptg-12 protein	X97303.1	6e-020	1
IpSkn00604	BM027841	RAB11B, member RAS oncogene family	NM_004218.1	1e-026	3
IpSkn00664	BM027844	Retinol binding protein	AJ236884.1	2e-023	1
IpSkn01915	BM027917	Rho GDI, GDP dissociation inhibitor	X52689.1	8e-020	1
IpSkn01196	BM027876	Ribonucleotide reductase	AB036063.1	1e-032	1
IpSkn02454	BM027928	Ribosomal protein L23	AF266222.1	e-112	4
IpSkn00169	BM027822	Ribosomal protein S17	NM_009092.1	3e-027	1
IpSkn01753	BM027911	Ribosomal protein S26	X63389.1	2e-041	3
IpSkn00035	BM027817	Ring finger protein 3	NM_006315.1	3e-047	1
IpSkn01287	BM027883	RNA binding protein p37	AB046618.1	1e-053	1
IpSkn01214	BM027878	Ryudocan/syndecan 2	NM_013082.1	3e-013	1
IpSkn00276	BM027830	Sarcoplasmic/ER calcium ATPase	U65229.1	4e-055	1
IpSkn01147	BM027872	Septin 2	AF179995.1	2e-062	1
IpSkn00256	BM027828	Seryl-tRNA synthetase	AF297553.1	2e-011	1
IpSkn00904	BM027857	Sid329, mouse homolog of Sop2p-like protein	NM_019767.1	8e-008	1
IpSkn00340	BM027833	Small zinc finger-like protein	AF150107.1	2e-043	1
IpSkn01539	BM027900	SPARC, acidic calcium-binding glycoprotein	U25721.1	3e-092	3
IpSkn01525	BM027899	S-phase kinase-associated protein 1A	NM_006930.1	2e-034	1
IpSkn00754	BM027850	TA2 t-complex polypeptide 1	AF164028.1	e-179	1
IpSkn01269	BM027881	Taurine transporter	AB006986.1	1e-062	1
IpSkn01063	BM027868	TGF-beta receptor interacting protein 1	U36764.1	8e-014	1
IpSkn00347	BM027834	Thioredoxin	NM_019913.1	1e-010	2
IpSkn00900	BM027856	Thrombospondin 2	NM_011581.1	5e-036	1
IpSkn00926	BM027859	Titin	NM_003319.1	1e-005	2
IpSkn00197	BM027825	Transcription factor IIE	Z14131.1	1e-024	1
IpSkn02518	BM027932	Transcription factor Smad2	AF229022.1	e-116	1
IpSkn01741	BM027910	Transcriptional regulator SIN3a	NM_011378.1	3e-019	1
IpSkn01619	BM027906	Translation initiation factor 3 subunit 3	NM_003756.1	3e-029	1
IpSkn02528	BM027933	Translation initiation factor 3 subunit 6	Q64252	e-140	1
IpSkn00694	BM027846	Translation initiation factor 3 subunit 9	NM_003751.1	2e-061	1
IpSkn01206	BM027877	TRAP-complex gamma subunit	Z14030.1	5e-060	1
IpSkn00638	BM027843	Tropomyosin alpha	AF180892.1	e-140	1
IpSkn01066	BM027869	Tropomyosin beta	Z66490.1	e-105	1
IpSkn01544	BM027901	Troponin C	AF180890.1	e-172	2
IpSkn00749	BM027849	Troponin I	U20111.1	3e-035	7
IpSkn00415	BM027836	Troponin T	AF180889.1	1e-078	1
IpSkn01335	BM027886	Twisted gastrulation protein	AJ297392.1	2e-035	1
IpSkn01885	BM027916	Tyrosine kinase 9-like protein	NM_007284.1	1e-008	1
IpSkn01584	BM027903	U2 small nuclear RNA	M12856.1	9e-017	1
IpSkn01509	BM027898	Ubiquitin specific protease 9	NM_004652.1	2e-041	3
IpSkn00818	BM027853	Ubiquitin-conjugating enzyme 7	NM_009456.1	3e-080	1
IpSkn00305	BM027831	Ubiquitin-conjugating enzyme 9	AF128240.1	4e-098	1
IpSkn01180	BM027875	Ubiquitin-conjugating enzyme E2D 1	NM_003338.1	8e-058	1

over five times. In spite of the fact that the vast majority of known genes were sequenced only once, gene expression in the skin tissue of channel catfish is highly polarized. A small number of genes accounted for a large proportion of tran-

Table 4
New channel catfish genes identified by BLASTX similarity comparison and their expressions

