

PLANT & FOREST

# Construction of a full-length cDNA library from *Typha laxmanni* Lepech. and *T. angustifolia* L. from an EST dataset

Subin Im<sup>1,†</sup>, Ho-Il Kim<sup>2,†</sup>, Dasom Kim<sup>1</sup>, Sang Heon Oh, Yoon-Young Kim<sup>3</sup>, Ja Hyeong Ku<sup>1</sup>, Yong Pyo Lim<sup>1,\*</sup>

<sup>1</sup>Department of Horticulture, Chungnam National University, Daejeon 34134, Korea

<sup>2</sup>KISTI ReSEAT, Seoul 02456, Korea

<sup>3</sup>Department of Forest Biodiversity, Korea National Arboretum, Pocheon 11186, Republic of Korea

\*Corresponding author: [yplim@cnu.ac.kr](mailto:yplim@cnu.ac.kr)

†These authors contributed equally to this work.

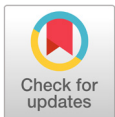
## Abstract

Genus *Typha* L. (*Typhaceae*; Cattail in common) is one of the hydrophytic plants found in semi-aquatic regions. About nine to 18 species of the genus exist all over the world. In Korea, the most commonly found cattail species are *T. laxmanni* and *T. angustifolia*. The aim of this study was to prepare a cDNA library and sequences and analyze expressed sequence tags (ESTs) from these species, *T. laxmanni* and *T. angustifolia*. In the case of *T. laxmanni*, we observed that 715 out of 742 ESTs had high quality sequences, whereas the remaining 27 ESTs were low quality sequences. In this study, we identified 77 contigs, 393 unassembled clones and 65.7% singletons. Furthermore, in the case of *T. angustifolia*, we recorded 992 high quality EST sequences, and by excluding 28 low quality sequences from among them, we retrieved 120 contigs, 348 unassembled clones and 48.9% singletons. The basic local alignment search tool (BLAST) and Kyoto encyclopedia of genes and genomes (KEGG) database results enabled us to identify the functional categories, i.e., molecular function (16.5%), biological process (22.2%) and cellular components (61.3%). In addition, between these two species, the no hits and anonymous genes were 4.2% and 11.7% and 6.2% and 11.2% in *T. laxmanni* and *T. angustifolia*, respectively, based on the BLAST results. The study concluded that they have certain species-specific genes. Hence, the results of this study on these two species could be a valuable resource for further studies.

**Keywords:** cDNA Library, Gene Ontology, *Typha angustifolia* L., *Typha laxmanni* Lepech

## Introduction

식물은 고착생활을 하므로 생장에 불리한 환경 변화로부터 이동을 할 수 없다(Dita et al., 2006). 이러한 환경 변화를 크게 두 가지로 비생물학적 스트레스(건조, 염, 고온, 저온, 홍수, 증금속 및 오존과 같은 오염물질에 의한 스트레스 등)와 생물학적 스트레스(세균, 곰팡이, 바



### OPEN ACCESS

**Citation:** Im S, Kim HI, Kim D, Oh SH, Kim YY, Ku JH, Lim YP. 2018. Construction of a full-length cDNA library from *Typha laxmanni* Lepech. and *T. angustifolia* L. from an EST dataset. Korean Journal of Agricultural Science. <https://doi.org/10.7744/kjoas.20180047>

**DOI:** <https://doi.org/10.7744/kjoas.20180047>

**Received:** March 12, 2018

**Revised:** June 19, 2018

**Accepted:** July 3, 2018

**Copyright:** © 2018 Korean Journal of Agricultural Science



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

이러스 감염에 의한 스트레스 등)로 나눌 수 있다. 이러한 환경을 극복하기 위해, 식물은 진화를 통해 다양한 전략을 발전시켜왔고, 스트레스에 대한 피해를 최소한으로 하도록 하는 특정한 메커니즘이 발달되었다(Ingram and Bartels, 1996; Thomashow, 1999).

