

## 벼 미숙종자의 발현유전자 구조특성분석

윤웅한<sup>1</sup> · 이강섭<sup>1</sup> · 이정숙<sup>1</sup> · 한장호<sup>1</sup> · 김창국<sup>1</sup> · Shoshi Kikuchi<sup>2</sup> · Kouji Satoh<sup>2</sup> · 김진아<sup>1</sup> · 이정화<sup>1</sup> · 이태호<sup>3</sup> · 김용환<sup>1</sup>  
<sup>1</sup>국립농업과학원, <sup>2</sup>일본농업생물자원연구소, <sup>3</sup>명지대학교

### Structural analysis of expressed sequence tags in immature seed of *Oryza sativa* L.

Ung-Han Yoon<sup>1</sup> · Gang-Seob Lee<sup>1</sup> · Jung-Sook Lee<sup>1</sup> · Jang-Ho Hahn<sup>1</sup> · Chang-Kug Kim<sup>1</sup> · Shoshi Kikuchi<sup>2</sup> ·  
Kouji Satoh<sup>2</sup> · Jin-A Kim<sup>1</sup>, Jeong-Hwa Lee<sup>1</sup> · Tae-Ho Lee<sup>3</sup> · Yong-Hwan Kim<sup>1\*</sup>

<sup>1</sup>Genomics Division, National Academy of Agricultural Science, RDA, Suwon, Korea

<sup>2</sup>Genome and Biodiversity Research Division, National Institute of Agrobiological Science, Tsukuba, Japan

<sup>3</sup>Bioscience and Bioinformatics Division, MyongJi University, Yongin, Korea

**Abstract** Rice (*Oryza sativa*) is the most important staple crop in Korea. With its small genome size of 389Mb, rice is a model plant for genome research. We analyzed expressed sequence tag (EST) clones from immature seeds of rice (cv. *Ilpum*) at 20 days after heading. The 25,668 EST clones were clustered by using SeqMan program and 7,509 clones were selected as unique clones. We compared the 7,509 unique genes with KOME database including the 32,127 FL-cDNA in rice. Finally, 4,990 clones were homologous and 2,519 clones non-homologous to FL-cDNA clones. In addition, we mapped the 7,509 cDNA clones by using TIGR rice pseudomolecule version 5. Ultimately, 7,347 clones were matched to be significant clones related to the TIGR rice pseudomolecules, but 162 clones were unmapped. For the clustering of orthologous group genes, we further analyzed the 7,509 EST clones from immature seeds using NCBI clusters of orthologous groups database. Among the clones, 4,968 clones were categorized into information storage and processing, cellular processes and signaling, metabolism and poorly characterized genes, proportioning 799 (14.89%), 1,536 (28.3%), 1,148 (21.2%) and 1,936 (35.7%) clones to the previous four categories, respectively.

## 서론

벼는 세계인의 주곡이며 게놈의 크기가 작아 유전자기능을 연구하기위한 모델 작물로 이용되고 있다. 국제 벼 염색체 완전해독 연구 (IRGSP 2005)가 2004년에 완료됨과 동시에 국제적으로 벼 유용유전자 기능분석과 활용기반 기술 확보를 위하여 염색체 구조해석 및 비교유전체연구 (Ammiraju et al. 2008; Paterson et al. 2008)와 대량 유전자 기능분석연구 (Jung et al. 2008; Nakamura et al. 2007)가 경쟁적으로 이루어지고 있다. 벼의 주요 농업형질관련 유전자의 기능을 해석하기 위하여 발현유전자 분리 및 확보가 필수적이므로 다양한 품종에서 대량발현 유전자 염기서열분석 및 완전장 유전자 분리 연구가 수행되고 있다. 벼의 대량발현유전자 염기서

열분석 및 완전장 cDNA 분리에 관한 연구를 일본 농업생물자원연구소의 Kikuchi 박사그룹 (Kikuchi et al. 2007)에서 수행하였다. 완전장 cDNA 분리는 벼 염색체 염기서열 완전해독 연구에 사용된 *japonica* 품종인 *Nipponbare*를 이용하여 조직부위별 및 발현시기별로 완전장 cDNA를 합성하고 581,446개 cDNA를 분석하여 32,127개의 완전장 cDNA를 분리하였으며 KOME DB (<http://cdna01.dna.affrc.go.jp/cDNA/>)를 통하여 연구자들에게 발현유전자 정보를 공개하고 있다. 일본은 이러한 완전장 cDNA 정보를 이용하여 벼 염색체의 유전자구조 해석과 유전자 발현조절부위인 *cis-element* 탐색 (Doi et al. 2008) 및 전사체의 발현유전자 annotation 분석 (The Rice annotation Project 2008) 등에 적극 활용하고 있으며 RAP-DB (<http://rapdb.lab.nig.ac.jp/>)를 통해 유전자 해석정보를 공개하고 있다. 아울러 대량 기능분석 연구를 위한 벼 및 애기장대의 과발현체 작성 등의 연구 재료로 사용하고 있다. 미국의 게놈연구소에서는 벼 염색체 염기서열분석 정보와 발현유전자 정보를 기반으로 2006