Clone #	Accession number	Putative identification	Homologous sequence	Probability	Frequency
IpSkn00928	BM027964	Adenylate kinase 3	NP_037542.1	6e-061	2
IpSkn00241	BM027940	Adipophilin	Q99541	5e-007	1
IpSkn02527	BM028020	ADP-ribosylation factor-like 4	NP_005729.1	6e-083	1
IpSkn00406	BM027945	Alcohol-steroid dehydrogenase	CAB02087.1	4e-005	1
IpSkn01730	BM027997	Alpha-1,6-mannosyl-glycoprotein beta-1, 2-*N*-acetylglucosaminyltransferase	A57044	1e-009	1
IpSkn00106	BM027937	AMP deaminase 1 (isoform M)	NP_000027.1	2e-018	1
IpSkn00052	BM027934	Antigen thy1	1103300A	9e-004	1
IpSkn01383	BM027979	ATP synthase oligomycin sensitivity conferral protein	Q06647	8e-066	2
IpSkn01541	BM027988	B-cell receptor-associated protein 31	S49265	4e-037	1
IpSkn00363	BM027943	Beta-1, 6-*N*-acetylglucosaminyltransferase	AAC52925.1	1e-012	1
IpSkn01361	BM027977	BRCA1 associated protein	NP_006759.1	4e-030	1
IpSkn01289	BM027974	CC chemokine CCL1	AAF17560.1	2e-004	1
IpSkn01015	BM027967	Chandra protein	AAG09739.1	7e-009	3
IpSkn02244	BM028012	Chloride intracellular channel protein 2	O15247	2e-083	1
IpSkn01387	BM027980	ClpX-like protein	CAA06933.2	1e-017	1
IpSkn01362	BM027978	Cofilin 2	NP_031714.1	7e-029	2
IpSkn00425	BM027947	Collagen alpha 1 type XI	S28791	3e-010	1
IpSkn01684	BM027995	Collagen alpha 1 type XII	NP_031756.1	2e-062	2
IpSkn00629	BM027951	Collagen alpha 3 type VI	P15989	3e-017	1
IpSkn00620	BM027950	Complement C1q A chain precursor	NP_057075.1	2e-026	2
IpSkn01059	BM027969	Complement C1q B chain precursor	I49560	5e-020	1
IpSkn01810	BM028001	Complement C4A	BAB03284.1	3e-019	1
IpSkn01018	BM027968	Cytochrome Bc1 complex chain J	1BCC	9e-019	1
IpSkn00639	BM027952	Cytochrome c oxidase subunit IV isoform 1	AAF79933.1	1e-049	1
IpSkn00921	BM027963	Cytochrome c oxidase subunit IV isoform 2	AAF79934.1	8e-019	1
IpSkn00947	BM027965	Cytochrome c oxidase subunit VIIB precursor	P56393	7e-013	1
IpSkn00426	BM027948	Cytochrome P450	AAD54014.1	1e-020	1
IpSkn02585	BM028023	Cytokine-inducible SH2 protein 3	JC5761	1e-056	1
IpSkn01747	BM027998	Dap1b	AAF66958.1	2e-023	1
IpSkn01872	BM028003	Dipeptidyl-peptidase IV	CAA70136.1	3e-031	1
IpSkn01165	BM027972	Endobrevin	AAD33595.1	1e-018	1
IpSkn00917	BM027962	Epididymal secretory protein E1 precursor	Q15668	1e-007	1
IpSkn02154	BM028011	Ethanolamine kinase	NP_061108.2	4e-023	1
IpSkn00836	BM027961	Fibrinogen-like protein	NP_006673.1	1e-012	1
IpSkn01474	BM027983	Galectin-4	NP_037107.1	9e-035	4
IpSkn01933	BM028005	Galectin-6	AAC27244.1	2e-018	1
IpSkn00650	BM027955	Glutaryl-CoA dehydrogenase	AAB24225.1	7e-018	1
IpSkn01991	BM028008	Glycosyltransferase AD-017	NP_060916.1	3e-090	1
IpSkn01796	BM028000	Growth factor receptor bound protein 10	Q13322	3e-032	1
IpSkn02529	BM028021	GRPE protein homolog 2 precursor	AAC31364.1	6e-039	1
IpSkn01711	BM027996	Heme-binding protein	NP_055135.1	6e-017	2
IpSkn02381	BM028015	Human seven transmembrane protein TM7SF3	NP_057635.1	1e-004	1
IpSkn01487	BM027985	Interferon-induced transmembrane protein 2	Q01629	6e-010	1
IpSkn00814	BM027960	J domain protein 1	BAA94963.1	2e-044	1
IpSkn02125	BM028010	Kinase suppressor of ras	NP_038599.1	2e-007	1
IpSkn01349	BM027976	LAFPTPase	CAA10199.1	2e-062	1
IpSkn01677	BM027994	Leptin receptor gene-related protein	O89013	2e-034	1
IpSkn00640	BM027953	Leucine zipper protein	BAA05376.1	1e-022	1
IpSkn01389	BM027981	M protein	AAB84054.1	2e-005	1
IpSkn00215	BM027939	MAT8 protein	CAA63606.1	2e-012	1
IpSkn01966	BM028007	Matrix Gla protein	AAD28354.1	3e-006	1
IpSkn01926	BM028004	Mcl-1a	AAF66961.1	2e-033	1
IpSkn01009	BM027966	Myomesin 2	NP_032690.1	8e-013	1
IpSkn00417	BM027946	Myosin X	NP_062345.1	7e-021	1
IpSkn01635	BM027992	Natural killer-tumor recognition protein	NP_005376.2	3e-032	1
IpSkn00055	BM027935	Noggin	AAD09176.1	4e-018	1
IpSkn00388	BM027944	Nuclear factor, interleukin 3 regulated	NP_005375.1	6e-036	1
IpSkn01846	BM028002	Numb-binding protein LNXp80	T09457	2e-011	1
IpSkn00696	BM027957	p53 apoptosis-associated target	AAF64306.1	2e-025	1

