Comprehensive identification and characterization of abiotic stress and hormone responsive glycosyl hydrolase family 1 genes in *Medicago truncatula*

Junfeng Yang

Institute of Botany Chinese Academy of Sciences

Lin Ma

Institute of Animal Science, CAAS

Wenbo Jiang

Institute of Animal Science, CAAS

Yu Yao

Institute of medicinal plant development

Yuhong Tang

Noble Research Institute LLC

Yongzhen Pang (✉ pangyongzhen@caas.cn)

Institute of Animal Science <https://orcid.org/0000-0002-2336-7476>

Research article

Keywords: *Medicago truncatula*, BGLU genes, Abiotic stresses, Phytohormones, Defense compounds, Expression profiling

Posted Date: March 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-17102/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Plant Physiology and Biochemistry on January 1st, 2021. See the published version at <https://doi.org/10.1016/j.plaphy.2020.11.046>.

Abstract

Background: β -glucosidases (BGLUs) hydrolyze the β -D-glycosidic bond with retention of anomeric configuration. BGLUs were associated with many aspects of plant physiological processes, in particular biotic and abiotic stresses through the activation of phytohormones and defense compounds. However, no comprehensive and systematic investigation on the stress- or hormone-responsive BGLU proteins had ever been reported in plant.

Results: In this study, total 51 BGLU genes of the glycoside hydrolase family 1 were identified in one of the model legume plant *Medicago truncatula* genome, and they were classified into five distinct clusters. Sequence alignments revealed several conserved and characteristic motifs among these MtBGLU proteins. Analyses of their putative signal peptides and N-glycosylation site suggested that the majority of MtBGLU members have dual targeting to the vacuole/chloroplast. Many regulatory elements possibly related with phytohormones and/or abiotic stresses were identified in MtBGLU genes. Moreover, Microarray and qPCR analyses showed that these MtBGLU genes exhibited distinct expression patterns in various tissues, and in response to different abiotic stress and hormonal treatments. Notably, MtBGLU21, MtBGLU22, MtBGLU28, and MtBGLU30 in cluster I were dramatically activated by NaCl, PEG, IAA, ABA, SA and GA3 treatments.

Conclusion: Collectively, our genome-wide characterization, evolutionary analysis, and expression pattern analysis of MtBGLU genes suggested that BGLU proteins play crucial roles in response to various abiotic stresses and hormonal cues in *M. truncatula*. This systematic analysis provided valuable information for the functional characterization and utilization of these MtBGLU genes in improving stress tolerance in *M. truncatula* and/or other plant species.

Background

Plants are sessile organisms that can not escape from the adverse circumstances as well as herbivore or pathogen attack. As a result, plants have evolved to protect themselves and adapted to various biotic and abiotic stresses by collectively synthesizing a plethora of defense compounds [1]. Many of these compounds are in glucosylated forms, such as cyanogenic-, isoflavone-, and hydroxamic acid-glucosides. Glucosylation could increase the solubility and stability of these compounds, so as to make them suitable for storage in the vacuole or organelle, and further protect plants from deleterious effects on its own defense system [2, 3]. These glucosylated compounds are activated by a group of enzymes called β -glucosidases (EC 3.2.1.21), which belong to the glycoside hydrolase (GH) family with 166 subfamilies. Among them, β -glucosidases catalyze the hydrolysis of terminal, non-reducing β -D-glucosyl residues to release β -D-glucose, and they are categorized into glycoside hydrolase families GH1, GH3, GH5 or GH116 [4]. Among these four subfamilies, GH1 is the largest one and most plant β -glucosidases have been characterized from GH1 so far.

Hydrolysis of the β -D-glycosidic bond by β -glucosidases (BGLUs) of GH1 involves two steps (glycosylation and deglycosylation) with retention of anomeric configuration. This process requires participation of two well-conserved active glutamate residues that located in the characteristic and highly conserved peptide motifs TF/LNEP (acid/base catalyst) and I/VTENG (nucleophilic) [5]. These two glutamate residues are considered to be the key amino acids for β -glucosidase [6, 7]. Furthermore, it has been shown that most BGLU enzymes are targeted to plastid, vacuole, cytosol or apoplast, and they are accumulated mainly in seedlings and young tissues of plants [8, 9].

Several BGLU enzymes were elucidated to be associated with diverse biological functions, including cell wall remodeling, lignification and plant secondary metabolites [10, 11]. Os4BGLu12 from *Oryza sativa* has been shown to hydrolyze cell wall β -glucan-derived oligosaccharides during development or wounding [12]. In *Arabidopsis thaliana*, two stem-specific β -glucosidases BGLU45 and BGLU46 involved in the hydrolysis of monolignol glucoside coniferin, which is the storage form of monolignols. Glucoside coniferin in *Arabidopsis* could be used when lignins need to be synthesized de novo resistance to various stresses [13]. In *Glycine max*, the release of free isoflavones from their conjugates is catalyzed by GmICHG, a BGLU enzyme exclusively localized in the cell wall of root hairs [2].

More importantly, substantial evidences also indicate that BGLU genes play important roles in the activation of defense chemicals response to stresses, and in the release of active phytohormones [14]. For example, AtBGLU26 has been demonstrated to be important to plant defense against broad-spectrum fungi in *Arabidopsis* [15]. AtBGLU23 from *Arabidopsis* is an ER body-localized β -glucosidase, which is implicated to be involved in the establishment of symbiosis with the endophytic fungus *Piriformospora indica*, by preventing it from overgrowing in roots and triggering a defense response [16]. In white spruce, expression of Pgbglu-1 led to the accumulation of acetophenone aglycons, which was proved to be a constitutive defense mechanism against spruce budworm [17]. In *Arabidopsis*, the production of active ABA from ABA-GE under abiotic stress conditions were catalyzed by β -glucosidases AtBGLU18 and AtBGLU33 [18, 19]. In addition, a few BGLUs in rice have also been demonstrated to hydrolyze salicylic acid glucoside [20], and gibberellic acid glucose ester [21].

