

배추의 조직 특이적 발현유전자 데이터베이스

유희주¹ · 박신기² · 오미진² · 황현주² · 김남신³ · 정희¹ · 손성한² · 박범석² · 문정환^{2*}

¹가톨릭대학교 생명과학과, ²국립농업과학원 유전자분석개발과, ³한국생명공학연구원 국가생명연구자원정보센터

The *Brassica rapa* Tissue-specific EST Database

Hee-Ju Yu¹, Sin-Gi Park², Mijin Oh², Hyun-Ju Hwang², Namshin Kim³, Hee Chung¹, Seong-Han Sohn²,
Beom-Seok Park², and Jeong-Hwan Mun^{2*}

¹Department of Life Sciences, The Catholic University of Korea, Bucheon 420-743, Korea

²Department of Agricultural Biotechnology, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-707, Korea

³Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea

Abstract. *Brassica rapa* is an A genome model species for *Brassica* crop genetics, genomics, and breeding. With the completion of sequencing the *B. rapa* genome, functional analysis of the genome is forthcoming issue. The expressed sequence tags are fundamental resources supporting annotation and functional analysis of the genome including identification of tissue-specific genes and promoters. As of July 2011, 147,217 ESTs from 39 cDNA libraries of *B. rapa* are reported in the public database. However, little information can be retrieved from the sequences due to lack of organized databases. To leverage the sequence information and to maximize the use of publicly-available EST collections, the *Brassica rapa* tissue-specific EST database (BrTED) is developed. BrTED includes sequence information of 23,962 unigenes assembled by StackPack program. The unigene set is used as a query unit for various analyses such as BLAST against TAIR gene model, functional annotation using MIPS and UniProt, gene ontology analysis, and prediction of tissue-specific unigene sets based on statistics test. The database is composed of two main units, EST sequence processing and information retrieving unit and tissue-specific expression profile analysis unit. Information and data in both units are tightly inter-connected to each other using a web based browsing system. RT-PCR evaluation of 29 selected unigene sets successfully amplified amplicons from the target tissues of *B. rapa*. BrTED provided here allows the user to identify and analyze the expression of genes of interest and aid efforts to interpret the *B. rapa* genome through functional genomics. In addition, it can be used as a public resource in providing reference information to study the genus *Brassica* and other closely related crop crucifer plants.

Additional key words: crop, gene expression, gene function, public database, unigene

서 언

배추과(Brassicaceae)는 약 350속 3,700여 종을 포함하는 현화 식물 중 다섯 번째 큰 과로서 특히 배추, 양배추, 흑겨자, 무 등이 속하는 배추족(Brassiceae tribe) 식물들은 채소, 양념, 사료, 또는 유지 작물로 개발되어 전세계적으로 널리 재배되고 있다. 다양한 배추과 식물들이 유전학 및 유전체학과 육종을 위하여 연구되고 있는데, 이들 가운데 배추(*Brassica rapa*)는 유채의 한 쪽 조상일 뿐만 아니라 배추속

(genus *Brassica*) 식물의 A genome 모델로서 배추과 식물 유전 및 유전체 연구의 중심 대상이 되어 왔으며 2011년 8월 다국적 유전체 해독 컨소시엄에 의해 배추의 유전자 영역 약 284Mb에 대한 해독이 완료되었다(The *Brassica rapa* genome sequencing project consortium, 2011). 배추 유전체는 Illumina의 GAII 차세대 염기서열 분석기기로 생산된 약 72X의 짧은 서열들을 컨티그(contig)로 연결한 후 약 20만 개 BAC 말단 서열과 통합하여 스캐폴드(scaffold)로 조립되었다. 조립 서열에 대한 유전자 예측(annotation) 결과 41,174

*Corresponding author: munjh@rda.go.kr

※ Received 14 October 2011; Accepted 4 November 2011. 본 연구는 차세대바이오그린21사업(과제번호: PJ008019)의 지원에 의해 수행되었음.

개의 단백질을 암호화하는 염기서열이 발견되었다. 배추의 유전체 해독이 완료됨에 따라 유전자의 기능을 밝히고 유용 유전자를 발굴하여 신품종 개발 등 육종에 활용하기 위한 기능 유전체 연구가 활발히 진행될 것으로 기대된다.

