Korean J. Ichthyol. 19(4), 257~265, 2007

Expressed Sequence Tags of Expression Profiles of Olive Flounder (*Paralichthys olivaceus*) Testis

Jeong-Ho Lee, Jong-Hyun Kim, Jae Koo Noh, Hyun Chul Kim,
Young-Ok Kim[1], Woo-Jin Kim[1], Kyu-Won Kim[2] and Kyung-Kil Kim*

Genetics and Breeding Research Center, National Fisheries Research and Development Institute, Geoje 656-842, Korea
[1]Biotechnology Research Institute, National Fisheries Research and Development Institute, Busan 619-900, Korea
[2]Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151-742, Korea

We constructed a cDNA library of testis from olive flounder (*Paralichthys olivaceus*) and a total of 248 expressed sequence tag (EST) clones were generated. In order to understand the molecular compositions of the olive flounder testis organs, the expression profiles of the identified clones in the cDNA library were analyzed. Gene annotation procedures and homology searches of the sequenced ESTs were locally done by BLASTX for amino acid similarity comparisons. Of the 248 EST clones, 156 ESTs showed significant homology to previously described genes while 92 ESTs were unidentified or novel. Comparative analysis of the 156 identified ESTs showed that 6 (3.8%) clones were representing 5 unique genes identified as homologous to the previously reported olive flounder ESTs, 100 (64.1%) clones representing 94 unique genes were identified as orthologs of known genes from other organisms, and orthologs were established for 50 (32.1%) clones representing 44 genes of known sequences with unknown functions. Furthermore, the testis library showed a more even distribution of cDNA clones with relatively fewer abundant clones that tend to contribute redundant clones in EST projects; thus, the testis library can supply more unique and novel cDNA sequences in olive flounder EST project.

Key words : Expressed sequence tags (ESTs), expression profile, olive flounder (*Paralichthys olivaceus*), testis

Introduction

Expressed sequence tags (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). Currently, the number of fish-related ESTs in the public databases is still small compared with mammalian sequences and there are relatively few tissue-specific cDNA libraries (Ton *et al.*, 2000). ESTs have been shown to be an excellent and proven method of identification and characterizing novel genes. Furthermore, comprehensive information on steady state mRNA levels

*Corresponding author: k48722@momaf.go.kr

is not known for most known fish transcripts (Virlon *et al.*, 1999).

The olive flounder (*Paralichthys olivaceus*) is one of the most significant fish species in Korea and Japan due to human interests in aquaculture and fisheries. The olive flounder stocks currently used in aquaculture have become appreciably different to their wild counterparts and it seems likely that genetics will play an increasingly more important role in achieving further improvements in the performance of the brood stocks (Lie *et al.*, 1994). Traits that may be amenable to genetic improvement include growth, delayed maturity, sex determination, disease resistance and temperature tolerance. Despite this interest, relatively little information is available about olive flounder genes and their sequences with less than 100 typical nuclear-encoded olive flounder gene sequences currently deposited in the international DNA sequence databases in the dbEST (Boguski *et al.*, 1993). This lack of knowledge may represent one obstacle to the effective use of genetics in aiding both olive flounder aquaculture and conservation activities. Therefore, in order to increase the current database of olive flounder genes, we investigated olive flounder gene identification and expression analysis project following the EST-based strategy (Adams *et al.*, 1991) now commonly used for the identification of large numbers of genes in species of interest (Marra *et al.*, 1998).

In this study, the 5′-terminal DNA sequences of 248 cDNA clones were analyzed and used to search the international protein databases for significant homology to genes previously identified in other species. The results showed that orthologs for 156 genes plus 92 unknown ESTs were established from the olive flounder testis cDNA library. These ESTs will be useful for comparative genomics by determination of their orthologous counterparts through evolution, for mapping by PCR analysis using radiation hybrid panels, and for identification of polymorphic markers in genes of known functions (type I markers).

Materials and Methods

Olive flounders were obtained from Koje Hatchery of National Fisheries Research and Development Institute (NFRDI) and maintained in 6 tons flow-through tank at 12 ± 1°C under a natural photoperiod. Testis tissues from 10 fish were collected and cut into as small pieces as possible. Pooled testis tissues 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).

Total RNA was extracted using the TRIzol reagent (Gibco BRL Life Technologies Ltd., Renfrewshire, UK), and mRNA was enriched by oligo-(dT) cellulose chromatography using the Poly A Tract mRNA isolation kit (Promega, WI, USA). cDNA synthesis was carried out using an oligo-$(dT)^{18}$ primer for the reverse transcription of approximately 5 μg of mRNA, and the library was constructed by directional cloning approach using Stratagene's Uni-ZAP XR cloning systems. Single-pass sequencing of the 5′-termini of selected testis cDNA clones in phagemid form was performed using the ABI 3100 automatic DNA sequencer (PE Applied Biosystems, CA, USA) and the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems).