In addition to their hydrolysis activity, a few BGLUs also showed transglucosidase activities in anthocyanin synthesis in carnation or delphinium [22], which indicates that BGLU also have the function of synthesizing gluco-conjugates in plant. All these studies suggest that BGLUs activated storage forms of glyco-conjugates against various stresses, and the glycosylation and deglycosylation process provides an efficient system to modulate the homeostasis of plant metabolism compounds.

As the genome of many plant species has been sequenced, the GH1 family members have been identified in plants, such as 47 BGLU genes in *A. thaliana* [23], 40 in *O. sativa* [24] and 26 in *Z. mays* [25]. As a fast-emerging legume model plant, *M. truncatula* showed the advantages of small diploid genome, self-pollination, short growing period and high genetic transformation efficiency. More importantly, the genome of *M. truncatula* has been sequenced [26] and the gene expression atlas (MtGEA) has also been publicly available, which makes the functional genome studies convenient. However, members of the

BGLU family in *M. truncatula* have hardly been investigated and their functions remain unclear, except four MJ-inducible BGLUs that were functionally characterized and shown to be active towards (iso)flavonoid glycosides [27]. It is, therefore, important to take an in-depth understanding of how *M. truncatula* regulates individual BGLU members in response to developmental and environmental signals, especially abiotic stresses and hormones.

In the present study, we identified and characterized all 51 MtBGLU genes from GH1 family and divided them into five main clusters. The comprehensive analyses on these MtBGLUs including sequence features, phylogenetic relationships, gene structures, protein motifs, chromosomal localization, synteny analysis, and cis-acting elements. We also investigate the tissue-specific expression and dynamic expression patterns of MtBGLUs in response to various abiotic stresses and hormones. This study provides valuable clues for further investigations aiming at the functional characterization of MtBGLUs gene family members, and their utilization for the genetic improvement of legume plants for resistance against environmental stresses.

Results

Identification of *BGLU* genes in the *M. truncatula* genome

A total of 51 candidate *MtBGLU* genes were obtained by homology search and domain confirmation, and they were designated as *MtBGLU1* to *MtBGLU51* based on their location on the chromosome, but four of them (*MtBGLU48-51*) were not assigned to any of the eight chromosomes (Table S2).

To avoid potential confusion, information on four previously published *MtBGLUs* are provided in Table S3. Physico-chemical properties of *MtBGLU* genes, including amino acid numbers, molecular weights, signal peptides, isoelectric points, GraV_y, N-gly site and possible subcellular localization were listed in Table 1. The majority of MtBGLU proteins, except 9 MtBGLUs (*MtBGLU5*, 15, 16, 17, 18, 25, 45, 46 and 48), were predicted to have signal peptides ranging from 17 to 33 amino acids in length, which would target them to the secretory pathway. The lengths of the predicted precursor proteins varied from 315 aa (*MtBGLU49*) to 1030 aa (*MtBGLU33*), which correspond to protein molecular weights of 35.8 to 115.7 kDa. The length of mature polypeptides vary from 286 to 617 amino acids, corresponding to MW 32.3 to 70.3 kDa. All MtBGLUs contain one to five N-linked glycosylation sites. Isoelectric points of the predicted proteins ranged from 5.30 to 9.53. The GRAVY ranged from -0.578 to 0, indicating that these MtBGLU proteins are all hydrophilic proteins. The predicted subcellular locations showed that all BGLU proteins from *M. truncatula* were located in the vacuole or chloroplast, except *MtBGLU33* that was located in the extracellular or Golgi.

Multiple sequence alignment, phylogenetic analysis, and classification of *MtBGLU* genes

The phylogenetic relationship of the MtBGLU proteins was examined by multiple sequence alignment analysis, which showed high conservation among each other (Fig. S1). All MtBGLU protein sequences contain the key amino acids involved in enzymatic catalysis: catalytic acid/base (WI/T/VTF/L/VNEP)

and nucleophilic glutamate residues (ITENG), except for MtBGLU49 that either is a non-functional enzyme or has a catalytic mechanism different from all the other MtBGLUs. Sequence-based phylogenetic analysis between *Medicago* and *Arabidopsis* showed that these proteins were grouped into 7 distinct clusters. Among these 51 MtBGLU proteins, 25 belong to cluster I, 8 to cluster III and 6 to clusters II, IV and V, respectively (Fig. 1). Clusters I, III, IV and V contain members from *Medicago* and representatives from *Arabidopsis*, clusters AtI and AtII with members only from *Arabidopsis*, and cluster II with only members from *Medicago*.

Gene structure, conserved domain and motif pattern of *MtBGLU* genes

In order to better understand the evolution of the GH1 family members in *M. truncatula*, we conducted the exon/intron structures of all the identified *MtBGLU* genes. As shown in Fig. 3B, all *MtBGLU* genes possessed 6 to 17 exons (five with 10 or less exons, twelve with 11 to 12 exons, twenty five with 13 exons, and nine with 14 or more exons). For analysis of the BGLU domain, the shortest conserved amino acid length is 223 aa for MtBGLU49, the longest is 503 aa for MtBGLU47, and the average length of conserved amino acid among MtBGLUs is 457 amino acid (Fig. 2B, Fig. S2).

To identify the conservative structure of MtBGLU proteins, 10 motifs were constructed through the MEME motif analysis, it showed that the majority of the MtBGLU proteins (88.2%) contained all these 10 motifs (Fig. 2C). The lengths of these conserved motifs ranged from 13 to 50 amino acids (Fig. S3). However, six MtBGLU proteins lacked the complete combination of the conserved motifs, including MtBGLU25/MtBGLU46 (lacked only motif 4), MtBGLU2 (lacked motif 4 and 6), MtBGLU45 (lacked motif 2 and 4), MtBGLU30 (lacked motif 3, 8, 9 and 10), and MtBGLU49 (lacked motif 3, 6, 7, 8, 9 and 10). All these gene structure and conserved motif analyses demonstrated that the BGLUs of GH1 family in *M. truncatula* are highly conservative.

