J. Microbiol. Biotechnol. (2008), **18**(9), 1510–1517

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Generation of Expressed Sequence Tags for Immune Gene Discovery and Marker Development in the Sea Squirt, *Halocynthia roretzi*

Kim, Young-Ok[1]*, Hyun Kook Cho[2], Eun-Mi Park[1], Bo-Hye Nam[1], Young Baek Hur[3], Sang-Jun Lee[1], and JaeHun Cheong[2]

[1]*Biotechnology Research Institute, National Fisheries Research and Development Institute, Busan 619-902, Korea*
[2]*Department of Molecular Biology, Pusan National University, Busan 609-735, Korea*
[3]*Research Center for Aquaculture Environment, Tongyoung 650-943, Korea*

Received: December 13, 2007 / Accepted: April 30, 2008

Expresssed sequence tag (EST) analysis was developed from three cDNA libraries constructed from cells of the digestive tract, gonad, and liver of sea squirt. Randomly selected cDNA clones were partially sequenced to generate a total of 922 ESTs, in which 687 unique ESTs were identified respectively. Results of BLASTX search showed that 612 ESTs (89%) have homology to genes of known function whereas 75 ESTs (11%) were unidentified or novel. Based on the major function of their encoded proteins, the identified clones were classified into ten broad categories. We also identified several kinds of immune-related genes as identifying novel genes. Sequence analysis of ESTs revealed the presence of microsatellite-containing genes that may be valuable for further gene mapping studies. The accumulation of a large number of identified cDNA clones is invaluable for the study of sea squirt genetics and developmental biology. Further studies using cDNA microarrays are needed to identify the differentially expressed transcripts after disease infection.

Keywords: Sea squirt, *Halocynthia roretzi*, expressed sequence tags (ESTs), immune gene, marker

The identification of the genes expressed within the cells of a given tissue is a basic step in the determination of gene function and the analysis of tissue physiology. One efficient approach to the characterization of gene transcripts is the expressed sequence tag (EST). ESTs are generated by large-scale sequencing of randomly picked clones from cDNA libraries constructed from mRNA isolated at a particular development stages and/or tissue. Although the possible functions of many genes can be deduced by homologies to known genes in the databases, the functions of most emerging genes and regulatory sequences are unknown [34]. Sequence information from ESTs could also be used in many other applications such as the generation of physical maps of chromosomes, the development of genetic linkage maps [30], the discovery of polymorphic markers by microsatellite tagging [23, 33], the identification of single nucleotide polymorphism [14], and the development of microarrays for functional genomics.

Sea squirts *Halocynthia roretzi* are sessile marine animals, ubiquitous throughout the world, with most living in shallow water and surviving by filtering plankton and nutrient materials from sea water. The sea squirt is commercially cultured in the southern coast of Korea. Large numbers of sea squirts in the farming areas have died during the last 20 years, causing serious losses in the aquaculture industry. The costs incurred by the softness syndrome of sea squirts in Korea in 2006 were estimated to amount to nearly $30 million. The dying sea squirts usually exhibited thinning of the tunics and meat and eventually becoming quite soft and dissolving. It has not been known the causes involved in this disease. Until now, researchers have proposed several possible disease-inducing factors, including environmental pollutants, microorganism infection, and abnormal temperature changes. Researches of the host defense mechanisms-associated genes in the sea squirt should be valuable for analyzing gene function during both disease pathogenesis and defense, and also for developing a molecular marker relevant to disease resistance in *Halocynthia roretzi* farming.

In this paper, as a component of a transcriptome analysis of the sea squirt, we have constructed three cDNA libraries from the liver, digestive tract, and gonad of sea squirt. Then, we have analyzed 922 EST clones and immune-related genes in the ESTs. This information can be used to identify the molecular mechanism for disease resistance of cultured sea squirt.

*Corresponding author
Phone: 82-51-720-2455; Fax: 82-51-720-2456;
E-mail: yobest@gmail.com

MATERIAL AND METHODS

Tissue Preparation

Sea squirt *Halocynthia roretzi* was obtained from the Aquaculture Environment Institute, National Fisheries Research and Development Institute (NFRDI). Liver, digestive tract, and gonad tissues were collected and cut into as small pieces as possible. Pooled tissues were rapidly frozen with liquid nitrogen and ground with a mortar/pestle.

Construction of Sea Squirt cDNA Libraries

mRNAs were isolated from liver, digestive tract, and gonad cells using a PolyA Tract mRNA isolation kit (Promega, WI, U.S.A.). cDNA synthesis was carried out using a cDNA synthesis kit with oligo-(dT) primer. The cDNA library was constructed in Uni-ZAP XR vectors according to the manufacturer's instruction (Stratagene, U.S.A.). All primary libraries were amplified, and aliquots of each amplified library were stored at both 4°C and −70°C.

Plasmid Preparation and Sequencing

Conversion of the recombinant Uni-ZAP XRs into pBluescript phagemids was carried out by *in vivo* excision according to the manufacturer's instructions (Stratagene, U.S.A.). Plasmid DNA was prepared by the alkaline lysis method [32] using the Qiagen Spin Column Mini-plasmid kits. Single-pass sequencing of 5'-termini of selected cDNA clones in phagemid form was performed using the ABI 3100 automatic DNA sequencer (PE Applied Biosystem, CA, U.S.A.) and the ABI Prism Big Dye Terminatior Cycle Sequencing Ready Reaction kit (PE Applied Biosystems).