Bioinformatic analysis was conducted to determine gene identities using GeneMaster software (Ensoltek, Korea). Briefly, vector sequences were removed and database search were limited to ESTs$>$100 bp in length. ESTs were then assembled in clusters of contiguous sequences (contig) using ICAtools program (Parsons, 1995). Gene annotation procedures and homology searches of the sequenced ESTs have been locally done by BLASTX for amino acid similarity comparisons (Altschul *et al.*, 1997). Matches with e-values less than 1.0×10^{-3} were considered to be significant. 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.

Results and Discussion

1. EST sequencing, clustering and similarity search of ESTs

A total of 248 randomly selected clones were single-pass sequenced from the 5′ end, resulting

Table 1. General characteristics of olive flounder testis ESTs

Total cDNA sequenced[a]	248
Average insert size[b]	1.8 ± 0.4 kb
Average EST length	642 bp
EST clusters[c]	16
Singletons[d]	216
Redundancy[e]	9.4%
ESTs with e-value $< 1 \times 10^{-3}$ (matched)	156 (62.9%)
ESTs with e-value $\geq 1 \times 10^{-3}$ (unknown)	92 (37.1%)

[a]Length of sequence used for comparison after editing (inserts <100 base pairs were excluded).
[b]The average insert size was calculated for 20 randomly selected cDNA clones.
[c]ESTs with 90% or greater identity over a 100 bp region were clustered together forming 16 EST clusters.
[d]216 sequences did not sufficiently match any sequence in the data set to allow assembly.
[e]Redundancy=The number of genes/total ESTs.

in the characterization of cDNA clones that were longer than 100 bp after elimination of vector sequence. The number of clones sequenced from the cDNA library, the average size of inserts, and the redundancy of the obtained sequences, are given in Table 1. The average insert size was estimated to be 1.8 ± 0.4 kb by PCR amplification of inserts from 20 randomly selected clones. We used the assembly program ICAtools software (Parsons, 1995) to organize the redundant ESTs into overlapping contigs. The results showed that the 248 testis ESTs were composed of 16 clusters and 216 singletons, suggesting that the overall redundancy of the library was 9.4%.

Due to the comparative lack of identified fish genes and as protein sequences have been shown to be more suitable to detect homology over long periods of evolutionary time (Pearson, 1997), a protein-based homology strategy was used to screen for homologous genes in the international databases. Gene annotation procedures and homology searches of the sequenced ESTs have been locally done by BLASTX for amino acid similarity comparisons. BLASTX analysis was performed using the substitution matrix BLOSUM62 (Henikoff and Henikoff, 1992). The ESTs with significant similarities ($E < 1 \times 10^{-3}$) to known proteins were evaluated to determine if the significant similarities were caused by simple amino acid matches (Table 2). Of the 248 clones, 156 (62.9%) were identified as orthologues of known genes from other organisms. The remaining 92 (37.1%) clones could not be identified by similarity comparisons ($E \geq 1 \times 10^{-3}$); most of them are expected to derive from hitherto uncharacterized or novel genes, whereas some of them may be due to the fact that the homology is too low to detect in the sequenced region. Among the 156 EST clones, 6 (3.8%) clones were representing 5 unique genes identified as homologous to the previously reported olive flounder ESTs, 100 (64.1%) clones representing 94 unique genes were identified as orthologs of known genes from other organisms (Fig. 1). 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. Fifty (32.1%) clones representing 44 unique genes showed significant similarities to known sequences of unknown functions from model systems such as *Homo sapiens, Mus musculus, Rattus norvegicus, Drosophila melanogaster, Anopheles gambiae* and *Caenorhabditis elegans*. 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, comparative functional genomics will allow annotation to these orthologous genes.

2. Expression profile and gene identification

Expression profiles of the EST clones identified from the olive flounder testis is shown in Fig. 2. Of the 248 EST clones identified by BLASTX, 216 (87.1%) were singletons. Although redundancy will increase as the number of sequenced clones increases, the high percentage of singletons indicated that the complexity and coverage of this olive flounder testis cDNA library were good. Among 156 identified distinct known genes, 130 genes (83.3%) were sequenced only once. The remaining 26 genes (16.7%) were sequenced 2 times. These results suggest that the testis library has a more even distribution of cDNA clones with relatively fewer abundant clones that tend to contribute redundant clones in EST projects; thus, the testis library can supply more unique and novel cDNA sequences in a olive flounder EST project.