(continued overleaf)

Table 4 (*continued*)

Clone #	Accession number	Putative identification	Homologous sequence	Probability	Frequency
IpSkn02359	BM028013	Pancreatic secretory trypsin inhibitor	P00996	4e-009	1
IpSkn01657	BM027993	Phosphatidylserine-binding protein	NP_004648.1	1e-013	1
IpSkn01065	BM027970	Phospholemman precursor	P56513	4e-004	1
IpSkn00242	BM027941	Placental protein 11	NP_032928.1	8e-025	2
IpSkn00742	BM027958	Pleckstrin 2	NP_038766.1	2e-016	1
IpSkn01494	BM027986	Predicted osteoblast protein	NP_055703.1	7e-041	1
IpSkn02375	BM028014	Prenylated RAB acceptor 1	AAD17296.1	3e-023	1
IpSkn01260	BM027973	Prolidase	BAB11685.1	1e-020	1
IpSkn00644	BM027954	Proteinase activated receptor 2 precursor	Q63645	1e-005	2
IpSkn01629	BM027991	Rat Chp, a homologue of the GTPase Cdc42Hs	AAC69198.1	8e-004	1
IpSkn01480	BM027984	Rhophilin	AAD31273.1	2e-008	1
IpSkn01789	BM027999	SARA protein	T17457	2e-005	1
IpSkn00350	BM027942	Selenoprotein P	NP_062065.1	3e-042	1
IpSkn01949	BM028006	Sex-regulated protein janus-a	AAF80759.1	3e-016	1
IpSkn02459	BM028016	SLA/LP autoantigen	AAG00491.1	1e-075	1
IpSkn01086	BM027971	Small inducible cytokine A14	Q16627	9e-005	1
IpSkn02479	BM028017	Small inducible cytokine A19	NP_006265.1	9e-009	3
IpSkn02577	BM028022	Small inducible cytokine B	NP_032625.1	8e-005	1
IpSkn00679	BM027956	Sorcin	NP_003121.1	4e-071	2
IpSkn00143	BM027938	Stefin C	P35478	2e-008	7
IpSkn02497	BM028018	SUMO-1-specific protease	AAF04852.1	2e-008	1
IpSkn01566	BM027990	*Takifugu rubripes* gag polyprotein	AAC33525.1	9e-007	1
IpSkn01439	BM027982	Tetracycline transporter-like protein	NP_001111.1	1e-036	1
IpSkn02063	BM028009	Tetratricopeptide repeat domain 4	NP_004614.1	3e-022	1
IpSkn00752	BM027959	Threonyl-tRNA synthetase	NP_003182.1	8e-040	1
IpSkn00063	BM027936	Transmembrane protein (63kD)	NP_006816.1	2e-015	1
IpSkn01542	BM027989	Tumor suppressor protein p53	AAD34212.1	3e-014	1
IpSkn01523	BM027987	Ubiquinol cytochrome c reductase (complex III, subunit II)	NP_003357.1	2e-062	1
IpSkn01308	BM027975	Ubiquinol cytochrome c reductase (complex III, subunit X)	CCBO17	2e-015	1
IpSkn00507	BM027949	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2	NP_064409.1	3e-010	2
IpSkn02587	BM028024	Zinc finger protein C3H-3	AAD24209.1	6e-028	1
IpSkn02513	BM028019	Zinc finger transcription factor	AAC05500.1	2e-022	1