식물의 적응력을 이용하여 토양이나 수자원, 공기 중으로부터 오염물질을 제거하거나 유해하지 않도록 처리하는 방법을 phytoremediation이라고 한다. 식물 종에 따라 내성 정도의 차이가 있지만(Parida and Das, 2005), 일반적인 토양오염 처리 방법에 비하여 phytoremediation은 안전하고 저렴하며 조작이 간단하다. 수질 정화 목적으로 갈대속(*Phragmites* Adans.), 억새속(*Miscanthus* Andersson.), 부들속(*Typha* L.) 등의 수생식물을 조성하기도 한다(Kim and Byoung, 1998; Kang and Gyu, 1999). 이 중, 부들속 식물은 부들과(*Typhaceae* Juss.)에 속하는 다년생 초본으로 전 세계에 약 9 - 18종이 분포한다고 알려져 있다(Heo, 2009). 주로 온대와 열대의 습지에 자라며 굵은 뿌리줄기와 잎자루가 없는 선상의 긴 잎을 가지고 있는 것이 특징이다(Heo, 2009). 부들은 번식이 빠르고 오염에 대한 저항성이 강하며 단위면적당 생물량이 많기 때문에 흡수하는 염류의 총량도 상대적으로 많다. 우리나라에는 부들(*Typha orientalis* L.), 애기부들(*T. angustifolia* L.), 큰잎부들(*T. latifolia* L.), 꼬마부들(*T. laxmanni* Lepech.) 총 4종이 자생하고 있는 것으로 알려져 있다(Korea National Arboretum and The Plant Taxonomic Society of Korea, 2007).

현재 의학적, 산업적 중요한 model 종들은 시퀀스 기술의 발전으로 WGS (Whole Genomic Sequencing)를 통해 전체 염기서열이 밝혀졌다. 하지만 WGS는 다양한 기반연구와 더불어 많은 비용과 시간이 필요하므로 여러 생물에 적용하는데 어려움이 있다. 또한 genome의 duplication이나 junk region 같은 요소의 증가로 염기서열의 분석이 어려우며 genome size가 발현되는 유전자의 전체 수를 반영하지 않는다는 연구결과가 있다(Thangadurai, 2004). 이를 극복하기 위해 유전자의 발현에 직접적으로 관여하는 정보만을 모아 프로파일링하는 방법으로 Expressed Sequence Tags (ESTs) 분석이 Whole Genome Sequencing (WGS) 대안으로 사용되고 있다(Rudd, 2003). EST는 일반적으로 poly (A)+ RNA로부터 얻어진 cDNA를 일컫는다. EST는 mRNA의 전사량을 반영하며 이는 특정 환경에 따른 발현 양상을 분석함으로써 유전자의 선발 및 분석에 효과적으로 이용된다(Park and Jang, 2010).

본 연구에서는 대전 인근에 자생하는 꼬마부들(*T. laxmanni* Lepech.)과 애기부들(*T. angustifolia* L.)의 EST library를 제작하여 분석한 결과를 토대로 기초적인 유전정보를 제공함으로써 다양한 분야에 활용하는데 목적을 두었다.

## Materials and Methods

### 실험재료

꼬마부들과 애기부들의 종자는 대전 근교의 습지에서 자생하는 식물체의 이삭으로부터 채취하였다. 물 속에 충분히 침지 시킨 이삭을 강한 힘으로 휘저어 종자를 분리하여 음건 후 4°C 냉장고에 보관하였다. 발아율 증가를 위해 Kim et al. (2005)의 방법으로 Sodium hypochlorite (NaOCl) 침지 처리하였고, 이를 충남대학교 실습당(36°22' 3.64" N, 127°21' 19.85" E)에서 재배, 수확하여 실험재료로 사용하였다.