In the testis library, the *hsp90β* clone is one of the most abundant clones. As a member of the heat shock protein (HSP) family, HSP90 is a component of the inactive and metastable hetero-oligomeric structure of steroid receptors and func-

Table 2. List of identified ESTs from testis cDNA of olive flounder

Clone no.	Putative identification	Closest species	Accession no.	E-value	I[a]	F[b]
germarium3-E08	RIKEN cDNA	*Mus musculus*	XP_109638	0.00E+00	99	1
germarium2-B08	RACK	*Oreochromis niloticus*	O42249	0.00E+00	97	1
germarium2-B12	40S ribosomal protein S2	*Ictalurus punctatus*	Q90YS3	0.00E+00	97	1
germarium3-F07	40S ribosomal protein S3	*Ictalurus punctatus*	Q90YS2	0.00E+00	97	1
germarium2-G02	muscle actin	*Lethenteron japonicum*	BAB19361	0.00E+00	97	1
germarium3-A06	E2 binding protein	*Mus musculus*	XP_134935	0.00E+00	97	1
germarium3-E11	chaperonin-containing T-complex protein	*Danio rerio*	AAM34673	0.00E+00	93	1
germarium1-D06	unknown	*Homo sapiens*	AAH09256	0.00E+00	90	1
germarium3-G11	cullin 3	*Mus musculus*	NP_057925	0.00E+00	89	1
germarium2-H02	COP9 homolog subunit 5	*Homo sapiens*	NP_006828	0.00E+00	87	1
germarium1-F10	eIF4E-1	*Danio rerio*	NP_571808	0.00E+00	85	1
germarium2-G11	frount	*Homo sapiens*	NP_079120	0.00E+00	85	1
germarium1-B07	hypothetical protein	*Homo sapiens*	AAH03081	0.00E+00	83	1
germarium3-G04	ribosomal protein L3	*Homo sapiens*	NP_000958	0.00E+00	83	1
germarium1-A12	prolyl endopeptidase	*Mus musculus*	NP_035286	0.00E+00	82	1
germarium2-F04	cytochrome c oxidase subunit II	*Paralichthys olivaceus*	NP_037585	2.00E-99	84	1
germarium3-F08	agCP1265	*Anopheles gambiae*	EAA08053	3.00E-98	99	2
germarium3-F04	cytochrome c oxidase subunit III	*Paralichthys olivaceus*	NP_037588	7.00E-98	87	2
germarium1-G05	smooth muscle cell associated protein 5	*Homo sapiens*	NP_110426	1.00E-97	80	1
germarium3-G07	ribosomal protein L7	*Ictalurus punctatus*	AAK95131	6.00E-97	79	1
germarium1-E07	adenylate kinase	*Cyprinus carpio*	P12115	2.00E-95	91	1
germarium2-A12	solute carrier family 27	*Mus musculus*	NP_036107	2.00E-94	84	1
germarium3-A09	60S ribosomal protein L9	*Rattus rattus*	P17077	5.00E-94	88	1
germarium2-G09	ATP synthase alpha chain	*Xenopus laevis*	P08428	6.00E-93	96	1
germarium3-D09	isocitrate dehydrogenase	*Macaca fascicularis*	Q28479	6.00E-92	74	1
germarium2-F12	Ras-related protein Rab-11B	*Discopyge ommata*	P22129	1.00E-90	97	1