Chromosomal distribution and synteny analysis of *MtBGLU* genes

The *MtBGLU* genes were unevenly distributed on seven chromosomes (chr1-7) except chromosome 8 (Fig. 3). In addition, one gene (*MtBGLU48*) localized on scaffold0108 and three (*MtBGLU48~ MtBGLU51*) on scaffold0110 (Fig. 3). Around one third of them were present on chromosome 4, and only two or three genes located on chromosome 1, 5 and 6.

By performing MCscan, we defined the tandem duplication and segmental duplication of *MtBGLU* genes in *M. truncatula* genome, which contribute to the formation of gene family during the process of evolution. Five *MtBGLU* gene pairs (*MtBGLU8/9*, *MtBGLU15/16*, *MtBGLU33/34*, *MtBGLU43/44* and *MtBGLU49/50/51*) were identified as tandem duplication, and they were localized on chromosomes 2, 3, 4, 7 and scaffold 0110. In addition to tandem duplication events, two segmental duplication events with four *MtBGLU* genes (*MtBGLU8/26* and *MtBGLU32/35*) located on chromosome 2, 4, 5 were also identified. All above results inferred that both tandem duplication and segmental duplication events played an important driving force for *MtBGLU* gene evolution, and the former played a predominant role.

Furthermore, three comparative syntenic maps of *M. truncatula* associated with two representative plant species *Arabidopsis* (dicot) and rice (monocot) were constructed to illustrate the evolution mechanism of *MtBGLU* gene family (Fig. 4). A total of 4 *MtBGLU* genes showed a syntenic relationship with those in *Arabidopsis* and rice, respectively (Additional file 1). Seven and five orthologous pairs were found between *Arabidopsis* and rice, respectively. *MtBGLU1* and *MtBGLU33* were found to be associated with two collinear gene pairs in *Arabidopsis*, which implied that these two genes may play an important role among GH1 gene family during evolution. Besides, another two syntenic pairs (*MtBGLU1* and *MtBGLU35*) were identified between *M. truncatula* and *A. thaliana/O. sativa*, which indicated that these collinear pairs may have already existed before the divergence of ancestral genes.

In addition, Ka/Ks analysis of the *MtBGLU* gene pairs were analyzed in order to better understand the effect of evolutionary stress on the formation of *MtBGLU* gene family (Additional file 1). All tandem and segmental duplicated *MtBGLU* gene pairs, and the orthologous *MtBGLU* gene pairs had Ka/Ks value of less than 1, and these results clearly indicated that *MtBGLU* genes of the GH1 family might have undergone strong purifying selective pressure during evolution.

Analysis of *cis*-acting elements in the promoter regions of *MtBGLU* genes

To further explore potential regulatory mechanism of *MtBGLU* gene under hormone and stress responses, the 2.0-kbp upstream promoter regions of *MtBGLU* genes were submitted into PlantCARE to scan for the *cis*-acting elements. Two types of *cis*-acting elements, phytohormone responsive and abiotic and biotic stress-responsive elements, were detected and presented in Fig. 5. Firstly, ten hormone-responsive elements were widely presented in their promoter regions, including TGA-element, AuxRR-core (auxin responses), GARE, P-box, TATC-box (gibberellin responses), TGACG-motif, CGTCA-motif (MeJA responses), ABRE (Abscissic acid responses), ERE (ethylene responses) and TCA-element (salicylic acid responses). Seventy-five of these elements were ABRE, ERE, TGACG-motif and CGTCA-motifs, among the phytohormone responsive clusters. Secondly, six abiotic and biotic stress-responsive elements were detected, namely TC-rich repeats, W-box (stress responses), WUN motif (wound responses), MBS (drought inducible), LTR (low-temperature responses) and ARE (anaerobic induction) (Fig. 5). Additionally, many light-responsive regulatory elements and MYB binding sites were found in the promoter regions of majority of the *MtBGLU* genes (Additional file 2). Various *cis*-acting elements identified in the promoters of *MtBGLU* genes implied that they might be involved in the response to various stresses and hormone treatments by participating in distinct regulatory processes.

Analysis of the transcription levels of *MtBGLU* genes from microarray data

BGLUs are known to play an important role in plants' response to environmental stresses [10, 11]. However, few *BGLUs* were documented to be involved in these stresses in *M. truncatula* [27]. In the present study, seven genechip data sets of *M. truncatula* were retrieved from MtGEA Web Server, including salinity, drought, limit N, bacteria and fungus as well as YE, MeJA and NAA treatments. A total of 82 probes corresponding to 36 *MtBGLU* genes (70.6%) were identified. One representative probe for each

gene was selected for expression analysis, and the expression level of the representative probes are relatively close to the average value of several probes (Additional file 3).

Our expression analyses indicated that many *MtBGLU* genes were induced in response to these stresses, especially under NaCl, PEG and NAA treatments (Fig. 6A, B, H and Additional file 3). However, minor activation effects on *MtBGLUs* were detected by limit N, bacteria and fungus treatments (Fig. 6 C, D, E and Additional file 3). In order to further screen the responsive genes under different stress treatments, the average induction times per treatment period (total induction times/number of treatment periods) of *MtBGLU* genes were calculated under different stress treatments (Fig. 7B and Additional file 3). It showed that, among *MtBGLU* family members, the expression levels of six genes (*MtBGLU14*, 21, 22, 26, 28 and 30, in the red box) is greatly up-regulated under these treatments compared with the remaining *MtBGLUs*. Furthermore, we also counted the number of treatments for each *MtBGLUs* that were up-regulated by more than two fold, which is consistent with the above results for the six genes (Fig. 7).

