



# Identification and Expression Analysis of BURP Domain-Containing Genes in *Medicago truncatula*

Yuan Li<sup>1†</sup>, Xue Chen<sup>1†</sup>, Zhu Chen<sup>1</sup>, Ronghao Cai<sup>2</sup>, Hongmei Zhang<sup>1</sup> and Yan Xiang<sup>1,2\*</sup>

<sup>1</sup> Laboratory of Modern Biotechnology, School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, China, <sup>2</sup> Key Laboratory of Crop Biology of Anhui Province, School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, China

## OPEN ACCESS

### Edited by:

Karen M. Vasquez,  
The University of Texas at Austin, USA

### Reviewed by:

Juan Caballero,  
Universidad Autónoma de Querétaro,  
Mexico  
Mingqi Zhou,  
University of Florida, USA

### \*Correspondence:

Yan Xiang  
xiangyanahau@sina.com

† These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 20 January 2016

**Accepted:** 25 March 2016

**Published:** 13 April 2016

### Citation:

Li Y, Chen X, Chen Z, Cai R, Zhang H  
and Xiang Y (2016) Identification  
and Expression Analysis of BURP  
Domain-Containing Genes  
in *Medicago truncatula*.  
*Front. Plant Sci.* 7:485.  
doi: 10.3389/fpls.2016.00485

BURP domain-containing proteins belong to a newly identified protein class that is unique to plants and plays an important role in plant development and metabolism. Although systematic characterization of BURP domain-containing proteins have been carried out in many species, such as rice, poplar and maize, little is known about BURP domain-containing proteins in *Medicago*. In this study, multiple bioinformatics approaches were employed to identify all the members of BURP family genes in *Medicago*. A complete set of 39 BURP family genes were identified. These genes have diverse structures and were distributed on chromosome 1–8 except 7. According to phylogenetic analysis, these BURP family genes could be classified into eight classes. Motif and exon-intron organization, stress-related *cis*-elements in promoter regions and microarray analysis of *MtBURP*s were also performed. Furthermore, transcript level analysis of *MtBURP* genes in response to drought stress revealed that all of the 39 BURP genes were regulated by drought stress. The results of this study reveal a comprehensive overview of the *Medicago* BURP gene family and provide the first step toward the selection of *MtBURP* genes for cloning and functional analysis of the BURP gene family in *Medicago truncatula*.

**Keywords:** abiotic stress, bioinformatics, BURP genes, *Medicago truncatula*

## INTRODUCTION

The BURP domain is a conserved C-terminal protein domain whose name is based on four typical members, BNM2, USP, RD22, and PG1 $\beta$ . BURP domain-containing proteins have so far only been found in plants, suggesting that their functions may be plant specific. BNM2 is a protein specifically expressed during microspore embryogenesis in oilseed rape (Boutilier et al., 1994); USPs, which are expressed during the early stages of zygotic (Bassüner et al., 1988) and *in vitro* embryogenesis (Chesnokov et al., 2002) in field bean; RD22 is a dehydration-responsive protein in *Arabidopsis thaliana* (Yamaguchi-Shinozaki and Shinozaki, 1993), and PG1 $\beta$ , a non-catalytic  $\beta$ -subunit of the polygalacturonase isozyme 1, which is expressed during the ripening of tomato (*Solanum lycopersicum*) (Zheng et al., 1992).

BURP domain-containing proteins have several conserved modules: an N-terminal hydrophobic domain with a presumptive transit peptide; a short conserved segment or other

short segment; an optional segment consisting of repeated units, which is unique to each member of the BURP-domain-containing family, and the C-terminal BURP domain (Hattori et al., 1998). BURP is a motif of approximately 230 amino acids and usually contains several highly conserved sequences, including two phenylalanine (F) residues, two cysteine (C) residues and four repeated cysteine-histidine (CH) motifs, and generally conforms to the sequence as CHXCHX<sub>23–27</sub>CHX<sub>23–26</sub>CHX<sub>8</sub>W (where X is any amino acid residue) (Ding et al., 2009).

BURP domain-containing proteins have been found in many species, such as rice, soybean, maize, and sorghum (Ding et al., 2009; Xu et al., 2010; Gan et al., 2011). A systematic analysis of BURP genes has been carried out in each of these species. Based on phylogenetic analysis of the putative BURP domain-containing proteins in soybean and other plant species, the members of the BURP family are classified into the BNM2-like, USP-like, RD22-like and PG1 $\beta$ -like subfamilies (Granger et al., 2002). Differences among the BURP domain-containing proteins mainly occur in a variable region between the hydrophobic signal sequence and the BURP domain (Ding et al., 2009). This region includes a short conserved segment and an optional segment consisting of repeated units which are absent in BNM2-like proteins. PG1 $\beta$ -like proteins are distinguished from the members of other subfamilies by the presence of a sequence segment containing a 14-amino-acid sequence (Zheng et al., 1992). Both the USP-like and RD22-like proteins are distinguished from other subfamilies by the presence of a sequence segment containing approximately 30 amino acids, but the RD22-like proteins have a segment consisting of repeated units of approximately 20 amino acids, while the USP-like proteins do not contain such a segment (Granger et al., 2002).

Many BURP proteins have been isolated from various plants species, but their expression patterns are diverse with some functions still unknown. BURP domain-containing proteins play an important role in maintaining normal plant metabolism or development (Shao et al., 2011; Tang et al., 2014). VfUSP, an abundant non-storage seed protein from field bean (*Vicia faba* L.) with unknown functions, is expressed during the early stages of zygotic embryogenesis (Bassüner et al., 1988) and *in vitro* embryogenesis (Chesnokov et al., 2002), respectively. PG1 $\beta$ , the non-catalytic  $\beta$ -subunit of the polygalacturonase isozyme (PG) from ripening tomato (*Solanum lycopersicum* L.), plays a significant role in regulating pectin metabolism by limiting the extent of pectin solubilization and depolymerization (Zheng et al., 1992; Liu et al., 2014). ASG1 is a gene that specifically expressed during early embryo sac development in apomictic gynoecea of *Panicum* (Chen et al., 1999), but it is not expressed in sexual gynoecea of *Panicum* (Chen et al., 1999). *AtUSPL1* was found in cellular compartments such as Golgi cisternae, dense vesicles, prevacuolar vesicles and protein storage vacuoles in the parenchyma cell of cotyledons and thus may play a role in seed development (Tiedemann et al., 2009). BNM2-like was originally found to be expressed in microspore-derived embryos of *Brassica napus*, but recent studies have shown that BNM2-like is confined to seeds and is localized to the protein storage vacuoles (Treacy et al.,

1997). The mRNA encoding *SCB1* (seed coat BURP-domain protein 1), is specifically found within the seed coat during the early stages of soybean seed development. *SCB1* helps regulate the formation of the soybean seed coat by governing the differentiation of seed coat parenchyma cells (Batchelor et al., 2002). *VvBURP1* belongs to the BURP domain-containing protein family. This protein is similar to the BNM2-like proteins, and the expression profiles of *VvBURP1* in different grapevine organs are close to those in soybean (Fernandez et al., 2007). *OsRAFTIN1*, an anther-specific protein in rice (*Oryza sativa* L.), transports sporopollenin from tapetum to developing microspores via Ubisch bodies (Wang et al., 2003).

Many BURP domain-containing proteins are also stress induced. The expression of RD22 is induced by drought, abscisic acid (ABA) and salt stress, and mechanisms underlying this regulation have been well-studied (Urao et al., 1993). Four genes from *Bruguiera gymnorrhiza* (*BgBDC1*, 2, 3, and 4) encode BURP proteins are similar to RD22 and may function in responses to abiotic stresses such as drought stress and ABA induction, as well as biotic stresses such as plant vegetative and reproductive development, plant diseases and insect damage (Harshavardhan et al., 2014). The expression of *SAL13-2* and *SAL15-4a*, putative BURP protein genes, is induced by aluminum stress in soybean (Ragland and Soliman, 1997). The transcript level of *BnBDC1*, a shoot-specific gene encoding a BURP protein in oilseed rape, is up regulated by salt, ABA and osmotic stresses and down regulated by salicylic acid. Thus this gene may be involved in both abiotic and biotic stresses (Shunwu et al., 2004). The over expression of *SAL13-2* increases salt tolerance in yeast, but signal peptide deletion weakens salt tolerance (Ragland and Soliman, 1997). These reports suggest that BURP family genes might be crucial not only for plant development but also for response and adaptation to stresses.

*Medicago* is an annual flowering plant known as “the king of the forage grass.” Compared with other leguminous plants, the diploid *Medicago truncatula* is a fast-growing plant with a small genome, and it has therefore become a model plant for DNA sequence analysis (Endre et al., 2002; Stacey et al., 2006). However, *Medicago* production is threatened by drought and other environmental stresses. Although BURP genes have been systematically characterized in rice, maize and poplar, an extensive analysis of BURP genes in *Medicago* has not been reported. As the genome sequence of the *Medicago* is complete, it is possible to analyze the entire family of *Medicago* BURP proteins. In the current study, 39 putative genes of the BURP family were identified. To discover the functions of all the members, these genes were subjected to phylogenetic analysis, structural analysis, microarray analysis and expression profile analysis. In addition, the transcript levels of all 39 genes under drought stress were investigated. The results presented in this study showed that the expression of most of the *Medicago* BURP genes is non-tissue-specific but stress-responsive. This research will enable systematic analysis of the functions of the *MtBURP* gene family in the future.

## MATERIALS AND METHODS

### *In Silico* Identification of Putative BURP Proteins in *Medicago truncatula*

First, the *M. truncatula* protein database was downloaded from <http://www.medicago-hapmap.org/home/view>. The Hidden Markov Model (HMM) profile of the BURP domain (PF03181) was then obtained from the Pfam database<sup>1</sup> (Finn et al., 2006; Ding et al., 2009). This HMM profile was then employed as a query to identify all BURP-containing sequences by searching BURP domain sequence against the *M. sativa* protein database using BlastP ( $P$ -value = 0.001). Finally, all of the primary candidates were identified by performing complete sequence alignment with MEGA v4.0 (Tamura et al., 2007) and checked manually to remove any overlapping sequences. And all of the non-redundant BURP genes were subjected to further analysis.

### Chromosomal Localization of *Medicago truncatula* BURP Genes

A chromosome location image of *M. truncatula* BURP genes was generated with MapInspect software<sup>2</sup> according to their starting positions on the *M. truncatula* chromosomes. The starting positions of *M. truncatula* BURP genes were obtained from <http://www.medicago-hapmap.org/genome>. The genes were designated *MtBURP1* to *MtBURP39* according to the gene's position, from the top to the bottom, on *M. truncatula* chromosomes (1–8).

### Phylogenetic Analysis

To study the genetic relationships among the BURP domain-containing proteins and infer the evolutionary history of this gene family, the multiple sequence alignment of all predicted *Medicago* BURP protein sequences was performed with Clustal X2.0 software using default parameters. Then, based on this alignment, phylogenetic trees were constructed using Clustal X2.0 with the Neighbor-Joining (NJ) method, and bootstrap analysis was conducted using 1,000 replicates (Hu et al., 2010). A phylogenetic tree was initially constructed using the complete *MtBURP* protein sequences and the full-length protein sequences from rice, sorghum, maize, soybean, 4 BURP proteins in *Arabidopsis* (*AT1G49320.1*, *AT1G70370.1*, *AT1G23760*, and *AT1G60390.1*) and 4 host BURP members (*AtRD22*, *Vf-USPs*, *BNM2* and *Le-PGIβ*). Classification of the *MtBURP* genes was then performed according to their phylogenetic relationships with corresponding representative BURP genes.

