**In silico** identification, phylogeny and expression analysis of expansin superfamily in *Medicago truncatula*

Ying Liu, Jun Zhang, Wei Li, Changhong Guo and Yongjun Shu

Key Laboratory of Molecular Cytogenetics and Genetic Breeding of Heilongjiang Province, College of Life Science and Technology, Harbin Normal University, Harbin, China

**ABSTRACT**
Expansins are important components of plant cell walls, which are involved in the process of cell wall loosening under low extracellular pH. By using a combinational method for homology search and protein domain analysis, a total of 42 expansin genes were identified from *Medicago truncatula* genome in this study. They were divided into four families, based on sequence alignment and phylogenetic analysis. Gene duplication events were identified in the expansins superfamily, especially in the extension of α-expansin family. By analysis of RNA-sequencing data from National Center for Biotechnology Information, the expansin (EXP) genes expressed during tissues development were characterized. Meanwhile, lots of cis-acting regulatory DNA elements in the EXP superfamily were identified, which were mainly related to plant growth and development processes. The results presented in this study are expected to facilitate further research works on this gene superfamily and provide new insights about the molecular mechanisms of expansins in *M. truncatula*.

**KEYWORDS**
*Medicago truncatula*; expansins; evolution; gene duplication; tissues development

**Introduction**
Expansin is a kind of plant cell wall protein which releases the mechanical tension of the cell wall polymers and regulates the relaxation and extension of plant cells.

This protein mainly regulates the cell wall expansion through breaking non-covalent bonds during plant growth and development.[3] Usually, expansins' length in plants range from 250 to 275 amino acids, and they contain two domains (named domain I and domain II) and a signal peptide. Domain I is homologous to glycoside hydrolase family 45 (GH45), whereas domain II is distantly related to group-2 grass pollen allergens, which bind to cell wall polysaccharides. According to the nomenclature of expansins in plants, they are divided into four families. They are named as α-expansin (EXPA), β-expansin (EXPB), expansin-like A (EXLA) and expansin-like B (EXLB).[3]

Functional analyses of expansin superfamily have shown that they play important roles in many plant growth and developmental processes by modifying the wall structure and elongating the cell wall. Expansins are widely found to be involved in plant cell growth,[4–7] root development,[8,9] leaf growth,[8–11] seed germination,[12,13] fruit ripening [5,14] and other developmental processes.[15] Recently, expansins have also been investigated and characterized as participating in response processes to drought, salt and pathogen stresses.

Alfalfa (*Medicago sativa* L.) has become an important forage crop worldwide because of its high nutritional value, wide adaptation and high yield potential. *Medicago truncatula* is an annual legume species with a small diploid genome and easy transformation, which makes it an excellent model organism for legume crops and for functional genomics researches of alfalfa. Its genome sequencing was completed and released,[16] which would promote function investigations of genetics and genomics in *M. truncatula*.

In this study, the expansins' genes from *M. truncatula* genome were identified by sequence search and domain analysis. Meanwhile, characteristics of expansins superfamily, including classification, phylogeny, expression and promoter, were investigated. Our results provided valuable information for revealing the molecular function mechanisms of the expansin genes and for promoting their utilization in legume crop genetic improvement.
Materials and methods

Identification of expansin genes (MtEXP) in M. truncatula

The expansins sequences of Arabidopsis and rice (AtEXPB1, AtEXLA1, AtEXLB1, OsEXPB1, OsEXPB2, OsEXPB15, OsEXLA1 and OsEXLB1) were collected from the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/). [17] The M. truncatula genome and protein sequences were downloaded from the J. Craig Venter Institute (JCVI) website (M. truncatula Genome Project v4.0, http://www.jcvi.org/medicago/). [16] Arabidopsis and rice expansins were used as basic local alignment search tool (BLAST) [18] queries against M. truncatula genome with expected values of $\leq$1E-5. All hits were retrieved and were examined using hidden Markov models (HMMs) [19] with Pfam [20] numbers PF03330 and PF01357, the cut-off value was set as 1E-2. The positive proteins were identified as putative expansins (MtEXP), and their genomic information, including chromosome location and distributions of introns, was retrieved from M. truncatula genome website. We aligned these putative expansins to Arabidopsis and rice EXP proteins and classified them based on the sequences alignment.