배추 유전자의 기능 분석을 효율적으로 수행하기 위해서는 유전체 서열뿐만 아니라 다양한 생물정보 및 실험 데이터들이 필요하다. 배추 유전체의 유전자 예측과 기능 연구를 위한 필수 연구 재료로서 발현유전자(expressed sequence tags, ESTs) 분석이 진행되어 왔다(Mun et al., 2010). 2011년 5월 현재 미국의 국립생물정보센터(NCBI)에는 국립농업과학원에서 등록한 127,144개 등 총 147,217개의 발현유전자가 등록 공개되어 있다. 한편, 해독 완료된 배추 유전체의 단백질 암호화 유전자와 발현유전자를 BLAST 분석할 경우 유전체 유전자 모델과 일치하지 않는 발현유전자 서열이 존재한다(The *Brassica rapa* genome sequencing project consortium, 2011). 1회 염기서열 결정(single pass)에 따른 유전자 서열의 부정확성 또는 키메라 클론(chimera clone)의 생성 등 발현유전자 분석 방법의 문제점에도 불구하고 유전체 유전자 모델에서 예측되지 않는 발현유전자가 존재한다는 점은 배추 구조 및 기능 유전체 연구를 위해 발현유전자 분석이 지속적으로 필요하다는 점을 제시한다. 특히 조직 특이성 등 발현 특성에 따른 유전자의 분류와 발굴은 프로모터 등 유전자 지적재산권을 확보하는 출발점이 되고 있다. 그러나 공개된 발현유전자 데이터는 각 서열별로 단편적이며 조직 특이성 등 발현 특성에 대한 정보가 체계적으로 정리되어 있지 않다. 그 결과 생물정보 분석 기법을 이용하여 서열 정보를 직접 발굴(data mining)할 수 없는 일반 연구자들의 경우 발현유전자 정보의 활용을 통한 정확한 후보 유전자의 발굴에 어려움이 많다.

본 연구에서는 배추 발현유전자를 단일 조합 유전자서열(unigene)로 정리하고, TAIR와 MIPS, UniProt 데이터베이스에 대한 탐색, gene ontology(GO) 분석과 통계분석을 수행하여 배추의 조직 특이적 발현유전자를 발굴하고, 이를 연구자들이 쉽게 사용할 수 있는 공공 데이터베이스로 개발하였다. 배추의 조직 특이성 발현유전자 데이터베이스(The *Brassica rapa* tissue-specific EST database, BrTED)는 크게 발현유전자 서열 처리 및 정보 검색 단위와 조직 특이성 유전자 분석 단위로 구성되어 있다. 각 단위는 상호 연결되어 웹 기반 검색 시스템으로 사용자들에게 제공된다. 한편 BrTED를 활용하여 발굴한 조직 특이성 발현유전자를 Reverse Transcription-Polymerase Chain Reaction(RT-PCR)로 확인함으로써 데이터베이스의 실효성을 검증하였다.

배추 발현유전자 정보

본 연구에서는 NCBI에 등록되어 있는 배추 발현유전자 147,217개를 사용하였다. 이중 127,144개 발현유전자는 국립농업과학원에서 다양한 조직과 발달 단계, 그리고 병 또는 스트레스 처리 후에 제작한 배추 cDNA library 28개로부터 개발되었으며, 나머지 20,073개 발현유전자는 NCBI dbEST (<http://www.ncbi.nlm.nih.gov/nucest>)에 보고된 서열로서 11개의 library로부터 개발된 것이다(Table 1).

발현유전자 단일 조합 유전자서열 제작과 기능 예측

단일 조합 유전자서열 제작은 StackPACK v2.2 프로그램(South African National Bioinformatics Institute)을 이용하여 98% 일치도 기준으로 조립하였다. 배추 발현유전자 단일 조합 유전자서열 세트의 기능을 예측하기 위하여 BLAST 프로그램(Altschul et al., 1990)을 이용하여 애기장대의 TAIR 유전자 모델(<http://www.arabidopsis.org>), GO 데이터, MIPS의 Functional Category(FunCat)에 비교 분석하였다(Carbon et al., 2009; Ruepp et al., 2004). 또한 UniProt 및 EC peptide 데이터와 비교하여 단일 조합 유전자서열이 암호화하는 단백질 정보를 추론하였고 EC 번호에 따라 유전자를 분류하였다(Scheer et al., 2011; UniProt Consortium, 2009).

배추 단일 조합 유전자서열 조직 특이성 분석

조직 특이성 분석은 특정 조직에서 유전자의 세포 내 평균 발현 양 보다 특이적으로 높게 발현되어 조직의 특성을 결정지을 수 있는 유전자를 선별하는 것을 목적으로 하기 때문에 무작위적인 방식의 발현유전자 서열 분석에 대한 확률 값을 계산해야 한다. 따라서 공통서열(tandem consensus) 단일 조합 유전자서열은 library의 식물체 조직 정보를 대상으로 p value 0.01 수준에서 포아송 분포(Poisson distribution)를 이용한 Audic's test(Audic and Claverie, 1997)를 수행하여 조직 특이 유전자를 선별하였다. 한편 단독서열(singleton)은 library의 조직 정보를 이용하여 조직 특이성을 직접 정의하였다.

데이터베이스 개발 환경

상기한 방법으로 분석한 모든 데이터를 효율적으로 보여 주기 위하여 Intel® Xeon® CPU E5430 2.66 GHz와 32 GB RAM 사양의 서버에 Red Hat 4.3.2-7의 운영체계를 이용하였다. 데이터베이스는 Mysql V.5.1.32를 사용하여 제작하였고, 데이터베이스-웹 브라우저의 연결과 웹 브라우저 상의

Table 1. Summary of *B. rapa* ssp. *pekinensis* ESTs used in this study.