EST sequences have been submitted to the dbEST and GenBank databases (accession numbers from EY456956 to EY457877).

Bioinformatic Analysis

Bioinformatic analysis was conducted to determine gene identities using GeneMaster 3.0 software (Ensoltek, Korea). Briefly, vector sequences were removed and database searches were limited to ESTs >100 bp in length. ESTs were then assembled in clusters of contiguous sequences (contig) using the ICAtools program [31]. Gene annotation procedures and homology searches of the sequenced ESTs have been locally done by BLASTX for amino acid similarity comparisons [1]. Matches were considered to be significant only when the probability (P) was less than 1×10^{-3} using BLASTX with all parameters at the defaults. All ESTs that were not identified as orthologs of known genes were designated as unknown EST clones. Phylogenetic analysis of immunoglobin-binding protein and peroxiredoxin-like protein was analyzed with the CLUSTALX and MEGA3.1 programs.

RT-PCR

Total RNA samples were extracted from the tissues using TRIzol reagent (Invitrogen). Subsequently, first-strand cDNA synthesis was carried out using the Advantage RT-PCR Kit (BD Bioscience). Primers were designed based on the cDNA sequences. The forward and reverse primers for peroxiredoxin-like protein, immunoglobin-binding protein, and β-actin were 5'-GCGAGAATTGGTCAACCT-GCTCCA-3' and 5'-TAAATGAGAGTATTCAGAATCGGTCGAG-3', 5'-ATCGATCGCATGGTCAAAGATGCT-3' and 5'-CTTCTCCA-TCCCTTCAGCGCCCTG-3', and 5'-GTACGTTGCCATTCAAG-CTG-3' and 5'-AGTCCTTACGGATATCGACG-3', respectively. The PCR conditions were as follows: 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, with a final step of 72°C for 7 min. The amplified PCR products were analyzed on a 1.5% agarose gel containing ethidium bromide (100 ng/ml).

Identification of Microsatellite-Containing cDNAs

EST sequences were screened for di-, tri-, and tetranucleotide microsatellite repeats using the MICAS (Microsatellite Analysis Server) program (http://210.212.212.7/MIC/index.html). Clones containing microsatellites were identified by determination of a minimal number of repeats in the microsatellite sequences: dinucleotide, five repeats; trinucleotide, four repeats; tetranucleotide, three repeats. Single nucleotide repeats were not included because they are not useful as polymorphic markers. Some cDNA clones contained more than one type of repeat, in which case these clones were categorized according to the longest repeats.

RESULTS AND DISCUSSION

EST Sequencing and General Characteristics

We have constructed cDNA libraries, consisting of 1.3×10^{7}–3.0×10^{7} pfu, from polyadenylated fractions of sea squirt liver, digestive tract, and gonad mRNA. The number of clones in the constructed cDNA libraries was believed to be sufficient to encompass the predominantly expressed mRNA within these organs. The average length of the insert cDNA fragment was found to be 1.5 kb, in a range

Table 1. General characteristics of sea squirt *Halocynthia roretzi* liver, digestive tract, and gonad ESTs.

	cDNA library		
	Liver	Digestive tract	Gonad
Primary library size (pfu/ml)	3.0×10^{7}	2.2×10^{7}	1.3×10^{7}
Average length of insert (kb)	1.5±0.6	1.3±0.5	1.7±0.5
Average of sequenced length (bp)	700	730	680
Number of clones sequenced	477	253	192
Number of unigenes	324	200	163
Singetons	283	182	148
Clusters	41	18	15
Redundancy (%)	32	21	15

from 0.5 to 3.0 kb. In total, 992 clones were selected randomly from the liver, digestive tract, and gonad cDNA libraries, respectively. The nucleotide sequences of these clones were determined using T3 primer, and then the GenBank database was searched for homologous sequences. A summary of the identified genes can be seen in Table 1. The initial ESTs were grouped into 687 consensus sequences, composed of 74 clusters (liver 41, digestive tract 18, gonad 15) and 613 singletons (liver 283, digestive tract 182, gonad 148). This suggests that the library had an overall redundancy of 22%. BLASTX comparisons established that 612 (89%) of the clones were orthologs of known genes, and the remaining 75 (11%) clones were unidentifiable *via* similarity comparisons ($E \geq 1 \times e^{-3}$). In addition, among the 687 EST clones, 34 (5%) unique genes were identified as homologs of previously reported sea squirt ESTs, and 653 (95%) genes were identified as orthologs of known genes from other organisms. These results suggest that EST analysis is an efficient and fast method for gene discovery of research subjects.