germarium1-C09	signal peptidase complex	*Homo sapiens*	NP_055115	1.00E-89	91	1
germarium1-E01	unknown	*Homo sapiens*	AAK39521	9.00E-89	81	1
germarium2-F07	solute carrier family 25 member 5 protein	*Danio rerio*	AAM34660	1.00E-88	93	1
germarium3-A12	60S ribosomal protein L18	*Oreochromis mossambicus*	Q9I836	5.00E-86	88	1
germarium3-F09	signal sequence receptor, gamma	*Homo sapiens*	NP_009038	1.00E-85	94	1
germarium1-A08	cytochrome c oxidase subunit I	*Paralichthys olivaceus*	NP_037584	1.00E-83	70	1
germarium1-G06	non-ATPase regulatory subunit 13	*Homo sapiens*	XP_043220	1.00E-79	84	1
germarium1-C03	deoxyhypusine synthase (DHS)	*Mus musculus*	XP_134414	3.00E-76	84	1
germarium3-B01	F-box and leucine-rich repeat protein 3A	*Homo sapiens*	NP_036290	1.00E-74	86	1
germarium2-E12	NADH-ubiquinone oxidoreductase	*Bos taurus*	P34943	2.00E-74	62	1
germarium1-G12	KIAA1911 protein	*Homo sapiens*	XP_088331	4.00E-73	61	1
germarium3-A01	HMG-1	*Oncorhynchus mykiss*	AAA58771	6.00E-72	77	1
germarium3-C02	pleckstrin 2	*Mus musculus*	NP_038766	8.00E-72	61	1
germarium1-E09	nucleoside-diphosphate kinase	*Homo sapiens*	NP_003542	6.00E-71	59	1
germarium3-C04	endothelial differentiation-related factor 1	*Homo sapiens*	NP_003783	1.00E-69	87	2
germarium3-D06	hypothetical protein	*Homo sapiens*	CAC29493	2.00E-69	65	1
germarium3-B03	ribosomal protein S7	*Takifugu rubripes*	P50894	1.00E-68	98	1
germarium2-E04	alpha tubulin	*Notothenia coriiceps*	AAG15322	7.00E-68	100	1
germarium1-F06	cell division protein kinase 2	*Carassius auratus*	P43450	2.00E-67	92	1
germarium2-C12	T05F1.1.p 1	*Homo sapiens*	XP_052629	3.00E-66	70	1
germarium1-A11	ubiquitin carboxy-terminal hydrolase L1	*Mus musculus*	XP_124529	5.00E-66	59	1
germarium1-C08	DEAD/H box polypeptide 15	*Homo sapiens*	NP_001349	2.00E-65	93	1
germarium3-H09	GABA receptor associated protein	*Gillichthys mirabilis*	AAG13318	1.00E-64	98	1
germarium2-E01	ras homolog 9 (RhoC)	*Homo sapiens*	XP_124152	6.00E-63	100	1
germarium2-C08	steroid 5alpha-reductase	*Rattus norvegicus*	A34239	3.00E-62	53	1
germarium3-D03	KIAA1414 protein	*Homo sapiens*	BAA92652	3.00E-61	65	1
germarium1-F04	NADH dehydrogenase subunit 6	*Paralichthys olivaceus*	NP_037593	2.00E-60	80	1
germarium2-H11	Arf79F-P1	*Drosophila melanogaster*	NP_476955	8.00E-60	81	1
germarium3-E09	profilin 2	*Mus musculus*	NP_062283	1.00E-59	78	1