Library	Tissue source	EST number	NCBI accession
KBAY	Anther, young anther	1,859	Ex015357 - Ex017215
KBCD	Whole plant, cold Treated	6,732	Ex017216 - Ex023947
KBCG	Callus, developing callus	11,847	Ex023948 - Ex035794
KBFL	Floral bud, > 2 mm in size	10,332	Ex035795 - Ex046126
KBFS	Floral bud, < 2 mm in size	9,102	Ex046127 - Ex055228
KBLS	Whole plant, salt treated	6,894	Ex055229 - Ex062122
KBLW	Non-photosynthetic mature leaf	2,379	Ex062123 - Ex064501
KBPL	Mature pollen	3,587	Ex064502 - Ex068088
KBPS	Seedling, 1 week old	4,174	Ex068089 - Ex072262
KBRT	Root, 1 month old	4,682	Ex072263 - Ex076944
KBSP	Silique, 1 to 10 days after pollination	6,226	Ex076945 - Ex083170
KBSQ	Silique, 10 to 25 days after pollination	933	Ex083171 - Ex084103
KBST	Floral stem, bolting	1,577	Ex084104 - Ex085680
KCOV	Ovule, before pollination	421	Ex115626 - Ex116046
KCOW	Ovule, 5 to 10 days after pollination	1,225	Ex116047 - Ex127271
KFFB	Floral bud, open flower	8,771	Ex085681 - Ex094451
KFPC	Leaf, <i>Pectobacterium carotoborum</i> infected	3,708	Ex094452 - Ex098159
KFRT	Root, mixture of 1, 3, and 7 weeks old	3,555	Ex098160 - Ex101714
KFSD	Mature seed and 2 days old germinating seed	4,689	Ex101715 - Ex106403
KFYP	Young plant, 3 weeks old	8,408	Ex106404 - Ex114811
KHCT	Cotyledon, in greening stage	2,733	Ex117272 - Ex120004
KHLD	Defected leaf	2,037	Ex120005 - Ex122041
KHLM	Mature green leaf	3,423	Ex122042 - Ex125464
KHLW	Non-photosynthetic mature leaf	2,541	Ex125465 - Ex128005
KHOS	Ovule and silique	2,390	Ex128006 - Ex130395
KHRT	Root, mixed stage and treatment	8,265	Ex130396 - Ex138660
KLPS	Seedling, 1 week old, etiolated	3,840	Ex138661 - Ex142500
NRFB	Floral bud and open flower, normalized	814	Ex114812 - Ex115625
12405	Heading leaf, folding stage	2,363	CA991436 - CA992590 DN191369 - DN192576
1392	Guard cell	536	AT000431 - AT000966
16161	Leaf	1,438	CO749247 - CO750684
16490	Root	2,062	CV432063 - CV434123, CV468724
16541	Flower	137	CV523155 - CV523240, CV544312 - CV544362
16551	Seedling, NaCl treated	2,729	CV544363 - CV547091
16632	Seedling	746	CV650357 - CV651102
17359	Leaf, one-month old	6,282	DN960311 - DN966599
1846	Flower bud	397	AT001683 - AT001881, AT002057 - AT002257
8791	Etiolated seedling	3,382	BG543030 - BG545028, BQ790644 - BQ792026
1902	Flower bud	1	AF086818
Total		147,217	

데이터 구현은 각각 PHP V.5.2.9와 Apache v.2.0을 사용하였다. 데이터베이스 웹 사이트는 <http://BrTED.rma.kr>이다.

조직 특이적 발현유전자 선발의 실효성 검증

데이터베이스를 이용하여 선발한 조직 특이적 유전자의 발현 양상을 검증하기 위하여 RT-PCR을 수행하였다. 지부

배추의 결구기 잎과 뿌리, 그리고 꽃 등 3개 조직의 시료 약 5g을 각각 채취하여 액체질소 하에서 동결 분쇄한 후 Hybrid-R kit(GeneAll, Seoul, Korea)을 이용하여 전체 RNA를 추출하였다. 각 조직의 약 1 μ g의 RNA로부터 Reverse Transcription System(Fermentas, Burlington, Canada)를 이용하여 cDNA를 합성하였다. 합성한 cDNA 1 μ L(0.05 μ g)에 각각 5 μ M의 상류와 하류 프라이머, 10X PCR buffer, 2mM의 dNTP 혼합액, 0.5U DNA polymerase로 만든 PCR 혼합액을 넣고 ddH₂O로 최종 반응 부피를 30 μ L로 하였다. PCR은 프라이머의 조건에 따라서 Tm 값을 조절하였으며 유전자의 증폭 양에 따라 20-24 cycles로 반응하였다. RT-PCR에 사용된 프라이머는 Pickprimer 프로그램(Chung et al., 2011)을 이용하여 제작하였다. RT-PCR의 유전자 대조구로 배추 actin(*BrACT*)을 사용하였으며, 사용된 프라이머 서열은 BrACT-F는 5'-ATATCCGATCGAGCACGGTA-3'와 BrACT-R로서 5'-CGCCCACTAGCGTAAAGAGA-3'이다.