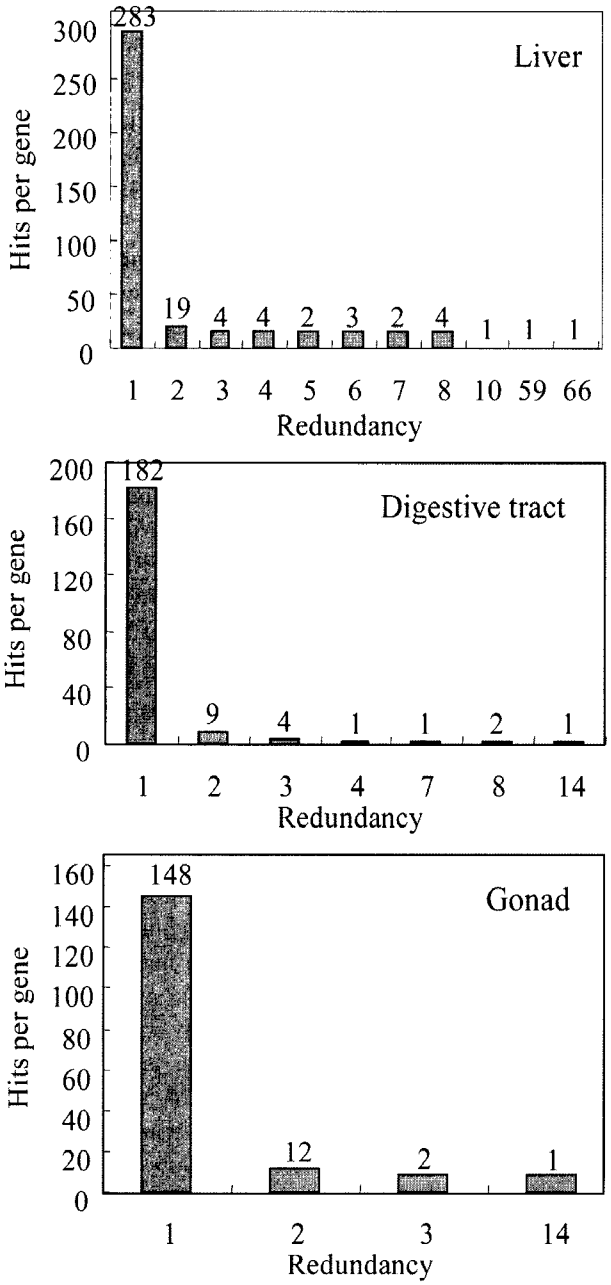

Fig. 1. Expression profiles and sequencing redundancy in the analysis of ESTs from the sea squirt liver, gonad, and digestive tract tissues.

Expression Profile of the Genes

The expression profiles of the known genes identified in the sea squirt liver, digestive tract, and gonad are provided in Fig. 1. Among 687 identified genes, 613 genes (89.2%) were sequenced only once; 41 genes (6.0%) 2–5 times; 15 genes (2.2%) 6–9 times; and 5 genes (0.7%) 10 times or

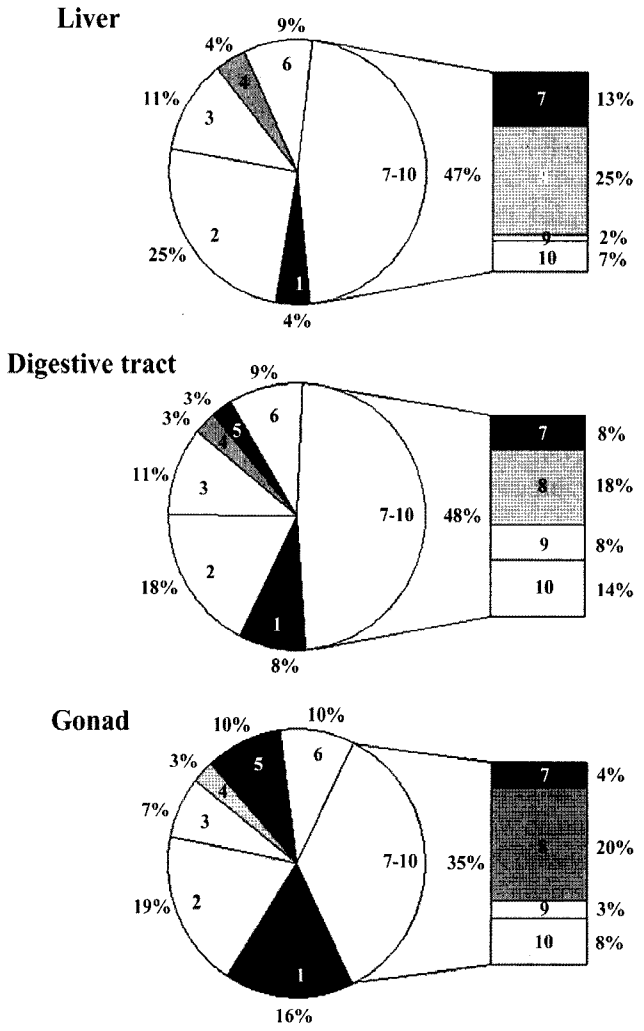

Fig. 2. Distribution of ESTs by functional classes.
Functional classes include (1) cell structure and motility; (2) metabolism; (3) cell defense and stress response; (4) cellular transport and transporter; (5) cell division and development; (6) cell signaling and communication; (7) transcription; (8) protein synthesis and function; (9) energy; and (10) unclassified proteins.

more. Although redundancy will increase as the number of sequenced clones increases, the high percentage of singletons indicated that the complexity and coverage of these libraries are good. The most abundantly expressed gene in the liver was unknown protein (n=66) and putative ubiquitin-conjugating enzyme (n=59), accounting for 20% and 18% of the 324 sequenced EST clones, respectively. Together, the most abundantly expressed genes in the digestive tract and gonad were unknown protein (n=14), cytoplasmic actin type 4 (n=8), *N*-acetylated alpha-linked acidic dipeptidase 2-like (n=7), and UDP-*N*-acetylmuramate-alanine ligase (n=14). The expression profile in squirt liver, then, was more polarized than the expression profile in other tissues.