Table 2. Continued

Clone no.	Putative identification	Closest species	Accession no.	E-value	I[a]	F[b]
germarium1-B09	HSPC176	*Homo sapiens*	AAF29139	2.00E-59	78	1
germarium2-B02	aldehyde reductase 1	*Rattus norvegicus*	NP_036630	8.00E-59	66	1
germarium2-A09	peroxiredoxin 3	*Homo sapiens*	NP_006784	7.00E-58	79	1
germarium1-E04	histone H2A.F/Z variant	*Homo sapiens*	NP_036544	1.00E-56	99	2
germarium2-E08	ISCU2	*Mus musculus*	XP_132287	6.00E-56	74	1
germarium1-E08	transformer-2b	*Oryzias latipes*	BAC06514	2.00E-55	93	1
germarium1-D09	nucleoside diphosphate kinase homolog 5	*Homo sapiens*	XP_063164	4.00E-55	42	1
germarium1-B10	RIKEN cDNA	*Mus musculus*	XP_127600	1.00E-54	45	1
germarium3-A04	testis specific protein A14	*Homo sapiens*	NP_061188	3.00E-54	55	2
germarium3-A08	human general transcription factor Tfib	*unknown*	1TFB	8.00E-54	83	1
germarium1-C05	P7D11	*Xenopus laevis*	BAB79596	2.00E-52	52	2
germarium3-B04	pendulin	*Oreochromis niloticus*	AAD51751	2.00E-51	67	1
germarium3-G01	TBP-associated factor 13	*Mus musculus*	NP_005636	1.00E-47	97	2
germarium1-E11	unknown (protein for MGC:41163)	*Mus musculus*	AAH27526	2.00E-47	60	1
germarium2-D08	Ras-related protein O-Krev	*Discopyge ommata*	P22123	2.00E-47	95	1
germarium3-A05	deoxycytidine kinase	*Mus musculus*	NP_031858	9.00E-47	58	1
germarium3-G06	NADH dehydrogenase subunit 4	*Paralichthys olivaceus*	NP_037591	2.00E-46	71	1
germarium3-G05	RIKEN cDNA	*Mus musculus*	NP_077153	3.00E-46	61	1
germarium1-A07	dynein light chain 2; RIKEN cDNA	*Mus musculus*	NP_080832	1.00E-45	96	1
germarium3-E03	transforming protein B-myb	*Xenopus laevis*	S33643	2.00E-45	72	1
germarium3-D10	ornithine decarboxylase antizyme	*Danio rerio*	Q9YI98	2.00E-45	73	1
germarium1-H03	testis specific protein	*Rattus norvegicus*	NP_620220	2.00E-45	54	1
germarium3-H07	TdT binding protein	*Homo sapiens*	NP_443183	7.00E-45	61	1
germarium3-G12	RIKEN cDNA	*Homo sapiens*	XP_059468	1.00E-44	51	1
germarium3-D08	diaphanous homolog 3	*Homo sapiens*	NP_112194	3.00E-44	56	1
germarium3-B07	hypothetical protein	*Homo sapiens*	NP_112577	3.00E-44	92	2
germarium2-C01	BBP-like protein 2 isoform b	*Homo sapiens*	NP_079417	2.00E-41	56	1
germarium1-E12	RIKEN cDNA	*Mus musculus*	AAH26136	4.00E-41	76	1
germarium3-E02	cytochrome b5	*Homo sapiens*	1513199A	1.00E-40	65	1
germarium3-A11	33 kDa Inner dynein arm light chain	*Strongylocentrotus purpuratus*	Q26630	2.00E-37	50	1
germarium1-F02	CG12314 gene product	*Homo sapiens*	XP_113904	3.00E-36	50	1
germarium1-F01	RIKEN cDNA	*Mus musculus*	NP_077152	7.00E-36	83	1
germarium3-E05	hypothetical protein	*Homo sapiens*	AAH13194	3.00E-35	59	1
germarium1-C10	homeodomain-interacting protein kinase 2	*Homo sapiens*	AAG35710	8.00E-34	42	1
germarium1-A06	hypothetical protein	*Homo sapiens*	NP_061911	4.00E-33	44	1
germarium2-A01	heat shock protein 90 beta	*Platichthys flesus*	CAC27523	9.00E-33	93	2
germarium2-D12	RIKEN cDNA 4921507O14	*Mus musculus*	XP_130840	3.00E-32	31	1
germarium3-G10	protein phosphatase 1	*Rattus norvegicus*	NP_002700	2.00E-29	93	1
germarium3-E04	unnamed protein product	*Homo sapiens*	BAB71463	2.00E-28	34	1
germarium3-D04	RIKEN cDNA	*Mus musculus*	XP_109996	2.00E-28	39	1
germarium3-E10	COP9 homolog, subunit 6	*Mus musculus*	XP_132437	3.00E-28	95	1
germarium2-H07	beta-carotene 15, 15-dioxygenase	*Danio rerio*	NP_571873	1.00E-27	79	1
germarium2-B01	hypothetical protein	*Homo sapiens*	CAB66805	1.00E-26	43	1
germarium2-C04	NADH dehydrogenase	*Homo sapiens*	NP_002480	2.00E-26	72	1
germarium3-G09	acetylhydrolase Ib-alpha subunit	*Xenopus laevis*	AAK52334	2.00E-25	98	1
germarium3-D07	SPTR	*Mus musculus*	BAB31913	5.00E-25	50	1
germarium3-C05	beta-carotene 15, 15'-dioxygenase	*Gallus gallus*	CAB90825	8.00E-25	52	1
germarium1-F08	RW1 protein	*Mus musculus*	XP_123646	3.00E-24	38	1
germarium3-B05	proteolipid protein 2	*Mus musculus*	NP_062729	3.00E-24	40	1
germarium3-H06	Sad1/unc-84-like protein 2	*Homo sapiens*	Q9UH99	5.00E-24	56	1
germarium1-G03	growth arrest-specific 11	*Homo sapiens*	NP_001472	2.00E-23	42	1
germarium1-G10	protein transport protein	*Canis familiaris*	P38384	4.00E-22	77	1
germarium2-C03	SPTR	*Mus musculus*	XP_134222	8.00E-21	38	2
germarium3-C07	TBC domain containing protein	*Mus musculus*	BAB30630	9.00E-21	61	1
germarium3-F10	endoplasmic reticulum protein 1	*Homo sapiens*	AAH29067	9.00E-21	75	1