결과 및 고찰

발현유전자 단일 조합 유전자서열 제작과 조직 특이성 분석

배추의 조직 특이적 발현유전자 데이터베이스 제작에 사용한 데이터와 분석 방법 및 프로그램, 데이터베이스 구성 등 기본 전략은 Fig. 1과 같다. EST는 발현된 유전자의 단편으로 길이가 짧고 중복적인 서열을 다수 포함하고 있다. 따라서 각 조직에서 발현되는 특정 서열 조각들로부터 중복 서열을 제거하고, 중첩되는 서열을 조립하여 보다 긴 서열로 가공한 단일 조합 유전자 서열로 정리하였다. 배추는 전체 유전체가 3배수화되어 있기 때문에 paralog 유전자간 서열 유사도가 높다(Mun et al., 2009). 이 점을 고려하여 중첩되는 서열간 서열 일치도 98% 수준에서 StackPACK 프로그램을 사용하여 EST 클러스터링(clustering)과 어셈블리(assembly)를 제작하였다(George, 2001). 그 결과 공통서열 14,779개와 단독서열 9,183개 등 총 23,962개의 단일 조합 유전자서열을 얻었다. 전체 단일 조합 유전자서열과 발표된 배추 유전체 유전자 모델을 BLASTN(E⁻⁵⁰, query coverage > 50%) 비교한 결과 90%의 단일 조합 유전자서열이 유전자 모델에 매핑되었으나 약 10%의 단일 조합 유전자서열은 유전자 모델과 일치하지 않아 현재까지 해독 완료된 유전체 이외의 영역에 존재하는 유전자를 반영하는 것으로 추정된다(Table 2). 제작된 단일 조합 유전자서열 데이터 세트의 유전자 정보 기능 예측을 위하여 TAIR의 애기장대 유전자와 UniProt 데이터베이스를 BLASTX(E⁻¹⁰, query coverage > 50%) 검색한 결과, 약 88%의 단일 조합 유전자서열이 이

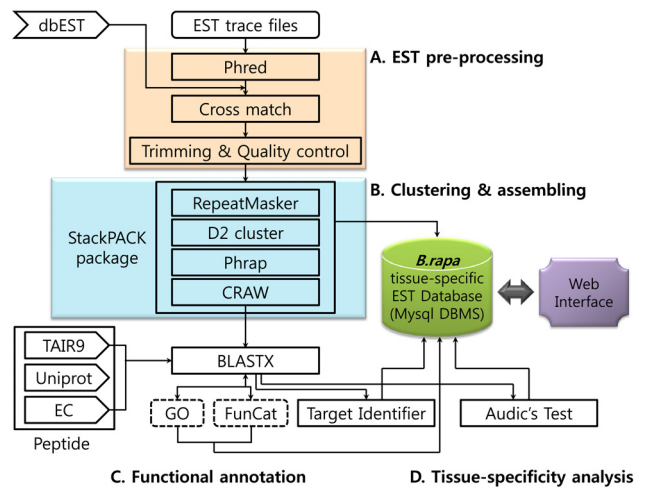


Fig. 1. Schematic diagram of work flow to construct BrTED. A. Sequences were extracted from dbEST and were subjected to quality control screening (vector, *E. coli*, polyA, T, or CT removal, minimum length = 100 bp, < 3% N). B. Cleaned EST sequences were assembled using the StackPACK package. Contig sequences were generated based on two or more ESTs that overlap for at least 40 bases with at least 98% sequence identity. These strict criteria help minimize the creation of chimeric contigs. These contigs were assigned a Tentative Consensus (TC) unigenes. ESTs that did not assemble into contigs were defined as singletons. C. Putative function of unigenes was annotated based on BLAST search, gene ontology mapping, and FunCat analysis against databases of plant reference genomes. D. Tissue-specificity of TCs was defined by Audic's test at p value 0.01.

Table 2. Number of unigenes clustered from EST collections. Unigenes were BLAST compared with data sets including *B. rapa* annotated gene model, TAIR, UniProt, and EC databases.

	Tandem Consensus	Singleton	Total
Unigene number	14,779	9,183	23,962
Gene model match ^z	13,970	7,617	21,587
TAIR match ^y	13,862	7,202	21,064
UniProt match ^y	13,845	7,162	21,007
EC match ^y	5,721	2,284	8,005

^zBLASTN search at E-value cutoff E⁻⁵⁰ and alignment coverage > 50%.

^yBLASTX search at E-value cutoff E⁻¹⁰ and alignment coverage > 50%.

미 알려진 유전자와 동일한 기능을 하는 것으로 조사되었으며 약 33%는 EC 데이터베이스에 따라 효소 단위 단백질 번호를 부여할 수 있었다(Table 2). 유전자 기능이 예측되지 않은 약 12%의 단일 조합 유전자서열은 배추 특이적인 유전자일 것으로 추정되며 이들에 대한 연구가 더 진행되어야 할 것으로 판단된다.