Functional Group of ESTs

Based on the major functions of their encoded proteins, the identified clones have been classified into ten broad categories, as follows: cell structure and motility, metabolism, cell defense and stress response, cellular transport and transporter, cell division and development, cell signaling and communication, transcription, protein synthesis and function, energy, and unclassified proteins. The distribution of identified clones from the liver, digestive tract, and gonad DNA libraries is shown in Fig. 2. Generally, clones associated with metabolism and protein synthesis and function were all represented in high percentages. Inversely, clones associated with cellular transport and energy were represented at relatively low percentages. Clones associated with transcription, unclassified proteins, and cell structure were displayed at high percentages in liver, digestive tract, and gonad, respectively.

Genes Potentially Involved in Defense Mechanisms

Another objective of this study involved the identification of cDNA clones in sea squirt liver, digestive tract, and gonad that could be investigated for immune system activity. In this research, 23 clones showed similarities to genes potentially involved in cell defense (Table 2). Among the clones, 20 clones were previously unreported in the sea squirt *Halocynthia roretzi* and 3 clones showed high identity with those of reported *H. roretzi* [27].

The putative amino acid sequence deduced from one cDNA clone, squirt-hepa111, was identified as the decay accelerating factor (DAF), a complement system regulator. The complement system, which is composed of more than 30 soluble and membrane-bound proteins, plays an important role in innate host defense [28, 29]. As inappropriate regulation of the complement system can lead to significant damage of host tissues [10], a number of membrane-bound complement regulatory proteins are active, such as DAF, a glycosylphosphatidylinositol-anchored membrane protein

Table 2. Putative cell defense and immune-related genes isolated from liver, digestive tract and gonad cDNA libraries of sea squirt.

Clone no.	Genes	Closest species	Identity (%)
squirt-hepa28	Cu/Zn-superoxide dismutase	*Mediterranean fruit fly*	34
squirt-hepa38	Ficolin 3	*Halocynthia roretzi*	81
squirt-hepa111	Decay accelerating factor	*Cavia porcellus*	38
squirt-hepa135	Galectin family xgalectin-IIIb	*Xenopus laevis*	33
squirt-hepa140	Natural killer cell receptor P1-34-like protein	*Boltenia villosa*	37
squirt-hepa142	Glutathione-requiring prostaglandin D synthase	*Herdmania curvata*	43
squirt-hepa150	Immunoglobin-binding protein	*Boltenia villosa*	78
squirt-hepa163	Glutathione peroxidase	*Haemonchus contortus*	47
squirt-dig4	Glutathione-S-transferase-like protein	*Galleria mellonella*	39
squirt-dig21	Transferrin	*Halocynthia roretzi*	100
squirt-dig44	ABC-type multidrug transport system permease	*Mycoplasma gallisepticum* R	27
squirt-dig49	Multidrug resistance protein 2	*Fusobacterium nucleatum* subsp. *vincentii* ATCC 49256	28
squirt-dig87	QM protein	*Pinctada fucata*	73
squirt-dig97	Peroxiredoxin-like protein	*Ciona intestinalis*	72
squirt-dig164	Alpha-2-macroglobulin precursor	*Gallus gallus*	35
squirt-gonad1-E06	Mitochondrial cytochrome *c* oxidase subunit Vb	*Bos taurus*	31
squirt-gonad2-F10	Programmed cell death 4 isoform 2	*Homo sapiens*	64
squirt-gonad3-E02	Toll interacting protein variant	*Homo sapiens*	38
squirt-gonad3-E05	Heat-shock 70kDa protein 9B precursor variant	*Homo sapiens*	63
squirt-gonad3-F02	DNA mismatch repair protein PMS2	*Homo sapiens*	75
squirt-gonad3-F06	Nuclear NF-kappaB activating protein	*Mus musculus*	34
squirt-gonad4-B6	Galectin	*Haemonchus contortus*	33
squirt-gonad4-G1	Ficolin 1	*Halocynthia roretzi*	96

that restricts complement activation by inhibiting the formation of C3 convertases in both the classical and alternative pathways [24, 25].

Alpha-2-macroglobulin (squirt-dig164), a 720-kDa glycoprotein consisting of four identical 180-kDa subunits, is among one of the earliest protease inhibitors discovered in the testis [9]. In the systemic circulation, alpha-2-macroglobulin is a major acute-phase protein produced by hepatocytes, inhibiting all classes of proteases *via* an entrapment mechanism [35]. It also plays a role as a carrier protein and regulator for various growth factors, polypeptide hormones, and cytokines [5, 15, 19].

Ficolins (squirt-gonad4-G1 squirt-hepa38) of animal lectins prevent infection by pathogens through the innate immune system, in which they aggregate microorganisms and in some cases act as an opsonin. Ficolins are a group of proteins that consist of collagen-like and fibrinogen-like domains and were originally identified as TGF-β1-binding proteins on porcine uterus membranes [17, 18]. Ficolin 1 has elastin-binding [11], corticosteroid-binding [8], and *Escherichia coli*-binding [26] activities, suggesting that it may be a multifunctional protein. The collagen-like and fibrinogen-like domains of ficolin are likely to be responsible for its opsonic and GlcNAc-binding activities, respectively [22]. These immune-related genes should be valuable for analyzing gene function during both disease pathogenesis and disease defense.