Phylogenetic analysis of MtEXP genes

All of the putative MtEXP protein sequences were aligned using ClustalW with default parameters, [21] and phylogenetic trees of all MtEXP proteins were generated using Molecular Evolutionary Genetics Analysis (MEGA, v4.0) with the neighbour-joining method using the following parameters: Poisson correction, pairwise deletion and 1000 bootstraps. [22]

Chromosomal locations and gene duplication of MtEXP genes

The DNA and coding sequences of the MtEXP genes were collected from the M. truncatula genome database. They were blasted against each other and gene pair with similarity of the aligned regions more than 85% was identified as gene duplication event. Then, position information for all these MtEXP genes was retrieved to investigate their chromosomal distribution, and genomic distributional plot was created using the Circos software (http://circos.ca/). [23] In addition, the duplicated genes were linked with coloured lines.

In silico expression analysis of MtEXP genes

Transcriptome data of tissues development of M. truncatula were downloaded from NCBI database (http://www.ncbi.nlm.nih.gov, SRX099057—SRX099062). These data included six tissues: root, nodule, blade, bud, seedpod and flower. The details are described by Young et al. [16] All the high-throughput sequences were mapped to M. truncatula genome using TopHat, [24] and gene expressions were accessed by Cufflinks. [25] The expression data of MtEXP genes were retrieved, and they were analysed, clustered and visualized on Matlab platform (R2012a) using the Bioinformatics Toolbox.

Cis-regulatory element analysis of MtEXP genes

For cis-regulatory element analysis, 1000 bp sequences upstream from the translational start sites (TSS) of the putative MtEXP genes were retrieved from the M. truncatula genome. Then, these sequences were subjected to Plant cis-acting Regulatory DNA Elements (PLACE) database (http://www.dna.affrc.go.jp/PLACE/signalscan.html) [26] to identify cis-regulatory elements and those motifs with clearly more than 200 copies were collected for functional annotation analysis.

Results and discussion

Identification and classification of MtEXP genes in M. truncatula

There were 42 members of expansins in total, identified from M. truncatula genome by using BLAST and HMM search, named as MtEXP01—MtEXP42 (Table 1S in the Online Supplemental data). These MtEXP proteins’ lengths varied from 179 to 286 amino acids. Most of MtEXP proteins had more than 240 amino acids (aa), which was consistent with other plants. Three MtEXP proteins (MtEXP26, MtEXP30 and MtEXP41) had less than 240 aa – they had 239, 199 and 179 aa, respectively. Among these MtEXP genes, 35 members (35/42, 83.3%) contained from 1 to 3 introns, whereas the other 7 MtEXP genes had harboured more than 4 introns. Based on the alignments with Arabidopsis and rice EXP genes, MtEXP genes were classified into four families (EXPB, EXPA, EXLA, EXLB). EXPB, including 31 members (31/42, 73.8%), was the largest family. EXPB family contained six members. EXLB family was the smallest family with only one member. The remaining four members belonged to the EXLA family.

Expansins are plant cell wall-loosening proteins, which regulate the cell enlargement by a pH-associated mechanism. They are widely distributed in plant cells throughout the plant world, ranging from green algae to land plants. They play important roles in the regulation of plant growth and development processes, such as root development, leaf growth, fruit ripening, seed germination, penetration of pollen tubes, etc. In this study,
we identified 42 MtEXP genes from *M. truncatula* genome by bioinformatics analysis, which was similar to previous reports in *Arabidopsis*,[3] rice,[27] soybean [28] and poplar,[29] and indicated that expansin genes are highly conservative in plants.[17,30]

**Phylogenetic analysis of MtEXP genes**

Based on the alignment of full length sequences of MtEXP proteins, a phylogenetic tree (Figure 1) was constructed to determine the evolutionary relationship of MtEXP superfamily in detail. The phylogenetic tree showed that MtEXP genes were classified into four families, named as EXPA, EXPB, EXLA and EXLB, which was consistent with alignment results of other plants (*Arabidopsis* and rice).[3] According to the phylogenetic tree, the families EXLA and EXLB were firstly branching from MtEXP ancestors, EXPA and EXPB families were gradually diverging and shaping the MtEXP genes in *M. truncatula*. The EXPA family had 31 members, which was the same as the alignment of *Arabidopsis* and rice. They were divided into six groups, which was similar to the evolutionary situation in *Arabidopsis* and rice. These results suggested that expansins are very conservative in plants.