\*Corresponding author Tel 031-299-1651 Fax 031-299-1657

E-mail: yhkim@korea.kr

년 TIGR Pseudomolecules ver. 4를 작성하고 벼 유전자 명명을 행하여 24,799개의 유전자를 분석 (31,737개 유전자 모델)하여 정보를 웹사이트 (<http://rice.plantbiology.msu.edu>)에 공개하고 있으며 (Ouyang et al. 2007) 이후 TIGR Pseudomolecules ver. 5에서는 28,706개의 유전자분석 (33,882개 유전자 model) 정보를 공개하고 있다. 중국의 Bin Han 박사그룹 (Bin Han et al. 2008)은 *indica* 품종인 *Guangluai 4*와 *Minghui 63*을 이용하여 40,000개의 발현유전자를 분석하였으며 20,000여개의 완전장 cDNA를 분리하여 RICD DB를 통해 정보를 공개하고 있다 (<http://www.ncgr.ac/ricd>). 최근까지 GenBank에 1,248,660개의 벼 발현유전자가 등록되어있으며 unigene으로는 40,739개의 유전자를 포함하고 있다. 이러한 벼 발현유전자 정보와 유전자들을 이용하여 세계각국은 농업상 중요한 형질 유전자의 대량기능분석 (Nakamura et al. 2007) 및 유전자분리에 적극 활용하고 있다. 특히 벼의 종자는 인간에게는 영양원으로서 벼 자체로서는 생장을 위한 저장원으로 사용되어지므로 종자 형질관련 유전자 발현분석과 유전자분리 (Shomura et al. 2008; Song et al. 2007)에 많은 연구가 이루어지고 있다. 그러나 실제 벼 종자 발현 유전자 분석 및 정보 종합화 연구는 미흡한 실정이다.

본 연구는 미숙종자 발달단계의 cDNA 합성과 미숙종자내 발현 유전자의 대량 염기서열 분석을 통하여 종자 내 발현유전자의 분석과 발현유전자 정보종합화를 목적으로 수행 하였다. 이러한 발현정보의 종합화는 벼 및 식물전반의 기능유전체를 연구하는 연구자들에게 보다 유익한 정보를 제공할 수 있을 것으로 기대한다.

## 재료 및 방법

### 벼 미숙종자 유전자은행 제작 및 구조분석

본 연구의 공시품종은 일품벼이었고, 출수 후 20일 경과한 미숙 종자를 재료로 사용하였다. 출수 후 20일이 지난 미숙종자를 채취하여 액체질소에 넣어 냉동시킨 시료를 이용하여 RNA를 분리하였다. Total RNA 분리는 Trizol 용액(Invitrogen Life Technologies)를 이용하여 분리한 후 PolyA Tract mRNA 분리 시스템 (Promega)으로 Poly(A) mRNA를 분리하였다. cDNA는 ZAP-cDNA Gigapack III Gold 클로닝 키트 (Clontech)를 이용하여 합성하였으며, 합성된 cDNA는 Uni-ZAP XR 벡터에 클로닝 하였다. 합성된 벼 미숙종자 cDNA 은행의 클론을 집단으로 in vivo excision하여 개별 클론을 -70°C에 보관하고 96well 밀리포어 멀티스크린 (Millipore MANANLY50, MAFBNOB50)을 이용하여 대량으로 플라스미드 DNA를 분리하였다.

염기서열은 각 클론의 5'말단을 BigDye terminator ver 3.1 (Applied Biosystems)과 SK 프라이머 (5'-cgctctagaactagtgatc-3')를 이용하여 염기서열분석 반응을 수행한 후 ABI3100 염기서열분석기 (Applied Biosystems)를 사용하여 결정하였다.

### Unigene 선별 및 *Nipponbare* 완전장 cDNA와 비교분석

각 클론의 발현유전자 염기서열을 결정한 후 SeqMan 프로그램 (DNASTAR)을 사용하여 벡터 염기서열부분과 불명확한 말단 염기서열 및 cDNA가 삽입되지 않은 클론을 제거하고 염기서열 정보분석을 실시하였다. 미숙종자에서 발현하는 unigene은 SeqMan 프로그램을 이용하여 100bp씩 95%의 상동성으로 분석을 하여 얻어진 singleton 유전자와 cluster된 contig 유전자로부터 선별하였다.

염기서열의 상동성은 국립농업과학원 농업생명공학정보센터 (<http://www.niab.go.kr>)에 설치된 webblast 및 NCBI의 BLAST (<http://www.ncbi.nlm.nih.gov>)을 이용하여 분석하였다.