scripts in the skin. The most abundantly expressed gene in the skin was ictacalcin accounting for almost 4.8% of the 1909 EST clones sequenced. This high expression level is more than double the representation of the most abundantly expressed beta-actin gene (~2%) in the head kidney (Cao et al., 2001). The other most abundantly expressed genes included creatine kinase (1.2%), keratin type I (1.2%), dopamine receptor (0.9%), and six ribosomal protein genes: S2 (2.0%), L11 (1.6%), L41 (1.6%), S20 (1.1%), S9 (1.0%), and L35 (0.8%). Together, the 10 most abundantly expressed genes in the skin accounted for 16.2% of all clones (Table 1). This expression profile in the channel catfish skin is more polarized than in its brain or head kidney, where the top ten most abundantly expressed genes accounted for 13.2 and 9%, respectively (Ju et al., 2000; Cao et al., 2001). Part of this polarization was also attributed to the high levels of expression of ictacalcin and ribosomal protein genes in the skin, presumably due to higher levels of translational activities (Patterson et al., 2002; Karsi et al., 2002).

When combined into functional groups, the most abundantly expressed genes in the channel catfish skin were calcium-binding proteins (Table 2). Partly this was due in part to the most abundantly expressed ictacalcin, but S100-like calcium binding protein was also expressed at a high level (0.7%). Other highly expressed groups were ribosomal proteins, cytoskeleton genes, kinases and phosphatases, receptors, and mitochondrial genes (Table 2).

3.3. Newly established orthologs from the ESTs of the channel catfish skin

New orthologs were established for 119 unique ESTs by BLASTN searches (Table 3). Those whose identities could not be determined by BLASTN searches were subjected to BLASTX searches by which an additional 91 new orthologs were established (Table 4). These genes represent newly identified genes in channel catfish that show sequence similarities to known genes from other organisms. They should be useful for the development of cDNA microarrays for functional genomic research in catfish, and perhaps in other fishes as well. The proportion of known genes obtained in the skin was comparable to that from channel catfish brain and head kidney (Ju et al., 2000; Cao et al., 2001). The fact that the

majority of EST clones could be identified by similarity comparisons suggests that high-quality EST analysis is an efficient way for gene annotation in less-well studied species. Similar to the situation in the brain and head kidney, 54 (2.8%) clones representing 51 unique genes showed significant similarities to known sequences of unknown functions from model systems such as *Homo sapiens*, *Mus musculus*, *Caenorhabditis elegans*, *Bos Taurus*, *Macaca fascicularis*, and *Arabidopsis thaliana* (Table 5). Although functions are not yet known, their conservation in fish demonstrated the existence of many gene families through evolution. Once a gene is characterized in any one of these species, compara-

Table 5
Channel catfish genes orthologs to known ESTs of unknown functions from other organisms