Analysis of the Induced Genes in the Diseased Sea Squirt Immunoglobulin-Binding Protein

Immunoglobulin-binding protein (Bip) is a member of the 70 kDa stress protein family that resides within the lumen of the ER. Initially discovered bound to unassembled or free immunoglobulin heavy chain [12], Bip has since been found to associate stably with malfolded proteins [16] and transiently with partially folded or incompletely assembled protein subunits [2]. A comparison of the deduced amino acid sequence of sea squirt Bip with sequences of Bip from other species is shown in Figs. 3A and 3B. Using the BioEdit program, the deduced amino acid sequence of the EST sequence of sea squirt Bip had 41.8%, 42.5%, 42.5%, 43.8%, 43.8%, 44.4%, 46.4%, 45.1%, and 46.4% identity with that of *Mus musculus* (NP_071705), *Macaca fuscata* (BAE79724.1), *Homo sapiens* (AAF13605.1), *Gallus gallus* (NP_990822.1), *Xenopus laevis* (AAB08760.1), *Danio rerio* (AAT68067), *Oncorhynchus mykiss* (BAD90025.1), *Crassostrea gigas* (BAD15288.1) and *Pocillopora damicornis* (BAD89540.1), respectively. Phylogenetic analysis indicated that Bip of sea squirt *Halocynthia roretzi* was closer to that of *Pocillopora damicornis* than to that of other species. The Bip of *Mus musculus* was the farthest from the Bip of sea squirt among the above-mentioned species. To discriminate between the differential expression of the Bip gene in normal and diseased sea squirt, we assessed its

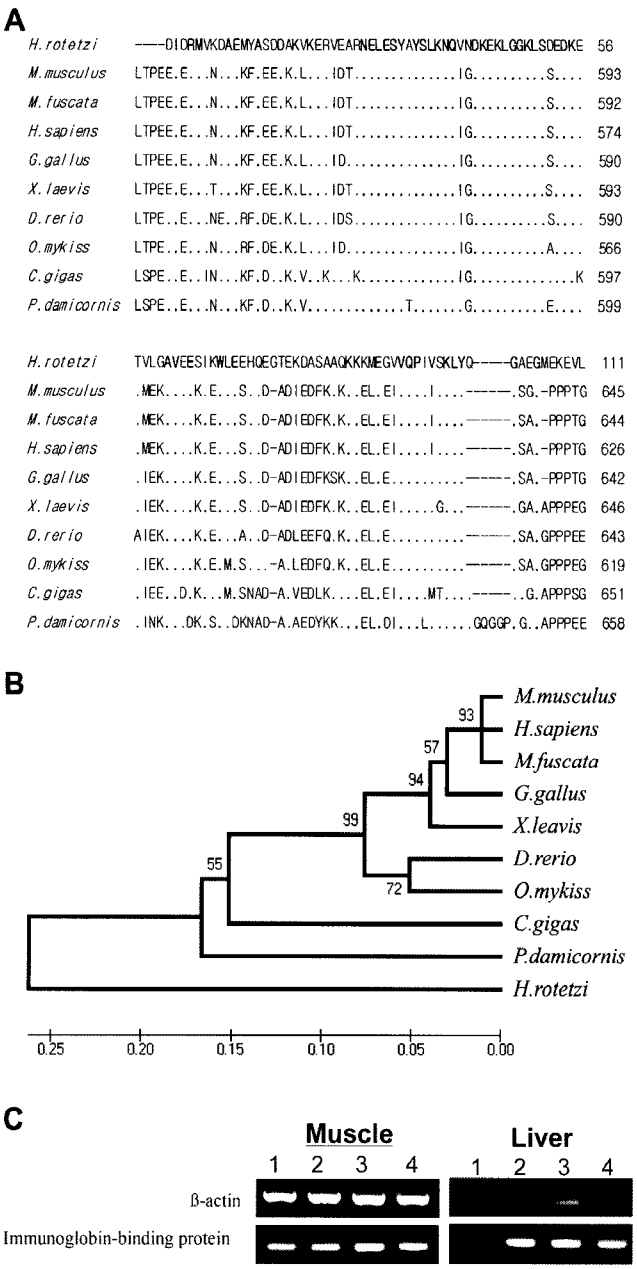

Fig. 3. Alignment and phylogenetic analysis of the deduced amino acid sequence of the sea squirt immunoglobin-binding protein (Bip) EST sequence with other species Bip sequences. **A.** Number on the top-right of alignment is the deduced amino acid of the Bip EST sequence. The remaining numbers refer to positions found in the sequences. Identity sequences are displayed by dots (...) and gaps used to maximize the alignment are shown by dashes (—). **B.** Phylogenetic analysis based on the deduced amino acid sequence of Bip from sea squirt and other species using the neighbor-joining method in MEGA 3.1. **C.** RT-PCR assay of Bip gene expression in the muscle and liver between normal sea squirt (1) and diseased ones at each stage showing softness syndrome: early (2), middle (3), and terminal (4). β-Actin was used as the control.

expression in muscle and liver tissues by RT-PCR (Fig. 3C). The Bip transcripts were highly expressed in muscle and liver tissues of the diseased sea squirt than the normal ones. This result suggests that Bip may be related to the immune mechanisms of sea squirts.