Table 2. Continued

Clone no.	Putative identification	Closest species	Accession no.	E-value	I[a]	F[b]
germarium3-F01	unnamed protein product	*Homo sapiens*	BAC05158	7.00E-20	82	1
germarium1-B12	RIKEN cDNA	*Mus musculus*	AAH04765	2.00E-18	41	1
germarium2-D05	RIKEN cDNA	*Homo sapiens*	XP_114126	7.00E-18	43	2
germarium3-B06	chromosome 11 open reading frame 5	*Homo sapiens*	NP_055020	9.00E-18	32	1
germarium1-A03	lyncein	*Bos taurus*	CAA76943	2.00E-17	50	1
germarium2-E03	G protein pathway suppressor 2	*Homo sapiens*	NP_004480	2.00E-17	48	1
germarium3-D05	sentrin/SUMO-specific protease 7	*Homo sapiens*	NP_065705	5.00E-17	67	1
germarium2-G04	hypothetical protein	*Homo sapiens*	XP_058897	1.00E-16	37	1
germarium3-H01	U2 snRNA	*Xenopus laevis*	CAA24954	1.00E-16	65	1
germarium2-E10	hypothetical protein	*Homo sapiens*	NP_115688	4.00E-15	67	1
germarium1-F09	SH3-containing protein	*Homo sapiens*	NP_057093	7.00E-15	63	1
germarium3-F06	ataxin-1 ubiquitin-like interacting protein	*Mus musculus*	NP_277068	1.00E-14	57	1
germarium1-B01	carbonic anhydrase	*Platichthys flesus*	AAC64172	1.00E-14	87	1
germarium3-F12	hypothetical protein	*Homo sapiens*	XP_059540	8.00E-14	41	1
germarium1-G09	translation initiation factor eIF-4 gamma	*Homo sapiens*	S49172	4.00E-13	53	1
germarium1-D03	CG8607 gene product	*Drosophila melanogaster*	NP_648100	9.00E-13	48	1
germarium1-C04	hypothetical protein	*Homo sapiens*	T50634	8.00E-11	46	1
germarium3-C01	SPTR	*Mus musculus*	BAB24507	7.00E-10	51	1
germarium1-G11	5OT-EST protein	*Mus musculus*	XP_130464	8.00E-10	67	1
germarium3-A10	protein phosphatase 1	*Homo sapiens*	NP_116222	5.00E-09	29	2
germarium1-B08	transcription factor Tcf3b	*Danio rerio*	AAD41491	5.00E-09	42	1
germarium2-B10	14S cohesin RAD21 subunit	*Xenopus laevis*	AAC26809	6.00E-09	42	1
germarium3-B09	NaDC-2	*Xenopus laevis*	AAB97879	1.00E-07	69	1
germarium1-H08	cytochrome c oxidase	*Mus musculus*	NP_031776	1.00E-06	37	1
germarium3-C03	agCP4537	*Anopheles gambiae*	EAA14698	2.00E-05	35	1
germarium1-F12	casein kinase II, alpha chain (CK II)	*Gallus gallus*	P21868	3.00E-05	60	1
germarium2-G01	agCP8854	*Anopheles gambiae*	EAA14543	4.00E-05	67	1
germarium1-G02	KPL2 protein	*Rattus norvegicus*	NP_072142	7.00E-05	51	1
germarium3-F03	myosin	*Mizuhopecten yessoensis*	BAB00612	7.00E-05	25	1
germarium3-H04	Y39B6A.n.p	*Caenorhabditis elegans*	NP_507706	1.00E-04	23	2
germarium3-B12	PE_PGRS family protein	*Mycobacterium tuberculosis*	NP_335725	2.00E-04	31	1
germarium1-E03	SPTR	*Mus musculus*	XP_133552	2.00E-04	49	1
germarium2-C10	synovial sarcoma associated SS18-delta	*Mus musculus*	AAL17749	8.00E-04	85	1

[a]Identity (%), [b]Frequency of the clones in the sequenced pool.

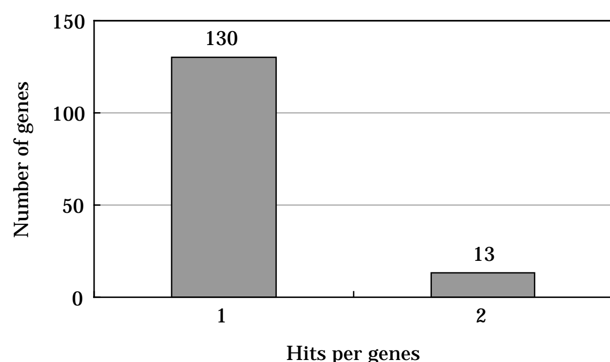

Fig. 1. Expression profiles and sequencing redundancy in the analysis of ESTs from the olive flounder (*Paralichthys olivaceus*) testis. Among 156 identified distinct known genes, 130 genes (83.3%) were sequenced only once; The remaining 26 genes (16.7%) were sequenced 2 times.