배추 단일 조합 유전자서열에 대한 조직 특이성 분석 결과 공통서열 단일 조합 유전자서열 중 2,289개의 유전자가 조직 특이적인 것으로 조사되었다($p < 0.01$). 특히 꽃과 뿌리, 장각과(silique) 그리고 저온 스트레스 처리를 한 조직에

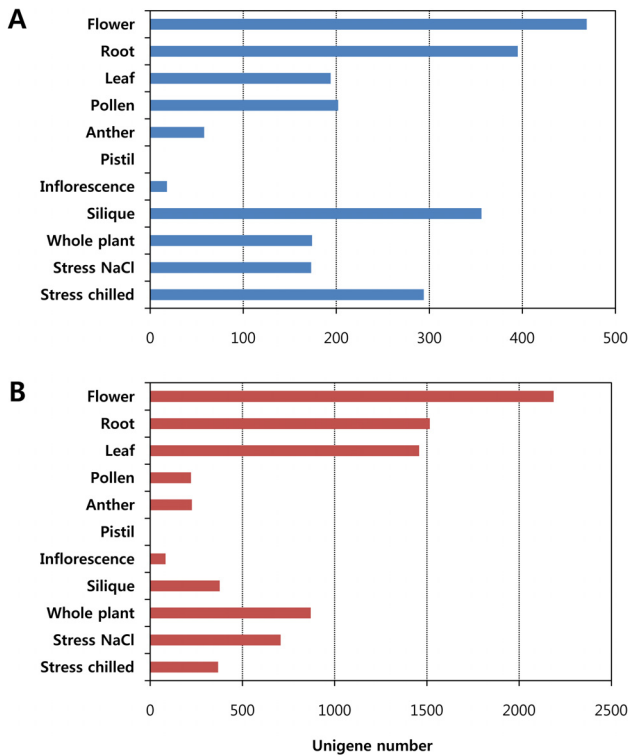


Fig. 2. Distribution of tissue-specific unigenes per plant tissue or experimental treatment. A. Tandem consensus unigene ($p < 0.01$). B. Singleton.

서 조직 특이성 유전자가 많이 발견되었으나 암술과 화서는 조직 특이성 유전자가 거의 발견되지 않았다(Fig. 2A). 한편 단독서열 9,183개의 경우도 공통서열과 비슷한 양상의 조직별 유전자 빈도를 나타냈다. 단독서열은 발현유전자 조립 과정에서 다른 서열들과 중복되지 않는 독특한 서열이기 때문에 특정 조직 특이적인 발현유전자로 여겨진다. 그러나 발현유전자가 추가 분석될 경우 공통서열로 조립한 후 조직 특이성 통계분석을 수행해야 정확한 조직 특이성을 예측할 수 있을 것으로 판단된다(Fig. 2B).

배추 조직 특이적 발현유전자 데이터베이스

배추 조직 특이적 발현유전자 데이터베이스는 메인 웹 페이지 5개와 서브 웹 페이지 4개로 구성하였다. 메인 페이지는 Introduction, Statistics, Tissue Specificity, Download, Contact 페이지로 이루어져 있다(Fig. 3). Introduction 페이지는 데이터베이스의 목적과 제공되는 배추 단일 조합 유전자서열 데이터가 어떻게 제작되었는지 설명한다. Statistics 페이지는 단일 조합 유전자서열의 조직 특이성 분석에 대한 통계 분석 결과를 보여준다. Tissue Specificity 페이지는 공통서열 단일 조합 유전자서열과 단독서열 세트 각각에 대하여 조직 특이적 유전자를 조직별로 검색할 수 있게 하였다. 이때 공통서열 단일 조합 유전자서열의 경우 조직 별로 p -value 기