Because of the close relationship between gene expression level and gene function, the expression profiles of *MtBGLUs* in eight tissues (root, shoot, leaf, vegetative bud, stem, petiole, 20-day-old seed, flower and pod) from microarray were investigated (Fig. 8A, B and Additional file 3). It showed that *MtBGLU* genes showed different transcription levels in various tissues. Most *MtBGLU* genes in cluster I had specifically low transcription level in roots; instead, genes in cluster II were highly expressed in all other tissues except roots. *MtBGLUs* in other three clusters (III, VI and V) had distinct expression patterns in different tissues, without obvious consistency. In general, the expression profiling of *MtBGLU* genes varies greatly, suggesting they might be involve in different functions.

Validation of the expression profile of *MtBGLU* genes by qPCR

Among all 51 *MtBGLU* genes, ten of them (*MtBGLU14*, 16, 18, 19, 21, 22, 26, 28, 30 and 34) were highly induced by various stresses, and two of them (*MtBGLU43* and *MtBGLU44*) were relatively highly expressed in various tissues (Fig. 7, 8B), based on available microarray dataset. To valid this result, these 12 representative *MtBGLU* genes were further validated by qPCR analysis.

The expression levels of these 12 representative *MtBGLU* genes were carried out with six another set of tissues: stems, roots, leaves, flowers, pods and seeds (20-day-old). It was revealed that *MtBGLU28*, *MtBGLU30* and *MtBGLU34* were expressed differently in all tissues. *MtBGLU21* and *MtBGLU22* were preferentially expressed in pods, and *MtBGLU16*, 18, 19 and 26 showed the highest transcript abundances in roots. Instead, *MtBGLU43* and *MtBGLU44* were expressed in all tissues except roots.

To confirm the expression changes of these 12 *MtBGLU* genes under various abiotic stresses and hormones, two representative stresses (NaCl and PEG) and four hormones (IAA, ABA, SA, GA3) treatments were mined (Fig. 9, 10 and Fig. S4). Overall, most *MtBGLU* genes were strongly induced by multiple treatments. Under NaCl and PEG stresses, *MtBGLU14*, 19, 21, 22, 26, 28 and 30 were significantly up-regulated to a certain level, which is consistent with the above microarray data (Fig. 7). *MtBGLU18* and *MtBGLU44* were obviously down-regulated during the late treatment stages. However, *MtBGLU16*,

MtBGLU34 and *MtBGLU43* seemed to have no significant response under these stresses before 72 h of NaCl treatment (Fig. 9).

In IAA treatment, the expression levels of six *MtBGLUs* were greatly activated, including *MtBGLU14*, 19, 21, 22, 28 and 30. Interestingly, the expression levels of four genes (*MtBGLU16*, 18, 34 and 43) were only up-regulated after 3 h. However, *MtBGLU26* and *MtBGLU44* experienced no significant increment in response to IAA supplement (Fig. 10). For ABA treatment, the expression levels of *MtBGLU19*, 21, 22, 26, 28, 30 and 34 genes were markedly promoted, whereas ABA showed no evident effect on the expression levels of *MtBGLU14*, 16, 43 or 44 genes. Additionally, *MtBGLU18* was remarkably diminished after 24 h ABA treatment. Under SA treatment, *MtBGLU14*, 19, 21, 22, 28 and 30 genes were dramatically activated, which were opposite to other six genes (*MtBGLU16*, 18, 26, 34, 43 and 44) that were significantly suppressed by SA. With regard to GA3, the expression levels of *MtBGLU19*, 21, 22, 28 and 30 genes were evidently increased, while *MtBGLU16* and *MtBGLU18* genes were inhibited by GA3 treatment to some extent. Besides, GA3 had no significant activating effect on the expression levels of *MtBGLU14*, 34, 43 or 44 genes.

Collectively, qPCR analyses of 12 representatives *MtBGLU* genes during various treatments were strongly paralleled with their expression pattern obtained from genechip data (Fig.7, Fig. S4). qPCRs analyses also confirmed that four *MtBGLU* genes (*MtBGLU21*, 22, 28 and 30) were frequently and strongly activated by various treatments, and they are most likely key *BGLU* genes involved in response to stress and hormone stimuli in *M. truncatula*.

Discussion

β -glucosidases play diverse and important roles in response to developmental and environmental cues in plants, including roles in recycling of cell-wall oligosaccharides, lignification, activation of phytohormones, defense, secondary metabolism [10, 14]. It is worth noting that BGLUs participates in most of the above processes by an effective mechanism that hydrolyze (activate) inactive and stable glucoside compounds to defense plant against various stresses. So far, the genome-wide investigation of GH1 gene families have been carried out predominantly in model plants *A. thaliana*, *O. sativa* and *Z. mays* [23–25]. However, no comprehensive characterization of stress- or hormone-responsive BGLU was reported at gene family level. In the current study, therefore, a search for BGLU genes in the *M. truncatula* genome brought about the identification of 51 BGLU gene members, which were characterized by a set of sequence analyses (phylogenetic tree, gene structure and motifs composition, gene distribution and synteny relationship), as well as gene expression patterns in response to various abiotic/biotic stresses and hormone treatments.

