

멸종위기 어류 어름치 *Hemibarbus mylodon* (Cypriniformes)로부터 조직별 EST library 제작 및 발현 유전자 탐색

방인철, 임윤희, 조영선¹, 이상윤¹, 남윤권^{1*}
순천향대학교 해양생명공학과, ¹부경대학교 양식학과

Survey of Expressed Sequence Tags from Tissue-Specific cDNA Libraries in *Hemibarbus mylodon*, an Endangered Fish Species

In-Chul Bang, Yoon Hee Lim, Young Sun Cho¹, Sang Yoon Lee¹, Yoon Kwon Nam^{1*}

Department of Marine Biotechnology, Soonchunhyang University, Asan 336-745, Korea
¹Department of Aquaculture, Pukyong National University, Busan 608-737, Korea

Representative cDNA libraries were constructed from various tissue sources of *Hemibarbus mylodon*, an endangered freshwater fish species in Korea, for the mining of expressed sequence tags (ESTs). Randomized and non-normalized EST analysis was performed with 7 unidirectional cDNA libraries generated from brain, intestine, kidney, liver, muscle, ovary or testis. Of 3,383 ESTs in total, the number of singleton was 2,029, and 333 contigs containing 1,354 ESTs were assembled (percent of unigene = 70.0%). Abundantly expressed gene transcripts and broad clustering of putative gene function were tissue-specific in general, and the redundancy was also variable among those libraries. Over half of *H. mylodon* ESTs were matched with orthologues from other teleosts among which zebrafish gene sequences were the most frequent in those matches. This initial setting of EST libraries achieved in the present study would be a fundamental basis for the banking of gene resources from this endangered fish species.

Keywords: *Hemibarbus mylodon*; EST analysis, tissue-specific cDNA libraries

서 론

한반도 고유 어종인 천연기념물 제 238 및 259호 어름치 (*Hemibarbus mylodon* Berg)는 최근 지속적인 인간활동으로 인한 서식지 파괴와 환경 악화 등으로 인해 그간 자연 자원 보호 정책에도 불구하고 그 개체수가 급격히 감소하고 있으며(Byeon et al., 1996), 이미 많은 한반도 수계에서 어름치 집단의 절멸 위험이 가속화되고 있다. 따라서 현재 서식지의 수동적인 자연 보호 정책만으로는 어름치의 멸종 위기를 효과적으로 대처해 나가기 어렵고 보다 능동적인 복원 전략과 기술 개발이 시급한 실정이다(Jang et al., 2003; Kim et al., 2007). 멸종 위기에 처한 생명체로부터 생물 기능을 암호화하고 있는 유전정보를 대량 수집하고 이의 체계적인 유전자 자원 관리 시스템의 개발은 미래 지속 이용 가능한 생명 기능을 보존한다는 측면에서 중요시 된다. 뿐만 아니라 멸종위기 종의 생물 기능을 분자 또는 세포 수준에서 이해함으로써 복원을 위한 유용 원천 정보를 제공한다

는 측면에서도 매우 중요시 된다. 특히 멸종 위기종의 장외 복원(*ex situ* restoration)을 위해서는 해당 어종의 인공 종묘 생산 기술의 확립은 물론 이들의 안정적인 생물 생산을 위한 다양한 양식 기법들의 활용이 요구되며, 나아가 종 보존을 위한 순계의 확립과 유지 기술 역시 접목되어야 한다. 따라서 이러한 목표를 효과적으로 달성하기 위해서는 생물 종의 생명 기능을 암호화하고 있는 유전자 발현 정보의 수집이 필수적이다.

Expressed sequence tag (EST) 분석 기술은 생명체의 genome 으로부터 특정 시점에 발현하는 유전자 정보를 대량 수집할 수 있는 기법이며(Adams et al., 1991), 본 기술은 유전체 지도 작성(genome mapping) 및 발현 유전자의 다형성 확보 등에도 매우 유용하게 사용된다(Marra et al., 1998; Serapion et al., 2004). 또한, 특정 생물 종의 서로 다른 조직들로부터 발굴된 EST 정보들은 해당 조직들의 기능과 특징 평가(tissue expression profiling)에도 유용하게 활용될 수 있다(Douglas et al., 2007).

이에 본 연구에서는 멸종 위기 어류 어름치의 유전자 banking 과 효과적인 인공 증식 및 사육기술을 개발하기 위한 첫 단계로서 어름치의 다양한 조직들로부터 발현 유전체 정보 수집을

*Corresponding author: yoonknam@pknu.ac.kr

위한 EST cDNA library들을 구축하고 발현 유전자의 탐색을 실시하고자 하였다.