### RNA 추출 및 cDNA library 제작

대전 근교에서 채집한 종자로 발아시켜 재배한 꼬마부들과 애기부들 개체로부터 cDNA library 제작을 위하여 total RNA를 분리하였다. 부들은 조직 내에 polyphenol 또는 polysaccharide의 함량이 높아 일반적으로 사용

되는 Trizol (acid guanidinium thiocyanate-phenol-chloroform) 방법으로 추출이 되지 않아 Rapid cetyltrimethylammonium bromide (CTAB) Method (Gambino et al., 2008)를 이용하여 total RNA를 추출하였다. cDNA library 제작을 위해 1 µg의 poly (A<sup>+</sup>) RNA를 In-Fusion<sup>®</sup> SMARTer™ cDNA library construction kit (Clontech, USA)를 이용하여 제조사의 지침에 따라 수행하였다(Okayama and Berg, 1982). 제작된 cDNA library의 각 clone들은 고체 LB + x-gal + IPTG + Ampicillin배지에서 배양하고 Blue/White selection법에 따라 선발하였다. 선발된 각각의 clone은 96 deep plate에서 배양 후 AccuPrep<sup>®</sup> 96 plasmid extraction kit (Bioneer, Daejeon, Korea)를 사용하여 제조사의 지침에 따라 plasmid DNA를 추출하였다.

## 염기서열 분석 및 BLAST N/X search

제작된 cDNA library는 국내 염기서열 분석 회사(Bioneer, Daejeon, Korea)에 염기서열 결정 및 분석 의뢰하여 결과 서열을 얻었다. 얻어진 결과 서열은 CodonCode Aligner Ver. 3.7.1 (CodonCode Corporation, Barnstable, USA)을 사용하여 Quality 확인 및 Vector trim을 수행하고 Contig assembling 후 분류되지 않은 서열은 singleton으로 처리하였다. 얻어진 서열을 BLAST2GO (<http://blast2go.bioinfo.cipf.es/>) 프로그램을 사용하여 각각의 서열의 상동성 탐색 및 gene ontology의 분석을 하였다. BLASTX의 cut off 값은 10<sup>-5</sup>로 설정하였고 mapping은 E-value 10<sup>-5</sup>, annotation cut off 값은 45, GO weight 값은 10으로 설정하였다(Conesa et al., 2005).

분석된 서열들은 유사도 검색을 위하여 NCBI (<http://www.ncbi.nlm.nih.gov/>)의 단백질 데이터베이스(nr)와 BLASTX를 수행하여 가장 유사도가 높은 결과를 찾아 선택하였다. 또한 gene ontology protein database와 BLASTX를 수행한 후, BLAST2GO 프로그램을 이용하여 기능을 찾아, 그 결과를 세 가지(Biological process, Cellular component, Molecular function)로 분류하였다. 각각의 서열을 대한 BLAST 검색을 통하여 기존의 database에 상동 서열이 존재하지 않는 것은 no hit, 그리고 기능을 알 수 없는 유전자와 일치하거나 기능이 분류되지 않은 유전자와 상동성이 높은 경우 unknown gene 그리고 나머지는 known으로 분류하여 각 종별로 분석을 하였다.