일품벼 유래 7,509개 발현유전자와 일본 농업생물자원연구소에서 분리한 32,127개의 *Nipponbare* 완전장 cDNA와의 비교분석 연구는 BLASTN 분석을 통해 공통 발현유전자를 선별하고 아울러 공통 발현유전자에 속하지 않는 일품 벼 유래의 발현유전자들을 분리하였다.

### 벼 염색체 mapping

미숙종자에서 발현하는 7,509개 unigene의 각 염색체별 위치는 벼 12개 염색체 염기서열정보를 이용하여 결정하였다. 벼 염색체 염기서열정보는 TIGR Pseudomolecule ver. 5 내 염색체 서열 정보를 기반으로 하였으며, BLASTN 분석을 통해 미숙종자 unigene 염기서열 길이 유사성에 대하여 95% 이상의 상동성을 나타내는 염색체 내 지역을 판별하여 해당 지역을 unigene 서열의 염색체 내 위치로 결정하였다.

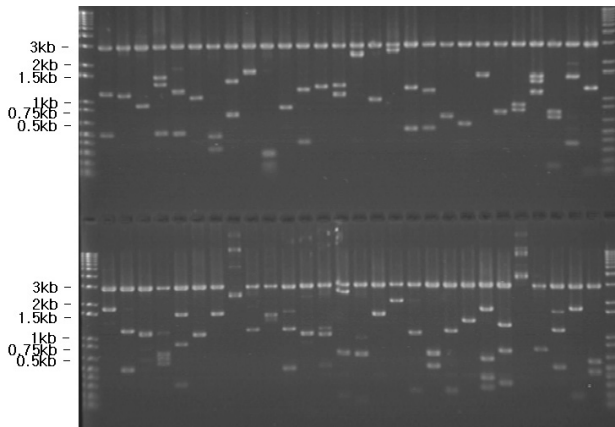
### NCBI COG DB 기반 유전자 기능분류

미숙종자로부터 얻은 7,509개의 unigene에 대한 발현유전자 기능분류는 NCBI에서 Arabidopsis 등 7개 주요 진핵생물 단백질 서열로부터 구축된 COG (Clusters of Orthologous Groups of proteins) DB인 KOG DB를 기반으로 하여 수행하였다. 먼저, BLASTX 분석을 통해 KOG DB 내 서열 중, unigene 서열 각각에 대해 가장 높은 유사성을 가지는 단백질 서열을 검색하였다. 이때 검색 조건으로 e-value가 1.0e-20 이하인 것들만을 추출하여 유사서열이 아닌 것이 나올 확률을 최소화하였다. 다음으로 이렇게 얻어진 KOG DB내 단백질 유사서열의 분류정보를 바탕으로 unigene 서열의 기능 분류를 수행하였다.

## 결과 및 고찰

### 벼 미숙종자 유전자은행 제작 및 구조분석

미숙종자 100ng의 cDNA를 1 $\mu$ g의 Uni-ZAP 벡터에 삽입한 후 Gigapack III Gold Packaging Extract에 Packaging하여 cDNA 은행을 제작한 결과 2.5X10<sup>6</sup> pfu/ml의 packaged phagemids를 얻었다. 벼의 유전자가 32,000개 정도임을 감안할 때 본 연구에서 제작된 미숙종



**Figure 1.** Insert DNA pattern of cDNA clones from immature seed of Ilpumbyeo. cDNA clones were digested with *EcoRI* and *XhoI*

자 유전자은행은 전체 유전자의 약 78배 양을 나타내어 미숙종자 전체 유전자를 포함하고 있다고 생각되어진다. 대량 발현유전자분석을 위하여  $1 \times 10^5$ 의 phagemids를 in vivo excision 시킨 후 33,792 콜로니를 개별 선발하여 균주는  $-70^\circ\text{C}$ 에 보관하고, 96well 멀티스 크린을 이용하여 DNA를 분리하여  $-20^\circ\text{C}$ 에 저장하였다. 각각 분리된 DNA를 *EcoRI*, *XhoI* 제한효소로 처리한 후 1.2% 아가로스 겔에 전기영동하여 삽입 단편을 확인 한 결과 0.75kb - 3.5kb의 단편을 포함하고 있는 것으로 나타났다(그림 1). 최근 각종 동식물에서 CAP trapper 방법 (Carnini et al. 1996; Kawai et al. 2001; Kikuchi et al. 2003; Seki et al. 1998)을 이용하여 1-10kb 길이의 완전장 cDNA를 제작하여 유전체 해독과 유전자 기능 연구를 위하여 많이 사용하고 있다. 그러나 이 방법을 이용하여 유전자은행을 제작하면 많은 기술과 비용이 소요된다. 따라서 본 연구에서는 일반적인 방법인 poly(A) RNA를 분리 한 후 oligo(dT) 프라이머와 역전사효소를 이용하여 cDNA를 합성하여 양질의 미숙종자 발현유전자는 얻을 수 있었다.