재료 및 방법

어류 시료

어름치 체세포 조직을 대상으로 한 EST cDNA library 구축에 사용된 어름치 시료(평균 어체중 130 g)는 문화재청의 문화재 훼손 허가를 취득하고 남한강 및 임진강 상류(강원도 평창군 평창읍 평창강, 경기도 연천군 군남면 임진강)에서 채집한 개체들이었다. 난소 및 정소 EST library의 구축은 개체들은 동일 수계에서 채집된 350 g 및 전장 40 cm 내외의 성숙 개체들을 이용하였다. 각 조직의 cDNA library 구축은 조직당 6마리씩을 사용하였다.

핵산 분리 및 cDNA library 구축

조직들로부터 total RNA를 분리하기 위해서 뇌(brain), 소화관(intestine), 신장(kidney), 간(liver), 근육(muscle), 난소(ovary) 및 정소(testis) 조직을 해부학적 기법에 의해 적출하고 TriPure reagent (Roche, Germany)를 이용하여 총 RNA를 추출하였다. 모든 방법은 제조사의 권고에 따라 실시하였다. Formaldehyde-MOPS 한천 전기영동과 흡광도 측정을 통해 추출된 총 RNA의 integrity, 순도 및 양을 평가하였다. DNase I (Promega, USA) 처리를 통해 RNA 분리 과정 중 오염될 수 있는 DNA 분자를 제거하였고 이 후 Qiagen RNeasy kit (Qiagen, Germany)를 이용하여 다시 순수 분리하였다. 역시 동일 제조사의 mRNA 분리 kit (Qiagen mRNA isolation kit)를 이용하여 제조사의 권고 방법대로 poly (A+) RNA 분획만을 다시 재 분리한 후 전기영동상으로 rRNA 제거 효율을 확인하였다. 최종 분리된 5 µg의 poly (A+) RNA를 조직별로 준비한 후 cDNA library 구축에 이용하였다.

cDNA library는 lambda Zap cDNA synthesis kit (Stratagene, USA)를 이용하여 단일 방향 삽입(unidirectional cloning) 방식을 통해 실시하였다. 제조사의 권고 방법에 의거하여 제1차 및 제2차 cDNA strands를 합성하였고, 클로닝을 위한 말단 조각, 제한효소 처리, 합성 cDNA 길이의 선택 분리(> 600 bp) 및 UniZap XR vector (Stratagene, USA)와의 융합 반응을 실시하였다. Gigapack gold packaging extract (Stratagene, USA)를 이용하여 *in vitro* packaging을 수행한 후 primary library의 titer를 측정하고 대장균 XL blue MRF' strain을 이용하여 amplified library를 조직별로 구축하였다. ExAssist helper phage (Stratagene, USA)와 대장균 SOLR strain을 이용하여 pBluescript SK (-) phagemid 벡터(Stratagene, USA)로의 *in vivo* excision을 실시하였다. 제작된 각 cDNA library로부터 무작위로 선발한 library 당 48개의 재조합 클론을 대상으로 pBluescript SK (-)의 multi-cloning site (MCS) 양 말단에 위치하는 SK 및 T7 primer를 이

용, PCR을 통해 각 library에 포함된 insert cDNA의 평균 길이를 평가하였다.

Expressed sequence tag 분석

제작된 7종의 EST library들로부터 무작위로 선발한 재조합 클론들을 대상으로 종래의 일반적인 alkaline lysis 방법을 이용한 플라스미드 DNA 분리를 수행하였으며, 염기서열 분석을 위해서 벡터의 MCS 말단 쪽에 존재하는 SK 프라이머 부착 위치로부터 자동염기서열 분석기 ABI 3700 (Applied Biosystem, USA)을 통해 single pass sequencing을 실시하였다. 염기서열 분석 프로그램인 Sequencer (Gene Codes, USA)를 이용하여 1차 자료를 편집 및 교정하였으며 400 bp 길이 이상의 cDNA 염기서열을 갖는 클론들을 만을 재선별하여 NCBI GenBank의 BLAST 검색 (<http://www.ncbi.nlm.nih.gov/blast>)을 통해 어름치 EST들과 상동성을 보이는 orthologues들과의 유사도를 평가하였다. 아미노산 서열 수준에서 상동성 검색을 실시하기 위해서 translated BLAST (BLASTx option)를 실시하였으며 BLASTx 검색은 non-redundant DB 전체를 대상으로 하였고 각 query 염기서열 당 가장 높은 상동성이 관찰되는 50개의 유사 단백질들을 비교하여 그 기능을 유추하였다. 나머지 BLAST 검색 방법은 GenBank BLASTx의 default 값을 그대로 이용하였다. BLAST 검색 결과를 포함하는 htm page를 local database에 저장하고 각 검색 결과를 바탕으로 가장 높은 상동성을 보이는 best match 유전자들과의 e-value 평가 및 기능 유추 분석을 실시하였다.