## Results and Discussion

### 꼬마부들 및 애기부들의 cDNA library 제작

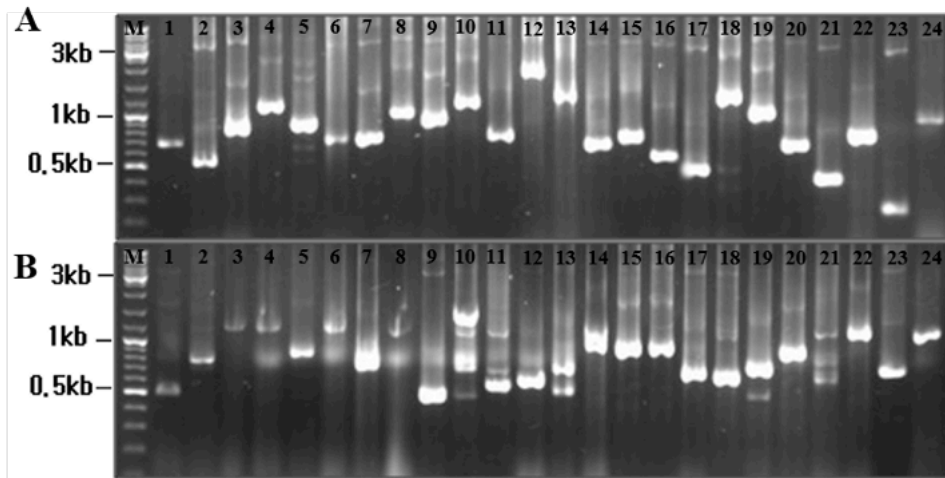
벡터와의 재조합 과정 및 선발 과정을 통해 얻은 꼬마부들 및 애기부들의 cDNA library로부터 임의의 clone을 선택하여 insert size를 확인하였다. 애기부들의 경우 임의로 선택된 clone들의 insert size의 확인 결과 약 0.5 - 2.5 kb 범위의 결과를 보였으며, 꼬마부들의 경우도 약 0.5 - 2.5 kb 범위의 insert size를 보였다. 이것은 cDNA library에 절편이 잘 삽입되었음을 보여준다(Fig. 1).

### Sequencing 및 BLAST N/X search

cDNA library의 clone의 수는 시료 채취 당시의 발현 유전자의 mRNA의 종류와 transcription의 양을 대변한다. 식물의 생육 당시 발현되는 유전자의 종류와 기능을 알기위해서는 고유한 cDNA library로부터 얻은 unigene의 수를 알아야 한다(Huang and Madan, 1999). 애기부들은 low quality서열 27개를 제외한 715개의 서열을 분석하여 77개의 contig와 363개의 unassembled clone을 얻어, 전체 서열 대비 65.7%의 singleton을 확보하였다. 꼬마부들은 28개의 low quality서열을 제외한 992개의 서열을 분석하여 120개의 contig와 348개

의 unassembled clone을 얻어 전체 대비 48.9%의 singleton을 확보하였다(Table 1). 2개 이상의 clone으로 assemble된것을 contig라하며, 해당 contig가 대표하는 유전자는 singleton의 유전자들보다 많은 수준으로 전사된 유전자라고 할 수 있다(White et al., 2000). 이를 통해 염기서열 분석을 위한 시료 채취 단계에 비취 애기부들의 경우 unigene의 10.76%에 해당하는 contig 유전자들은 영양생장기에 발현되는 유전자들이라고 볼 수 있다. 꼬마부들은 12.09%에 해당하는 contig들이 singleton의 유전자들 보다 영양생장에 더 많이 관련된 유전자라고 할 수 있다.

애기부들과 꼬마부들의 분석된 각 clone의 서열은 유사도 검색을 위하여 NCBI (<http://www.ncbi.nlm.nih.gov/>)의 단백질 데이터베이스(nr)와 BLASTX를 수행하여 가장 유사도가 높은 결과를 찾아 전체 EST를 'No hit', 'Known gene', 'Unknown gene'으로 구별 하였다. 애기부들과 꼬마부들의 유사도 분석결과, GeneBank에 등록된 어떤 유전자와도 상동성을 나타내지 않는 unigene이 각각 4.2%, 6.2%를 나타냈다. GeneBank에 등록된 유전자와 높은 상동성을 나타내는 경우는 각각 84.0%, 82.4%를 나타냈고(Table. 2), GeneBank에는 등록이 되어 있으나 기능이 알려지지 않은 유전자와 상동성이 높은 경우가 각각 11.7%, 11.2%를 나타냈다. 이중 'Unknown gene'은 각종의 특이적 유전자일 가능성이 높은 것으로 사료된다. 또한 상대적으로 높게 나온 'Known gene'의



**Fig. 1.** Identification of cDNA library inserts size. (A) *T. angustifolia* Lepech; M-size marker; lane 1to 24-PCR products from different clones. (B) *T. laxmanni*; M-size marker; lane 1 to 24-PCR products from different clones.