### Characterization of *MtBURP* Proteins

Information regarding the *MtBURP* sequences, such as the number of amino acids, molecular weights (MW) and PIs, was obtained using ProtParam<sup>3</sup>. Multiple sequence alignment of the *MtBURP* proteins was performed with Clustal X and calculated

using MEGA v4.0. Determinations of the signal peptides were performed with SignalP<sup>4</sup>.

### Motif Distribution and Exon-Intron Structure of *MtBURP* Proteins

The conserved motifs encoded by each *MtBURP* gene were investigated. Protein sequences were subjected to online Multiple Expectation Maximization for Motif Elicitation (MEME<sup>5</sup>). Parameters were set as follows: the occurrence of a single motif was zero or one per sequence; minimum motif width was 10; maximum motif width was 200 and maximum number of motifs to find was 15; all other parameters were default values. To predict the exon-intron structure of the *MtBURP* genes, comparison of the genomic sequences and their predicted coding sequences (CDS) was performed using GSDS<sup>6</sup> (Guo et al., 2007). To analyze whether the motifs in each putative BURP gene was implicated in stress responses, 2,000-bp genomic sequences upstream of the transcription start codon (ATG) were acquired from ([www.phytozome.net/search.php](http://www.phytozome.net/search.php)). PLACE<sup>7</sup>, an online database of plant *cis*-acting regulatory DNA elements (*cis*-element) (Higo et al., 1999), was employed to investigate putative *cis*-elements in the promoter regions of the *MtBURP* genes.

### Calculation of Ka/Ks Values

The synonymous (Ks) and non-synonymous (Ka) nucleotide substitution rates and the Ka/Ks values were calculated for each pair of duplicated *Medicago* BURP proteins. Protein sequences of the gene pairs were aligned using Clustal X (Thompson et al., 1997), and the results were used to guide the codon alignments by PAL2NAL<sup>8</sup> (Stacey et al., 2006). The generated codon alignments were subjected to computation of Ks and divergence levels (Ka/Ks ratios) using DnaSP software (version 5.10), and these values were then translated into divergence time in millions of years assuming a rate of  $6.1 \times 10^{-9}$  substitutions per site per year. The divergence time (T) was calculated as  $T = Ks/2\lambda$  ( $\lambda = 6.1 \times 10^{-9}$ ) (Juretic et al., 2005). A sliding window analysis of Ka/Ks ratios was performed with the following parameters: window size, 150 bp; step size, 9 bp.

### Microarray Analysis

The genome-wide microarray data were obtained from the *M. truncatula* Gene Expression Atlas (MtGEA) Project<sup>9</sup>. The expression data were gene-wise normalized. Hierarchical clustering analysis was conducted using clustering distance “correlation” (Pearson correlation) and the clustering method used “complete” (complete linkage method) in R (Robin et al., 2011). A heat map was generated in R using the pheatmap function (Robin et al., 2011).

<sup>4</sup><http://www.cbs.dtu.dk/services/SignalP/>

<sup>5</sup><http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>

<sup>6</sup><http://gsds.cbi.pku.edu.cn/>

<sup>7</sup><http://www.dna.affrc.go.jp/PLACE/signalscan.html>

<sup>8</sup><http://www.bork.embl.de/pal2nal/>

<sup>9</sup><http://mtgea.noble.org/v2/>

<sup>1</sup><http://pfam.sanger.ac.uk/>

<sup>2</sup>[http://www.plantbreeding.wur.nl/uk/software\\_mapinspect.html](http://www.plantbreeding.wur.nl/uk/software_mapinspect.html)

<sup>3</sup><http://web.expasy.org/protparam/>

**TABLE 1 | Sequence characteristics of 39 BURP genes identified in *Medicago truncatula*.**

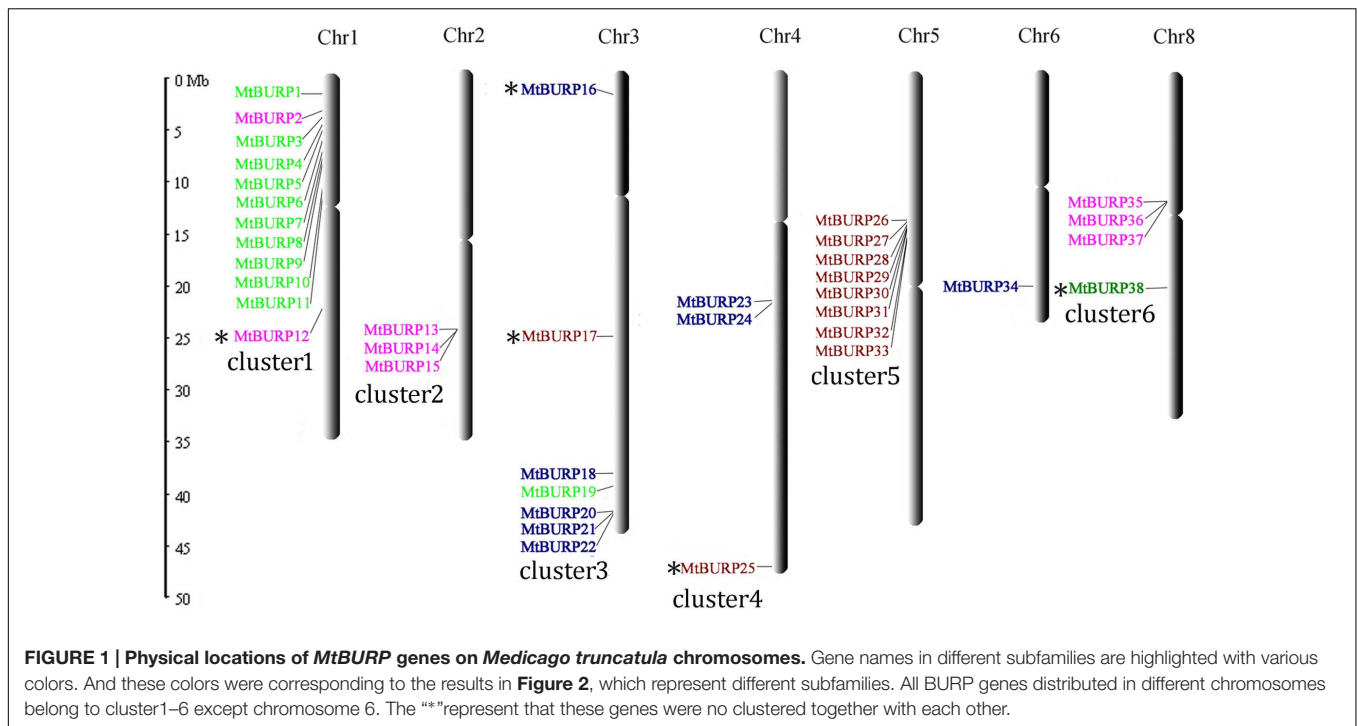
Gene name	Sequence ID	Deduced polypeptide			Chr	Chr. Locations	Signal P position	Cleavage Site
		Length (aa)	MW (kDa)	PI				
MtBURP1	Medtr1g007520.1	325	36.0253	8.7	1	0.989854	1–21	TNA-TM
MtBURP2	Medtr1g008420.1	217	24.5913	9.33	1	1.470671	1–21	THA-AL
MtBURP3	Medtr1g008450.1	313	34.7706	6.83	1	1.478971	1–21	THA-TL
MtBURP4	Medtr1g008460.1	305	34.2124	8.67	1	1.48444	No	
MtBURP5	Medtr1g008470.1	325	36.8425	8.06	1	1.491672	No	
MtBURP6	Medtr1g008500.1	385	41.4282	8.87	1	1.504798	1–21	THA-AL
MtBURP7	Medtr1g008510.1	326	36.5821	7.63	1	1.511292	1–21	TNA-AI
MtBURP8	Medtr1g008530.1	327	36.8692	6.86	1	1.517723	1–21	TNA-EI
MtBURP9	Medtr1g008560.1	330	37.0536	7.16	1	1.529702	1–21	TNA-AM
MtBURP10	Medtr1g008580.1	323	36.2235	6.64	1	1.537707	1–21	TNA-TV
MtBURP11	Medtr1g008590.1	325	36.0043	8.7	1	1.544253	1–21	TNA-VM
MtBURP12	Medtr1g045860.1	269	29.9221	6.45	1	13.74125	1–21	ISA-TT
MtBURP13	Medtr2g081590.1	273	30.5491	6.26	2	24.542149	1–23	SHA-SV
MtBURP14	Medtr2g081610.1	268	29.81	5.73	2	24.549179	No	
MtBURP15	Medtr2g081770.1	278	31.0876	5.52	2	24.61816	1–23	AFA-GE
MtBURP16	Medtr3g116380.1	307	34.7511	6.21	3	1.981704	1–26	SCA-RK
MtBURP17	Medtr3g078090.1	628	68.6082	9.41	3	24.944282	1–20	SIA-TI
MtBURP18	Medtr3g107830.1	264	30.0483	8.1	3	38.355243	No	
MtBURP19	Medtr3g109490.1	322	35.6941	8.11	3	39.318334	1–21	THA-AL
MtBURP20	Medtr3g116270.1	512	58.8735	6.63	3	41.93637	1–25	GYA-RD
MtBURP21	Medtr3g116320.1	565	64.3098	6.63	3	41.951485	1–24	ASC-AR
MtBURP22	Medtr3g116410.1	285	32.3102	6.7	3	42.003836	1–26	DQA-RE
MtBURP23	Medtr4g068620.1	128	14.3274	6.12	4	21.664271	No	
MtBURP24	Medtr4g069680.1	296	33.4937	8.16	4	21.786799	1–26	SRA-RE
MtBURP25	Medtr4g133840.1	190	20.9981	6.99	4	47.42306	No	
MtBURP26	Medtr5g033330.1	621	70.6707	9.62	5	13.95499	No	
MtBURP27	Medtr5g033360.1	225	25.8113	9.22	5	13.960123	16-Jan	CDG-RK
MtBURP28	Medtr5g033410.1	406	46.3597	5.98	5	13.97349	No	
MtBURP29	Medtr5g033420.1	226	27.0057	6.92	5	13.975735	No	
MtBURP30	Medtr5g033430.1	121	14.4494	5.93	5	13.977282	No	
MtBURP31	Medtr5g033440.1	200	23.6538	4.95	5	13.979471	No	
MtBURP32	Medtr5g033450.1	305	33.8686	4.55	5	13.981007	1–22	CFA-QE
MtBURP33	Medtr5g034320.1	630	68.7653	8.57	5	14.438111	1–23	ISA-GD
MtBURP34	Medtr6g086530.1	209	23.3262	8.84	6	20.258771	No	
MtBURP35	Medtr8g045890.2	523	59.6125	5.95	8	11.990199	1–24	SKS-GE
MtBURP36	Medtr8g045970.1	452	51.7236	5.68	8	12.020602	No	
MtBURP37	Medtr8g046000.1	236	25.7667	6.59	8	12.033004	1–17	ALA-GI
MtBURP38	Medtr8g075930.1	87	9.6231	6.94	8	20.582426	No	
MtBURP39	AC235667_9.1	325	37.2305	7.07	0	20.440642	1–21	INA-SQ

## Plant Material and Treatments

The plant material used in this analysis were young *M. truncatula* seedlings, which were grown in a growth chamber with a continuous 30°C temperature, a photoperiod of 12 h/12 h, 80  $\mu\text{molm}^{-2}\text{s}^{-1}$  photon flux density and 50% relative humidity were used. For drought stress, the seedling leaves were treated with 20% PEG-6000 (polyethylene glycol). Leaves of the stress-treated plants were collected at time intervals of 0, 2, 4, 8, 12, and 24 h. Control seedlings were not subjected to stress treatment (Zhou et al., 2012, 2014). After harvest, the materials were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for further analysis.