각 library들로부터 확보된 EST 분포 특성 및 redundancy를 조사하기 위해 Sequencer를 이용하여 contig 작성을 수행하였다. Contig 작성을 위해 minimum match 및 overlapping bp를 각각 90% 및 50 bp로 하여 assembly를 수행하였으며 각 조직별로 작성된 contig의 수, contig에 포함된 EST수 및 singleton의 수를 조사함으로써 각 조직에서 많은 빈도로 중복 발현하는 유전자의 종류와 그 빈도를 추정하였다. 아울러 본 연구에서 발굴된 어름치 EST 전체에 대한 동일한 contig assembly를 수행하여 함께 비교하였다.

결과 및 고찰

cDNA library 구축 및 평가

어름치 조직별로 구축된 cDNA library의 primary titer, amplified titer 및 excised titer를 Table 1에 나타내었다. Table 1에서 보듯이 본 연구에서 구축된 조직별 cDNA library의 primary size는 모두 최소 5×10^5 pfu/mL 이상으로 나타났으며 아울러 amplified 및 excised stock의 titer 역시 각기 4.6×10^{10} pfu/mL 및 5×10^7 cfu/mL 이상으로 나타나 다양한 발현 유전자 정보를 포함하는 충실한 cDNA library들이 구축되었음을 알 수 있었다. 뿐만 아니라 각 cDNA library로부터 무작위로 취한 각

Table 1. Summary for the construction and characteristics of EST libraries in *Hemibarbus mylodon*

	Tissue source						
	Brain	Intestine	Kidney	Liver	Muscle	Ovary	Testis
Primary titer (pfu/mL)	5.8×10^5	1.0×10^6	1.2×10^6	1.4×10^6	1.8×10^6	1.8×10^6	5.0×10^5
Amplified titer (pfu/mL)	4.6×10^{10}	1.2×10^{11}	3.8×10^{11}	4.0×10^{11}	2.5×10^{11}	8.6×10^{11}	1.0×10^{11}
Excised titer (cfu/mL)	4.5×10^5	4.8×10^5	4.0×10^5	2.9×10^6	4.2×10^5	3.0×10^5	4.0×10^5
Recombinant: background ratio	120:1	98:1	105:1	109:1	130:1	105:1	98:1
Average insert length (kb)	1.6	1.4	1.7	1.6	1.5	1.4	1.3
No. EST collected	591	512	355	505	375	540	505
E-value clustering (%)							
No match	10.7	5.5	21.6	10.9	15.2	7.2	9.7
$E > 10^{-10}$	33.7	24.8	34.2	13.1	11.5	12.2	17.6
$10^{-50} < E \leq 10^{-10}$	15.1	19.9	16.5	17.6	8.0	20.0	18.2
$10^{-100} < E \leq 10^{-50}$	17.8	22.3	16.5	26.7	20.8	34.1	26.9
$E \leq 10^{-100}$	22.8	27.5	11.3	31.7	44.5	26.5	27.5
Broad functional clustering (%)							
Information storage	28.0	18.4	19.0	11.1	9.0	21.0	10.4
Cellular processes	33.1	29.6	23.0	24.9	47.6	15.0	13.6
Metabolism	16.9	34.0	38.7	54.0	31.4	43.9	48.0
Unclassified	22.0	18.0	19.3	10.0	12.0	20.1	28.0

48개 클론의 insert 길이를 PCR (SK/T7 primer)로 분석한 결과 그 평균 길이가 1.3~1.7 kb로 나타나 size fraction 역시 효과적으로 이루어진 것으로 나타났다.

Total redundancy 및 고빈도 발현 유전자 분석

본 연구를 통해 확보한 어류 EST clustering 분석 결과의 요약은 역시 Table 1 그리고 Fig. 1에 나타내었다. 조직 별로 뇌(591 ESTs), 장(512 ESTs), 신장(355 ESTs), 간(505 ESTs), 근육(375 ESTs), 난소(540 ESTs) 및 정소(505 ESTs)를 분석하여 총 3,383개의 발현 유전자 정보를 확보하였다. 어류 총 EST를 대상으로 contig assembly를 수행한 결과 333개의 contig가 형성되었으며 333개의 contig는 총 1,354 EST들을 포함하는 것으로 나타났다. 따라서 contig에 포함되지 않는 singleton은 2,029개로서 전체 평균 redundancy는 약 40%, 그리고 unigene (singleton과 contig 수의 합)의 비율은 69.8% (2362/3330)로 나타나 비교적 높은 빈도의 유전자 확보율을 나타내었다.