tions as a molecular chaperone. It has been reported in rats that *hsp90* is highly expressed in primordial germ cells and continues to be expressed in both male and female premeiotic germ cells (Ohsako *et al.*, 1995). Thus, it is likely that *hsp90* plays an important role in development of germ cells. Another clone, germarium2-B08, showed homology to RACK (receptor of activated protein kinase C) of tilapia with identities of 97%. PKC has been reported to be involved in spermatogonium proliferation *in vitro* and it may act as the physiological regulators of spermatogonium development *in vivo* (Li *et al.*, 1997). In addition, two kind of testis-specific protein are thought to play apparent roles in the olive flounder testis. The putative amino acid sequences deduced from two cDNA clone, germarium1-H03 and germarium3-

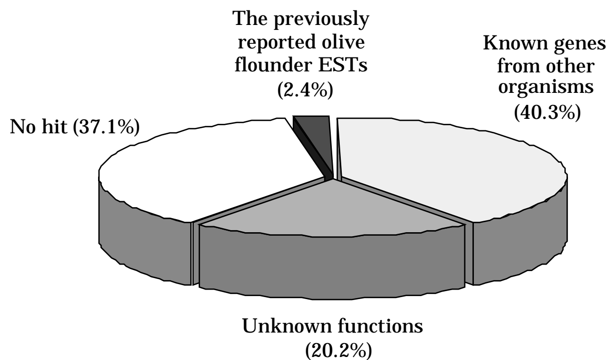

Fig. 2. Classification of 248 ESTs from the testis cDNA library of *Paralichthys olivaceus*. ESTs with BLASTX $E < 1 \times 10^{-3}$ were classified: the previously reported olive flounder ESTs, known genes from other organisms and known sequences of unknown functions. ESTs with $E \geq 1 \times 10^{-3}$ were classified as novel.

A04, were identified as testis-specific protein. The alignment revealed that these putative sequences showed homology to testis-specific protein of human and rat, with identities of 55% and 54%, respectively. 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, comparative functional genomics will allow annotation to these orthologous genes.

3. Microsatellite containing genes

Microsatellites represent another class of genetic markers. They are short tandem repeats of 2～6 nucleotides. Markers generated from these repeats are known to be highly polymorphic because of length variation of these repeats, and consequently, they exhibit multiple alleles, which makes them very informative for genetic studies. Polymorphism can be evaluated by PCR, which requires only a small amount of DNA, and allele sizing can be achieved with fluorescent primers and an automatic sequencer which assures reliability of the results. Microsatellites are commonly believed to occur primarily in noncoding DNA. However, surveys of other cDNA libraries have indicated that up to 8% of clones may contain microsatellites (Khan *et al.*, 1992). A high percentage of clones containing repetitive sequences have been observed from eight olive flounder cDNA libraries. Among 248 random clones, 22 unique EST clones have obvious repetitive sequences, including di-, tri-, penta- and hexa-nucleotide repetitive. Single nucleotide repeats were not included since they are not very useful for polymorphic markers. Table 3 details the numbers of microsatellite containing ESTs found in the olive flounder testis. These microsatellites can be potentially useful for genomic mapping if they are polymorphic. It was reported 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. Type I markers can serve as anchor points to allow the transfer of gene map information from a 'map-rich' model species to a 'map-poor' species, such as fish, and also facilitate physical mapping. Markers with known functions can also provide gene candidates for production traits. Thus, ESTs are important components in genome mapping projects (e.g., human, Adams *et al.*, 1991; rice, Aliyeva *et al.*, 1996; catfish, Liu *et al.*, 1999). 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.

Table 3. Analysis of microsatellite regions identified ESTs from testis cDNA of olive flounder. Numbers in parentheses are percentages of the number of microsatellites

No. of cDNA clones	No. of microsatelites					
	Total (%)	di	tri	tetra	penta	hexa
248	22 (8.9)	9	8	1	2	2

In conclusion, this study reports an expressed sequence tag (EST)-based gene identification analysis of 248 ESTs derived from olive flounder testis cDNA library. These ESTs will be useful for comparative genomics by determination of their orthologous counterparts through evolution, for mapping by PCR analysis using radiation hybrid panels, and for identification of polymorphic markers in genes of known functions (type I markers). Therefore, the testis has a potentially larger diversity of transcript profile and thus a

more efficient gene discovery could be carried out in a testis cDNA library.

Acknowledgements

This work is funded by a grant from the National Fisheries Research and Development Institute (RP-2007-AQ-064).