Peroxiredoxin-Like Protein

Cells possess multiple pathways to detoxify peroxides. It is reasonable to assume that each peroxide-removing enzyme plays a particular role in cell protection against oxidative stress, which can be defined by its cellular location, substrate specificity, and expression pattern among other factors. Peroxidases can oxidize different substrates such as glutathione and thioredoxin, which can also be relevant for cell signaling since these thiols are involved in modulation of transcriptional factors and other regulators [3, 7, 21]. Alignmnent and phylogenetic analysis of the amino acid sequences of peroxiredoxin-like protein from sea squirt and other animals are shown in Figs. 4A and 4B. Using the BioEdit program, the deduced amino acid sequence of sea squirt peroxiredoxin-like protein showed 68.2% identity with *Brugia malayi* (AAK07634), 68.2% identity with *Dirofilaria immitis* (AAC38831), 68.2% identity with *Litomosoides sigmodontis* (AAG10102), 68.2% identity with *Onchocerca volvulus* (AAC48312.1), 66.7% identity with *Globodera rostochiensis* (CAB48391), 69.9% identity with *Ciona intestinalis* (NP_001027810), 63.1% identity with *Paralichthys olivaceus* (AAY25400), and 69.0% identity with *Haliotis discus hannai* (AAZ22925). Alignment of the deduced amino acid sequence and phylogenetic analysis revealed that the sea squirt peroxiredoxin-like protein is highly related to that of *Ciona intestinalis* among the above-stated animals, which is consistent with the systematic relationship of these species. Analysis of gene expression using RT-PCR indicated that the peroxiredoxin-like protein gene was highly induced in muscle and liver of the diseased sea squirt compared with the normal ones (Fig. 4C).

The distinct difference in gene expression of BIP and peroxiredoxin-like protein between the normal sea squirt and diseased ones showing softness syndrome [5] suggest that both genes play important roles in the immune response of sea squirt to softness syndrome and also may be used for a developing molecular marker related to disease resistance in *Halocynthia roretzi* farming.

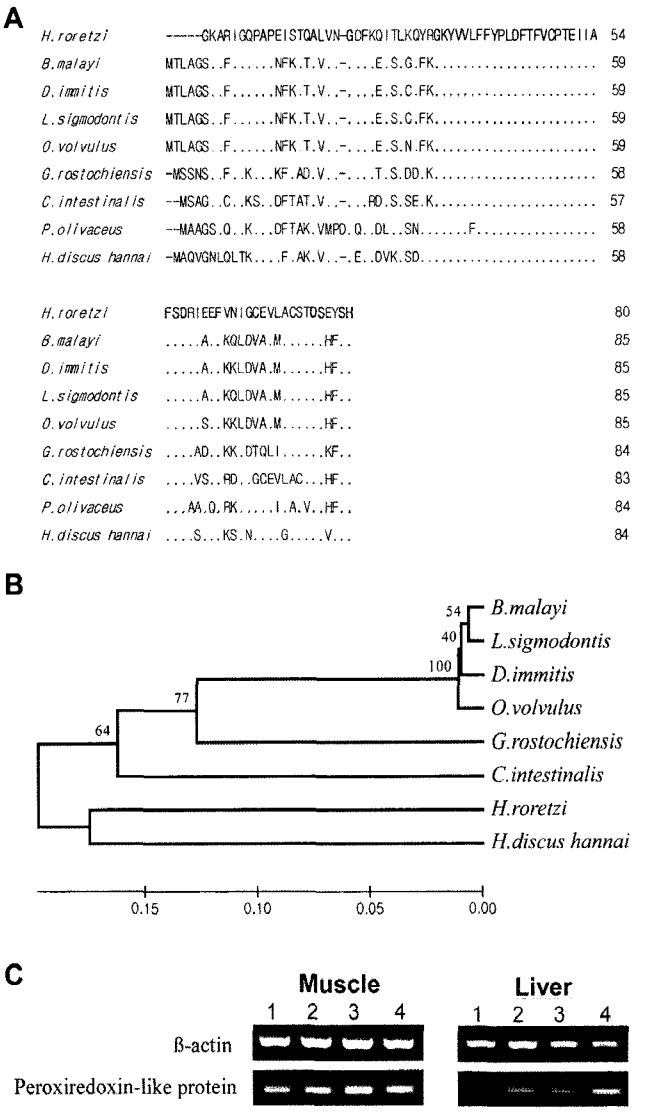

Fig. 4. Alignment and phylogenetic analysis of the deduced amino acid sequence of the sea squirt peroxiredoxin-like protein EST sequence with that of other organisms.

A. Number on the top-right of alignment is the deduced amino acid of the peroxiredoxin-like protein EST sequence. The remaining numbers allude to positions found in the sequences. Identity sequences are displayed by dots (...) and gaps used to maximize the alignment are shown by dashes (—). **B.** Phylogenetic analysis is based on the deduced amino acid sequence of peroxiredoxin-like protein from sea squirt and other organisms using the neighbor-joining method in MEGA 3.1. **C.** RT-PCR assay of peroxiredoxin-like protein gene expression in the muscle and liver between normal sea squirt (1) and diseased ones at each stage showing softness syndrome: early (2), middle (3), and terminal (4). β-Actin was used as the control.

EST-Derived Microsatellite Marker

During the last decade, microsatellites or SSRs (simple sequence repeats) have proven to be useful markers in animal genetic research and have been used for marker-assisted breeding purposes. Unfortunately, development of microsatellite markers is expensive, labor intensive, and time consuming if microsatellites are being developed from genomic libraries [13]. The cost of mining EST libraries is far lower than other traditional methods, and microsatellite development from ESTs has been successful in EST data mining [6, 20]. After screening the ESTs from the libraries, several sequenced cDNA clones harbored microsatellite markers (Table 3). In the expressed sequences, there were only one type of dimeric repeat motif [(AT)n]. Trimeric and tetrameric repeat motifs are relatively higher than dimeric repeats. When polymorphic, these microsatellite markers may prove useful for genomic mapping. The primary objective of this study was to develop EST resources. Therefore, we did not attempt to characterize the polymorphisms of these microsatellite clones.