Multi-sequence alignments revealed that all deduced *MtBGLU* protein sequences shared high sequence identity with each other (Fig. S1). Each *MtBGLU* protein sequence, except *MtBGLU49*, contains the glutamate residues in the motifs WI/T/VTF/L/VNEP and ITENG. In addition, the length of the predicted β -glucosidases polypeptide in *M. truncatula* vary from 441 to 640 aa, except *MtBGLU49* (315 aa) and

MtBGLU33 (1030 aa, which has a Triose Phosphate Translocator domain) (Table 1). This is consistent with *A. thaliana*, *O. sativa* and *Z. mays* that vary from 490 to 613 aa, 458 to 647 aa, and 468 to 570 aa, respectively [23–25]. Notably, it is important to distinguish full-length BGLUs, which is a functional enzyme with a typical sequence length approximately 450 aa [25], from pseudoproteins or truncated proteins. For example, three MtBGLUs (Medtr2g022730, Medtr5g069350, Medtr8g038650) contain a Glyco_hydro_1 conservative domain, but they only encode 212, 191 and 162 aa, respectively, which are less than half in size as compared with all other BGLU isozymes. Results from MEME analysis also elucidated that most BGLU family members in either *Medicago* or *Arabidopsis* contain all 10 motifs (Fig. S5). Therefore, all other 50 MtBGLUs (except MtBGLU49) with appropriate active glutamic acids and typical sequence length seemed to possess the potential to produce catalytically active β -glucosidases, and exhibited a feature with highly conserved protein during evolution. And it is meaningful to deepen our knowledge on how plants regulate these individual members by releasing metabolites to modulate particular physiological process.

Further, subcellular localization and the conditions under which BGLUs come into contact with their physiological substrates determine their biological function. Some BGLUs may be targeted to specialized cellular organelles, like endoplasmic reticulum bodies in *Brassicaceae* [28]. The cellular locations of others are supposed to be standard cell compartments or apoplast, although it may be specific regions of a compartment, such as specific sections of the cell wall or peroxisome [15, 29]. In *M. truncatula*, each BGLU member (except MtBGLU33) has dual targeting to either vacuole or chloroplast. Besides, each BGLU member contains at least one predicted N-glycosylation site, and the process of glycosylation facilitates the storage and transport of inactive enzymes, which correspond to our prediction that a large proportion of them catalyze the hydrolysis through the secretory pathway (Table 1). These location results suggest that MtBGLUs can access the substrates that are inside the chloroplast (e.g. in vivo release of active hormones) or in the vacuole (e.g. cyanide, isoflavones, coumarins, and hydroxamic acids) [30–32]. In some plants, the defense glycosides and β -glucosidases are additionally separated at cellular level. In sorghum (*Sorghum bicolor* Moench), the cyanogenic glucoside dhurrin is stored exclusively in the epidermal tissue, while the corresponding BGLUs, are stored in chloroplasts in the mesophyll cells [33]. The compartmentalization provide separate two-component defense system in plant that each separate component is chemically idle while contact with each other to utmost release the toxic defense compounds upon tissue disruption.

Phylogenetic analysis of MtBGLU proteins revealed a close relatedness between *M. truncatula* and *A. thaliana* (Fig. 1). For example, MtBGLU6 and MtBGLU7, were clustered with AtBGLU45 and AtBGLU46 in cluster IV, which hydrolyzed lignin precursors in *Arabidopsis* [13], suggesting that they may be involved in lignin metabolism in *Medicago*. AtBGLU6 and AtBGLU10 in cluster V catalyzed the synthesis of flavonoid glycosides in *Arabidopsis* as glucosyltransferase [34, 35], indicating the six MtBGLUs in the same clusters might have similar transglucosidase function. Notably, cluster I contain nearly half (25/51) of MtBGLU family members, but only six members of *Arabidopsis*. The genes of this cluster are mainly expressed in roots that accumulate naturally isoflavonoids [35], which participate in plant-microbe interactions, such as symbiotic or defensive mechanisms against pathogens infection. In *G. max*, the release of free

isoflavones from their conjugates is hydrolyzed by a BGLU [2]. In *Medicago*, four BGLUs participated in MJ signal induced hydrolysis of stored isoflavone glucosides for phytoalexin medicarpin accumulation [27]. Therefore, the BGLUs of cluster I in *Medicago* may have similar roles in the hydrolysis of isoflavonoid glucosides to defensive aglycons in response to stresses. Besides, 22 out of 47 *Arabidopsis* BGLUs were grouped in two large clusters, including myrosinases in cluster At I and ER body-localized BGLUs in cluster At II. Actually, orthologous genes of these two clusters also present in other plants of Brassicales order, but not any closely related *Medicago* counterparts (At I ~ II in Fig. 1), indicating they appear to have diverged before *Arabidopsis* and *Medicago*. Similarly, six *Medicago* members in cluster II have no any closely related *Arabidopsis* homologs, indicating that these genes have diverged before *Arabidopsis* and *Medicago*.

By combining gene expression, phylogenetic and synteny analysis, some valuable knowledge was acquired about the biological function of MtBGLU genes that involved in specific physiological process. For example, MtBGLU43 exhibited the same expression profile to its orthologs MtBGLU44, which showed relatively high expression levels in various tissues except for roots, indicating that they may be involved in important and specific physiological processes of the aerial part in *Medicago*. MtBGLU16 was specially expressed in roots with extremely low expression levels in other detected tissues, which is consistent with its ortholog MtBGLU15 with the highest expression level in roots. Interestingly, their closest ortholog gene was AtBGLU42 in *Arabidopsis*, which is expressed predominantly in root epidermal cells, where it regulates the rhizobacteria-induced systemic resistance and modulates iron deficiency responses [37]. This result indicated that MtBGLU15 and MtBGLU16 may share similar function in resistance to root-colonizing bacteria in *Medicago*. Another orthologous pair MtBGLU32 and MtBGLU35 was homology with Os3BGLu6 that hydrolyzed disaccharides and hydrophobic glycosides in rice [38], suggesting the potential function of MtBGLU32 and MtBGLU35 in participating carbohydrate metabolic process.