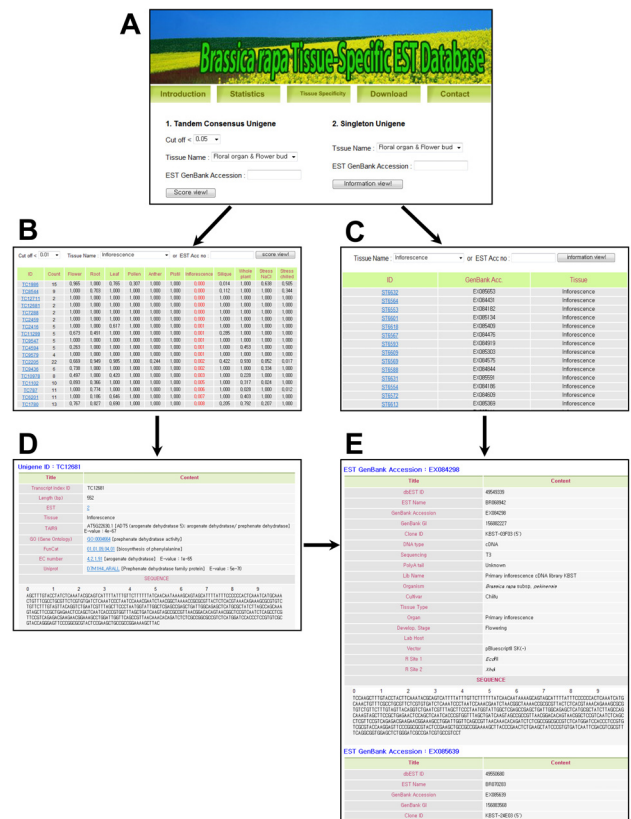


Fig. 3. Web page work flow of BrTED (<http://BrTED.ma.kr>). A. Tissue-specific unigene search main page. B. Tandem consensus (TC) unigene search result page. C. Singleton unigene search result page. D. TC unigene information page. E. EST unigene information page. Tissue-specific unigenes of each tissue can be retrieved from the database based on p -value cutoff for TC unigenes or library information of singleton unigene (A). The search result of TC (B) or singleton unigenes (C) is presented on independent page. Information of each unigene candidate is provided to user by clicking the gene ID (D) or EST information page (E) displays detailed information of unigene including sequence and functional annotation results.

준 값(0.05, 0.01, 0.005)에 따라 사용자가 조직 특이적 단일 조합 유전자서열을 검색할 수 있게 했고, 단독서열의 경우 각 서열의 조직 정보를 이용하거나 GenBank accession으로 직접 검색할 수 있게 하였다. 검색 결과는 서브 웹 페이지로 이동하여 각 조직 별 특이적 단일 조합 유전자서열에 대한 유전자 주석과 기능 예측 정보, 클러스터링된 EST들의 정보를 확인할 수 있게 하였다. 또한 GO와 FunCat, EC number, UniProt 데이터베이스에 연결되게 하여 보다 자세한 정보를 열람할 수 있게 구성하였다. Download 페이지는 데이터베이스를 통해 분석한 모든 데이터를 내려 받을 수 있게 하였다. Contact 페이지는 사용자들이 관련 사항에 대하여 문의할 수 있는 메일 서비스를 제공하게 하였다. 상기한 모든 정보를 이용할 수 있는 데이터베이스는 웹사이트 <http://BrTED.ma.kr> 를 통해 이용할 수 있다.

데이터베이스를 이용한 발현유전자의 선발과 RT-PCR 분석

배추의 뿌리, 잎, 그리고 꽃에서 특이적으로 발현되는 공통 서열 단일 조합 유전자서열 29개를 $p < 0.01$ 수준에서 선발하였고 이들 유전자들의 조직 특이성 발현 양상을 RT-PCR을 이용하여 조사하였다(Table 3 and Fig. 4). 뿌리 특이적 발현이 예상되는 10개의 유전자들 가운데 8개는 모두 뿌리 특이적인 발현을 나타냈다. 또한 2개 유전자(TC775와 TC1114)는 뿌리에서 가장 강한 발현을 보였으며, 잎에서도 약하게 발현되는 것을 확인할 수 있었다(Fig. 4A). 잎 특이적 발현이 예상된 9개의 유전자들은 모두 잎에서 가장 높은 발현을 보였으나 4개 유전자(TC900, TC2185, TC2691, 그리고 TC1003)가 꽃에서 약하게 발현되었다(Fig. 4B). 이 유전자들은 아마도 꽃의 꽃받침 잎에서 발현되거나 서열이 유사한

paralog들이 목표 조직 이외의 조직에서 발현되는 것으로 추정된다. 이에 반해 꽃에서 선발된 10개 유전자들의 경우 잎과 뿌리에서 선발된 유전자들과 달리 모두 꽃에서만 특이적으로 발현되는 것을 알 수 있었다(Fig. 4C). 이상의 결과는 BrTED를 이용할 경우 특정 조직에서 발현되는 유전자를 성공적으로 선발할 수 있음을 나타낸다.

본 연구를 통해 개발된 BrTED는 공공 데이터베이스로서 웹을 통해 연구자들에게 제공된다. 따라서 연구자들이 자유롭게 데이터베이스의 정보를 이용함으로써 배추 유전자를 손쉽게 선발하여 배추 기능 유전체 연구를 촉진할 수 있고, 나아가 배추를 포함한 근연 배추속 작물의 유전학, 유전체학 및 육종 연구의 효율을 더욱 높일 수 있을 것으로 기대한다.

Table 3. List of genes and primers used in RT-PCR analyses.