Among the many possibilities and applications of the findings of this study, these identified clones may prove

Table 3. Microsatellite-containing cDNA clones from the three cDNA libraries of sea squirt.

Clone name	Nucleotide repeat unit	Microsatellite repeats	Identified genes
seasquirt-dig-30	Dinucleotide	(TA)6	Atp6v0c protein
seasquirt-dig-56	Dinucleotide	(AT)5	Plasma membrane H+-ATPase
seasquirt-dig-72	Tetranucleotide	(AACA)3	Transmembrane 4 superfamily member 3
seasquirt-dig-153	Trinucleotide	(AGC)5	Unknown
seasquirt-dig-172	Tetranucleotide	(TTGT)3	Unknown
seasquirt-dig-170	Tetranucleotide	(TTGT)3	Hypothetical protein
seasquirt-dig-166	Dinucleotide	(AT)9	Type I restriction enzyme S protein (fragment)
seasquirt-dig-112	Trinucleotide	(CAT)5	Hypothetical protein
seasquirt-dig-128	Trinucleotide	(CTT)4	Similar to Protein KIAA0152 precursor
squirt-gonad1-D03	Trinucleotide	(AGA)4	Troponin T
squirt-gonad2-H05	Trinucleotide	(GAA)4	Muscle actin
squirt-gonad3-E07	Trinucleotide	(ATC)4	Hypothetical protein DDB0190606
squirt-gonad4-B9	Dinucleotide	(AT)5	Unknown
squirt-gonad4-C3	Trinucleotide	(GAG)4	Reticulon
squirt-gonad4-H9	Trinucleotide	(GGA)5	TPA: RTN-A1
seasquirt-hepa-79	Dinucleotide	(TA)7	Similar to zinc finger protein 341
seasquirt-hepa-95	Tetranucleotide	(TGCG)3	Ferric-chelate reductase, putative
seasquirt-hepa-170	Trinucleotide	(TTA)4	Hypothetical protein XP_531486
seasquirt-hepa-163	Trinucleotide	(ATC)4	Glutathione peroxidase
seasquirt-hepa-100	Trinucleotide	(CTT)5	Unknown
seasquirt-hepa-173	Tetranucleotide	(AATA)3	Ribosomal protein L4
seasquirt-hepa-235	Trinucleotide	(GAT)4	Selenium-dependent salivary glutathione peroxidase
seasquirt-hepa-204	Tetranucleotide	(TGTT)3	Actin, cytoplasmic 1
seasquirt-hepa-229	Dinucleotide	(TA)7	Similar to SQTR9367
seasquirt-hepa-262	Trinucleotide	(ATC)4	Hypothetical protein
seasquirt-hepa-199	Dinucleotide	(TA)6	Elogation factor 1 beta
seasquirt-hepa-309	Tetranucleotide	(CATG)3	Unknown
seasquirt-hepa-402	Tetranucleotide	(AACA)3	Transmembrane 4 superfamily member 3 variant
seasquirt-hepa-403	Tetranucleotide	(TTGT)3	Unknown

useful in the selection of tissue-specific or cell-type-specific markers, the isolation of full-length clones and gene promoters, and gene expression pattern and gene function analyses.

References

1. 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 program. *Nucleic Acids Res.* **25:** 3389–3402.
2. Blount, P. and J. P. Merlie. 1991. BIP associates with newly synthesized subunits of the mouse muscle nicotinic receptor. *J. Cell Biol.* **113:** 1125–1132.
3. Carmel-Harel, O. and G. Storz. 2000. Roles of the glutathione- and thioredoxin-dependent reduction systems in the *Escherichia coli* and *Saccharomyces cerevisiae* responses to oxidative stress. *Annu. Rev. Microbiol.* **54:** 439–461.
4. Cho, H. K., B. H. Nam, H. J. Kong, H. S. Han, Y. B. Hur, T. J. Choi, Y. H. Choi, W. J. Kim, and J. H. Cheong. 2008. Identification of softness syndrome-associated candidate genes and DNA sequence variation in the sea squirt, *Halocynthia roretzi*. *Mar. Biotechnol.* **10:** 447–456.
5. Chu, C. T., D. S. Rubenstein, J. J. Enghild, and S. V. Pizzo. 1991. Mechanism of insulin incorporation into α2-macroglobulin: Implication for the study of peptide and growth factor binding. *Biochemistry* **30:** 1551–1560.
6. Cordeiro, G., M. R. Casu, C. L. McIntyre, J. M. Manners, and R. J. Henry. 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.* **160:** 1115–1123.
7. Delaunay, A., A. D. Isnard, and M. B. Toledano. 2000. H_2O_2 sensing through oxidation of the Yap1 transcription factor. *EMBO J.* **19:** 5157–5166.
8. Edgar, P. F. 1995. Hucolin, a new corticosteroid-binding protein from human plasma with structural similarities to ficolins, transforming growth factor-ß1-binding proteins. *FEBS Lett.* **375:** 159–161.
9. Fritz, I. B., P. S. Tung, and M. Ailenberg. 1993. Proteases and antiproteases in the seminiferous tubule, pp. 217–235. *In* L. D. Russell and M. D. Griswold (eds.). *The Sertoli Cell.* Cache River Press, Clearwater, FL.
10. Gasque, P. 2004. Complement: A unique innate immune sensor for danger signals. *Mol. Immunol.* **41:** 1089–1098.
11. Harumiya, S., A. Omori, T. Sugiura, Y. Fukumoto, H. Tachikawa, and D. Fujimoto. 1995. EBP-37, a new elastin-binding protein