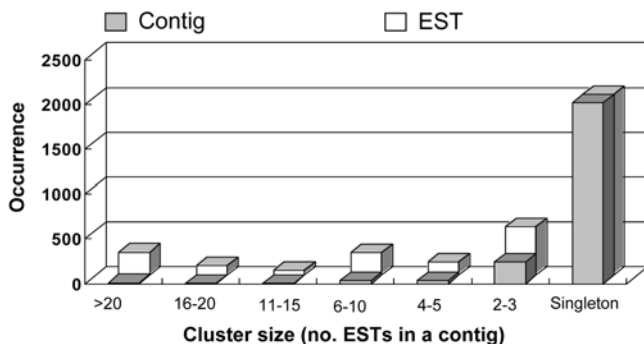


Fig. 1. Redundancy analysis of ESTs collected from *Hemibarbus mylodon*.

Assembly를 통해 형성된 333개의 contig의 분포를 포함하는 EST수에 따라 class I (> 20 ESTs), class II (16-20 ESTs), class III (11-15 ESTs), class IV (6-10 ESTs), class V (4-5 ESTs) 및 class VI (2-3 ESTs)로 구분하였다. Contig당 20개의 EST 이상을 포함하는 class I은 총 9개의 contig로 구성되어 있었으며(9/333 = 2.7%) 전체 EST의 7.7%에 해당하는 256개의 ESTs를 포함하고 있었다. 이들 contig들은 주로 대부분의 조직에서 발현하는 housekeeping 유전자들(cytochrome c oxidase subunit I 및 actin genes)과 조직특이적인 기능을 담당하는 유전자들(예, fertilization envelop outer membrane protein gene)로 대별되어 나타났다. Class II는 6개의 contigs (104 ESTs; 3.1%)를 포함하고 있었으며 Class III는 4개의 contig (1.2%)와 55개의 ESTs (1.6%)를 포함하는 것으로 관찰되었다. Class II 및 III에 속하는 유전자들 역시 beta-actin을 위시한 housekeeping 단백질을 암호화하는 그룹과 zona pellucida glycoprotein 3 (ZP3) 등 비교적 조직 특이적인 기능을 담당하는 단백질을 암호화하는 그룹으로 형성되어 있었다. 아울러 아직 정확한 생물학적 기능이 규명되어 있지 않은 unknown 또는 hypothetical protein들 역시 포함되어 있어 향후 이들 신규 유전자의 생화학적 및 생물학적 기능 분석의 필요성을 시사하였다. Class IV 및 class V의 경우 조직간 overlapping EST를 포함하는 contig보다는 특정 단일 조직 내 발현하는 transcripts로 구성된 contig들을 많이 포함하고 있었으며 전체 형성된 contig 중 10.5% (35 contigs; class IV) 및 10.2% (34 contigs; class V)를 차지하였고 총 EST의 7.5% (class IV) 및 7.4% (class V)를 차지하는 것으로 나타났다. 마지막으로 2~3개의 EST를 포함하는 class VI는 어류 EST에서 형성된 대부분의 contig인 245 (74%)개의 contigs를 포함하고 있었고 이에 속하는 총 EST 수는 539개로 나타났다(Fig. 1).

Table 2. Abundantly expressed genes in *Hemibarbus mylodon* tissues in EST survey

Tissue	Abundantly expressed genes	Incidence (%)
Brain	Unknown or hypothetical protein	1.0
	Beta-actin	1.0
	Ependymin	1.0
Intestine	Chitinase	6.4
	Apolipoprotein A-I	2.3
	Apolipoprotein Eb	1.2
Kidney	Cytochrome oxidase subunit I	1.3
	Unknown or hypothetical protein	1.0
	Beta-integrin	1.0
Liver	Apolipoprotein A-I	6.9
	Serpina protein	2.4
Muscle	Alpha skeletal actin	7.2
	Aldolase	5.9
	Creatine kinase	5.3
Ovary	Fertilization envelop outer membrane protein	5.2
	Zona pellucida 3 (ZP3)	4.1
	Zona pellucida 2 (ZP2)	3.0
Testis	Cytochrome oxidase subunit I	2.2
	Creatine kinase	1.4
	Tubulin alpha	1.2