**Table 1.** Summary of cDNA library construction in *Typha angustifolia* and *Typha laxmanni*.

Species	Total ESTs	Excluded by quality	Total remaining	Contigs	Unassembled clones	% of singletons
<i>Typha laxmanni</i> Lepech.	742	27	715	77	393	65.7
<i>Typha angustifolia</i> L.	1,020	28	992	120	348	48.9

ESTs, Expressed Sequence Tags.

**Table 2.** Classification of functional annotations of all ESTs based on BLASTX analysis.

Species	No hit (%)	Known gene (%)	Unknown gene (%)
<i>Typha laxmanni</i> Lepech.	4.2	84.0	11.7
<i>Typha angustifolia</i> L.	6.2	82.4	11.2

ESTs, Expressed sequence tags.

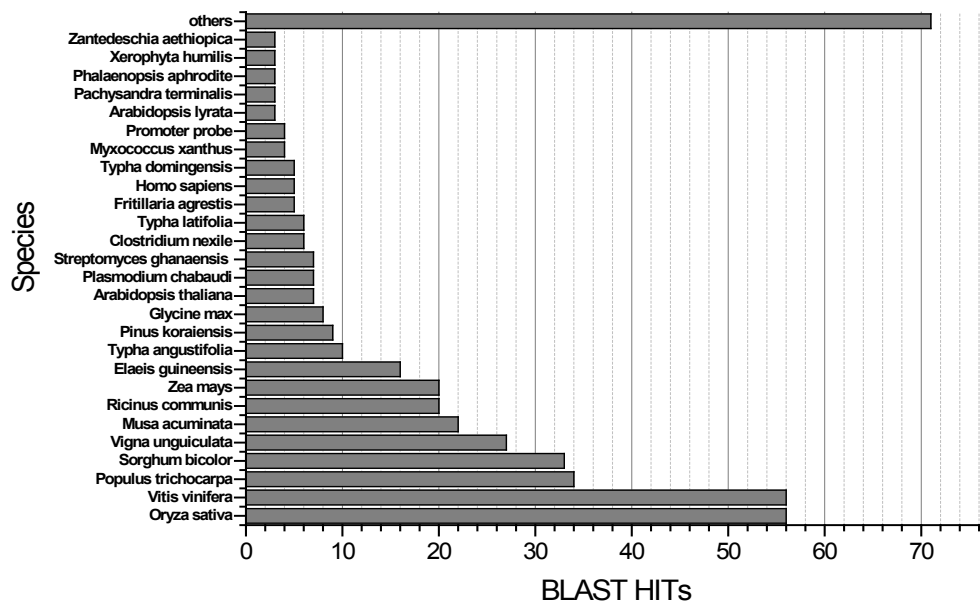
비율은 모델 식물체 또는 다양한 종의 유전체 분석자료 (EST 또는 next generation sequencing)의 축적에 따른 결과로 판단된다(Kim et al., 2011).

## Funcat analysis 및 GO 분석

애기부들의 Funcat analysis 및 Gene ontology 분석 결과 전체 742개의 서열에 대한 BLASTX 분석결과 *Oryza sativa* (8%, 56/672), *Vitis vinifera* (8%, 56/672), *Populus trichocarpa* (5%, 34/672), *Sorghum bicolor* (5%, 33/672)의 순으로 높은 일치도를 보였다(Fig. 2). 상위의 상동성을 보인 *Oryza sativa*의 경우 GeneBank의 EST database (<https://www.ncbi.nlm.nih.gov/nucest>)에서 전체 분석된 EST database가 1,281,057개이다. 이는 같은 비율의 상동성을 보인 *Vitis vinifera* 경우 511,695개의 EST data보다 2.5배 많음에도 불구하고 동일한 값을 나타낸 것을 보아 애기부들의 유전적 특성이 *Vitis vinifera*와 유전적으로 밀접한 관계임을 보임을 알 수 있다. 하지만 애기부들은 외떡잎 식물이며 *Vitis vinifera*은 쌍떡잎식물로 분류학적 유연관계가 밀접하지 않음에도 불구하고 높은 상동성을 보이는 이유에 대한 연구가 더 필요한 실정이다. 하위의 상동성을 가진 종들에는 사람, 세균 같은 plantae가 아닌 종들이 있으며, 이는 같은 기능을 하는 유전자일지라도 생물 종에 따라 유전자의 기능이 다르게 작용한다는 것을 보여주는 결과로, 이 유전자는 애기부들의 특이적 유전자일 것으로 추정된다(Boo et al., 2007). 하지만 본 연구에서 사용된 clone은 full-length clone이 아닌 EST 단편을 이용하여 얻은 결과로서 full-length cDNA 확보를 통한 신뢰성을 높일 필요가 있다.