The functional roles of several MtBGLU genes which were associated with various abiotic stresses and hormones were also investigated. Salt tolerance and drought tolerance are important traits for *Medicago* improvement [39, 40]. According to the genechip data under NaCl and drought treatments, the expressions of several MtBGLU genes were highly induced (Figs. 6 and 7,). The expression pattern of seven genes (MtBGLU14, 19, 21, 22, 26, 28 and 30) were further verified by qPCR analyses (Fig. 9). The remaining six genes (except MtBGLU19) in cluster I have phylogenetically closest relationship with BGLUs of other plants. Their close related gene AtBGLU15 of *Arabidopsis* is essential for the degradation of flavonol glucosides during the process of recovering from nitrogen deficiency and low temperature [41]. In rice, their closest homologous gene Os4BGLu12, is a wound-induced and functions in hydrolyzing cell-wall derived disaccharides and glycosides [42]. Their another homologous gene in *Crocus sativus*, CsBGLu12 was highly induced in response to UV-B, dehydration, NaCl, methyl jasmonate, and abscisic acid treatments through accumulation of antioxidant flavonols [7]. These results suggested the potential roles of MtBGLU14, 16, 21, 22, 28 and 30 in abiotic stresses-resistance in *Medicago*.

The expression level of MtBGLU19 was induced by NaCl stress, and its homolog AtBGLU6 in *Arabidopsis*, is a GH1-type transglucosidase required for flavonol 3-O-gentiobioside 7-O-rhamnoside accumulation [43],

thus MtBGLU19 might catalyze the synthesis of flavonoid glycosides under NaCl treatment in *Medicago*. Considerable evidence showed that BGLU genes were involved in the activation of inactive storage form of phytohormones glucosides, like ABA, SA and GA [18, 20, 21, 44]. MtBGLU19, MtBGLU21, MtBGLU22, MtBGLU28 and MtBGLU30 were both induced by IAA, ABA, SA and GA3 treatments, implying their potential roles in conferring various signaling network.

Consisted with previous studies in other plant species [14], our current research showed that several MtBGLU genes were differentially expressed following various abiotic stresses and hormone treatments, highlighting the diverse roles of BGLU genes in plants defense. It is noteworthy that four genes (MtBGLU21, MtBGLU22, MtBGLU28 and MtBGLU30) in cluster I are strongly induced not only by abiotic stresses but also by hormone treatments, indicating their important role against stresses in *Medicago* GH1 family (Fig. S4). Coincidentally, MtBGLU30 (MtG2) has been reported to be involved in turnover of formononetin glucoside during wound signal-induced medicarpin synthesis in *M. truncatula* [27]. Besides, some BGLU genes and their paralogues, like MtBGLU15/MtBGLU16 or MtBGLU43/MtBGLU44, shared similar expression patterns in response to different stresses treatments, suggesting that their function might be redundant (Fig. 6, 8).

In addition, cis-acting elements function as important molecular switches associated with transcriptional regulation of gene expression controlling various biological processes [45]. In the promoter regions of MtBGLU genes, several phytohormone- and stress-responsive regulatory elements were identified, which may be interacted with transcription factors to activate stress tolerance and produce chemical defense compounds in *Medicago*. In the promoter region of MtBGLU26, nine ABRE elements potentially involved in ABA regulation was identified, which might be associated with activation of MtBGLU26 under ABA treatment. The identified regulatory elements in *Medicago* will help in understanding their roles in various abiotic and biotic stresses. Overall, this study provides insight into potentially functional roles of MtBGLU genes during developmental and environmental signals, and further experiments will be useful for identification and characterization of key MtBGLU genes, such as protein structure, enzyme activity assay, subcellular localization and gene functions by analyzing loss-of-function mutants and transgenic plants.

Conclusion

In the present study, a comprehensive investigation of GH1 β -glucosidases in *M. truncatula* were carried out, and 51 full-length BGLU proteins were characterized and further classified into five main clusters, with highly identical amino acid sequence and motif compositions. Synteny analysis and phylogenetic comparison of BGLU genes from different plant species provided valuable clues about the evolutionary functions of MtBGLU genes. Moreover, analyses of their expression profiles in different tissues and in response to various treatments based on available microarray data and qPCR validation indicated that MtBGLUs played important roles in *Medicago* development and stress tolerance, especially MtBGLU21, MtBGLU22, MtBGLU28 and MtBGLU30. This work represents a new insight toward comprehensive characterization of stress- and hormone-responsive MtBGLUs by phylogenetic and gene expression

analysis, which also provide a valuable resource for better understanding on the biological roles of individual BGLU genes in *M. truncatula*.

Methods

Identification of *MtBGLU* genes of GH1 family in the *M. truncatula* genome

The genome sequences and deduced protein sequences of *M. truncatula* were downloaded from the *M. truncatula* genome website (<http://www.medicagogenome.org/>) [26]. BGLU protein sequences of *Arabidopsis* were downloaded from the TAIL database (<https://www.arabidopsis.org/>), which was utilized as the query file against the *M. truncatula* protein sequences with $e\text{-value} \leq 1e^{-3}$. After removing all of the redundant sequences, the output putative BGLU protein sequences were submitted to InterProScan (<https://www.ebi.ac.uk/interpro/search/sequence-search>), CDD (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>), Pfam (<https://pfam.xfam.org/>), SMART (<http://smart.embl-heidelberg.de/>) and InterProScan (<https://www.ebi.ac.uk/interpro/search/sequence-search>) to confirm the conserved BGLU domain. In total, 55 predicted protein sequences were curated manually by softberry (<http://linux1.softberry.com/>), eliminating three short gene fragments [*Medtr2g022730* (636bp), *Medtr5g069350* (573bp), *Medtr8g038650* (486bp)], and a cellulase-like protein (*Medtr4g122980*). Finally, 51 candidate *MtBGLU* genes were obtained and assigned based on their locations on chromosome (Table S2).