조직별 EST 출현 특성

조직별 redundancy는 12~55%로 다양하게 나타났으며 조직 간 서로 다른 EST 출현 특성을 나타내었다. 여타 조직에 비해 근육(redundancy = 54.4%)과 난소(redundancy = 38.5%) 조직에서 상대적으로 높은 redundancy를 보였고 신장 (12.3%) 및 뇌 (15.4%) 조직에서 가장 높은 EST의 다양성이 관찰되었다. 조직별 고빈도 발현 유전자의 종류는 Table 2에 나타내었으며 Table 2에서 보듯이 뇌(brain)의 경우 총 EST의 1%에 해당하는 유전자들로서 beta-actin 및 ependymin을 암호화하는 유전자들이 높은 빈도로 발현하였다. Beta-actin은 다양한 조직에서 발현되는 housekeeping 단백질로서 cytoskeletal 구조의 결정에 중요한 역할을 담당하는 것으로 알려져 있으며(Kim et al., 2007), ependymin은 신경의 재생 반응에 중요한 glycoprotein으로서 뇌의 기억 기능에 관여하며 어류의 학습 기능에 중요한 역할을 담당하는 것으로 알려져 있다(Suarez-Castillo et al., 2004; Suarez-Castillo and Garcia-Arraras, 2007). 또한 본 ependymin 단백질은 아미노산 및 유전자 구조가 경골 어류 군에서 매우 잘 보존되어 있는 것으로 알려지면서 최근 ependymin의 glycosylation domain을 이용한 경골 어류의 분자 진화 분석이 이루어진 바 있다(Suarez-Castillo et al., 2004). 아울러 뇌조직에서 아직 그 기능이 규명되지 않은 신규 단백질이 1%에 해당하는 비율로 발현하였으나 아직 본 단백질 및 유전자의 생화학적/생리학적 기작에 관해서는 추가 연구가 요구된다. 소화관의 경우 chitin을 분해하는 chitinase (6.4%)와 지질대사에 관여하는 apolipoprotein (Apo A-I 2.3% 및 Apo Eb 1.2%) 유전자들이 고 빈도로 검출되어 분해 및 소화기능을 담당하는 본 조직의 특성을 잘 보여주고 있다. 신장의 경우 역시 unknown 유

자 (1%)와 함께 cytochrome oxidase (1.3%) 유전자가 상대적으로 고 빈도로 발현하는 것으로 나타났으며, 간 조직의 경우 cholesterol 대사를 조절하는 주요 단백질인 Apo A-I이 6.9%로 높게 발현되었다(Lee et al., 2007). 근육 조직의 경우 근 섬유형성에 필수적인 skeletal muscle type의 alpha actin 단백질을 암호화하는 유전자가 7.2%로 매우 높게 발현하여 근육조직의 특성을 잘 반영하고 있다(Venkatesh et al., 1996; Venkatesh and Brenner, 1998; Moutou et al., 2001). 아울러 당 대사와 관련된 aldolase 및 ATP energy metabolism과 관련된 creatine kinase 역시 각각 5.9 및 5.3%로 높게 발현하고 있었다. 어름치 난소 조직의 경우 난 성숙 및 발달과 밀접한 조직특이적인 EST 분포 양상을 잘 보여 주고 있다. 난소 조직의 경우 fertilization envelope outer membrane을 coding하고 있는 transcript들이 5.2%로서 가장 높은 빈도로 관찰되었으며 아울러 역시 egg envelop protein의 일종인 ZP isoform들이 3~4%내외로 높게 발현하고 있었다(Wang and Gong, 1999; Liu et al., 2006). 반면 정소 (testis)의 경우 정소 특이적인 유전자들의 고 빈도 발현보다는 cytochrome oxidase (2.2%), creatine kinase (1.4%) 및 tubulin alpha (1.2%) 등 housekeeping 유전자들이 높은 빈도로 검출되었다.

EST의 기능별 대분류 분석

조직 별로 확보된 어름치 EST들을 대상으로 NCBI의 Clusters of Orthologous Groups of Proteins (COGs; <http://www.ncbi.nlm.nih.gov/COG/>) 및 gene ontology 검색(www.geneontology.org)에 의해 얻어진 정보를 이용하여 broad clustering을 수행하였다 (see Table 1). 그 결과 각각의 조직들은 서로 다른 clustering 양상을 보였으며 일반적으로 알려져 있는 각 조직의 생물학적 기능을 비교적 잘 대변하는 것으로 나타나 뇌조직의 경우 DNA 복제/repair, 전사(transcription) 관련 등 전반적인 'information storage and processing' 그룹이 여타 조직들에 비해 높게 나타났다. 반면 세포분열, protein turnover, cell motility/structure, ion transport/metabolism 및 signal transduction 등과 관련된 'cellular process' 그룹은 뇌, 소화관 및 근육조직의 EST에서 상대적으로 높게 나타났으며 이는 뇌의 경우 signal transduction 관련, 소화관의 경우 protein turnover 및 ion transport 관련 그리고 근육의 경우 cell motility 관련 등의 유전자들이 높은 빈도로 발현하기 때문으로 판단된다. 난소 조직의 경우 cell structure와 직접적으로 관련된 egg membrane protein gene이 높은 빈도로 발현하였으나, 그 종류가 매우 제한적이어서 본 cellular process 그룹에 높은 score를 얻지는 못하였다. 반면 energy production/conversion 및 당/단백질/지방/핵산의 대사와 관련되는 'metabolism' 그룹은 대부분의 모든 조직에서 높게 나타났으며 특히 간 조직에서 가장 높게 발현하는 것으로 관찰되어 (54%), 생체의 다양한 대사 조절에 관여하는 본 조직의 특성을 잘 반영하고 있다. 마지막으로 아직 기능이 알려져 있지 않은