애기부들로부터 얻어진 742개의 서열을 functional category에 따라 분류한 결과 molecular function 17.8%, biological process 26.5% 그리고 cellular component 55.7%의 비율로 나타났다. Gene ontology가 결정된 EST 서열을 BLAST2GO software를 사용하여 KEGG database 검색을 통하여 분석하였다. 전체 742개의 EST중 354개의 서열에서 대사산물합성 관련 유전자들이 확인되었다(Fig. 3).

꼬마부들의 Funcat analysis 및 Gene ontology 분석 결과 전체 1020개의 서열에 대한 BLASTX 분석결과

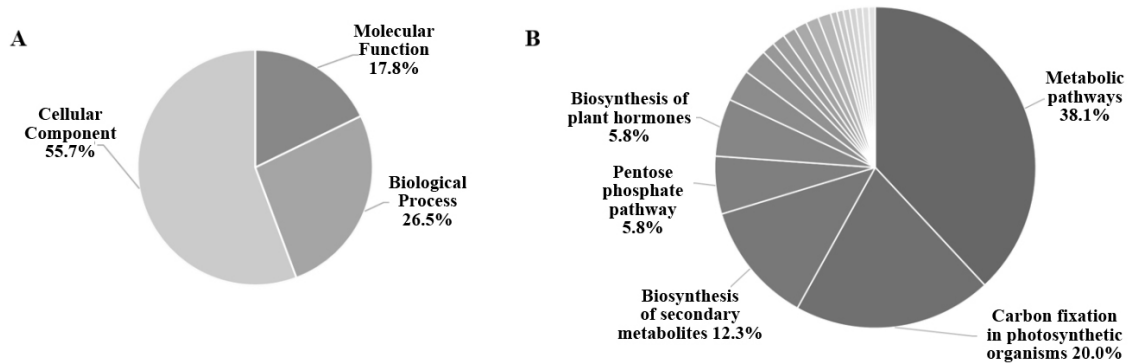


**Fig. 2.** Graphical representation of *Typha angustifolia* sequence similarity with top-hits plant species using BLASTX.