in human plasma: Structural similarity to ficolins, tranforming growth factor-β1-binding proteins. *J. Biochem.* **117:** 1029–1035.
12. Hass, I. G. and M. Wabl. 1983. Immunoglobulin heavy chain binding protein. *Nature* **306:** 387–389.
13. He, C., L. Chen, M. Simmons, P. Li, S. Kim, and Z. J. Liu. 2003. Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Anim. Genet.* **34:** 445–448.
14. He, G., R. Meng, M. Newman, G. Gao, R. N. Pittman, and C. S. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* **3:** 1–6.
15. Huang, S. S., P. O. Grady, and J. S. Huang. 1998. Human transforming growth factor-β-α2-macroglobulin complex is a latent form of transforming growth factor β. *J. Biol. Chem.* **263:** 1535–1541.
16. Hurtley, S. M., D. G. Bole, H. Hoover-Litty, A. Helenius, and C. S. Copeland. 1989. Interactions of misfolded influenza virus hemagglutinin with binding protein (Bip). *J. Cell Biol.* **108:** 2117–2126.
17. Ichijo, H., L. Ronnstrand, K. Miyagawa, H. Ohashi, C. H. Heldin, and K. Miyazono. 1991. Purification of transforming growth factor-β 1 binding proteins from porcine uterus membranes. *J. Biol. Chem.* **266:** 22459–22464.
18. Ichijo, H., U. Hellman, C. Wernstedt, L. J. Gonez, L. Claesson-Welsh, C. H. Helden, and K. Miyazono. 1993. Molecular cloning and characterization of ficolin, a multimeric protein with fibrinogen- and collagen-like domains. *J. Biol. Chem.* **268:** 14505–14513.
19. James, K. 1990. Interactions between cytokines and α2-macroglobulin. *Immunol. Today* **11:** 163–166.
20. Kantety, R. V., M. La Rota, D. E. Matthews, and M. E. Sorrells. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* **48:** 501–510.
21. Kuge, S., M. Arita, A. Murayama, K. Maeta, S. Izawa, Y. Inoue, and A. Nomoto. 2001. Regulation of the yeast Yap1p nuclear export signal is mediated by redox signal-induced reversible disulfide bond formation. *Mol. Cell. Biol.* **21:** 6139–6150.
22. Le, Y., S. H. Lee, O. L. Kon, and J. Lu. 1998. Human L-ficolin: plasma levels, sugar specificity, and assignment of its lectin activity to the fibrinogen-like (FBG) domain. *FEBS Lett.* **425:** 367–370.
23. Liu, Z., P. Li, A. Kocabas, Z. Ju, A. Karsi, D. Cao, and A. Patterson. 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. Commun.* **289:** 317–324.
24. Lublin, D. M. 2005. [Review] Cromer and DAF: Role in health and disease. *Immunohematology* **21:** 39–47.
25. Lublin, D. M., S. Kompelli, J. R. Storry, and M. E. Reid. 2000. Molecular basis of Cromer blood group antigens. *Transfusion* **40:** 208–213.
26. Lu, J., Y. Le, L. Kon, J. Chan, and S. H. Lee. 1996. Biosynthesis of human ficolin, an *Escherichia coli*-binding protein, by monocytes: Comparison with the synthesis of two macrophage-specific proteins, C1q and the mannose receptor. *Immunology* **89:** 289–294.
27. Makabe, K. W., T. Kawashima, S. Kawashima, T. Minokawa, A. Adachi, H. Kawamura, *et al.* 2001. Large-scale cDNA analysis of the maternal genetic information in the egg of *Halocynthia roretzi* for a gene expression catalog of ascidian development. *Development* **128:** 2555–2567.
28. Miwa, T. and W. C. Song. 2001. Membrane complement regulatory proteins: Insight from animal studies and relevance to human diseases. *Int. Immunopharmacol.* **1:** 445–459.
29. Morgan, B. P. 1999. Regulation of the complement membrane attack pathway. *Crit. Rev. Immunol.* **19:** 173–198.
30. Murakawa, K., K. Matsubara, A. Fukushima, J. Yoshii, and K. Okubo. 1994. Chromosomal assignments of 3'-directed partial cDNA sequences representing novel genes expressed in granulocytoid cells. *Genomics* **23:** 379–389.
31. Parsons, J. D. 1995. Improved tools for DNA comparison and clustering. *Comp. Appl. Biosci.* **1:** 603–613.
32. Sambrook, J., E. F. Frisch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Habor Laboratory Press, Cold Spring Habor, NY, U.S.A.
33. Serapion, J., H. Kucuktas, J. Feng, and Z. Liu. 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar. Biotechnol.* **6:** 364–377.
34. Somerville, C. and S. Somerville. 1999. Plant functional genetics. *Science* **285:** 380–383.
35. Sottrup-Jensen, L. 1989. Alpha-macroglobulins: Structure, shape, and mechanism of proteinase complex formation. *J. Biol. Chem.* **264:** 11539–11542.