References

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

Aliyeva, E., A.M. Metz and K.S. Browning. 1996. Sequences of two expressed sequence tags (EST) from rice encoding different cap-binding proteins. Gene, 21(180) : 221～223.

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

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

Boguski, M.S., T.M. Lowe and C.M. Tolstoshev. 1993. dbEST -database for expressed sequence tags. Nat. Genet., 4 : 332～333.

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

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

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

Henikoff, S. and J.G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U.S.A., 89 : 10915～10919.

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

Khan, A.S., A.S. Wilcox, M.H. Polymeropoulos, J.A. Hopkins, T.J. Stevens, M. Robinson, A.K. Orpana and J.M. Sikela. Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nat. Genet., 2 : 180～185.

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

Li, H., V. Papadopoulos, B. Vidic, M. Dym and M. Culty. 1997. Regulation of rat testis gonocyte proliferation by platelet-derived growth factor and estradiol: identification of signaling mechanisms involved. Endocrinology, 138 : 1289～1298.

Lie, O., A. Slettan, U. Grimholt, M. Lundin, M. Syed and I. Olsaker. 1994. Fish gene maps and their implications for aquaculture. Anim. Biotech., 5 : 209～218.

Liu, Z., G. Tan, P. Li and R.A. Dunham. 1999. Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus*. Biochem. Biophys. Res. Commun., 259 : 190～194.

Marra, M.A., L. Hillier and R.H. Waterston. 1998. Expressed sequence tags-ESTablising bridges between genomes. Trends Genet., 14 : 4～7.

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

Ohsako, S., D. Bunick and Y. Hayashi. 1995. Immunocytochemical observation of the 90 KD heat shock protein (HSP90): high expression in primordial and pre-meiotic germ cells of male and female rat gonads. J. Histochem. Cytochem., 43 : 67～76.

Parsons, J.D. 1995. Improved tools for DNA comparison and clustering. Comp. Appl. Biosci., 11 : 603～613.

Pearson, W.R. 1997. Identifying distantly related protein sequences. Comput. Appl. Biosci., 13 : 325～332.

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

Schuler, G.D. *et al.*, 1996. Genome maps 7. The human transcript map. Wall chart. Science, 274 : 547～562.

Ton, C., D.M. Hwang, A.A. Dempsey, H.-C. Tang, J. Yoon, M. Lim, J.D. Mably, M.C. Fishman and C.C. Liew. 2000. Identification, characterization, and mapping of expressed Sequence tags from an embryonic zebrafish heart cDNA library. Genome Res., 10 : 1915～1927.

Virlon, B., L. Cheval, J.M. Buhler, E. Billon, A. Doucet and J.M. Elalouf. 1999. Serial microanalysis of renal transcriptomes. Proc. Natl. Acad. Sci. U.S.A., 96 : 15286～15291.

Received : October 1, 2007
Accepted : November 22, 2007

ESTs (Expressed Sequence Tags)를 통한 넙치 (*Paralichthys olivaceus*) 정소의 유전자 발현 패턴 분석

이정호 · 김종현 · 노재구 · 김현철 · 김영옥[1]
김우진[1] · 김규원[2] · 김경길*

국립수산과학원 육종연구센터, [1]국립수산과학원 생명공학연구소,
[2]서울대학교 약학대학

본 연구에서는 넙치 (*Paralichthys olivaceus*) 정소에 대한 cDNA library를 제작하여 총 248개의 EST (Expressed sequence tag)를 분석을 하였다. 넙치 정소의 유전자 발현 패턴을 조사하기 위하여 염기서열의 유사성 분석을 한 결과 248개의 EST 중 156개의 EST는 이미 밝혀진 유전자와 유사성이 있는 것으로 나타났으며, 92개의 EST는 새로운 유전자로 밝혀졌다. 유전자의 기능이 밝혀진 250개의 EST 중 6개 (3.8%)의 EST는 이미 알려진 넙치 EST와 상동성이 있는 유전자로 확인되었고, 100개 (64.1%)의 EST는 다른 생물에서 알려진 유전자와 상동성이 있는 것으로 나타났다. 그러나 50개 (32.1%)의 EST는 전혀 기능이 알려지지 않은 새로운 유전자로 밝혀졌다. 이상의 결과에서 넙치 정소에서 발현되는 유전자는 다른 조직에 비해 일부 유전자의 상대적인 발현정도가 많지 않고, 대부분의 유전자가 골고루 발현함으로써 다양한 유전자의 발현패턴이 확인되었다. 따라서 넙치 정소에 대한 cDNA library는 특이하고 새로운 발현 유전자의 탐색에 좋은 재료로 사용될 것으로 추측된다.