Expansins were divided into four families, but the gene numbers of each family were quite different when compared with other plants. To study the EXPA family in depth, we compared the members of the EXPA family, and identified 17 pairs of gene duplications among them. These gene duplications, including segmental (10 pairs) and tandem duplications (7 pairs), were the main driving forces of EXPA family expansion in *M. truncatula* genome. Tandem duplications mainly conferred expansin gene clusters, and segmental duplications brought rapidly extension of MtEXP genes across chromosomes.

**Chromosomal location analysis of MtEXP genes**

Chromosomal location analysis showed that MtEXP genes were distributed throughout the whole genome. However, there were five MtEXP genes (MtEXP30, MtEXP31, MtEXP35, MtEXP36 and MtEXP37) located on supercontigs and not presented on chromosomes. The remaining MtEXP members (37/42, 88.1%) were distributed across 8 chromosomes (Figure 2) and each chromosome harboured from 1 to 9 MtEXP genes. Chromosome MtChr4 contained the most members (nine MtEXP genes), chromosome MtChr7 contained seven MtEXP genes, whereas MtChr6 contained only one MtEXP gene.

Interestingly, MtEXP genes were clustered in *M. truncatula* genome and formed MtEXP clusters, for example, MtChr2, MtChr4 and MtChr7 (Figure 2). By homologues search, 18 pairs of MtEXP genes were identified as gene duplications. Among these duplications, 8 pairs (8/18,
44.4%) were closely co-located in the same chromosomes, which formed MtEXP clusters. For example, MtEXP39–MtEXP41 were closely located on chromosome MtChr4 and produced an MtEXP cluster on MtChr4 (Figure 2). In addition, most of MtEXP duplications (17/18, 94.4%) were present in the EXPA family, which had mainly contributed to expansion of EXPA family in M. truncatula genome.

Expression analysis of MtEXP genes during tissues’ development

Transcriptome analysis is an effective method for detecting genes’ expression during plant growth and development. Consequently, high-throughput sequencing data of M. truncatula related to tissues development were downloaded from NCBI. Expression of MtEXP genes in six tissues, including root, nodule, blade, bud, seedpod and flower, were collected, and the results showed that most MtEXP genes were specifically expressed in some tissues. Based on their expression patterns, the MtEXP genes were clustered into four groups (Figure 3). Cluster A included 12 MtEXP genes (MtEXP01, MtEXP08, MtEXP13, MtEXP15, MtEXP17, MtEXP26, MtEXP27, MtEXP30, MtEXP35, MtEXP40, MtEXP41 and MtEXP42), which were mainly expressed in roots and nodules. This indicated that they may be participating in the root development. Cluster B contained 13 MtEXP genes (MtEXP03, MtEXP04, MtEXP11, MtEXP12, MtEXP18, MtEXP21, MtEXP22, MtEXP24, MtEXP29, MtEXP32, MtEXP33, MtEXP36 and MtEXP37), which were expressed in buds and seedpod tissues and were not present or lowly expressed in the other tissues. Cluster C contained four MtEXP genes (MtEXP06, MtEXP07, MtEXP19 and MtEXP39), which were only expressed in blade. Cluster D, included 13 members (MtEXP02, MtEXP05, MtEXP09, MtEXP10, MtEXP14, MtEXP16, MtEXP20,
MtEXP23, MtEXP25, MtEXP28, MtEXP31, MtEXP34 and MtEXP38), which were highly expressed in flowers.