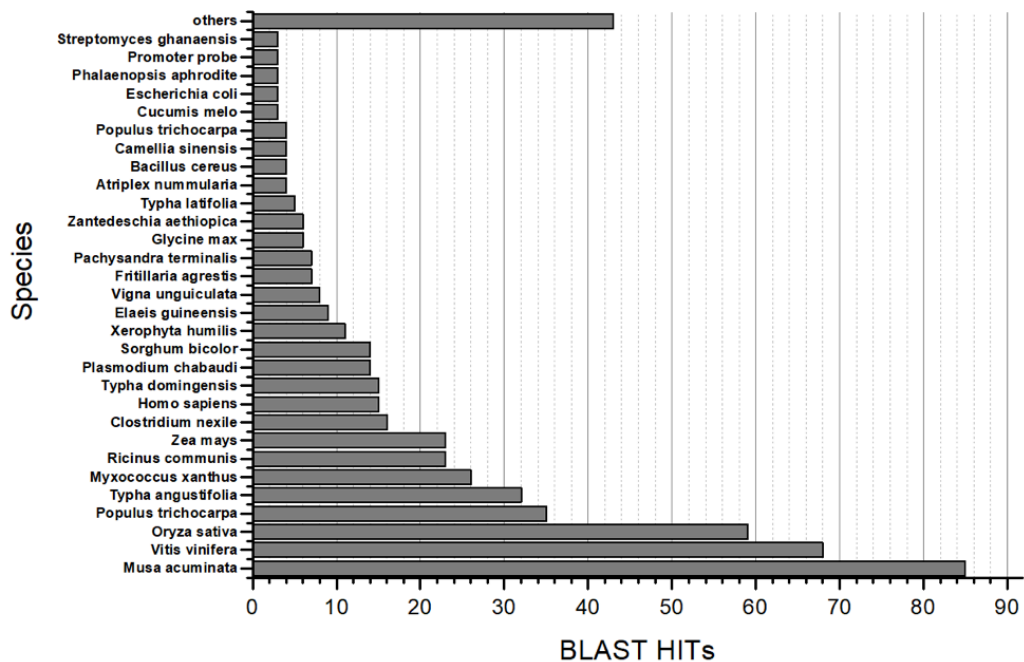


*Musa acuminata* (9.3%, 95/1,020), *Vitis vinifera* (6.5%, 67/1,020) 그리고 *Oryza sativa* (5.7%, 59/1,020)의 순으로 높은 일치도를 보였다(Fig. 4). 상위의 상동성을 보이는 *Musa acuminata*의 경우 GeneBank의 EST database (<https://www.ncbi.nlm.nih.gov/nucest>)에서 전체 분석된 EST database가 29,705개이며, 이는 *Vitis vinifera*의 511,695개, *Oryza sativa*의 1,281,057개에 비해 현저히 적은 수입에도 가장 높은 상동성을 보였다. 이는 *Musa acuminata*의 상동성이 꼬마부들과 유전적 특성이 다른 종들에 비해 가까움을 나타낸다.

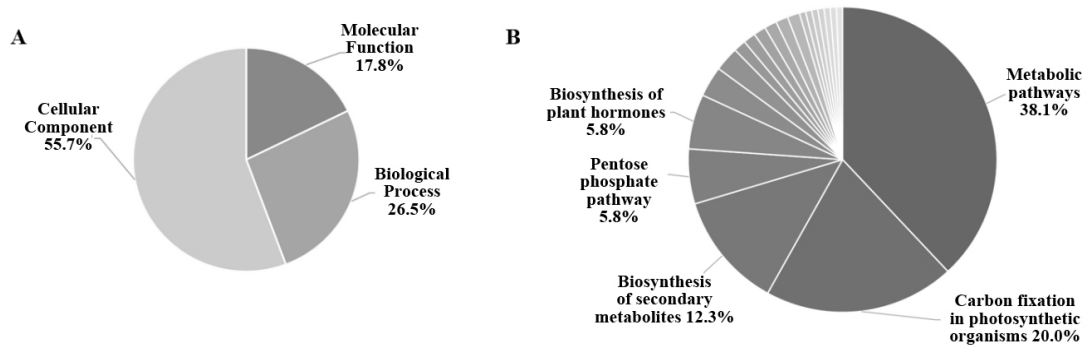
꼬마부들로부터 얻어진 1,020개의 서열을 functional category에 따라 분류한 결과 molecular function 16.5%, biological process 22.2% 그리고 cellular component 61.3%의 비율로 나타났다. Gene ontology가 결정된 EST 서열을 BLAST2GO software를 사용하여 KEGG database 검색을 통하여 분석하였다. 전체 1,020



**Fig. 3.** Classification of ESTs functions in *Typha angustifolia* with different databases (A) BLAST2GO and (B) Kyoto encyclopedia of genes and genomes (KEGG).



**Fig. 4.** Graphical representation of *Typha laxmanni* sequence similarity with top-hits plant species using BLASTX.



**Fig. 5.** Classification of ESTs functions in *Typha laxmanni* with different databases (A) BLAST2GO and (B) Kyoto encyclopedia of genes and genomes (KEGG).

개의 EST중 686개의 서열에서 대사산물합성 관련 유전자들이 확인되었다(Fig. 5).