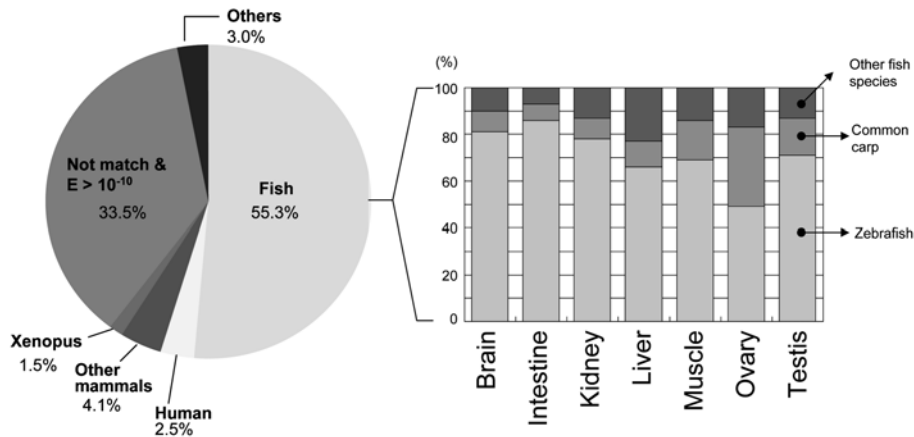


Fig. 2. Organism analysis of ESTs collected from different tissues of *Hemibarbus mylodon*. Organism source of the best matched orthologue from each BLASTx search was used to evaluate the frequency analysis.

단백질 등을 포함하는 'poorly characterized' 그룹의 경우 뇌 및 생식소 조직에서 상대적으로 높게 나타났으며(20~28%) 이는 현존하는 어류 유전자 database들이 뇌 및 생식소 기원의 유전자 수를 다른 조직 기원의 유전자들보다 적게 포함하고 있음과 동시에 어류가 아닌 다른 데이터베이스들 역시 본 조직들에서 발현하는 유전자들의 기능이 상대적으로 자세하게 규명되어 있지 못하기 때문으로 판단된다(Wang and Gong, 1999; Zheng and Gong, 2002; Douglas et al., 1999, 2007). 그러나 본 연구에서 활용된 functional clustering은 유전자 전체 open reading sequence가 아닌 EST 단편만을 이용한 결과이므로 향후 각 유전자에 대한 자세한 정밀 기능의 유추는 transcript 전체 염기서열의 확보는 물론 나아가 생물학적 *in vivo* 실험이 뒷받침되어야 할 것이다.

E-value에 따른 clustering 분석

각 조직 별 E-value에 의한 EST의 clustering은 다양하게 나타났다(see Table 1). EST의 e-value class를 no match를 포함한 총 5개 군으로 대별하여 분석할 경우 Brain은 각 class가 다양하게 분포되어 있으며 비교적 낮은 상동성을 보이는 $E > 10^{-10}$ 군이 33.7%로 가장 높게 나타났다. Intestine 및 kidney 역시 brain의 경우와 유사한 경향을 보임으로써 비교적 일정한 분포양상을 나타내었고 $E > 10^{-10}$ 군이 24.8% (intestine) 및 34.2% (kidney)로 높은 점유율을 보였다. 특히 kidney에서는 현존하는 GenBank 내 어떠한 클론과도 일치하지 않는 신규 transcript들이 다수 발굴되었음을 알 수 있다. 어류 liver에서 발현하는 유전자들은 비교적 기능이 잘 알려진 transcript들로 구성되어 있었으며 이에 상동성 여부 역시 $E < 10^{-50}$ 에 속하는 높은 상동성을 보이는 EST들이 전체 57% 이상을 차지하고 있음을 알 수 있었다. 또한 어류 근육 역시 liver와 유사한 경향을 보였으며 특히 어류 조직들 중 $E < 10^{-100}$ (GenBank 유전자와 매우 높은 상동성을 보이는 군)에 속하는 그룹이 다른 조직들에 비해 월등히

높게 나타났다. 두 종의 생식소 조직(난소 및 정소)은 서로 유사한 e-value 분포 양상을 보였다.