클론 각각의 염기서열 결정은 96well PCR plate를 이용하여 각각의 well에 200ng의 DNA, 5pmol의 SK 프라이머, Bigdye terminator sequence reaction 2ul, 5X buffer 1ul를 넣은 후 염기서열분석 반응을 시키고 ABI3100염기서열분석기를 이용하여 염기서열을 결정하였다. 33,792클론(96well×352 plate)의 염기서열을 결정한 후 서열을 교정하여 25,668 클론의 염기서열을 얻을 수 있었다. 부분 염기서열이 결정된 종자 형질 관련 유전자의 완전염기서열은 primer extension 방법을 이용하여 나머지 중간 부위의 염기서열을 결정하였고, 결정된 염기서열을 GenBank에 등록하였다.

**미숙종자 발현 unigene 선발 및 특성분석**

구조유전체연구가 수행됨과 동시에 동식물의 유전자 명명 및 기능을 밝히기 위한 기능유전체 연구 재료로 완전장 cDNA 분리와 발현유전자분석이 매우 중요시되고 있다. 벼의 발현유전자정보는

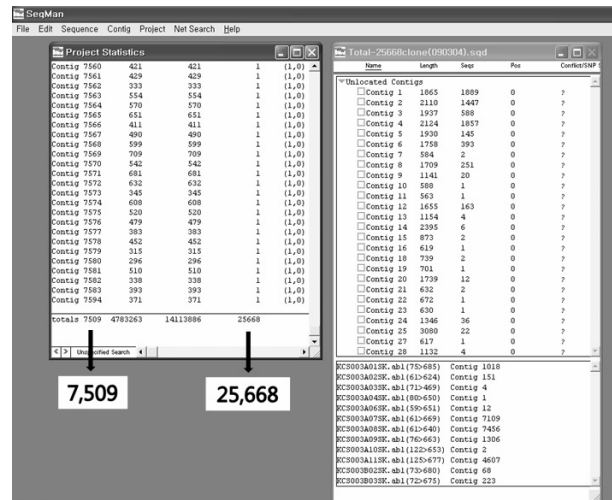
현재 NCBI dbEST (database of Expressed Sequence Tags)에 1,248,660개 (February 27, 2009) 수록되어 있으며 unigene으로는 40,739개를 나타내고 있다. 벼의 유전자는 32,000여개로 추정되므로 대다수의 유전자가 포함되어 있을 것으로 생각된다.

본 연구에서는 미숙종자 발현유전자 대량 염기서열 분석 및 기능분류를 위하여 미숙종자 발현 unigene을 분리하였다. 미숙종자에서 발현되는 25,668개 클론 중 중복클론을 제외한 unigene 분리를 위하여 DNASTAR의 SeqMan 프로그램을 이용하여 basecalling과 clustering을 실시한 결과 2개 이상의 클론이 중복된 contig 2,869개와 singleton 4,640개 등 미숙종자에서 발현하는 unigene 7,509개를 얻었다(표 1, 그림 2, 그림 3).

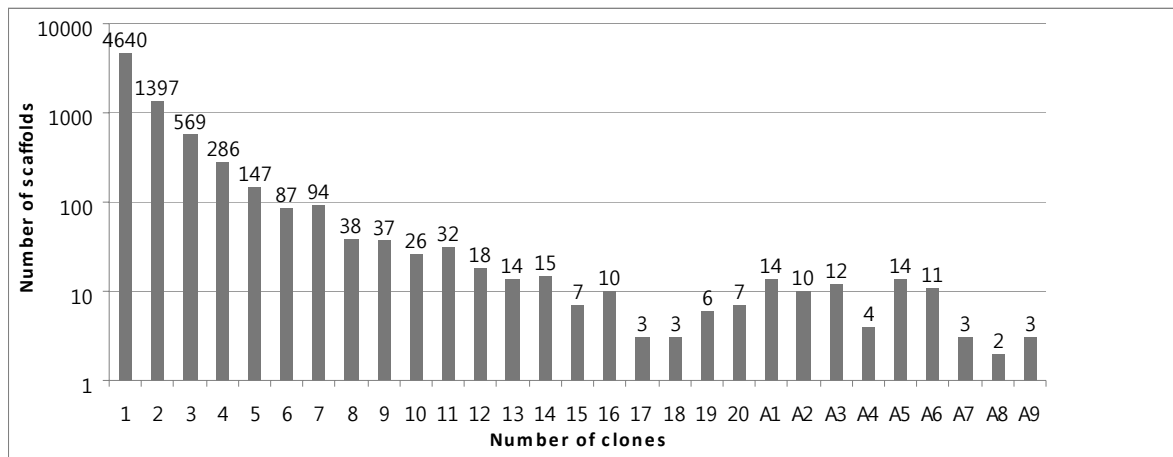
미숙종자 25,668개 클론으로부터 분석된 전체 염기서열 길이는 14,113,886bp였으며 1개 클론의 평균 길이는 550bp 정도로 확인되었다. 염기서열 분석의 성공률은 76%였으며 유전자 중복클론은 21,028개로 나타났다. 이와 같이 많은 클론들이 중복을 나타내는 이유는 미숙종자 형성시기에 조직 특이적으로 많이 발현하고 있는