Clone #	Accession number	Putative identification	Homologous sequence	Probability	Frequency
IpSkn01554	BM028058	*Arabidopsis thaliana* putative protein	CAB94144.1	2e-025	1
IpSkn00772	BM028042	*Bos taurus* 50 kDa protein	U04706.1	2e-013	1
IpSkn01518	BM028056	*Caenorhabditis elegans* hypothetical protein	T25472	2e-041	1
IpSkn01956	BM028068	*Caenorhabditis elegans* hypothetical protein	C13C4.5	9e-032	1
IpSkn02101	BM028073	*Caenorhabditis elegans* hypothetical protein	T21337	6e-021	1
IpSkn01545	BM028057	*Drosphila melanogaster* CG11110 gene product	AAF57438.1	4e-047	2
IpSkn00292	BM028034	*Drosphila melanogaster* CG13384 gene product	AAF52663.1	3e-007	1
IpSkn00278	BM028033	*Drosphila melanogaster* CG15016 gene product	AAF47889.1	9e-019	1
IpSkn01060	BM028050	*Drosphila melanogaster* CG7623 gene product	AAF55438.1	4e-022	1
IpSkn01959	BM028069	*Homo sapiens* chromosome 19 clone CTC-429P9	AC024075.4	3e-005	1
IpSkn01749	BM028059	*Homo sapiens* clone RP1-34M23	AL121988.1	3e-054	1
IpSkn00259	BM028032	*Homo sapiens* clone RP5-1162C3	AL133335.29	3e-007	1
IpSkn00504	BM028036	*Homo sapiens* DKFZP564B167 protein	NM_015415.1	3e-008	1
IpSkn01381	BM028054	*Homo sapiens* DKFZP566C243 protein	NM_015388.1	1e-033	1
IpSkn00178	BM028029	*Homo sapiens* EST homologue	CAA75443.1	9e-017	1
IpSkn02506	BM028075	*Homo sapiens* HCDI protein	NP_064580.1	9e-040	1
IpSkn02133	BM028074	*Homo sapiens* HSPC306	AAF28984.1	3e-005	1
IpSkn00518	BM028037	*Homo sapiens* hypothetical protein AF151083	NP_057587.1	6e-024	1
IpSkn00333	BM028035	*Homo sapiens* hypothetical protein AK001475	NP_061940.1	7e-059	1
IpSkn01917	BM028067	*Homo sapiens* hypothetical protein AL137199	CAB69909.1	8e-011	1
IpSkn01457	BM028055	*Homo sapiens* hypothetical protein AL365515	CAB97211.1	3e-028	1
IpSkn00184	BM028030	*Homo sapiens* hypothetical protein DKFZp564F052.1	T08691	8e-028	1
IpSkn01787	BM028062	*Homo sapiens* hypothetical protein FLJ10375	NP_060545.1	3e-049	1
IpSkn01029	BM028048	*Homo sapiens* hypothetical protein FLJ10482	NM_018107.1	9e-011	2
IpSkn02044	BM028071	*Homo sapiens* hypothetical protein FLJ10509	NP_060589.1	1e-052	1
IpSkn00092	BM028027	*Homo sapiens* hypothetical protein FLJ10788	NM_018221.1	2e-041	1
IpSkn01768	BM028061	*Homo sapiens* hypothetical protein FLJ20217	NP_060186.1	7e-009	1
IpSkn00870	BM028046	*Homo sapiens* hypothetical protein FLJ20479	NP_060308.1	2e-050	1
IpSkn00787	BM028043	*Homo sapiens* hypothetical protein FLJ20485	NM_019042.1	3e-011	1
IpSkn01094	BM028051	*Homo sapiens* hypothetical protein FLJ20624	NP_060376.1	3e-047	1
IpSkn00825	BM028045	*Homo sapiens* hypothetical protein PRO1580	NP_060972.1	1e-034	1
IpSkn01874	BM028065	*Homo sapiens* hypothetical protein S164	P49756	8e-006	1
IpSkn01751	BM028060	*Homo sapiens* KIAA0193 protein	NM_014766.1	3e-005	1
IpSkn00761	BM028041	*Homo sapiens* KIAA0332 protein	BAA20790.1	9e-010	1
IpSkn00799	BM028044	*Homo sapiens* KIAA0766 protein	AAD15420.1	5e-023	1
IpSkn01858	BM028063	*Homo sapiens* KIAA0782 protein	BAA34502.1	1e-005	1
IpSkn01863	BM028064	*Homo sapiens* KIAA0796 protein	BAA34516.1	2e-004	1
IpSkn02050	BM028072	*Homo sapiens* KIAA0819 protein	BAA74842.1	3e-028	1
IpSkn01037	BM028049	*Homo sapiens* KIAA0974 protein	BAA76818.1	5e-016	1
IpSkn00538	BM028038	*Homo sapiens* KIAA1408 protein	AB037829.1	1e-009	1
IpSkn02040	BM028070	*Homo sapiens* KIAA1593 protein	BAB13419.1	4e-011	1
IpSkn01340	BM028053	*Homo sapiens* locus AC004382	AAC24311.1	3e-038	1
IpSkn00049	BM028026	*Homo sapiens* locus AF007170	AAC39582.1	7e-036	1
IpSkn00039	BM028025	*Homo sapiens* sequence 232 from patent WO9946375	AX017997.1	2e-016	1
IpSkn00728	BM028039	*Homo sapiens* uncharacterized hematopoietic stem/progenitor cells protein MDS032	NM_018467.1	1e-004	2
IpSkn00958	BM028047	*Homo sapiens* uncharacterized hematopoietic stem/progenitor cells protein MDS033	NP_060938.1	1e-027	1
IpSkn00255	BM028031	*Homo sapiens* CGI-55 protein	AAD34050.1	4e-016	1
IpSkn01122	BM028052	*Macaca fascicularis* brain cDNA	AB047883.1	4e-018	1
IpSkn00741	BM028040	*Macaca fascicularis* hypothetical protein	BAB12304.1	9e-092	1
IpSkn01882	BM028066	*Mus musculus* brain cDNA clone MNCb-1308	AB041652.1	1e-011	1
IpSkn00176	BM028028	*Pseudomonas aeruginosa* hypothetical protein AE004696	AAG06067.1	3e-012	1