Based on the combination of MtEXP genes chromosomal location, we found that MtEXP genes involved in same process were closely located and this could be considered to confer the specific function of MtEXP genes clusters. For example, groups of co-located MtEXP genes (MtEXP13, MtEXP15, MtEXP17, MtEXP35, MtEXP40, MtEXP41 and MtEXP42), regulating the root growth and development of M. truncatula, were identified in chromosomes MtChr4 and MtChr5. MtEXP03, MtEXP04, MtEXP32 and MtEXP33, involved in seedpod development, were clustered on MtChr2, whereas MtEXP21, MtEXP22 and MtEXP24 were located on MtChr7. MtEXP genes (MtEXP06 and MtEXP07), participating in the process of bud development, were co-located on MtChr3 by a tandem duplication, whereas MtEXP genes (MtEXP05, MtEXP09 and MtEXP10), involved in flower development, were also clustered on MtChr3 by one tandem duplication (between MtEXP09 and MtEXP10) and segmental duplications from MtChr5. Overall, segmental and tandem duplications were found to be the main contributors to the evolution and diversification of MtEXP genes. The functions of MtEXP genes clustered on different chromosomes diverged. MtEXP genes involved in different processes, such as root, bud, flower development, etc., were co-located on various chromosome regions.

Cis-regulatory element analysis of MtEXP promoter

In plants, gene expression is regulated by transcription factors (TFs), which bind to cis-regulatory elements in promoters of target genes. Therefore, the promoter sequences of the MtEXP genes TSS were analysed by searching the PLACE database. The results showed that many similar cis-acting regulatory DNA elements associated with plant tissues development processes and stress responses were identified and characterized (Table 2S in the Online Supplemental data). For example, tissues development related motifs, including ROOTMOTIFAPX1, POLLEN1LELAT52, CACTFTPPCA1, GTGANTG10 and CAATBOX1, were abundant in promoter regions of MtEXP genes. Light regulative elements, including GT1CONSENSUS and GATABOX, were also abundantly present in MtEXP promoters. Finally, many important transcript factors binding sites, such as DOFCOREZM, MYCCONSENSUSAT, WRKY71OS, ARR1AT and EBOXBNAPA, were also overrepresented in the promoter regions of MtEXP, which may play important role in plant development and stress response.

In addition, there were also lots of TF binding elements identified in the promoters of MtEXP genes, including DOFCOREZM for Dof TF,[31] MYCCONSENSUSAT for MYC TF,[32] WRKY71OS for WRKY TF [33] and ARR1AT for ARF TF.[34] Many of them, such as WRKY, MYC and ARF, are important regulators of plant stress response. This indicated that MtEXP may also be involved in many stress response processes, but their functional mechanisms of stress response need to be further investigated in future.

Conclusions

In summary, 42 MtEXP genes were identified in M. truncatula, and their classification, phylogeny, expression pattern and promoters were investigated and characterized in silico. The MtEXP genes expression showed that these genes participated in various aspect of the regulation of plant tissues development and promoter analysis suggested they may be also playing important role in stress response. Our research provided the bases for revealing the molecular regulation of MtEXP genes in tissues development and stress response of M. truncatula.
Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by the Natural and Science Foundation of China [grant number 31302019], [grant number 31470571]; China Postdoctoral Science Foundation [grant number 2015M571430]; MOST 863 project [grant number 2013AA102607-5]; Heilongjiang Province Postdoctoral Science Foundation [grant number LBH-Z14126].

Supplemental data
Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/13102818.2015.1093919