**Table 1.** Summary of assembly and clustering of cDNA clones from immature seed of rice cv. Ilpumbyeo

Groups	Records
Number of initial clones	33,792
Number of sequenced clones	25,668
Total sequence length(bp)	14,113,886
Average length(bp)	550
Singletons	4,640
Clustering	2,869
A total Unigene	7,509



**Figure 2.** Identification of unigenes using SeqMan program. To estimate the level of redundancy, the Ilpumbyeo ESTs were clustered with each other using the SeqManII program (DNASTAR), which considers two sequences being originated from the same transcript when they have 95% nucleotide identity over a minimum of 100 bp. Poor quality sequence (<100 bp) and vector sequence were eliminated



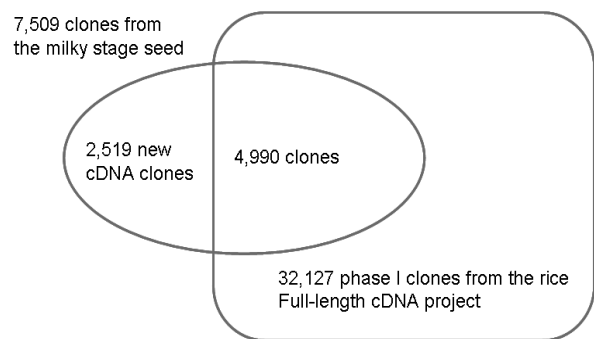
**Figure 3.** Cluster distribution of cDNA clones from immature seed of rice  
 A1 : 21-25, A2 : 26-30, A3 : 31-40, A4 : 41-50, A5 : 51-100, A6 : 101-200, A7 : 201-500, A8 : 501-1000, A9 : 1001-2000  
 A total of 25,668 EST clones were clustered using SeqMan program. Among them, 7,509 clones were identified as unique clones

종자 저장단백질인 glutelin과 prolamin 유전자와 전분 합성 관련 유전자의 발현이 매우 높았기 때문에 추정된다. Cluster 내 중복된 클론 수를 도표(그림 3)로 정리한 결과, 그중 1,000개 이상의 클론으로 구성된 cluster는 3개 있었으며 glutelin 유전자로 밝혀진 발현 유전자들은 구조를 분석한 후 GenBank에 등록 하였다(Accession Number EF122456, EF122460). 염기서열을 결정한 벼 미숙종자 발현유전자중 glutelin 유전자는 전체의 약 30%를 차지하고 있었다. 이러한 결과를 미루어 생각할 때 벼 유래 미숙종자 발현 유전자는 행을 제작할 경우에는 유전자의 중복을 피하기 위하여 glutelin 유전자를 최대한 제외시킨 유전자는 행 작성이 필요하다.

Umezawa 등 (2008)은 콩 유래 10종의 cDNA 은행으로부터 37,834개 클론을 분리하고 발현유전자분석을 수행하여 68,661개의 염기서열로부터 22,674개 cluster와 15,255개의 singleton을 분리 한 바 있으며, 이때 60개 이상의 클론을 포함하는 cluster는 10개미만으로 보고 하였는데, 본 연구의 결과와는 다소 차이가 있었다. 이러한 경향은 앞서 설명한 바와 같이 벼 미숙종자 유래 발현유전자에는 종자저장단백질 및 전분합성 유전자가 다량으로 포함되어 있기 때문에 추정된다.

**미숙종자 unigene과 Nipponbare 완전장 cDNA의 비교분석**

종자특이 발현유전자 분석은 일품벼 미숙종자 발현유전자 7,509개의 unigene분석 결과를 일본 농업생물자원연구소의 Kikuchi 그룹에서 분리한 Nipponbare의 32,127개 완전장 cDNA (FL-cDNA)와 비교하여 실시하였다. 일본의 Nipponbare 완전장 cDNA는 벼 여러 조직에서 제작한 발현유전자는행으로부터 각각의 클론을 분리 한 후 5'와 3'측의 염기서열을 결정하여 얻어진 581,446개의 결과로부터 최종 32,127개의 FL-cDNA를 확보하였다. 이들 FL-cDNA와 미숙종자 7,509개의 unigene을 이용하여 비교 분석을 한 결과 미숙종자