## Expression Analysis

Total RNAs of all the collected samples were extracted using the Trizol reagent (Bio Basic Inc, Canada). Subsequently, the RNAs were treated with DNase I (Promega, Madison, WI, USA) to remove contaminating genomic DNA. Approximately 1  $\mu\text{g}$  of each DNase-treated RNA samples was reverse transcribed by Prime Script RT Master Mix (Takara, Dalian, China) at 37°C for 20 min in a 20  $\mu\text{l}$  reaction sample. The amplification reactions were carried out using an SYBR Premix Ex Taq<sup>TM</sup> II (Perfect Real Time) (Takara, Dalian, China) according to manufacturer's instructions with three replicates (Zhou et al., 2012, 2014). The specific primers of each BURP gene (**Supplementary**



**Table S1**) were designed using Primer5.0, and their specificity was checked using information provided on the NCBI website. SEC (TC77416) (Kuppusamy et al., 2004), a constitutively expressed *Medicago* housekeeping gene, was used as reference for normalization. The PCR conditions were as follows: 95°C for 10 min, 40 cycles of 15 s at 95°C, 58°C for 1 min. Three technical replicates were taken in each biological replicate. The relative expression level was calculated as  $2^{-\Delta\Delta C_T}$  [ $\Delta C_T = C_{T,Target} - C_{T,CYP2}$ .  $\Delta\Delta C_T = \Delta C_{T,treatment} - \Delta C_{T,CK(0 h)}$ ]. The relative expression level ( $2^{-\Delta\Delta C_T,CK(0 h)}$ ) in the control plants without treatment was normalized to 1 as described previously. Statistical analysis was performed using SDS software 1.3.1 (Applied Biosystems).

## RESULTS

### Identification and Annotation of BURP Genes in *Medicago truncatula*

To identify all putative genes of BURP family in the *Medicago* genome, the HMM profiles of the BURP domain were employed as a query to search against the protein database of *M. truncatula* using the BlastP program. Through this approach, a total of 43 BURP domain-containing sequences were identified. A complete sequence alignment in MEGA 4.0 was used to identify 39 non-redundant BURP genes in *M. truncatula*. These 39 *MtBURPs* were entered into the Pfam and SMART databases to ensure that they all contain the BURP domain. The total number of BURP genes identified in *M. truncatula* (39) is greater than that in other species, such as poplar (18) (Shao et al., 2011), rice (17) (Ding et al., 2009), soybean (23) (Xu et al., 2010) and sorghum (11) (Gan

et al., 2011). Information about the putative *MtBURP* proteins, including the number of amino acids (length), MW, isoelectric point (PI) and physical location on the chromosomes, is listed in **Table 1**. The lengths of the *MtBURP* protein sequences range from 87 aa (*MtBURP38*) to 630 aa (*MtBURP33*), while the MWs of these *MtBURP* proteins range from 9.6231 kDa (*MtBURP38*) to 70.6707 kDa (*MtBURP26*). The lowest PI of the *MtBURP* proteins is 4.55 (*MtBURP32*), while the highest is 9.62 (*MtBURP26*).

### Chromosomal Locations of *MtBURP* Genes

Thirty-eight out of 39 BURP genes are distributed unevenly across all of the chromosomes of *M. truncatula* except for chromosome 7, while only one gene (*BURP39*) was mapped to as yet unattributed scaffolds (**Figure 1**). The number of *MtBURP* genes on each chromosome varies widely. The largest number of BURP genes was found on chromosome 1 (12 genes), followed by chromosome 5 (eight genes). By contrast, chromosome 6 contains only one BURP gene. Three genes are located on chromosomes 2, and three genes are located on chromosome 4. Seven genes are distributed on chromosome 3. Four genes were identified on chromosomes 8. These results show that the distribution pattern of each type of BURP gene is highly irregular. The authors also noticed that BURP genes on chromosome 2 and 5 clustered together with each other. The BURP genes on chromosomes 1, 3, 4, and 8 clustered together with each other except *MtBURP12*, *MtBURP16*, *MtBURP17*, *MtBURP25*, and *MtBURP38*. The names of the BURP genes (from *MtBURP1* to *MtBURP39*) were determined according to their positions on the chromosomes. This nomenclature system provides a unique identifier for each member in a gene family. Such a system has

previously been used in genome-wide analysis, including studies of the auxin response factor (ARF) gene in *Arabidopsis* (Nakano et al., 2006) and rice (Wang et al., 2007).

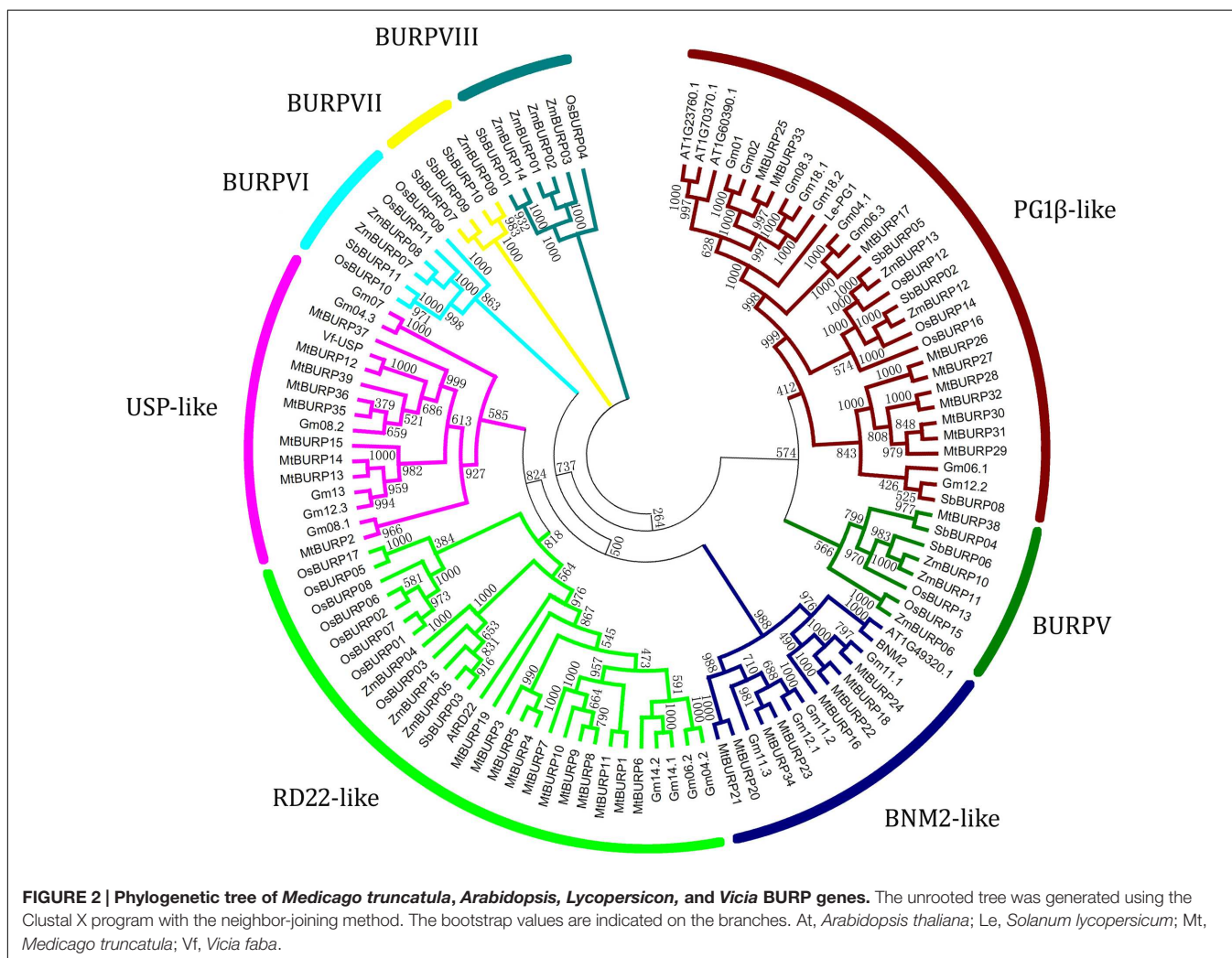
In *M. truncatula*, the 38 BURP genes are present in six clusters. According to Holub (2001), a gene cluster is a chromosomal region containing two or more genes within a 200-kb region. Each *M. truncatula* chromosome contains a BURP gene cluster, except for chromosome 6. A high-density region was identified, containing ten BURP genes (*MtBURP2–11*) on a 0.6 Mb region of chromosome 1; the authors therefore determined that region, is a hot spot of BURP gene distribution.

### Phylogenetic Analysis of *MtBURPs*

The abundance of *MtBURP* genes, compared to that in other plant species, may be derived from multiple gene duplication events, represented by a whole-genome duplication following multiple segmental and tandem duplications. To verify this hypothesis, a phylogenetic tree was constructed from the 113 BURP sequences, including 17 rice, 11 sorghum, 15 maize, 23 soybean, 39 *Medicago*, 4 BURP proteins in *Arabidopsis* (*AT1G49320.1*, *AT1G70370.1*, *AT1G23760* and *AT1G60390.1*)

and 4 host BURP members (*AtRD22*, *Vf-USPs*, *BNM2* and *Le-PG1β*) (Gan et al., 2011). Based on the domain composition and phylogenetic relationship of the protein sequences, the 113 plant BURP proteins could be divided into eight subfamilies: BNM2-like, RD22-like, USP-like, PG1β-like, BURPV, BURPVI, BURPVII, BURPVIII (Ding et al., 2009). Interestingly, all the members in the BURPVI and BURPVIII subfamilies are exclusively from rice, maize and sorghum, belonging to monocotyledon, which suggests that the main characteristics of this family in plants were generated before the dicot-monocot split. Likewise, the members of the BNM2-like subfamily and USP-like subfamily were all from dicotyledons, while RD22-like subfamily, BURPV subfamily and PG1β-like family contained members from both dicotyledon and monocotyledon species. Interestingly, BURPVII subfamily contained members from maize and sorghum that belong to the same family of plants. This suggested that the functions and evolution of the BURP genes might have some variation in dicotyledon and monocotyledon (Figure 2).