Best matched organism 분석

전체 어류 EST를 대상으로 상동성이 가장 높게 나타난 orthologues의 organism을 분석한 결과 organism 분석이 어려운 낮은 상동성 보유 클론들(no match 및 10^{-10} 이상의 e-value 클론들)을 제외할 경우 예상대로 대부분의 어류 EST들은 어류의 유전자들과 높은 상동성을 보이는 것으로 나타났다(Fig. 2). 조직 별로 그 정도의 일부 차이는 관찰되었으나 모든 조직에서 어류 유전자와 가장 많은 EST 클론의 best match가 이루어졌다. 어류 유전자와 상동성을 보이는 EST만을 다시 선별하여 종별 분석을 수행한 결과 각 조직에서 zebrafish 유전자와 가장 많은 match가 이루어졌는데 이는 현존하는 어류 EST database 중 zebrafish (*Danio rerio*) database가 여타 어종보다 월등히 많은 유전자 정보를 포함하고 있어 높은 확률로 match가 이루어졌기 때문으로 판단된다(<http://nibi.nlm.nih.gov/dbEST>). 어류 유전자와 상동성을 보이는 EST는 조직 별로 48~85% 정도가 zebrafish transcript와 best match를 이루었고 그 다음은 같은 잉어과 어류인 common carp (*Cyprinus carpio*)과 높은 상동성을 보였다.

본 연구를 통해 얻어진 발현 유전자들과 EST libraries들은 본 어종의 복원 시 요구되는 인공 증식 기술 등 여러 연구에 유용한 정보와 소재를 제공할 수 있으리라 기대된다. 요구되는 목적 유전자의 full-length cDNA 및 genomic 유전자 sequence의 클로닝에 유용한 바탕 정보를 제공케 될 것은 물론 인공 사육 시 유발될 수 있는 다양한 생물학적 스트레스를 검출하고 평가하기 위한 발현 분자 표지들로서 활용가능하다고 판단된다(Gong et al., 1997; Jenny et al., 2002; Savan and Sakai, 2002; Kono et al., 2003). 아울러 다양한 환경 스트레스 관련 EST들은 어류 서식지의 생물 환경을 평가할 수 있는 유용 생물 분자 지

표로 개발될 수 있을 것이며(Cho et al., 2006), 일부 simple repeat sequence 등을 포함하는 EST들은 transcribed genetic marker로서 어름치 계통 및 집단 분석에도 추후 활용될 수 있을 것으로 기대된다(Serapion et al., 2004). 앞으로 본 연구에서 확보된 조직별 발현 정보를 바탕으로 normalization 등을 통한 고품질의 EST gene bank 구축이 이루어져야 할 것이다.

요 약

멸종위기 천연기념물 어류 어름치(*Hemibarbus mylodon*)를 대상으로한 어름치 유전자 은행 구축 연구의 일환으로 뇌, 소화관, 근육, 간, 신장, 난소 및 정소 조직으로부터 expressed sequence tag (EST) library들을 구축하고 발현 유전자의 탐색을 실시하였다. EST 탐색을 통해 총 3,383개의 발현 유전자 염기서열 단편을 확보하였고 이들로부터 1,354개의 EST를 포함하는 총 333개의 contig들이 형성됨으로써 비교적 높은 빈도(69.8%)의 unigene 확보율을 나타내었다. EST의 조직 별 출현 양상은 orthologue들과의 상동성 정도 및 유추 기능의 대부분류를 기준으로 분석할 때 각 조직들은 서로 다른 특징을 나타내었다. 어름치에서 발굴된 EST들은 zebrafish의 유전자들과 가장 높은 match 빈도를 나타내었다. 본 연구를 통해 확보된 EST library들과 염기서열 정보는 본 종의 장외 복원을 위한 효율적인 인공증식 기술 개발에 유용한 기초 정보로 활용될 수 있을 것이다.

감사의 글

본 연구는 환경부 차세대 환경기술 개발 사업의 지원에 의해 이루어졌음.