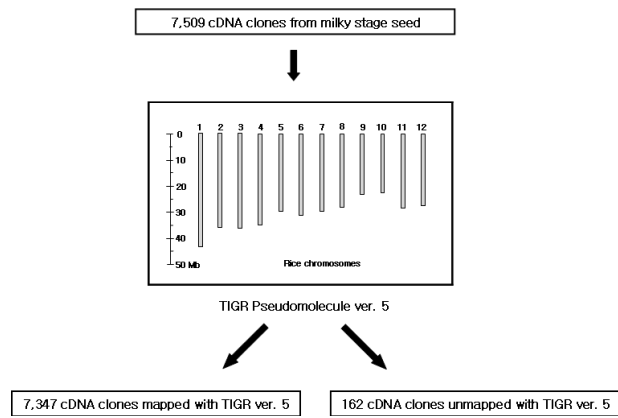


**Figure 4.** Comparison of 7,509 unigenes from immature seeds of rice cultivar Ilpumbyeo with the 32,127 FL-cDNA clones from rice cultivar Nipponbare

4,990개 클론은 일본에서 제작된 FL-cDNA에 포함되어 있었으며, 2,519개 클론은 FL-cDNA에는 포함되지 않는 유전자로 확인되었다(그림 4). 이러한 결과로 볼 때, 벼 미숙종자 발현유전자 은행에는 종자에서만 특이적으로 발현 하는 유전자가 다수 존재함을 확인할 수 있었다. 이러한 결과를 바탕으로 기존에 분리되지 않은 종자 고유 유전자를 이용한 과발현체 제작 등의 기능분석연구에 이용함으로써 고유 유전자 개발과 응용이 가능하리라 생각된다. 아울러 이러한 종자 특이 발현유전자 정보는 향후 벼 유전자 명명작업등에 적극 활용 가능하리라 생각된다.

**미숙종자 발현유전자의 벼 염색체 mapping**

TIGR 연구소에서는 3,450개의 벼 BAC/PAC의 염기서열 정보를 이용하여 370,637,721bp의 염색체 구조해석을 행한 후 TIGR Pseudomolecule ver. 5에서 전사인자를 포함하는 56,304개의 유전자의 유전자좌를 결정한 바 있다. 본 실험에서 선발된 벼 미숙종자



**Figure 5.** Schematic presentation of chromosome mapping of cDNA clones from immature seed with the TIGR rice pseudomolecule ver. 5

**Table 2.** Mapping of the 7,509 cDNA clones from immature seed of rice cv. Ilpumbyeo in the TIGR rice pseudomolecules ver. 5

Chromosome No.	TIGR gene loci*	Japonica FL-cDNA**	cDNA mapping
Chr. 1	6,662	4,026	1,090
Chr. 2	5,416	3,196	850
Chr. 3	5,604	3,569	983
Chr. 4	5,372	2,531	612
Chr. 5	4,645	2,313	649
Chr. 6	4,741	2,292	589
Chr. 7	4,507	2,183	587
Chr. 8	4,204	1,933	482
Chr. 9	3,436	1,605	370
Chr. 10	3,477	1,538	371
Chr. 11	4,224	1,685	383
Chr. 12	4,056	1,693	381
Unmapped			162
Total	56,304	28,564	7,509

\* The TIGR ver. 5 gene loci, excluding small gene models (<50 amino acids) (<http://rice.plantbiology.msu.edu/>)

\*\* Characterization of Japonica 32,775 FL-cDNA clones (Sato K and Kikuchi S, 2007)

발현유전자 7,509개의 unigene 정보를 TIGR Pseudomolecule ver. 5와 비교분석하여 미숙종자 발현유전자들의 벼 염색체내 위치를 결정하였다(그림 5).

TIGR Pseudomolecule ver. 5의 염색체 염기서열 정보를 기반으로 BLASTN 분석을 행하여 95%이상 상동성을 나타내는 미숙종자 유전자를 가지고 염색체별 위치를 결정한 결과 7,347개가 해당 염색체별로 mapping 되었으며 162개 클론은 mapping되지 않았다(표 2). 각 염색체별로 mapping되어진 유전자들의 수는 TIGR 유전자좌 및 Japonica FL-cDNA 수 (Sato et al. 2007)와 비슷한 경향으로

**Table 3.** Functional classification of unigenes by the COG database

Total Sequence Number:	25,668(7,509)
The Number of Sequence, Matched against COGs DB:	4,968
<b>INFORMATION STORAGE AND PROCESSING 799(14.8%)</b>	
J Translation, ribosomal structure and biogenesis	256
A RNA processing and modification	175
K Transcription	211
L Replication, recombination and repair	93
B Chromatin structure and dynamics	64
<b>CELLULAR PROCESSES AND SIGNALING 1,536(28.3%)</b>	
D Cell cycle control, cell division, chromosome partitioning	99
Y Nuclear structure	22
V Defense mechanisms	28
T Signal transduction mechanisms	384
M Cell wall/membrane/envelope biogenesis	62
N Cell motility	1
Z Cytoskeleton	100
W Extracellular structures	6
U Intracellular trafficking, secretion, and vesicular transport	295
O Posttranslational modification, protein turnover, chaperones	539
<b>METABOLISM 1,148(21.2%)</b>	
C Energy production and conversion	189
G Carbohydrate transport and metabolism	257
E Amino acid transport and metabolism	210
F Nucleotide transport and metabolism	59
H Coenzyme transport and metabolism	51
I Lipid transport and metabolism	177
P Inorganic ion transport and metabolism	82
Q Secondary metabolites biosynthesis, transport and catabolism	123
<b>POORLY CHARACTERIZED 1,936(35.7%)</b>	
R General function prediction only	564
S Function unknown	1372