The exon-intron structural analysis of the *MtBURP* genes was also performed by the authors. Each CDS of the *MtBURP* genes



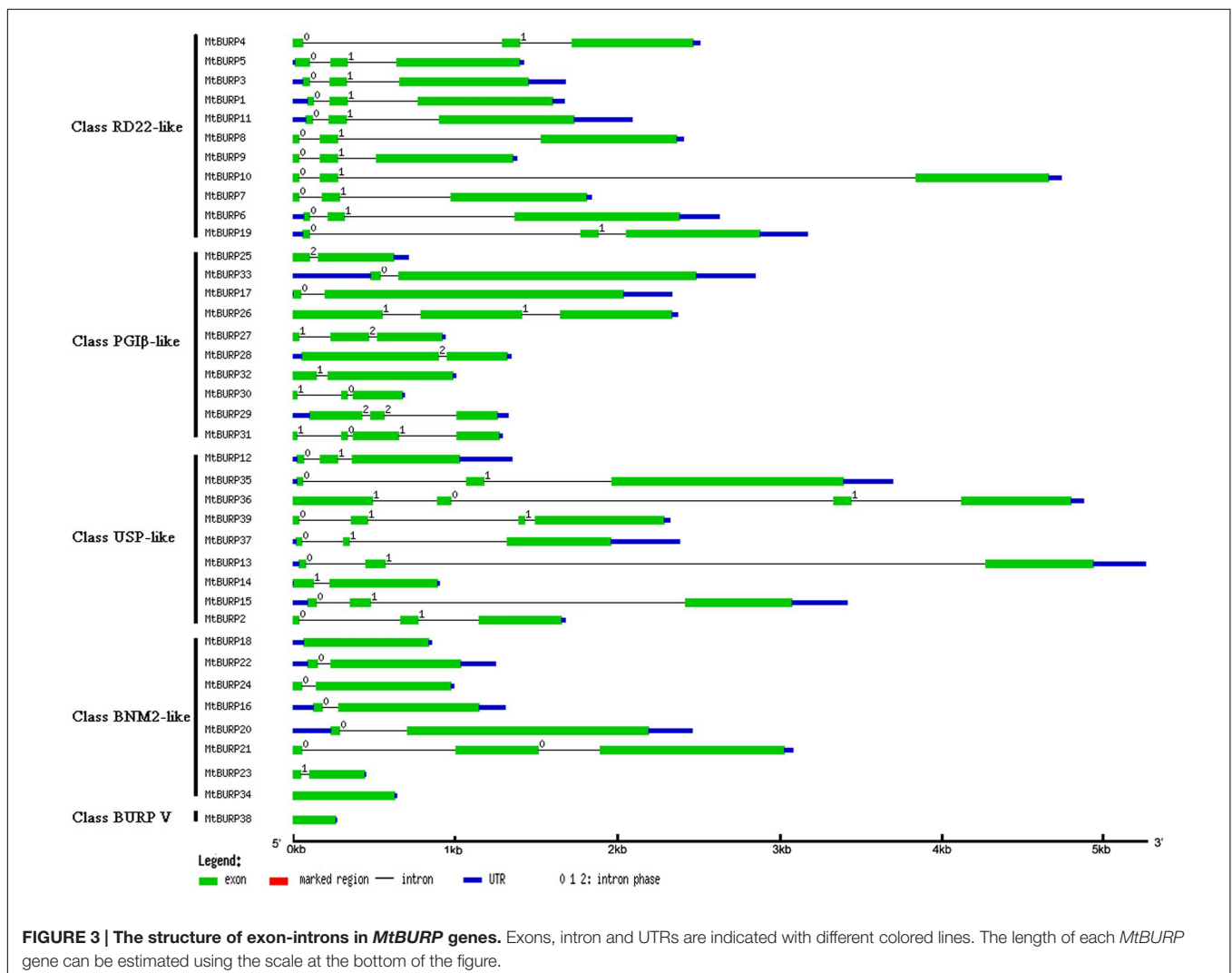
is disrupted by one or more exons (Figure 3), indicating that the BURP protein diversity. The number of exons in each gene varies widely among the *MtBURP* genes. The largest number is 4, such as *MtBURP31*, -36 and -39. By contrast, the *MtBURP38*, -18 and -34 contain only one exon.

In preliminary analysis, gene duplication and divergence events were found to be the main contributors to evolutionary momentum. A gene duplication event is commonly defined by the following criteria (Gu et al., 1998; Yang et al., 2008): (1) the alignment sequence covers  $\geq 70\%$  of the longer gene; (2) the similarity of the aligned regions  $\geq 70\%$ . Among the 39 *MtBURP* genes, the authors found that six gene pairs (*MtBURP4/MtBURP5*, *MtBURP1/MtBURP11*, *MtBURP28/MtBURP32*, *MtBURP13/MtBURP14*, *MtBURP18/MtBURP22*, and *MtBURP20/MtBURP21*) participated in gene duplication events.

### Strong Purifying Selection for BURP Genes in *Medicago*

Assuming that synonymous silent substitutions per site (Ks) occur at a constant rate over time, the dates of the large-scale

duplication events can be estimated (Maher et al., 2006). To explore the divergence fate of different duplicated BURP genes, the Ka (non-synonymous substitution rate) and the Ks (synonymous substitution rate) and the Ka/Ks ratios were calculated for each duplicated *MtBURP* gene pair. Generally, Ka/Ks ratios  $< 1$  are indicative of functional constraints with negative or purifying selection of the genes; a ratio = 1 indicates neutral selection and ratios  $> 1$  indicates accelerated evolution with positive selection (Juretic et al., 2005). Our study showed that the Ka/Ks ratios from 6 *Medicago* BURP duplicated gene pairs (Table 2) were less than 0.6 except *MtBURP4/MtBURP5*. This result suggests that the *Medicago* BURP gene family has evolved mainly under the influence of strong purifying selection pressure, with limited functional divergence occurring after duplication, and the BURP genes are slowly evolving at the protein level. Based on the divergence rate of  $6.1 \times 10^{-9}$  synonymous mutations per synonymous site per year as previously proposed for Legumes, the duplication that gave rise to the six duplication gene pairs were estimated to have occurred between 5.38 and 31.77 million years ago (Mya) (Table 2).



**TABLE 2 | Divergence between paralogous BURP gene pairs in *Medicago*.**

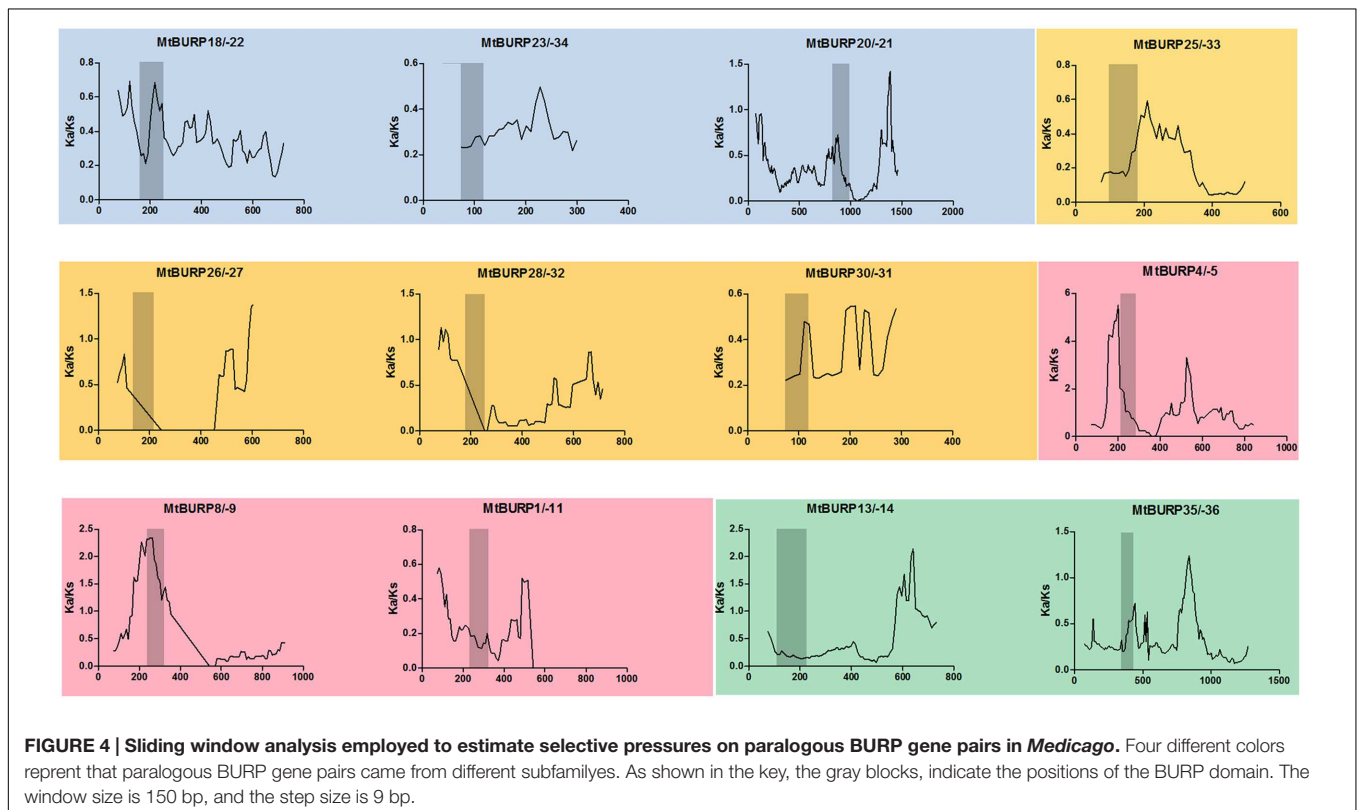
Group	No.	Paralogous pairs	Ka	Ks	Ka/Ks	Duplication date (MY)	Duplicate Type
RD22	1	MtBURP1-MtBURP11	0.017	0.066	0.2599	5.385245902	T
	2	MtBURP4-MtBURP5	0.084	0.136	0.6166	11.13114754	T
BNM2	4	MtBURP18-MtBURP22	0.101	0.296	0.3414	24.2704918	T
	5	MtBURP20-MtBURP21	0.124	0.346	0.3578	28.3442623	T
PG1 $\beta$	9	MtBURP28-MtBURP32	0.042	0.08	0.5267	6.581967213	T
USP	12	MtBURP13-MtBURP14	0.109	0.388	0.2804	31.7704918	T

To better understand the evolutionary constraints acting on this gene family, we compared all Ka/Ks values for 12 unambiguous pairs of BURP paralogs (including six duplicated gene pairs), since positive selection at a few individual codon sites can be masked by overall strong purifying selection, we performed a sliding-window analysis of Ka/Ks between each pair of BURP paralogs. As expected from the basic Ka/Ks analysis, sliding window analysis clearly showed that numerous sites/regions are under moderate to strong negative selection (Figure 4). As shown in Figure 4, the conserved domains of BURP, described as VXCHXCHX23-27CHX23-26CHX8W regions, are mainly subjected to strong purifying selection, with Ka/Ks ratios <1. Moreover, the domains of BURPs generally had lower Ka/Ks ratios (valleys) than the regions outside of them (peaks). There were a few exceptions to the generally low Ka/Ks ratios in domains. For instance, the gene pairs *MtBURP4*-5 and *MtBURP8*-9, revealed sites with Ka/Ks ratios >1 in the BURP domain, indicating positive selection in this region.