Table 6
Microsatellite-containing cDNA clones from the skin cDNA library[a]

Clone #	Gene identity	Microsatellite repeats
IpSkn00772	*Bos taurus* 50 kDa protein	$(GT)_{12}$
IpSkn00701	Calpain	$(TTA)_{12}$
IpSkn00110	Collagen, alpha 1 type V/XI	$(GTTT)_5$
IpSkn01810	Complement C4A	$(GAGAC)_4$
IpSkn02115	Dopamine receptor	$(CA)_{35}/(CA)_{25}$
IpSkn00292	*Drosphila melanogaster* CG13384 gene product	$(GT)_{11}$
IpSkn02130	EIG-1	$(AC)_{14}$, $(AC)_{10}$
IpSkn00881	gC1qBP, a glycoprotein that binds to C1q.	$(AAT)_8$
IpSkn01291	Glucose transporter 1A	$(ATTT)_5$, $(AT)_{14}$
IpSkn01749	*Homo sapiens* DNA sequence-clone RP1-34M23	$(GA)_9$
IpSkn01457	*Homo sapiens* hypothetical protein AL365515	$(TC)_{18}$
IpSkn00728	*Homo sapiens* uncharacterized hematopoietic stem/progenitor cells protein MDS032	$(CA)_{10}$
IpSkn00404	Ictacalcin	$(GTTT)_5$
IpSkn01336	Keratin type I	$(TGG)_6$
IpSkn01754	LINE-like DNA	$(CA)_{15}$
IpSkn00741	*Macaca fascicularis* hypothetical protein	$(CT)_{10}$
IpSkn00215	MAT8 protein	$(TTA)_8A(TTA)_6$
IpSkn01966	Matrix Gla protein	$(ATT)_6$
IpSkn01251	Oligophrenin-1	$(GT)_{10}$
IpSkn01004	Oracle 2 protein	$(CA)_8$
IpSkn01286	Profilin	$(CA)_{11}$, $(CA)_{22}$, $(CA)_{10}/(CA)_{13}$, $(CA)_{23}$, $(CA)_{10}$
IpSkn01576	Protein-tyrosine-phosphatase IF1	$(CA)_{16}$, $(CA)_{24}/(CA)_{11}$, $(CA)_{27}$
IpSkn01480	Rhophilin	$(GA)_9GT(GA)_8GT(GA)_{15}$, $(CA)_{15}$
IpSkn00737	Ribosomal protein S16	$(CT)_8/(CT)_9/(CT)_{10}/(CT)_{11}$
IpSkn01665	S100-calcium binding protein A14	$(CT)_{15}/(CT)_{16}/(CT)_{17}$
IpSkn01789	SARA protein	$(TG)_9(TC)_{21}$
IpSkn00350	Selenoprotein P	$(AC)_{25}$
IpSkn00347	Thioredoxin	$(AC)_{24}$, $(GT)_9$
IpSkn00752	Threonyl-tRNA synthetase	$(TC)_{26}$, $(TG)_8(AG)_7$
IpSkn01133	Translation initiation factor 5A	$(GA)_8$
IpSkn00563	Urokinase receptor	$(CA)_{24}/(CA)_{29}$
IpSkn00002	Unknown	$[(GT)4GA]_5$, $(AC)_{16}$, $(AC)_{20}$
IpSkn00012F	Unknown	$(ATT)_8$
IpSkn00019	Unknown	$(CAA)_7$
IpSkn00020	Unknown	$(ATT)_{18}$, $(CA)_{11}$
IpSkn00030	Unknown	$(GA)_{20}$
IpSkn00051	Unknown	$(GAA)_{11}$
IpSkn00054	Unknown	$(TAA)_{14}$
IpSkn00095F	Unknown	$(CA)_8$
IpSkn00097	Unknown	$(CT)_{18}$
IpSkn00119	Unknown	$(GTT)_7$
IpSkn00237	Unknown	$(AT)_{33}$
IpSkn00245	Unknown	$(CA)_{17}$
IpSkn00260	Unknown	$(TTA)_7(ATT)_6$
IpSkn00265	Unknown	$(ATTTT)_4$, $(AAC)_4$
IpSkn00287	Unknown	$(TA)_{29}$
IpSkn00345	Unknown	$(ATT)_8$
IpSkn00369	Unknown	$(GA)_8$
IpSkn00376	Unknown	$(ATT)_{10}/(ATT)_{14}$
IpSkn00399	Unknown	(GT)13
IpSkn00584	Unknown	$(AAATT)_4$
IpSkn00642	Unknown	$(TA)_{47}$
IpSkn00695	Unknown	$(GGA)_{10}$
IpSkn00714	Unknown	$(GT)_{18}$, $(ATT)_{11}$, $(ATT)_9$
IpSkn00753	Unknown	$(TC)_{14}$
IpSkn00767	Unknown	$(TA)_{48}$
IpSkn00780	Unknown	$(TA)_{10}$
IpSkn00784	Unknown	$(TA)_{21}$
IpSkn00800	Unknown	$(TA)_{33}$
IpSkn00845	Unknown	$(GA)_{35}$
IpSkn00856	Unknown	$(CA)_{28}$