mapping 되었으며 종자 특이적으로 mapping되어진 염색체는 나타나지 않았다. 이와 같은 결과는 미숙종자 발현유전자의 경우 염색체별로 특이적으로 위치하는 유전자좌 없이 고루 분포하고 있는 것으로 해석되어진다. TIGR Pseudomolecule ver. 5에 mapping 되지 않은 162개의 유전자는 아직까지도 벼 염색체구조가 완전히 결정되어지지 않고 95%의 완성도 (IRGSP 2005)로 구조결정이 되어있어 염색체 구조결정이 되지 않은 부위 내에 위치하는 유전자로 추정된다. 유전자 위치가 결정되지 않은 162개의 유전자는 향후 gap-filling을 통한 염색체 구조결정시에 기준점으로 사용되어 질수 있다.

### COGs 분석을 이용한 유전자 기능분류

미숙종자에 발현하는 25,668개의 발현유전자의 기능분류를 위하여 COG분석을 행하였다. 미숙종자 7,509개의 unigene 정보에 대해 NCBI KOG DB 기반의 기능분류 분석을 수행하였다. 미숙종자

의 기능분류 분석 결과는 정보축적 및 처리 기능관련유전자 14.8, 세포생리기능관련 유전자 28.3%, 대사관련 유전자 21.2% 그리고 기능미확인 유전자 35.7%의 비율을 나타내었다. 미숙종자 발현유전자의 기능분류 결과 전사 후 수정, 단백질 이동, 사페론 유전자 등이 가장 많이 발현하였으며 그 외 신호전달에 관련된 다수의 유전자들이 발현되었다. 이러한 결과는 미숙종자 형성시기에 많이 발현하고 있는 glutelin 등 저장단백질의 이동과 접합에 관여하는 유전자들이 많이 발현하고 있는 것과 일치한 결과를 나타낸다. 미숙종자의 기능분류를 통하여 얻어진 정보를 이용하여 중요 대사관련 유전자의 기능분석과 농업형질관련 유전자들의 대량 기능 분석 연구에 적극 활용 가능하리라 생각한다.

## 적 요

벼는 국내에서 가장 중요한 식량작물로 게놈 크기가 389Mb이며 식물 유전자 기능분석을 위한 모델 작물이다. 종자 발달과정에 관여하는 유전자의 기능을 분석하기 위하여 일품벼의 미숙종자(출수 후 20일)의 발현유전자의 염기서열을 결정하였다. 염기서열이 결정된 25,668개의 발현유전자를 SeqMan 프로그램을 이용하여 cluster를 분석하여 7,509개의 unigene을 분리하였다. 일본 농업생물 자원연구소의 완전장 cDNA 32,127 클론 (KOME)과 미숙종자 발현 cDNA 7,509개를 비교분석 한 결과 4,990개는 일본의 클론과 일치하였으며 2,519개 클론은 상동성을 나타내지 않았다. 미숙종자발현 7,509개 unigene을 TIGR Pseudomolecule ver. 5의 염색체에 mapping 한 결과 7,347개는 각 염색체에 mapping되었으며 162개는 mapping되지 않았다.

미숙종자발현 7,509개 유전자의 기능을 NCBI COG DB를 이용하여 분류한 결과 정보축적 및 처리 기능관련유전자 14.8%, 세포생리기능관련 유전자 28.3%, 대사관련 유전자 21.2% 그리고 기능미확인 유전자 35.7%의 비율을 나타내었다.

## 사 사

이 연구는 농촌진흥청 국립농업과학원 어젠더 2-7-11 “벼 변이 집단 및 생물정보를 이용한 유용 농업형질 유전자탐색” 과제 및 “벼배추의 생장 및 대사관련 유전자발현 네트워크구축” 과제 지원에 의해 수행되었다. 본 연구수행 중 미숙종자 발현유전자 DNA 분리 및 염기서열 분석은 양희은, 최은영 양이 수고해 주었다.