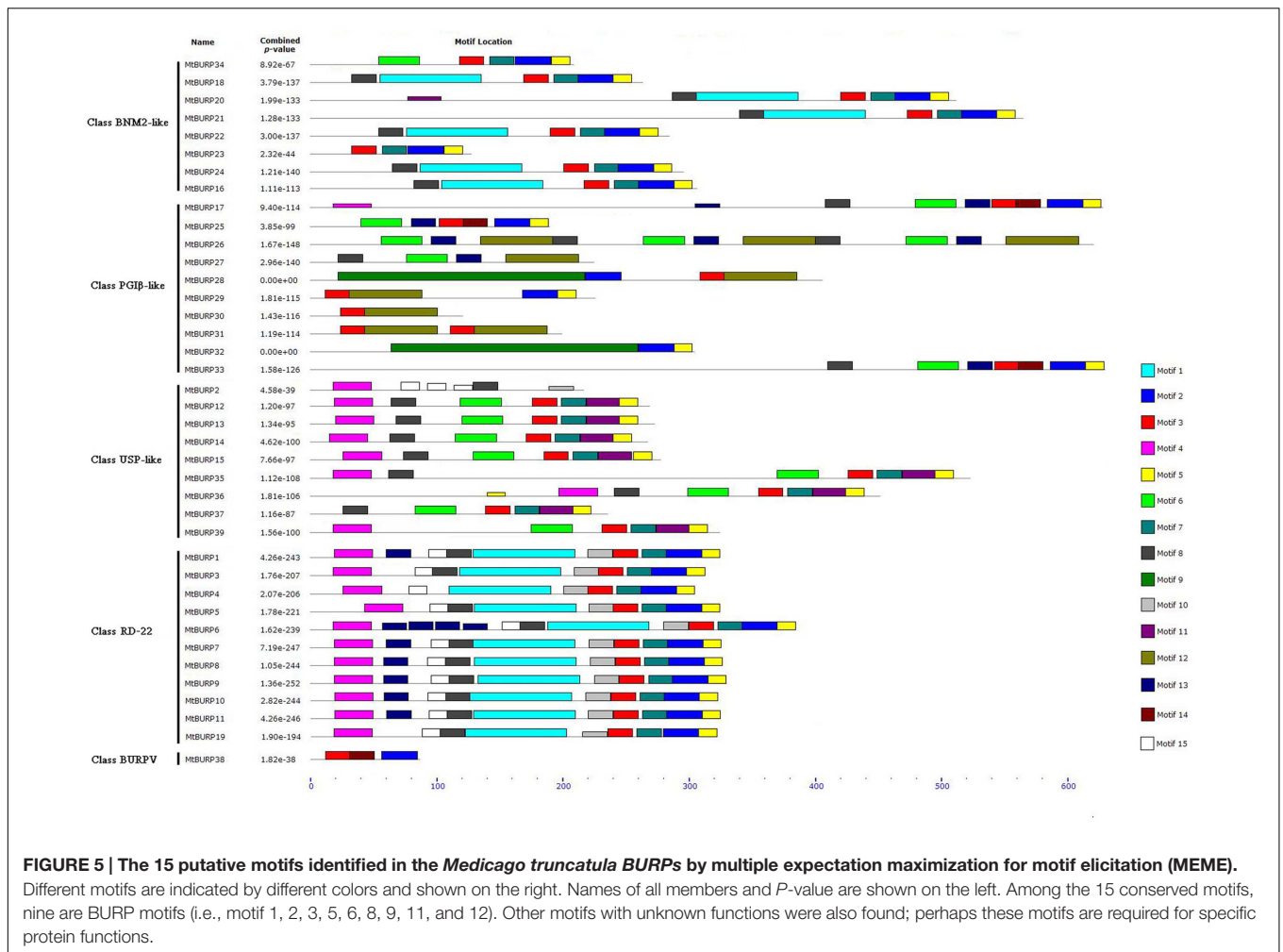
Divergence fate was also found between only paralogous pairs and duplicated gene pairs. While they all have the common characteristic, the Ka/Ks ratios of the conserved domains of BURP in duplicated gene pairs were all lower than paralogous pairs in the same subfamily. For example, *MtBURP14*-13 with ka/ks <0.5, while the same subfamily gene pairs, *MtBURP35*-36, the Ka/Ks value >0.5.

### Sequence Analysis of *MtBURP* Genes

To determine whether the *MtBURP* proteins contain signal peptides, the putative *MtBURP* protein sequences were searched in the SMART database. Twenty-five *MtBURP* proteins contain signal peptides, and their predicted cleavage sites vary. Fifteen conserved motifs (Figure 5) were identified in the *MtBURP* proteins using the MEME web server. Each putative motif obtained from MEME was annotated by searching the Pfam and SMART databases. Among the fifteen conserved motifs that were identified, nine are BURP motifs (i.e., motif 1, 2, 3, 5, 6,







8, 9, 11, and 12). Other motifs with unknown functions were also found; perhaps these motifs are required for specific protein functions. Detailed information about conserved amino acid sequences and lengths of the 15 motifs are shown in **Table 3**. Fifteen motif logo were showed as **Supplementary Figure S1**. Multiple protein sequence alignments were also performed with Clustal X (**Figure 6**). The alignment revealed the existence of several highly conserved amino acid residues, including two phenylalanines, one leucine, three glycines, two prolines, one cysteine, one serine, one valine, one threonine, one alanine and four CH motifs. Some proteins have low homology to the others, such as *MtBURP30* and *-32*. *MtBURP30* contains no amino acid residues. *MtBURP32*, has just one proline and one CH motif. In summary, the C-terminus of the *MtBURP* proteins can be described as VXCHX<sub>11</sub>CHX<sub>25</sub>AXCHX<sub>2</sub>TX<sub>16</sub>PGX<sub>7</sub>CH.

## Identification of Putative Stress-Related *cis*-Elements in the Promoters of the *MtBURP* Genes

To elucidate the possible regulatory functions of *MtBURP* genes in various stress responses, the authors identified putative stress-related *cis*-elements in the 2-kb promoter regions upstream of

the transcription start coden (ATG) of stress-responsive *MtBURP* genes (**Figure 7**). Three types of *cis*-elements, which are directly related to the low temperature-responsive element (LTRE) (Yamaguchi-Shinozaki and Shinozaki, 2005), ABA-responsive element (ABRE) (Narusaka et al., 2003; Yamaguchi-Shinozaki and Shinozaki, 2005) or dehydration-responsive element (DRE) (Narusaka et al., 2003), were detected by searching the promoter sequences against the PLACE database. All *MtBURP* genes from each group were analyzed. Surprisingly, the authors found that 39 *MtBURP* genes contained at least one putative ABRE, LTRE or DRE in their promoter regions. Moreover, some members contain multiple ABREs, LTREs and DREs in their promoter regions, such as *MtBURP2*, *-6* and *-19*. In addition, some members, such as *MtBURP16* and *-32*, have only one putative stress-related *cis*-element. Among the stress-related *cis*-elements, ABREs are more common than LTRE or DRE, which indicates that *MtBURP* genes may play an important role in regulating the level of ABA in the plant. As a few BURP family genes (such as RD22) have been reported for their responsiveness to drought stresses, the transcript levels of all the *MtBURP* genes were investigated in leaves under drought treatment in this study to obtain a complete

**TABLE 3 | Multiple expectation maximization for motif elicitation (MEME) motif sequences in predicted *MtBURP* proteins.**

Motif	Width	Conserved amino acid sequences
1	80	QRTPRNGAKFWPREVANSIPFSSNKLENILNYFSIKQGSPEAEIMKRTISECEANPIKGEEKLCVTSLESMDVFTTSLKLG
2	28	AVCHLDTSQWNPNHFAFQVLGVQPGTIP
3	19	NKWWMCHRENYPYAVFYCH
4	30	THAMLPPELYWKSMPWNTMPKAITDLLHP
5	14	VCHWIPQDHMMVWP
6	32	CCERPPARGETKYCATSLESMDFAISKLRN
7	19	TRVYMPLEGGVDGTRVKAI
8	19	NVALFFREKDLHHGTKMNM
9	195	GSPNGEHRHYEFGPGGQEKYADGEDSPNGEHKYEFGPSAQEKYTDDEGSPNGEHKYEFPGGAQENYTDGEGSPNGE HKYEFPGTQDKYTDGKGSYPYGEHSTNSDLAVPMGILLLSALSEFVLMFPIRTAFADILLLSARSEFVLMFPIRTAFADILH QEMFPFMLLYFCHYIPMGRFYEVEILDQKIMINQA
10	19	NESGLQQYVIKGVKKLGE
11	26	VCHHDTRGMNPDMLYELKVDPGTIP
12	57	HYIPMARFYEVEVLDLQKIMINQAVDVCHLNTSSWSRNHPTFLELGSTPGEIEWCHW
13	19	VDGSKGGVMVGRRKGYEGG
14	19	YIPMVRVYQVEILDQRIM
15	14	YNYAASETQLHDKQ

view of the stress responsiveness in the *Medicago* BURP family.