Tissue	Unigene	<i>A. thaliana</i> ortholog	Gene description	Forward primer	Reverse primer	Amplicon (bp)	
Root	TC2813	AT2G35980	YLS9 (yellow-leaf-specific gene 9)	acctcaaacatgggctgaac	agataagagcggctatgcca	144	
	TC6858	AT5G01320	pyruvate decarboxylase, putative	cttactggggagcagtgagc	ccattggcaatagtgacacg	131	
	TC147	AT4G12550	AIR1 (auxin-Induced in root cultures 1)	ttgcgctaattgtgctcaaac	tcacgcattcgtaaaccgta	243	
	TC5909	AT4G36430	peroxidase	aggacttgatgtcaccgacc	cactgacgatgtccaacacc	169	
	TC7639	AT5G44550	integral membrane family protein	cctctccatcgctattctcg	gtggggagatggagcattt	220	
	TC9261	AT3G14210	ESM1 (epithiospecifier Modifier 1)	gctagcaccattgggttgtt	caacatgatgagttcggcac	196	
	TC775	AT4G08950	phosphate-responsive protein	ccgcaagttagcctcaaaaag	ccaggataagctctttccc	377	
	TC1114	AT1G27730	STZ (salt tolerance zinc finger)	actaatggcgttgaggcta	gagcagacgtgagactccc	395	
	TC4154	AT2G38380	peroxidase 22	aagtttgtgctcggctggtc	tggttgctgctgattgctc	317	
	TC6922	AT4G19530	TIR-NBS-LRR disease resistance protein	gtgcttgatgagcttgga	tccgagcatgagaggtcttt	506	
	Leaf	TC1182	AT1G06190	ATPase	gaaagggctttttctacggg	ttctgtgcttccgatatg	384
		TC5240	AT1G69490	NAP transcription factor	aaaccagacatgtccaagc	cgagagcttcttgacacc	389
		TC900	AT2G15620	NIR1 (nitrite reductase)	gaaactcgaaaacgcaagc	atatccgccacttggacttg	352
		TC2694	AT1G08540	SIGB (sigma factor B) transcription factor	gttcgatgcaacaaggggt	atgttcttccacatgtga	374
		TC2185	AT1G58180	carbonic anhydrase family protein	tgttctaacttggcggctt	acatggttcaggaaatggga	439
		TC2691	AT4G34200	EDA9 (embryo sac development arrest 9)	ctctggatgctgggtattgt	tgaagtgcagcgttgaccaag	416
TC6710		AT3G45300	Isovaleryl-CoA-dehydrogenase	gaggaaataagtcgggcctc	cagtactgaatccggccatt	327	
TC689		AT5G60800	heavy-metal-associated domain-containing protein	aaaggtgtaatgggctgac	tcacgttctcatcgctgaac	313	
TC1003		AT1G69930	glutathione transferase	ggctacccccaaagaaagag	tttctgcttcttgaacg	495	
Flower		TC8229	AT1G75910	EXL4 (extracellular lipase 4)	ggtttctgtgtgcaaggt	tccgagggatgaacaaagtc	143
	TC3842	AT3G11980	MS2 (male sterility 2)	tgaagcttttcgactccgtt	atagtggaaggtgtttggc	126	
	TC10360	AT3G54340	AP3 (APETALA 3) transcription factor	tttgagcgcacactatgag	ttcgatctgattccaaggg	168	
	TC8218	AT5G13380	auxin-responsive GH3 family protein	ctgggtgatcgtcagggtt	tgttcgagcataacgcagtc	217	
	TC8188	AT1G74540	CYP98A8 (cytochrome P450)	acaaggtccatggtcgagtc	ctcgggtctgaactcttcc	112	
	TC10004	AT5G33370	GDSL-motif lipase	gtggattggtcggagaagaa	ccaccttcgacgttacgaat	417	
	TC10090	AT3G14040	exopolygalacturonase	tggcgtgaacttaatagggg	cgttgtggaccaagttgatg	448	
	TC851	AT1G20132	hydrolase	gacaagagaagggcgatcag	ttccactagtccggtccac	427	
	TC9943	AT3G55290	reductase (SDR) family protein	tcgaccctctcgaatctctt	acaccaccttggacaagc	354	
	TC10018	AT5G47330	palmitoyl protein thioesterase family protein	ttgttgaagatctcagggg	gaagaggtccgaaatcacca	398	

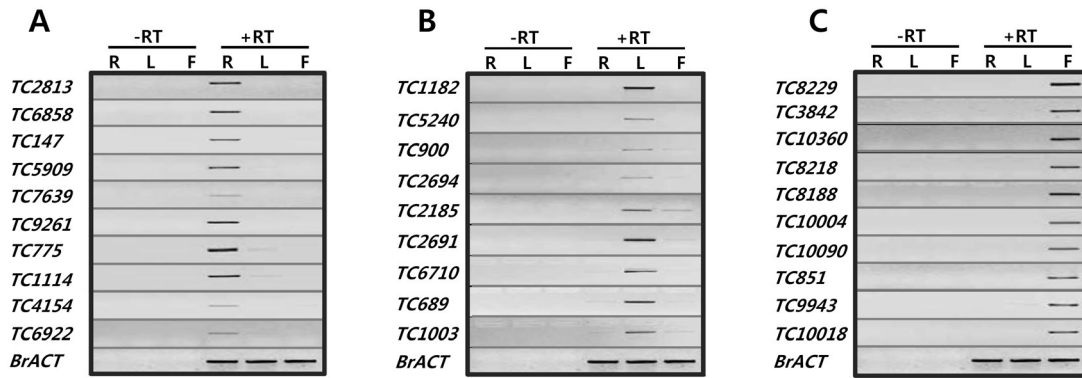


Fig. 4. Expression of tissue-specific unigene candidates in different types of tissues. R, root; L, mature leaf; F, flower. Unigene information, PCR primers, and the amplicon sizes of the investigated genes are summarized in Table 3. A. Root-specific unigene candidates. B. Leaf-specific unigene candidates. C. Flower-specific unigene candidates. An internal control *B. rapa* actin (*BrACT*) is approximately 250 bp in amplicon size. -RT, negative control of total RNA without reverse transcription reaction; +RT, cDNA after reverse transcription reaction.