Table 6 (*continued*)

Clone #	Gene identity	Microsatellite repeats
IpSkn00863	Unknown	$(TAA)_9$
IpSkn00876	Unknown	$(TC)_{13}$
IpSkn01035	Unknown	$(GA)_{17}$
IpSkn01056	Unknown	$(GT)_8$
IpSkn01067	Unknown	$(TA)_5A(AT)_5$
IpSkn01284	Unknown	$(CA)_{29}$
IpSkn01285	Unknown	$(TTTA)_5$, $(ATTT)_{10}$
IpSkn01327	Unknown	$(GTCT)_6$
IpSkn01348	Unknown	$(TA)_{19}TGTATGTG(TA)_{15}(GT)_8$
IpSkn01368	Unknown	$(ATT)_{13}$, $(TTA)_9$
IpSkn01392	Unknown	$(TTTTA)_4$
IpSkn01579	Unknown	$(AC)_{28}$
IpSkn01581	Unknown	$(TA)_{36}$
IpSkn01603	Unknown	$(CA)_{10}$, $(TTA)_{27}$
IpSkn01626	Unknown	$(TA)_{41}$
IpSkn01644	Unknown	$(TA)_{37}$
IpSkn01658	Unknown	$(ATT)_7$
IpSkn01687	Unknown	$(CA)_{10}$
IpSkn01710	Unknown	$(GT)_{14}$
IpSkn01765	Unknown	$(CA)_{12}$
IpSkn01766	Unknown	$(ATT)_7$
IpSkn01778	Unknown	$(CA)_{10}A(AC)_{23}/(CA)_{13}A(AC)_{19}/(CA)_{11}A(AC)_{22}$
IpSkn01780	Unknown	$(ATT)_8$
IpSkn01791	Unknown	$(CA)_{10}$
IpSkn01814	Unknown	$(TA)_{13}$
IpSkn02061	Unknown	$(CA)_{10}$
IpSkn02080	Unknown	$(GT)_{12}$
IpSkn02472	Unknown	$(CT)_{14}$

[a] Comma between the microsatellite repeats separates different microsatellites in the same cDNA, while slash separates same microsatellite repeats in different cDNAs.

tive functional genomics will allow annotation to these orthologous genes.

3.4. Microsatellite containing genes

Among 1909 sequenced cDNA clones, 89 unique EST clones harbor microsatellite sequences (Table 6). These microsatellites can be potentially useful for genomic mapping if they are polymorphic. We have found that targeting microsatellite regions within cDNAs is an efficient way to develop type I molecular markers representing genes of known functions (O'Brien, 1991). Because of the evolutionary conservation, mutation rates within gene-coding sequences are lower than those in non-coding genomic sequences. As a result, type I polymorphic markers are often more difficult to be identified. By tagging the highly polymorphic microsatellites to known genes, the efficiency for the development of type I markers can be dramatically enhanced. The major objective of this work was to develop EST resources and, therefore, we did not attempt to characterize polymorphism of these microsatellite clones. However, by cDNA alignment, we have observed seven unique polymorphic microsatellite loci represented by several cDNA clones that each showed different numbers of repeat units. Two alleles for dopamine receptor [$(CA)_{35}$/$(CA)_{25}$], two alleles for profilin [$(CA)_{11}$, $(CA)_{22}$/$(CA)_{13}$, $(CA)_{23}$], four alleles for ribosomal protein S16 [$(CT)_8$/$(CT)_9$/$(CT)_{10}$/$(CT)_{11}$], three alleles for S100-calcium binding protein A14 [$(CT)_{15}$/$(CT)_{16}$/$(CT)_{17}$], two alleles for urokinase receptor [$(CA)_{24}$/$(CA)_{29}$], two alleles for protein-tyrosine-phosphatase [$(CA)_{16}$, $(CA)_{24}$/$(CA)_{11}$, $(CA)_{27}$], two alleles were observed for unknown IpSkn00376 [$(ATT)_{10}$/$(ATT)_{14}$], and three alleles were observed for unknown IpSkn01778 [$(CA)_{10}A(AC)_{23}$/$(CA)_{13}A(AC)_{19}$/$(CA)_{11}A(AC)_{22}$]. Therefore, at least seven polymorphic ESTs were identified during EST analysis without additional bench work, adding to the benefits of EST analysis.