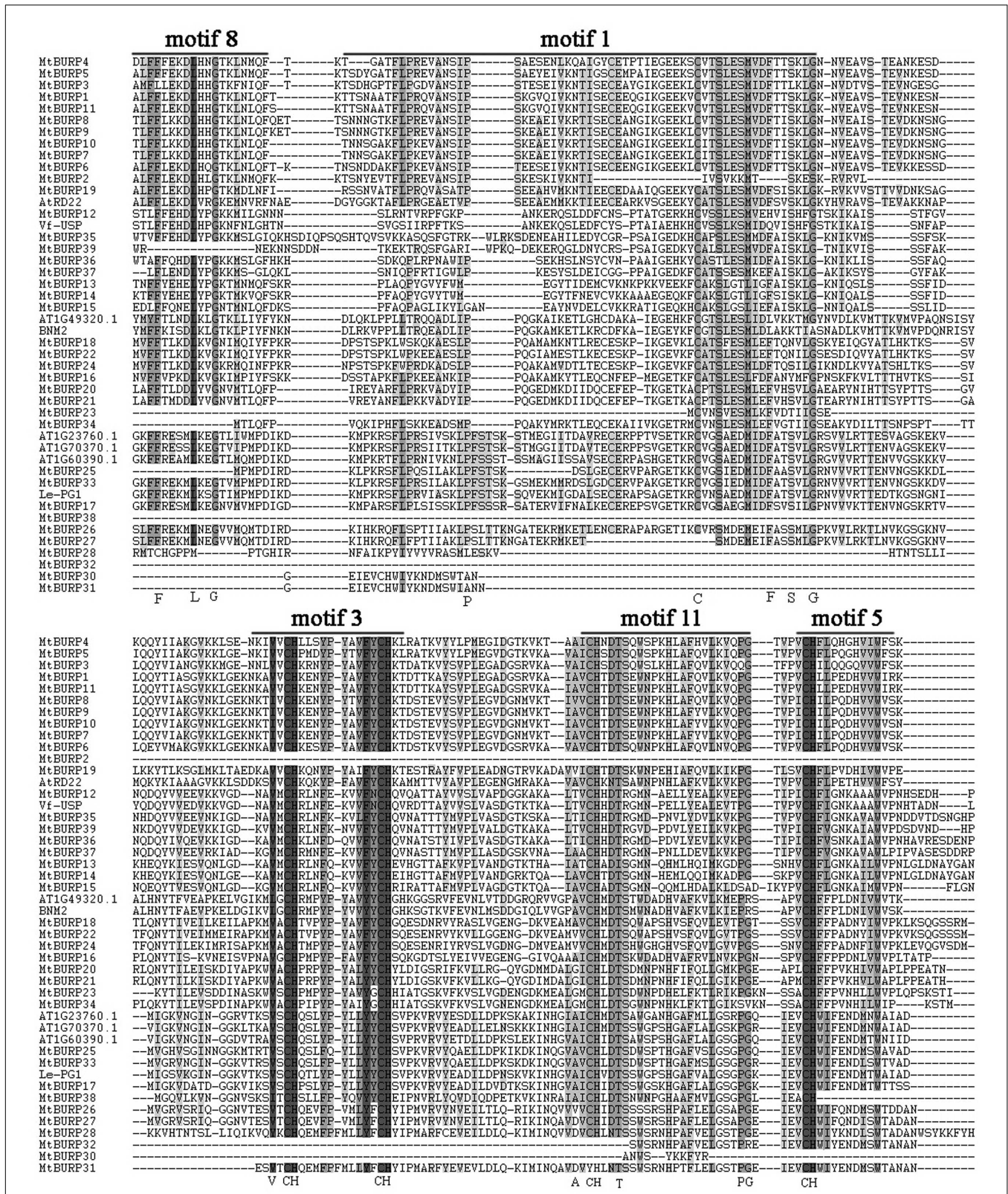
## Microarray Analysis

To gain insight into the expression patterns of *Medicago* BURP genes in various tissues, we searched the Gene Expression Atlas (MtGEA) of *Medicago*, this atlas provides high-resolution gene expression data from 4 diverse tissues, including aerial tissues (leaf, flower, stem), underground tissues (root). Because the expression profiles of 13 BURP genes (*MtBURP12*, *-15*, *-19*, *-21*, *-23*, *-28*, *-29*, *-30*, *-31*, *-32*, *-34*, *-38*, *-39*) weren't obtained in the database, we only examined the expression patterns of 26 BURP genes (Figure 8). Most *Medicago* BURP genes exhibit broad expression patterns (Figure 8). 26 *Medicago* BURPs are expressed in all of the four tissues (leaves, flowers, roots, stems). The heat map also revealed that the majority of *MtBURPs* showed preferential expression. Based on a hierarchical clustering analysis, 26 BURP genes were mainly clustered into four groups (A–D) (Figure 8). Group A showed partial expression in roots, group B in flowers, group C in stems and group D in leaves. 15 *MtBURPs* (*MtBURP16*, *-24*, *-36*, *-37*, *-13*, *-20*, *-35*, *-26*, *-27*, *-22*, *-18*, *-17*, *-5*, *-1*, and *-11*) showed marked high transcript abundance profiles in only a single tissue. Among the 26 *Medicago* BURP genes examined, six showed the highest transcript accumulation in flowers (*MtBURP3*, *-14*, *-18*, *-22*, *-26*, and *-27*), 7 in roots (*MtBURP13*, *-16*, *-20*, *-24*, *-35*, *-36*, and *-37*), one in stems (*MtBURP17*) and four in leaves (*MtBURP1*, *-11*, *-5*, and *-25*; Figure 7). Genes in different subfamilies have their primary abundant transcripts, for instance, BNM-like in flowers and stems, USP-like in flowers, stems and roots, RD22 in leaves, flowers and stems and PG1 in stems, flowers and leaves (Figure 8). These subfamily specific tissue expression patterns may be closely related to gene functions. The expression patterns of the paralogous pairs were also revealed by heat maps; paralogous pairs with high sequence

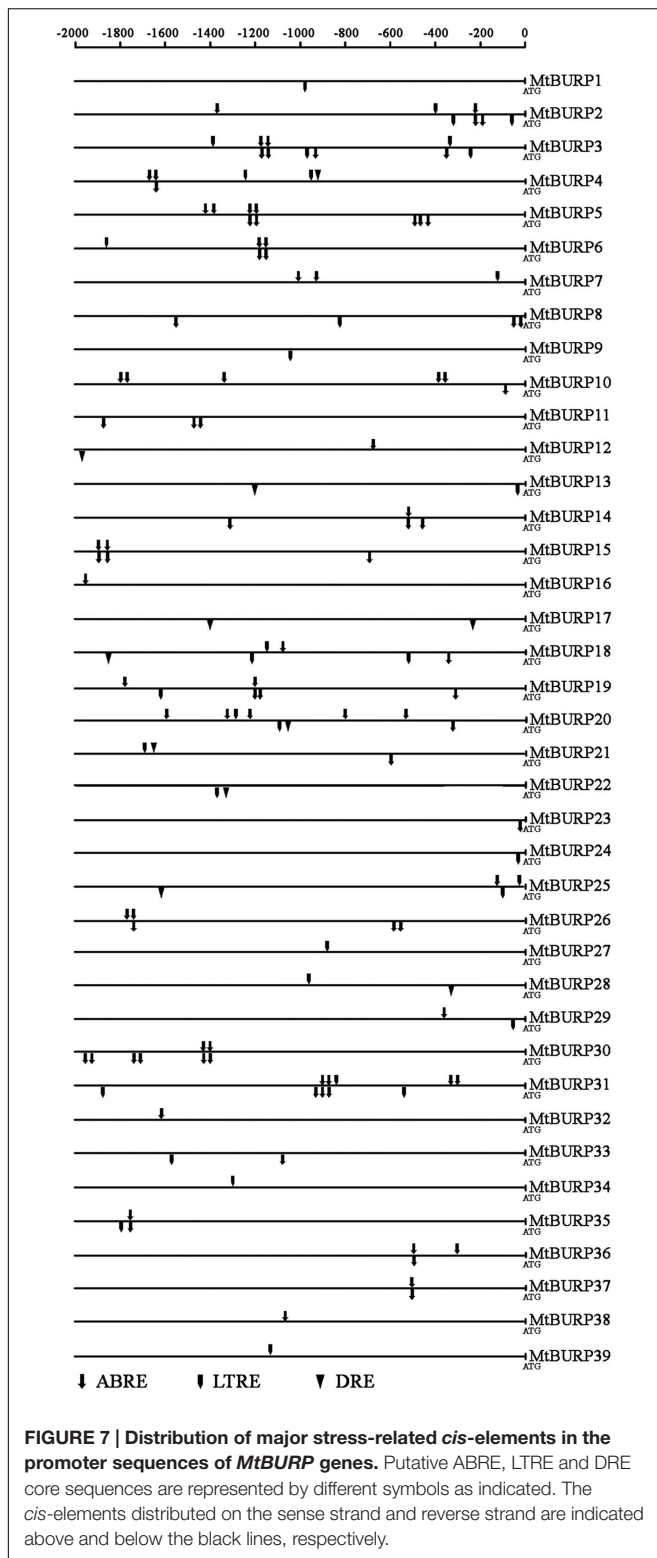
similarity have similar expression patterns. The best examples of this include *MtBURP18*–*22*, *MtBURP26*–*27*, *MtBURP35*–*36* and *MtBURP8*–*9*, which are strongly expressed in leaves, flowers and roots respectively, with little or no expression in other tissues. Expression divergence was also found in paralogous pairs. For example, *MtBURP14* is highly expressed in flowers, while its paralog, *MtBURP13*, is highly expressed in roots.

## Expression Levels of *Medicago* BURP Genes in Response to Drought Stress

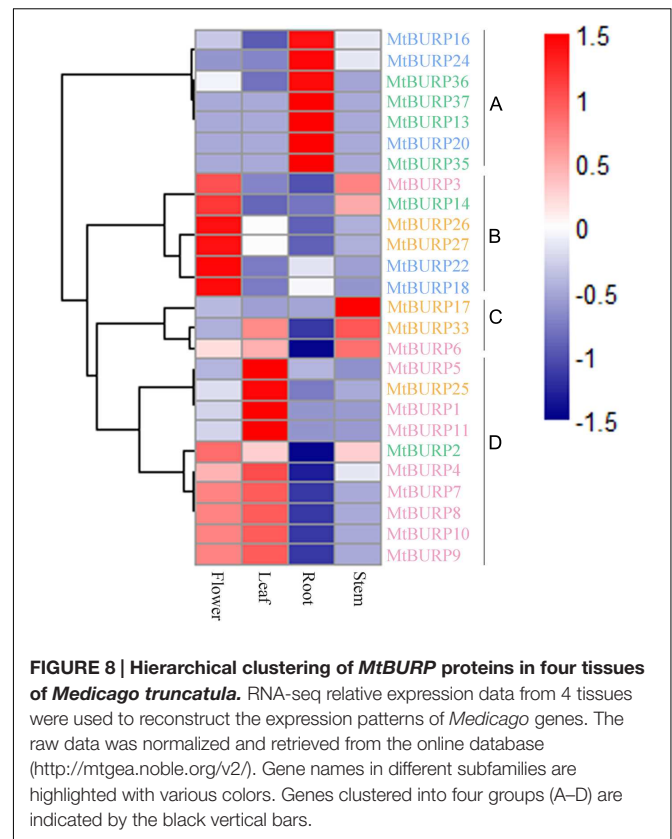
The BURP genes have been reported to affect the plant development, as well as to respond and adapt to stresses. To examine the drought stress responsiveness of the BURP family genes in *Medicago*, the expression patterns of the BURP genes in 3-week-old leaves under drought stress, one of the serious environmental stresses affecting *Medicago* production, was investigated. As shown in Figure 9, the expression levels of all the 39 BURP genes corresponding to early (0–8 h) or late (12 and 24 h) response to drought stress were compared with normal plants, but some differences were observed among these genes. For instance, only two genes (*MtBURP33* and *-28*) both included in PG1 $\beta$  were down-regulated and 10 genes (*MtBURP10*, *-25*, *-26*, *-5*, *-16*, *-19*, *-39*, *-11*, *-15*, and *-2*) were up-regulated by drought treatment through all the time points. Among the 10 up-regulated genes, four genes (*MtBURP5*, *-10*, *-11*, and *-19*) belong to RD22-like, two genes (*MtBURP25* and *-26*) to PG1 $\beta$ , one genes (*MtBURP16*) to BNM2-like and others (*MtBURP2*, *-15* and *-39*) to USP-like. Interestingly, *MtBURP11* from RD22-like subfamily and *MtBURP15* from USP-like subfamily reach the maximum at certain time point respectively, which is different to other up-regulated genes. *MtBURP11* slowly up-regulates in the time range 2–12 h, and reaches the maximum at 24 h, while *MtBURP15* reaches the maximum at 8 h, with a slight up-regulation in other time ranges. Notably, the authors found that every subclass family at least had one gene strongly up-regulated by drought



**FIGURE 6 | Multiple alignments of MtBURP domain-containing proteins.** The complete amino acid sequences of 39 MtBURP proteins and eight representative BURP proteins were aligned using Clustal X. Motif (n) represents the motifs identified in *Medicago truncatula* BURP proteins by MEME. Conserved domains are shaded in black.