## 인용문헌

Ammiraju JS, Lu F, Sanyal A, Yu Y, Song X, Jiang N, Pontaroli AC, Rambo T, Currie J, Collura K, Talag J, Fan C, Goicoechea JL, Zuccolo A, Chen J, Bennetzen JL, Chen M, Jackson S, Wing RA. (2008) Dynamic evolution of oryza genomes is revealed by comparative genomic analysis of a genus-wide

- vertical data set. *Plant Cell*. 20:3191-209
- Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, Muramatsu M, Hayashizaki Y, Schneider C. (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*. 37: 327-36
- Doi K, Hosaka A, Nagata T, Satoh K, Suzuki K, Mauleon R, Mendoza MJ, Bruskiewich R, Kikuchi S. (2008) Development of a novel data mining tool to find cis-elements in rice gene promoter regions. *BMC Plant Biol*. 8:20
- International Rice Genome Sequencing Project. (2005) The map-based sequence of the rice genome. *Nature*. 436: 793-800
- Jung KH, An G, Ronald PC. (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. *Nat Rev Genet*. 9:91-101
- Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A, Fukunishi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaka I, Saito T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R, Kadota K, Matsuda H, Ashburner M, Batalov S, Casavant T, Fleischmann W, Gaasterland T, Gissi C, King B, Kochiwa H, Kuehl P, Lewis S, Matsuo Y, Nikaido I, Pesole G, Quackenbush J, Schriml LM, Staubli F, Suzuki R, Tomita M, Wagner L, Washio T, Sakai K, Okido T, Furuno M, Aono H, Baldarelli R, Barsh G, Blake J, Boffelli D, Bojunga N, Carninci P, de Bonaldo MF, Brownstein MJ, Bult C, Fletcher C, Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann M, Hume DA, Kamiya M, Lee NH, Lyons P, Marchionni L, Mashima J, Mazzarelli J, Mombaerts P, Nordone P, Ring B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schönbach C, Seya T, Shibata Y, Storch KF, Suzuki H, Toyo-oka K, Wang KH, Weitz C, Whittaker C, Wilming L, Wynshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohtsuki S, Hayashizaki Y; RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*. 409:685-90
- Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T; Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K; RIKEN, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y, Yasunishi A. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science*. 301:376-9

- Lu T, Huang X, Zhu C, Huang T, Zhao Q, Xie K, Xiong L, Zhang Q, Han B. (2008) RICD: a rice indica cDNA database resource for rice functional genomics. *BMC Plant Biol.* 8:118
- Nakamura H, Hakata M, Amano K, Miyao A, Toki N, Kajikawa M, Pang J, Higashi N, Ando S, Toki S, Fujita M, Enju A, Seki M, Nakazawa M, Ichikawa T, Shinozaki K, Matsui M, Nagamura Y, Hirochika H, Ichikawa H. (2007) A genome-wide gain-of-function analysis of rice genes using the FOX-hunting system. *Plant Mol Biol.* 65:357-71
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR. (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 35(Database issue):D883-7
- Paterson AH, Freeling M, Sasaki T. (2005) Grains of knowledge: genomics of model cereals. *Genome Res.* 15: 1643-50.
- Rice Annotation Project. (2008) The Rice Annotation Project Database (RAP-DB). *Nucleic Acids Res.* 36(Database issue): D1028-33
- Satoh K, Doi K, Nagata T, Kishimoto N, Suzuki K, Otomo Y, Kawai J, Nakamura M, Hirozane-Kishikawa T, Kanagawa S, Arakawa T, Takahashi-Iida J, Murata M, Ninomiya N, Sasaki D, Fukuda S, Tagami M, Yamagata H, Kurita K, Kamiya K, Yamamoto M, Kikuta A, Bito T, Fujitsuka N, Ito K, Kanamori H, Choi IR, Nagamura Y, Matsumoto T, Murakami K, Matsubara K, Carninci P, Hayashizaki Y, Kikuchi S. (2007) Gene organization in rice revealed by full-length cDNA mapping and gene expression analysis through microarray. *PLoS ONE.* 2(11):e1235
- Seki M, Carninci P, Nishiyama Y, Hayashizaki Y, Shinozaki K. (1998) High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper. *Plant J.* 15:707-20
- Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M. (2008) Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet.* 40: 1023-8
- Song XJ, Huang W, Shi M, Zhu MZ, Lin HX. (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet.* 39:623-30
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 4:41
- Umezawa T, Sakurai T, Totoki Y, Toyoda A, Seki M, Ishiwata A, Akiyama K, Kurotani A, Yoshida T, Mochida K, Kasuga M, Todaka D, Maruyama K, Nakashima K, Enju A, Mizukado S, Ahmed S, Yoshiwara K, Harada K, Tsubokura Y, Hayashi M, Sato S, Anai T, Ishimoto M, Funatsuki H, Teraishi M, Osaki M, Shinano T, Akashi R, Sakaki Y, Yamaguchi-Shinozaki K, Shinozaki K. (2008) Sequencing and analysis of approximately 40,000 soybean cDNA clones from a full-length-enriched cDNA library. *DNA Res.* 15:333-46

(접수일자 2009년 3월 24일, 수리일자 2009년 4월 6일)