3.5. Conclusions

Transcriptome analysis is an efficient alternative to genomic sequencing analysis. Such analysis of overall transcripts of tissues and organs not only produces large numbers of ESTs, but also generates expression profiles by using non-normalized cDNA libraries. EST cataloging and profiling will provide the basis for functional genomics research (Mekhedov et al., 2000). In the present work, we established orthologs for 496 genes plus sequence tags for additional 478 unknown clones. This demonstrated the rapid discovery of large numbers of genes. These ESTs should be useful for functional genomics as well as comparative genomic studies and genome evolution. It is interesting to note that EST

analysis may be one of the most efficient ways for the development of polymorphic type I markers as well as through tagging of microsatellites existing within cDNAs. The ESTs will also be valuable molecular reagents for the production of microarrays. In particular, the application of cDNA microarrays may facilitate research attempting to answer questions concerning immune responses and other protective responses of channel catfish skin tissue upon infection of pathogens.

Acknowledgements

This project was supported by a grant from US Department of Agriculture National Research Initiative Competitive Grants Program (USDA-NRICGP) to Z.L. and R.D. (98-35205-6738), by Auburn University Competitive BioGrant (Biogrant J. Liu 99). We appreciate the support of Auburn University Department of Fisheries and Allied Aquacultures, College of Agriculture and the Vice President for Research for their matched funds to USDA National Research Initiative Equipment Grants to Z.L. (98-35208-6540, 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., Venter, J.C., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656.

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

Azam, A., Paul, J., Sehgal, D., Prasad, J., Bhattacharya, S., Bhattacharya, A., 1996. Identification of novel genes from *Entamoeba histolytica* by expressed sequence tag analysis. Gene 181, 113–116.

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

Buchmann, K., 1999. Immune mechanisms in fish skin against monogeneans-a model. Folia Parasitol. 46, 1–9.

Burke, J., Wang, H., Hide, W., Davison, D.B., 1998. Alternative gene form discovery and candidate gene selection from gene indexing projects. Genome Res. 8, 276–290.

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

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

Franco, G.R., Adams, M.D., Bento, S.M., Simpson, A.J.G., Venter, J.C., Pena, S.D.J., 1995. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. Gene 152, 141–147.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

Hudson, T.J., Stein, L.D., Gerety, S.S., Ma, J., Castle, A.B., Silva, J., Slonim, D.K., Baptista, R., Kruglyak, L., Xu, S.H., et al., 1995. An STS-based map of the human genome. Science 270, 1945–1954.

Ju, Z., Karsi, A., Kocabas, A., Patterson, A., Li, P., Cao, D., Dunham, R., Liu, Z.J., 2000. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. Gene 261, 373–382.

Karsi, A., Patterson, A., Feng, J., Liu, Z., 2002. Translational machinery of channel catfish: I. A transcriptomic approach to the analysis of 32 40S ribosomal protein genes and their expression, Gene, in review.

Lee, C.K., Weindruch, R., Prolla, T.A., 2000. Gene-expression profile of the aging skin in mice. Nat. Genet. 25, 294–297.

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

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

Patterson, A., Karsi, A., Feng, J., Liu, Z., 2002. Translational machinery of channel catfish: II. Complementary DNA and expression of the complete set of 47 60S ribosomal proteins, J. Mol. Evol., in review.

Sambrook, J., Frisch, E.F., Maniatis, T., 1989. Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., 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., 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., et al., 1996. Genome maps 7. The human transcript map. Wall chart. Science 274, 547–562.

USDA, 2000. Aquaculture Outlook, March 13, 2000. USDA Economic Research Service Publication #LDP-AQS-11.

Wang, K., Gan, L., Jeffery, E., Gayle, M., Gown, A.M., Skelly, M., Nelson, P.S., Ng, W.V., Schummer, M., Hood, L., Mulligan, J., 1999. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. Gene 229, 101–108.

Zhang, J., Madden, T.L., 1997. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. Genome Res. 7, 649–656.