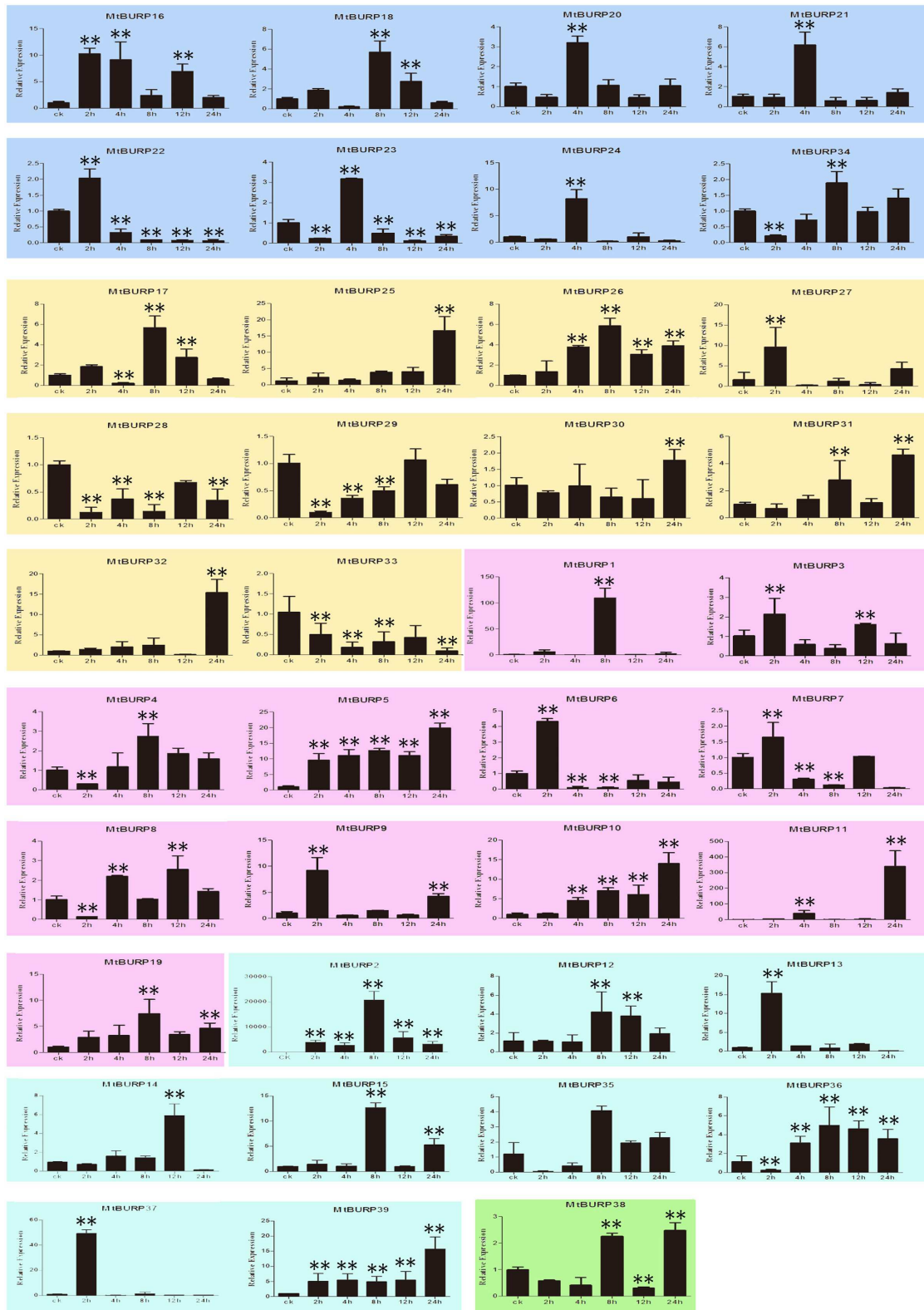
stress except subfamily BURPV. Interestingly, the differential expression patterns were also found among the same subfamily. Besides the above-mentioned 12 genes with regular up-regulation or down-regulation after drought treatment, other 27 genes also



showed responses to drought treatment. As shown in **Figure 9**, it should be noted that *MtBURP31*, -4, -36, -12, -35, and -8 were down-regulated at slight (0–8 h) drought stress but were dramatically up-regulated thereafter. Whereas *MtBURP22*, -6, -23, and -37 were up-regulated at slight drought stress but down-regulated at moderate (12 h) and severe (24 h) drought treatment. The authors also pay more attention to these genes of *MtBURP32*, -24, -20, -21, -14, -13, -1, -17, -27, -18, and -9 which showed major expression changes. They were expression peaked at one time point after drought treatment, and their relative expression were not all the same when they were expression peaked. The lowest is 4 (*MtBURP20*), while the highest is 150 (*MtBURP1*). Other time points showed minor changes might also have specific functions in *Medicago* under drought treatment. The highest expression levels of *MtBURP3* and -7 of RD22-like subfamily were found at 2 and 12 h after treatment, while those of *MtBURP34*, -30, -29, and -38 were observed later, which listed as follow, *MtBURP38* and -34 were found at 8 and 24 h, *MtBURP30* at 24 h and *MtBURP29* at 12 h after treatment.

## DISCUSSION

Preliminary analysis of BURP gene family has been performed in abundant plant as mentioned earlier. However, this family has not previously been studied in *Medicago*. In the present study, an overall analysis of the BURP gene family in *Medicago* was performed, including analysis of their phylogeny, chromosomal



**FIGURE 9 | Expression patterns of the 39 BURP genes under drought stress.** Relative expressions of BURP genes were examined by qRT-PCR and normalized by SEC expression. Different subfamilies were presented in different background color. The Y axis is the scale of the relative expression level. The X axis is the time course of drought stress treatment. Mean value and SDs were obtained from three biological and three technical replicates. Two asterisks (\*\*  $P < 0.01$ , Student's  $t$ -test) represent significant differences between the controls and treatments.

location, gene structure, conserved motifs and expression profiles. A total of 39 non-redundant BURP genes were identified in the *M. truncatula* genome, which is 1.7 times that of Soybean (Xu et al., 2010). This difference might be attributed to the fact that the identification of BURP genes in this study was based on the amino acid sequence using BlastP program, whereas the analysis in the previous report was based on the nucleotide sequence using TblastN searches. Moreover, more and more sequences had been assembled and introduced into the *Medicago* genome database, this might be another reason caused the different number of the BURP genes. The lengths of these sequences varied remarkably, suggesting a highly degree of complexity within the *MtBURP* gene family.

Gene duplication and divergence events are the main processes that multiply genetic material during evolution and selection (Pickett and Meeks-Wagner, 1995). The location of BURP genes in the *Medicago* genome may provide insights into the evolution of this gene family. In *M. truncatula*, 72% of *MtBURP* genes (28 out of 39) from six gene clusters are distributed on *Medicago* chromosomes. These gene clusters may have been derived from common ancestors respectively and may have been formed through a series of gene duplications. In *M. truncatula*, six sister pairs, including twelve BURP genes, have been identified. Gene duplication plays a major role in genomic rearrangement and expansion. Gene duplication events include tandem and segmental duplication events (Xu et al., 2010). Two or more genes located on the same chromosome result from tandem duplication events, while gene duplication between different chromosomes is a segmental duplication event (Xu et al., 2010). Interestingly, for all six sister pairs of *MtBURP* genes, including *MtBURP4/MtBURP5*, *MtBURP1/MtBURP11*, *MtBURP28/MtBURP32*, *MtBURP13/MtBURP14*, *MtBURP18/MtBURP22* and *MtBURP20/MtBURP21*, both genes in a pair are on the same chromosome. In addition, all six sister pairs have undergone tandem duplication; the putative proteins encoded by the genes share more than 80% similarity at the amino acid level. In our analysis, all duplicated BURP gene pairs were involved in tandem duplication events, but there were no segmentally duplicated gene pairs, indicating that tandem duplication rather than segmental duplication served as the most important driving force throughout the long period of *Medicago* BURP gene evolution.

The ratio of non-synonymous versus synonymous substitutions (Ka/Ks) is an indicator of history of selection acting on a gene or gene region (Cristina et al., 1996). The Ka/Ks ratios of the 12 paralogous pairs showed that these gene pairs were subjected to purifying selection. Furthermore, the relatively higher Ka/Ks ratios for the *MtBURP4/-5* and *MtBURP28/-32* gene pairs may suggest that they experienced rapid evolutionary diversification following duplication. The domains of BURPs generally had lower Ka/Ks ratios (valleys) than the regions outside of them (peaks) except the gene pairs *MtBURP4/-5* and *MtBURP8/-9*, the authors suggested that is consistent with functional constraint being dominant in these domains. By calculating the duplication dates for the duplication gene pairs, we concluded that all of the duplication events in the *MtBURP* gene family occurred between 5.38 and 31.77 Mya.

BURP domain-containing proteins are specific to plants (Wang et al., 2011). While this gene family has been identified in many plants, the functions of these genes are often unclear. Previous studies have suggested that BURP genes may play important roles in plant development and metabolism, and some BURP domain-containing proteins are responsive to stress treatment. To elucidate the functions of BURP genes in *M. truncatula*, stress-related *cis*-element analysis of the promoters of the *MtBURP* genes was performed, as well as Quantitative real-time PCR (qRT-PCR) analysis of the all *MtBURP* genes. From the RT-PCR analysis, the authors determined that the expression profiles of the BURP family in *Medicago* were complicated, suggesting that these genes performed a variety of physiological functions to help the plant adapt to environment challenges. The results showed that expression levels of all 39 genes were induced by drought stress, even though the induction of some genes was slight. It is noteworthy that *MtBURP19*, *-5*, *-11*, and *-10* were significantly up-regulated under drought stress. The author concluded that they may play an essential role in response to drought stresses. This conclusion was supported by the close relationship of these genes (*MtBURP19*, *-5*, *-11*, and *-10*) to *AtRD22*, which was previously identified as stress-response gene (Yamaguchi-Shinozaki and Shinozaki, 1993). The other six genes of this subfamily with minor expression changes may also have specific functions in *Medicago* under drought conditions (Xu et al., 2013). As reported, *AtRD22* transcript was largely restricted to the leaf, that of *AtUSPL1* was more prevalent in the root (Harshavardhan et al., 2014). We inferred this phenomenon was also exist in *Medicago*, as showed in **Figure 8**, three up-regulated genes (*MtBURP2*, *-15*, and *-39*) to USP-like, two genes (*MtBURP2*, *-15*) were just up-regulated obviously at 8 h. Two down-regulated genes, *MtBURP33* and *-28*, belong to PG1 $\beta$ -like subfamily may be play positive roles in response to other stress, such as abiotic condition and cold condition. Intriguingly, the results of qRT-PCR were not always consistent with those of the promoter sequence analysis. For example, the promoter analysis showed that *MtBURP2*, *-5*, *-10*, *-16*, *-19*, *-25*, and *-26* have no DRE elements detected in the promoter regions, but the real-time PCR results showed that all these were induced by drought treatment. This strongly indicated that some unidentified stress responsive *cis*-elements might play important roles in regulating the *Medicago* stress response. A comparison of amino acid sequences and regulatory characteristics of the *MtBURP* gene pairs suggests that divergence in regulation evolves more quickly than functional divergence (Li, 1983).