참고문헌

- Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merrill, 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.
- Byeon, H.-K., J.-S. Choi, S. R. Jeon, J.-K. Choi and B.-Y. Song, 1996. Fish fauna in the streams of Mt. Chiak. *Korean J. Environ. Biol.*, 14, 47–54.
- Cho, Y.S., S. Y. Lee, I. C. Bang, D. S. Kim and Y. K. Nam, 2006. Transcriptional response of major antioxidant enzyme genes to heat stress in mud loach (*Misgurnus mizolepis*). *J. Aquacult.*, 19, 157–165.
- Douglas, S.E., L.C. Knichkle, J. Kimball and M. Reith, 2007. Comprehensive EST analysis of Atlantic halibut (*Hippoglossus hippoglossus*), a commercially relevant aquaculture species. *BMC Genomics*, 8, 144.
- Douglas, S.E., J. W. Gallant, C. E. Bullerwel, C. Wolff, J. Munholland and M. E. Reith, 1999. Winter flounder expressed sequence tags: establishment of an EST database and identification of novel fish genes. *Mar. Biotechnol.*, 1, 458–464.
- Gong, Z., T. Yan, J. Liao, S. E. Lee, J. He and C. L. Hew, 1997. Rapid identification and isolation of zebrafish cDNA clones. *Gene*, 201, 87–98.
- Jang, M.-H., M. C. Lucas and G.-J. Joo, 2003. The fish fauna of mountain streams in South Korean national parks and its significance to conservation of regional freshwater fish biodiversity. *Biol. Conserv.*, 114, 115–126.
- Jenny, M. J., A. H. Ringwood, E. R. Lacy, A. J. Lewitus, J. W. Kempton, P. S. Gross, G. W. Warr and R. W. Chapman, 2002. Potential indicators of stress response identified by expressed sequence tag analysis of hemocytes and embryos from the American oyster, *Crassostrea virginica*. *Mar. Biotechnol.*, 4, 81–93.
- Kim, K. Y., S. Y. Lee, Y. S. Cho, I. C. Bang, D. S. Kim and Y. K. Nam, 2007. Characterization and phylogeny of two beta-cytoskeletal actins from *Hemibarbus mylodon* (Cyprinidae, Cypriniformes), a threatened fish species in Korea. *DNA Seq.* DOI: 10.1080/10425170701445691.
- Kono, T, M. Sakai and S. E. LaPatra, 2000. Expressed sequence tag analysis of kidney and gill tissues from rainbow trout *Oncorhynchus mykiss* infected with infectious hematopoietic necrosis virus. *Mar. Biotechnol.*, 2, 493–498.
- Lee, Y. H., J. K. Noh, K. Y. Kim, Y. S. Cho, Y. K. Nam and D. S. Kim, 2007. Characterization of mud loach (*Misgurnus mizolepis*) Apolipoprotein A-I: cDNA cloning, molecular phylogeny and expression analysis. *J. Aquacult.*, 20, 65–72.
- Liu, X., H. Wang, and Z. Gong, 2006. Tandem-repeated zebrafish zp3 genes possess oocyte-specific promoters and are insensitive to estrogen induction. *Biol. Reprod.*, 74, 1016–1025.
- Marra, M. A., L. Hiller and R. H. Waterston, 1998. Expressed sequence tags – ESTablishing bridges between genomes. *Trends Genet.*, 14, 4–7.
- Moutou, K. A., S. Socorro, D. M. Power, Z. Mamuris and A. V. Canario, 2001. Molecular cloning and sequence of gilthead sea bream (*Sparus aurata*) alpha-skeletal actin: tissue and developmental expression. *Comp. Biochem. Physiol.*, 130B, 13–21.
- Savan, R. and M. Sakai, 2002. Analysis of expressed sequence tags EST obtained from common carp, *Cyprinus carpio* L., head kidney cells after stimulation by two mitogens, lipopolysaccharide and concanavalin-A. *Comp. Biochem. Physiol.*, 131, 71–82.
- 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.
- Suarez-Castillo, E. C. and J. E. Garcia-Arraras, 2007. Molecular evolution of the ependymin protein family: a necessary update. *BMC Evol. Biol.*, 7, 23.
- Suarez-Castillo, E. C., W. E. Medina-Ortiz, J. L. Roig-Lopez and J. E. Garcia-Arraras, 2004. Ependymin, a gene involved in regeneration and neuroplasticity in vertebrates, is overexpressed during regeneration in the echinoderm, *Holothuria glaberrima*. *Gene*, 334, 1333–1343.

- Venkatesh, B. and S. Brenner, 1998. Genomic structure and sequence of the pufferfish (*Fugu rubripes*) gene encoding an actin-related protein. *Gene*, 211, 169–175.
- Venkatesh, B., B. H. Tay, G. Elgar and S. Brenner, 1996. Isolation, characterization and evolution of nine pufferfish (*Fugu rubripes*) actin genes. *J. Mol. Biol.*, 259, 655–665.
- Wang, H. and Z. Gong, 1999. Characterization of two zebrafish cDNA clones encoding egg envelop proteins ZP2 and ZP3. *Biochim. Biophys. Acta.*, 1446, 156–160.
- Zheng, S. and Z. Gong, 2002. Expressed sequence tag analysis of expression profiles of zebrafish testis and ovary. *Gene*, 294, 45–53.

원고접수 : 2007년 10월 11일

수정본 수리 : 2007년 11월 6일