초 록

배추는 배추속 식물의 A genome을 대표하는 모델로서 다양한 배추과 작물의 유전학 및 유전체학과 육종연구의 기반이 되는 중요한 작물이다. 최근 들어 배추 유전체 해독이 완료됨에 따라 유전체의 기능 연구가 보다 활발히 진행될 것으로 기대된다. 유전체 정보로부터 유전자의 구조를 예측하고, 기능을 분석하여 프로모터를 포함한 유용 유전자를 개발하기 위한 필수 재료로 이용되는 것이 다양한 조직 또는 실험 처리로부터 생성된 발현 유전자 데이터이다. 2011년 7월 현재 공공 데이터베이스에는 39개의 cDNA library로부터 분석된 147,217개의 배추 발현유전자가 보고되어 있다. 그러나 이들 발현 유전자들은 체계적으로 분석되거나 데이터베이스 형태로 정리되어 있지 않기 때문에 연구자들이 유전자 서열로부터 유용한 정보를 추출하여 사용하기 어려운 문제점이 있다. 따라서 해독 완료된 배추 유전체와 함께 발현 유전자 정보를 보다 잘 활용하기 위하여 배추의 조직 특이적 발현 유전자 데이터베이스인 BrTED를 개발하였다. 데이터베이스는 EST 서열 처리-정보 검색 단위와 조직 특이성 발현 특성 분석 단위로 이루어져 있으며, 각 정보들은 상호 연결되어 유기적인 검색 환경을 제공하게 하였다. BrTED는 23,962개의 단일 조합 유전자서열을 포함하고 있으며, 각 서열들의 유전자 주석과 암호화하고 있는 단백질의 기능을 동시에 제공한다. 또한 각 단일 조합 유전자서열들의 조직별 발현 특이성을 통계 분석을 통해 조사하여 연구자의 검색 기준에 따라 제공한다. BrTED의 실효성을 검증하기 위하여 데이터베이스를 통해 조직 특이적 발현 유전자 29개를 선발하고, 이들의 발현 특성을 RT-PCR로 확인한 결과, 선발한 유전자 모두 목표한 조직에서 특이적이거나

강한 발현을 보였다. BrTED는 조직 특이적 발현유전자를 신속하게 선발할 수 있는 공공 데이터베이스로서 배추의 기능 유전체 연구뿐만 아니라 근연 배추속 작물의 유전체학 유전체학 연구에 유용한 공공 연구 자료로 이용될 수 있을 것이다.

추가 주요어 : 작물, 유전자 발현, 유전자 기능, 공공 데이터베이스, 단일 조합 유전자서열

인용문헌

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Audic, S. and J.M. Claverie. 1997. The significance of digital gene expression profiles. *Genome Res.* 7:986-995.
- Carbon, S., A. Ireland, C.J. Mungall, S. Shu, B. Marshall, S. Lewis, AmiGO Hub, and Web Presence Working Group. 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics* 25:288-289.
- Chung, H., J.H. Mun, S.C. Lee, and H.J. Yu. 2011. Pickprimer: A graphic user interface program for primer design on the gene target region. *Kor. J. Hort. Sci. Technol.* 29:461-466.
- George, R.A. 2001. StackPACK clustering system. *Brief. Bioinformatics* 2:388-404.
- Mun, J.H., S.J. Kwon, T.J. Yang, Y.J. Seol, M. Jin, J.A. Kim, M.H. Lim, J.S. Kim, S. Baek, B.S. Choi, H.J. Yu, D.S. Kim, N. Kim, K.B. Lim, S.I. Lee, J.H. Hahn, Y.P. Lim, I. Bancroft, and B.S. Park. 2009. Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* 10:R111.
- Mun, J.H., S.J. Kwon, and B.S. Park. 2010. The strategy and current status of *Brassica rapa* genome project. *J. Plant Biotechnol.* 37:153-165.

- Ruepp, A., A. Zollner, D. Maier, K. Albermann, J. Hani, M. Mokejs, I. Tetko, U. Güldener, G. Mannhaupt, M. Münsterkötter, and H.W. Mewes. 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.* 32:5539-5545.
- Scheer, M., A. Grote, A. Chang, I. Schomburg, C. Munaretto, M. Rother, C. Söhngen, M. Stelzer, J. Thiele, and D. Schomburg. 2011. BRENDA, the enzyme information system in 2011. *Nucleic Acid Res.* 39:670-676l.
- The *Brassica rapa* Genome Sequencing Project Consortium. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43:1035-1039.
- UniProt Consortium. 2009. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* 37:D169-174.