## CONCLUSION

*Medicago truncatula* is a forage grass and a valuable herbal medicine (Crane et al., 2006). This plant is also useful for soil and water erosion control. However, due to inclement weather, the production of *M. truncatula* has been declining. Therefore, numerous stress-responsive genes from various protein families have been studied and used in genetic engineering to improve the stress tolerances of economic plants. In this study, the authors identified and characterized BURP genes in *M. truncatula*, this work should promote further studies of the

functions of stress-induced BURP domain-containing genes in this model species.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: YL, XC. Performed the experiments: YL, XC. Analyzed the data: YL, XC, ZC, RC, HZ. Contributed reagents/materials/analysis tools: YL, XC, ZC. Wrote the paper: YL, XC, ZC. Guided the experiment: YX.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (No. 31370561),

## REFERENCES

Bassüner, R., Bäumlein, H., Huth, A., Jung, R., Wobus, U., Rapoport, T. A., et al. (1988). Abundant embryonic mRNA in field bean (*Vicia faba* L.) codes for a new class of seed proteins: cDNA cloning and characterization of the primary translation product. *Plant Mol. Biol.* 11, 321–334. doi: 10.1007/BF00027389

Batchelor, A. K., Boutilier, K., Miller, S. S., Hattori, J., Bowman, L., Hu, M., et al. (2002). SCB1, a BURP-domain protein gene, from developing soybean seed coats. *Planta* 215, 523–532. doi: 10.1007/s00425-002-0798-1

Boutilier, K. A., Ginés, M. -J., Demoor, J. M., Huang, B., Baszczynski, C. L., Iyer, V., et al. (1994). Expression of the BnmNAP subfamily of napin genes coincides with the induction of *Brassica* microspore embryogenesis. *Plant Mol. Biol.* 26, 1711–1723. doi: 10.1007/BF00019486

Chen, L., Miyazaki, C., Kojimai, A., Saito, A., and Adachi, T. (1999). Isolation and characterization of a gene expressed during early embryo sac development in apomictic guinea grass (*Panicum maximum*). *J. Plant Physiol.* 154, 55–62. doi: 10.1016/S0176-1617(99)80318-6

Chesnokov, Y. V., Meister, A., and Manteuffel, R. (2002). A chimeric green fluorescent protein gene as an embryogenic marker in transgenic cell culture of *Nicotiana plumbaginifolia* Viv. *Plant Sci.* 162, 59–77. doi: 10.1016/S0168-9452(01)00532-5

Crane, C., Wright, E., Dixon, R. A., and Wang, Z. -Y. (2006). Transgenic *Medicago truncatula* plants obtained from *Agrobacterium tumefaciens*-transformed roots and *Agrobacterium rhizogenes*-transformed hairy roots. *Planta* 223, 1344–1354. doi: 10.1007/s00425-006-0268-2

Cristina, M., Sessa, G., Dolan, L., Linstead, P., Baima, S., Ruberti, I., et al. (1996). The *Arabidopsis* Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development. *Plant J.* 10, 393–402. doi: 10.1046/j.1365-313X.1996.10030393.x

Ding, X., Hou, X., Xie, K., and Xiong, L. (2009). Genome-wide identification of BURP domain-containing genes in rice reveals a gene family with diverse structures and responses to abiotic stresses. *Planta* 230, 149–163. doi: 10.1007/s00425-009-0929-z

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., and Kiss, G. B. (2002). A receptor kinase gene regulating symbiotic nodule development. *Nature* 417, 962–966. doi: 10.1038/nature00842

Fernandez, L., Torregrosa, L., Terrier, N., Sreekantan, L., Grimplet, J., Davies, C., et al. (2007). Identification of genes associated with flesh morphogenesis during grapevine fruit development. *Plant Mol. Biol.* 63, 307–323. doi: 10.1007/s11103-006-9090-2

Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., et al. (2006). Pfam: clans, web tools and services. *Nucleic Acids Res.* 34, D247–D251. doi: 10.1093/nar/gkj149

Gan, D., Jiang, H., Zhang, J., Zhao, Y., Zhu, S., and Cheng, B. (2011). Genome-wide analysis of BURP domain-containing genes in Maize and Sorghum. *Mol. Biol. Rep.* 38, 4553–4563. doi: 10.1007/s11033-010-0587-z

Granger, C., Coryell, V., Khanna, A., Keim, P., Vodkin, L., and Shoemaker, R. C. (2002). Identification, structure, and differential expression of members of a

Specialized research Fund for the Doctoral Program of Higher Education (No.20133418110005), Anhui Provincial Natural Science Foundation (No. 1308085MC36) and Anhui Agricultural University disciplinary construction Foundation (No.XKTS2013001).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00485>

**FIGURE S1 | Fifteen motif logo were created by online MEME.**

**TABLE S1 | A list of primer sequences of the 39 selected MtBURP genes for qRT-PCR analysis.**

BURP domain containing protein family in soybean. *Genome* 45, 693–701. doi: 10.1139/g02-032

Gu, Q., Ferrándiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125, 1509–1517.

Guo, A., Zhu, Q., Chen, X., and Luo, J. (2007). [GSDS: a gene structure display server]. *Yi Chuan* 29, 1023–1026. doi: 10.1360/yc-007-1023

Harshvardhan, V. T., Seiler, C., Junker, A., Weigelt-Fischer, K., Klukas, C., Altmann, T., et al. (2014). AtRD22 and AtUSPL1, members of the plant-specific BURP domain family involved in *Arabidopsis thaliana* drought tolerance. *PLoS ONE* 9:e110065. doi: 10.1371/journal.pone.0110065

Hattori, J., Boutilier, K., Campagne, M. L., and Miki, B. (1998). A conserved BURP domain defines a novel group of plant proteins with unusual primary structures. *Mol. Gen. Genet. MGG* 259, 424–428. doi: 10.1007/s004380050832

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297

Holub, E. B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2, 516–527. doi: 10.1038/35080508

Hu, R., Qi, G., Kong, Y., Kong, D., Gao, Q., and Zhou, G. (2010). Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. *BMC Plant Biol.* 10:145. doi: 10.1186/1471-2229-10-145

Juretic, N., Hoen, D. R., Huynh, M. L., Harrison, P. M., and Bureau, T. E. (2005). The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res.* 15, 1292–1297. doi: 10.1101/gr.4064205

Kuppusamy, K. T., Endre, G., Prabhu, R., Penmetsa, R. V., Veereshlingam, H., Cook, D. R., et al. (2004). LIN, a *Medicago truncatula* gene required for nodule differentiation and persistence of rhizobial infections. *Plant Physiol.* 136, 3682–3691. doi: 10.1104/pp.104.045575

Li, W. -H. (1983). “Evolution of duplicate genes and pseudogenes,” in *Evolution of Genes and Proteins*, eds M. Nei and R. K. Koeh (Sunderland, MA: Sinauer), 14–37.

Liu, H., Ma, Y., Chen, N., Guo, S., Liu, H., Guo, X., et al. (2014). Overexpression of stress-inducible OsBURP16, the  $\beta$  subunit of polygalacturonase 1, decreases pectin content and cell adhesion and increases abiotic stress sensitivity in rice. *Plant Cell Environ.* 37, 1144–1158. doi: 10.1111/pce.12223

Maier, C., Stein, L., and Ware, D. (2006). Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res.* 16, 510–519. doi: 10.1101/gr.4680506

Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol.* 140, 411–432. doi: 10.1104/pp.105.073783

Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., et al. (2003). Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydration and high-salinity stresses. *Plant J.* 34, 137–148. doi: 10.1046/j.1365-313X.2003.01708.x

- Pickett, F. B., and Meeks-Wagner, D. R. (1995). Seeing double: appreciating genetic redundancy. *Plant Cell* 7, 1347–1356. doi: 10.1105/tpc.7.9.1347
- Ragland, M., and Soliman, K. (1997). Sali5-4a and Sali3-2, two genes induced by aluminum in soybean roots. *Plant Physiol.* 114, 395.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. doi: 10.1186/1471-2105-12-77
- Shao, Y., Wei, G., Wang, L., Dong, Q., Zhao, Y., Chen, B., et al. (2011). Genome-wide analysis of BURP domain-containing genes in *Populus trichocarpa*. *J. Integr. Plant Biol.* 53, 743–755. doi: 10.1111/j.1744-7909.2011.01068.x
- Shunwu, Y., Zhang, L., Zuo, K., Li, Z., and Tang, K. (2004). Isolation and characterization of a BURP domain-containing gene BnBDC1 from *Brassica napus* involved in abiotic and biotic stress. *Physiol. Plant.* 122, 210–218. doi: 10.1111/j.1399-3054.2004.00391.x
- Stacey, G., Libault, M., Brechenmacher, L., Wan, J., and May, G. D. (2006). Genetics and functional genomics of legume nodulation. *Curr. Opin. Plant Biol.* 9, 110–121. doi: 10.1016/j.pbi.2006.01.005
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Tang, Y., Cao, Y., Qiu, J., Gao, Z., Ou, Z., Wang, Y., et al. (2014). Expression of a vacuole-localized BURP-domain protein from soybean (SALI3-2) enhances tolerance to cadmium and copper stresses. *PLoS ONE* 9:e98830. doi: 10.1371/journal.pone.0098830
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876
- Tiedemann, J., Rutten, T., Hillmer, S., Hinz, G., Zank, T., Manteuffel, R., et al. (2009). The BURP domain protein AtUSPL1 of *Arabidopsis thaliana* is destined to the protein storage vacuoles and overexpression of the cognate gene distorts seed development. *Plant Mol. Biol.* 71, 319–329. doi: 10.1007/s11103-009-9526-6
- Treacy, B. K., Hattori, J., Prud'homme, I., Barbour, E., Boutillier, K., Baszczynski, C. L., et al. (1997). Bnm1, a *Brassica* pollen-specific gene. *Plant Mol. Biol.* 34, 603–611. doi: 10.1023/A:1005851801107
- Urao, T., Yamaguchi-Shinozaki, K., Urao, S., and Shinozaki, K. (1993). An *Arabidopsis* myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell* 5, 1529–1539. doi: 10.1105/tpc.5.11.1529
- Wang, A., Xia, Q., Xie, W., Datla, R., and Selvaraj, G. (2003). The classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14487–14492. doi: 10.1073/pnas.2231254100
- Wang, D., Pei, K., Fu, Y., Sun, Z., Li, S., Liu, H., et al. (2007). Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene* 394, 13–24. doi: 10.1016/j.gene.2007.01.006
- Wang, P., Su, R., Zheng, J., Cheng, S., and Zhu, G. (2011). Molecular cloning and phylogenetic analysis of a novel BURP domain-containing gene from *Camellia sinensis*. *Afr. J. Biotech.* 10, 15470–15476. doi: 10.5897/AJB11.422
- Xu, B., Gou, J. Y., Li, F. G., Shanguan, X. X., Zhao, B., Yang, C. Q., et al. (2013). A cotton BURP domain protein interacts with -expansin and their co-expression promotes plant growth and fruit production. *Mol. Plant* 6, 945–958. doi: 10.1093/mp/sss112
- Xu, H., Li, Y., Yan, Y., Wang, K., Gao, Y., and Hu, Y. (2010). Genome-scale identification of Soybean BURP domain-containing genes and their expression under stress treatments. *BMC Plant Biol.* 10:197. doi: 10.1186/1471-2229-10-197
- Yamaguchi-Shinozaki, K., and Shinozaki, K. (1993). The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in *Arabidopsis thaliana*. *Mol. Gen. Genet.* MGG 238, 17–25.
- Yamaguchi-Shinozaki, K., and Shinozaki, K. (2005). Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci.* 10, 88–94. doi: 10.1016/j.tplants.2004.12.012
- Yang, S., Zhang, X., Yue, J. X., Tian, D., and Chen, J. Q. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* 280, 187–198. doi: 10.1007/s00438-008-0355-0
- Zheng, L., Heupel, R. C., and Dellapenna, D. (1992). The beta subunit of tomato fruit polygalacturonase isoenzyme 1: isolation, characterization, and identification of unique structural features. *Plant Cell* 4, 1147–1156. doi: 10.1105/tpc.4.9.1147
- Zhou, M., Wu, L., Liang, J., Shen, C., and Lin, J. (2012). Cold-induced modulation of CbICE53 gene activates endogenous genes to enhance acclimation in transgenic tobacco. *Mol. Breed.* 30, 1611–1620. doi: 10.1007/s11032-012-9744-5
- Zhou, M., Xu, M., Wu, L., Shen, C., Ma, H., and Lin, J. (2014). CbCBF from *Capsella bursa-pastoris* enhances cold tolerance and restrains growth in *Nicotiana tabacum* by antagonizing with gibberellin and affecting cell cycle signaling. *Plant Mol. Biol.* 85, 259–275. doi: 10.1007/s11103-014-0181-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Li, Chen, Chen, Cai, Zhang and Xiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.