References
[1] Mendre C, Sarrade V, Calas B. Continuous flow synthesis of peptides using a polyacrylamide gel resin (Expansin). Int J Pept Protein Res. 1992;39:278–284.
[2] McQueen-Mason SJ, Cosgrove DJ. Expansin mode of action on cell walls. Analysis of wall hydrolysis, stress relaxation, and binding. Plant Physiol. 1995;107:87–100.
[3] Sampedro J, Cosgrove DJ. The expansin superfamily. Genome Biol. 2005;6:242.
[4] Cho HT, Kende H. Expansins in deepwater rice internodes. Plant Physiol. 1997;113:1137–1143.
[5] Cosgrove DJ, Li LC, Cho HT, et al. The growing world of expansins. Pept Protein Res. 1992;39:278–284.
[6] Cosgrove DJ, Li LC, Cho HT, et al. The growing world of expansins. Pept Protein Res. 1992;39:278–284.
[7] Choi D, Lee Y, Cho HT, et al. Regulation of expansin gene transcription in the maize primary root at low water potentials. Plant Physiol. 2001;126:1471–1479.
[8] Guo W, Zhao J, Li X, et al. A soybean beta-expansin gene GmEXPB2 intrinsically involved in root system architecture responses to abiotic stresses. Plant J. 2011;66:541–552.
[9] Lu P, Kang M, Jiang X, et al. RhEXP4A, a rose expansin gene, modulates leaf growth and confers drought and salt tolerance to Arabidopsis. Plant. 2013;237:1547–1559.
[10] Kulesh BR, Kniazev AV, Nikonorov Iu M, et al. Role of the expansin genes NTEXP1A and NTEXP4A in the regulation of cell extension during tobacco leaf growth. Genetika. 2014;50:560–569.
[11] Chen F, Bradford KJ. Expression of an expansin is associated with endosperm weakening during tomato seed germination. Plant Physiol. 2000;124:1265–1274.
[12] Magneschi L, Kudahettige RL, Alpi A, et al. Expansin gene expression and anoxic coleoptile elongation in rice cultivars. J Plant Physiol. 2009;166:1576–1580.
[13] Harrison EP, McQueen-Mason SJ, Manning K. Expression of six expansin genes in relation to extension activity in developing strawberry fruit. J Exp Bot. 2001;52:1437–1446.
[14] Pezzotti M, Feron R, Mariani C. Pollination modulates expression of the PPAL gene, a pistil-specific beta-expansin. Plant Mol Biol. 2002;49:187–197.
[15] Young ND, Debell E, Oldroyd GE, et al. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature. 2011;480:520–524.
[16] Dal Santo S, Vannozzi A, Tornielli GB, et al. Genome-wide analysis of the expansin gene superfamily reveals grapevine-specific structural and functional characteristics. PLOS One. 2013;8:e62206.
[17] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–3402.
[18] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29–W37.
[19] Finn RD, Mistry J, Schuster-Böckler B, et al. Pfam: clans, web tools and services. Nucleic Acids Res. 2006;34: D247–D251.
[20] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–4680.
[21] Tamura K, Dudley J, Nei M, et al. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24:1596–1599.
[22] Kryzinski Mi, Schein JE, Birol I, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–1645.
[23] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–1111.
[24] Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–578.
[25] Higo K, Ugawa Y, Iwamoto M, et al. PLACE: a database of plant cis-acting regulatory DNA elements. Nucleic Acids Res. 1998;26:358–359.
[26] Hemalatha N, Rajesh MK, Narayanan NK. Genome-wide analysis and identification of genes related to expansin gene family in indica rice. Int J Bioinform Res Appl. 2011;7:162–167.
[27] Zhu Y, Wang S, Song W, et al. Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. BMC Plant Biol. 2014;14:93.
[28] Sampedro J, Carey RE, Cosgrove DJ. Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs. J Plant Res. 2006;119:9–21.
[29] Zhang W, Yan H, Chen W, et al. Genome-wide identification and characterization of maize expansin genes expressed in endosperm. Mol Genet Genomics. 2012;288:1061–1074.
[30] Yanagisawa S. The Dof family of plant transcription factors. Trends Plant Sci. 2002;7:555–560.
[31] Pattanaik S, Xie CH, Yuan L. The interaction domains of the plant Myc-like bHLH transcription factors can regulate the transactivation strength. Planta. 2008;227:707–715.
[33] Agarwal P, Reddy MP, Chikara J. WRKY: its structure, evolutionary relationship, DNA-binding selectivity, role in stress tolerance and development of plants. Mol Biol Rep. 2011;38:3883–3396.

[34] Matsui A, Mizunashi K, Tanaka M, et al. taSiRNA-ARF pathway moderates floral architecture in arabidopsis plants subjected to drought stress. Biomed Res Int. 2014; 2014:303451.