Sequence analyses and structural characterization of *MtBGLU* genes

Physical and chemical properties of MtBGLU proteins, including the number of amino acids, molecular weight, grand average of hydropathicity (GraV_y) and theoretical isoelectric points (pI), were calculated by using online ProtParam tools (<https://web.expasy.org/protparam/>). Signal sequences and N-glycosylation sites were predicted by SignalP (<http://www.cbs.dtu.dk/services/SignalP>) [46], and NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>), respectively. Cellular locations of MtBGLU proteins were predicted by the ProtComp v. 9.0 server (<http://www.softberry.com>). Sequence alignments and dendrograms in BGLU homeodomain sequences were analyzed by Jalview (http://www.jalview.org/Web_Installers/install.htm). Conserved motifs in MtBGLU protein sequences were identified by the MEME program (MEME- Suite version 5.1.0) (<http://meme-suite.org/>) [47] using the default settings, except the motif number was set to 10 and the width of minimum and maximum motif were changed to 10 and 200, respectively. The chromosomal localization of all *MtBGLUs* were acquired from the *M. truncatula* genome website. And the visualization of exon-intron positions was executed through Amazing Optional Gene Viewer software [48].

Analyses of chromosomal distribution and gene duplication of *MtBGLU* genes

Multiple Collinearity Scan toolkit (MCScanX) was adopted to analyze the gene duplication events, with the default parameters [49]. All *MtBGLU* genes were mapped to eight *M. truncatula* chromosomes based on physical location information from the *M. truncatula* genome database, followed by the analysis on

their intraspecific synteny relationship in *M. truncatula* using Amazing Gene Location software [48]. To exhibit the interspecific synteny relationship between *M. truncatula* and two other representative model plant species (*Arabidopsis* and rice), the syntenic maps were constructed using the Dual Synteny Plotter software [48]. Non-synonymous (ka) and synonymous (ks) values of *MtBGLU* homologous gene pairs were calculated using Simple Ka/Ks Calculator software [48].

Phylogenetic analysis and classification of the *MtBGLU* genes

The amino acid sequences of 47 *BGLUs* derived from *Arabidopsis* and 51 newly identified *MtBGLUs* were used for phylogenetic analysis. All of these sequences were firstly aligned by using ClustalX with the default parameters. Subsequently, an unrooted neighbor-joining phylogenetic tree was constructed based on the neighbor-joining method (with 1000 bootstrap replicates) using MEGA-X software.

Analysis of *cis*-acting elements in the promoter region of *MtBGLU* genes

The *cis*-acting elements were predicted from the 2 kb upstream promoter sequences of the *MtBGLU* genes which were uploaded to PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [50].

Analysis of the expression levels of *MtBGLU* genes from microarray data

We download all the Gene Chip data from *M. truncatula* Gene Expression Atlas (<https://Mtgea.noble.org/v3/>), which has been developed as a compendium or "atlas" of gene expression profiles for the *M. truncatula* genes. The selected genechip data covered all its major organ systems (roots, nodules, stems, petioles, leaves, vegetative buds, flowers, seeds and seed pods), from plants subjected to various abiotic and biotic stresses, and specific cell and tissue types. Amazing HeatMap software [48] was used to generate the heatmap.

Plant materials and treatments

The *M. truncatula* (cv. Jemalong A17) used in this study was initially obtained from the Noble Research Institute, identified by Beijing Botanic Garden. The seeds were further reproduced and stored in our lab. The stems, roots, leaves, flowers, pods (20-day old pods) and seeds (20-day old seeds) of mature *M. truncatula* plants, were collected separately for RNA extraction and used for qPCR analysis. To investigate the expression pattern of *MtBGLU* genes in response to various stress and hormonal treatments, seeds were germinated uniformly and transferred into the MS liquid medium, then sustained in a growth chamber with 16 h light/8 h dark. After the third leaf is fully expanded, the plants were respectively supplied with 0.1 mM NaCl, PEG, IAA, ABA, SA and GA3 to the MS liquid medium. The whole plants were collected at 0 h (as control), 3 h, 24 h, 48 h and 72 h after treatments, and immediately frozen in liquid nitrogen and stored at -80°C for subsequent analysis.

RNA extraction and gene expression analysis

Total RNAs were extracted using Easestep[®] Super total RNA Extraction kit (Promega, Shanghai, China) according to the manufacturer's instructions, first-strand cDNA synthesis was performed using Trans[®]Script One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China) per manufacturer's recommendations. The qPCR were carried out using a 2⁺RealStar Green Fast Mixture (GeneStar, Shanghai, China) and ABI 7500 real-time Detection System (Applied Biosystems, USA). The housekeeping gene of actin-related protein 4A gene was used as an internal control. The reaction was carried out as follows: 94°C for 30 s, followed by 40 cycles of 94°C, 5 s, 60°C, 34 s. Each reaction was performed in biological triplicates and the data from qPCR amplification was analyzed using $2^{-\Delta\Delta CT}$ method. The results were analyzed by means $\pm$ standard deviation (SD). The primer sequences used in this study were shown in details in Table S1.

Abbreviations

GH:glycosyl hydrolase; BGLU: β -glucosidase; MW:Molecular weight; PI:Isoelectric point; qPCR:Quantitative polymerase chain reaction; aa:Amino acid; ABA:Absciscic Acid; PEG:Polyethylene Glycol; SA:Salicylic acid; GA3:Gibberellic acid.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

Availability of data and material

All data of the current study is available in the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was sported by the National Nature Science Foundation of China (U1906201 and 31670305), and Fundamental Research Funds for Central Non-profit Scientific Institution (2019-YWF-YB-06). The funding body played no role in the design of the study, analysis and interpretation of data, and in writing the manuscript.

Authors' contributions

Experiments were performed by JFY, LM, WBJ, and YY. JFY and YHT analyzed the data. JFY drafted the manuscript. YZP supervised the experiments and finalized the manuscript. All authors read and approved the manuscript.

Acknowledgements

The authors would like to thank Dr. Guoan Shen from the Institute of Medicinal Plant Development, Beijing, for helpful comments.