본 연구를 통해 얻어진 애기부들과 꼬마부들은 EST 분석결과를 통해 얻어진 유전 정보들을 통하여 토양 및 수질오염에 대한 자정과 같은 산업적인 이용 등 다양한 분야의 기초자료로 활용될 것으로 기대된다. 애기부들과 꼬마부들에서 밝혀낸 유전자가 1,000개로 일반적으로 알려진 식물의 전 생육기간 동안 발현하는 유전자의 수는 15,000개에서 60,000개로 부족한 상황이다(Flavell, 1980). 이를 보완하기 위해 추가적인 유전자 확보를 위해 full-length cDNA 등 발현 유전체 정보의 확보가 필요하다. 또한 분석된 유전자들의 기능과 역할을 밝히기 위하여 추후 생리적 연구를 수행할 예정이다.

## Acknowledgements

본 연구는 “충남대학교 자체연구과제” 지원에 의해 수행되었다.

## References

- Boo KH, Kim DW, Cho SK, Jin SB, Kim JH, Lee HY, Riu KZ. 2007. Construction and profiling of a cDNA library from young fruit of satsuma mandarin. *Journal of Plant Biology* 50:403-409.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676.
- Dita MA, Rispaill N, Prats E, Rubiales D, Singh KB. 2006. Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. *Euphytica* 147:1-24.
- Flavell R. 1980. The molecular characterization and organization of plant chromosomal DNA sequences. *Annual Review of Plant Physiology* 31:569-596.
- Gambino G, Perrone I, Gribaudo I. 2008. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis* 19:520-525.
- Heo JA. 2009. Cultural and usable characteristics and absorbing ability of heavy metals in cattail (*Typha* spp.). Ph.D. Diss., Chungnam univ., Daejeon, Korea.
- Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Research* 9:868-877.

- Ingram J, Bartels D. 1996. The molecular basis of dehydration tolerance in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 47:377-403.
- Kang HC, Gyu JY. 1999. The structural characteristics in natural wetlands and fitted depth zones of *Phragmites Japonica*. *Journal of Korean Institute of Traditional Landscape Architecture* 17:191-200.
- Kim JK, Im SB, Choi SH, Lee JS, Roh MS, Lim YP. 2011. Construction of a full-length cDNA library from *Pinus koraiensis* and analysis of EST dataset. *Korean Journal of Agricultural Science* 38:11-16.
- Kim SY, Byoung BY. 1998. Studies on the photosynthetic characteristics of *Oenanthe stolonifera* DC. *The Korean Society for Horticultural Science* 6:64-65.
- Kim YJ, Heo JA, Hwang YS, Ku JH. 2005. Effect of sodium hypochlorite pretreatment, light intensity and depth of soil covering on germination of cattail (*Typha* spp.) seeds. *Weed & Turfgrass Science* 19:115-123.
- Korea National Arboretum and The Plant Taxonomic Society of Korea. 2007. Synonymic list of vascular plants in Korea. p. 367. Korea National Arboretum, Pocheon, Korea.
- Okayama H, Berg P. 1982. High-efficiency cloning of full-length cDNA. *Molecular and Cellular Biology* 2:161-170.
- Parida AK, Das AB. 2005. Salt tolerance and salinity effects on plants. *Ecotoxicology and Environmental Safety* 60:324-349.
- Park W, Jang YS. 2010. Construction and characterization of a cDNA library from the *Camelina sativa* L. as an alternative oil-seed crop. *Korean Journal of Crop Science* 55:151-158.
- Rudd S. 2003. Expressed sequence tags: Alternative or complement to whole genome sequences? *Plant Science* 8:321-329.
- Thomashow MF. 1999. Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. *Annual Review of Plant Biology* 50:571-599.
- Thangadurai S. 2004. The human genome project: The role of analytical chemists. *Analytical sciences* 20:595-601.
- White JA, Todd J, Newman T, Focks N, Girke T, de Ilarduya OM, Jaworski JG, Ohlrogge JB, Benning C. 2000. A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiology* 124:1582-1594.