References

1. VanEtten HD, Mansfield JW, Bailey JA, Farmer EE: **Two classes of plant antibiotics: phytoalexins versus "phytoanticipins"**. *Plant Cell* 1994, **6**(9):1191-1192.
2. Suzuki H, Takahashi S, Watanabe R, Fukushima Y, Fujita N, Noguchi A, Yokoyama R, Nishitani K, Nishino T, Nakayama T: **An isoflavone conjugate-hydrolyzing beta-glucosidase from the roots of soybean (*Glycine max*) seedlings: purification, gene cloning, phylogenetics, and cellular localization**. *J Biol Chem* 2006, **281**(40):30251-30259.
3. Jones P, Vogt T: **Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers**. *Planta* 2001, **213**(2):164-174.
4. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active enzymes database (CAZy) in 2013**. *Nucleic Acids Res* 2014, **42**:D490-495.
5. Czjzek M, Cicek M, Zamboni V, Bevan DR, Henrissat B, Esen A: **The mechanism of substrate (aglycone) specificity in beta-glucosidases is revealed by crystal structures of mutant maize beta-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(25):13555-13560.
6. Davies GJ, Henrissat B: **Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era**. *Biochemistry Society Transactions* 2002, **30**(2):291-297.
7. Baba SA, Vishwakarma RA, Ashraf N: **Functional characterization of CsBGlu12, a beta-glucosidase from *Crocus sativus*, provides insights into its role in abiotic stress through accumulation of antioxidant flavonols**. *J Biol Chem* 2017, **292**(11):4700-4713.
8. Kristoffersen P, Brzobohaty B, Hohfeld I, Bako L, Melkonian M, Palme K: **Developmental regulation of the maize *Zm-p60.1* gene encoding a beta-glucosidase located to plastids**. *Planta* 2000, **210**(3):407-415.
9. Forslund K, Morant M, Jorgensen B, Olsen CE, Asamizu E, Sato S, Tabata S, Bak S: **Biosynthesis of the nitrile glucosides rhodiocyanoside A and D and the cyanogenic glucosides lotaustralin and linamarin in *Lotus japonicus***. *Plant Physiol* 2004, **135**(1):71-84.
10. Ketudat Cairns JR, Esen A: **beta-Glucosidases**. *Cell Mol Life Sci* 2010, **67**(20):3389-3405.
11. Ketudat Cairns JR, Mahong B, Baiya S, Jeon JS: **beta-Glucosidases: Multitasking, moonlighting or simply misunderstood?** *Plant Sci* 2015, **241**:246-259.

12. Opassiri R, Maneesan J, Akiyama T, Pomthong B, Jin S, Kimura A, Cairns JRK: **Rice Os4BGlu12 is a wound-induced β -glucosidase that hydrolyzes cell wall- β -glucan-derived oligosaccharides and glycosides.** *Plant Sci* 2010, **179**(3):273-280.
13. Chapelle A, Morreel K, Vanholme R, Le-Bris P, Morin H, Lapierre C, Boerjan W, Jouanin L, Demont-Caulet N: **Impact of the absence of stem-specific beta-glucosidases on lignin and monolignols.** *Plant Physiol* 2012, **160**(3):1204-1217.
14. Morant AV, Jorgensen K, Jorgensen C, Paquette SM, Sanchez-Perez R, Moller BL, Bak S: **beta-glucosidases as detonators of plant chemical defense.** *Phytochemistry* 2008, **69**(9):1795-1813.
15. Bednarek P, Pislewska-Bednarek M, Svatos A, Schneider B, Doubsky J, Mansurova M, Humphry M, Consonni C, Panstruga R, Sanchez-Vallet A *et al*: **A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense.** *Science* 2009, **323**(5910):101-106.
16. Nakano RT, Pislewska-Bednarek M, Yamada K, Edger PP, Miyahara M, Kondo M, Bottcher C, Mori M, Nishimura M, Schulze-Lefert P *et al*: **PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum bodies and glucosinolates in *Arabidopsis thaliana*.** *Plant J* 2017, **89**(2):204-220.
17. Mageroy MH, Parent G, Germanos G, Giguere I, Delvas N, Maaroufi H, Bauce E, Bohlmann J, Mackay JJ: **Expression of the beta-glucosidase gene *Pgbetaglu-1* underpins natural resistance of white spruce against spruce budworm.** *Plant J* 2015, **81**(1):68-80.
18. Xu ZY, Lee KH, Dong T, Jeong JC, Jin JB, Kanno Y, Kim DH, Kim SY, Seo M, Bressan RA *et al*: **A vacuolar beta-glucosidase homolog that possesses glucose-conjugated abscisic acid hydrolyzing activity plays an important role in osmotic stress responses in *Arabidopsis*.** *Plant Cell* 2012, **24**(5):2184-2199.
19. Lee KH, Piao HL, Kim HY, Choi SM, Jiang F, Hartung W, Hwang I, Kwak JM, Lee IJ, Hwang I: **Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid.** *Cell* 2006, **126**(6):1109-1120.
20. Himeno N, Saburi W, Wakuta S, Takeda R, Matsuura H, Nabeta K, Sansenya S, Ketudat Cairns JR, Mori H, Imai R *et al*: **Identification of rice beta-glucosidase with high hydrolytic activity towards salicylic acid beta-D-glucoside.** *Biosci, Biotechnol, Biochem* 2013, **77**(5):934-939.
21. Hua Y, Sansenya S, Saetang C, Wakuta S, Ketudat Cairns JR: **Enzymatic and structural characterization of hydrolysis of gibberellin A4 glucosyl ester by a rice beta-D-glucosidase.** *Archives of Biochemistry and Biophysics* 2013, **537**(1):39-48.
22. Miyahara T, Takahashi M, Ozeki Y, Sasaki N: **Isolation of an acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase from the monocot *Agapanthus africanus*.** *J Plant Physiol* 2012, **169**(13):1321-1326.
23. Xu ZE-T, LL; Zeng, LH; Lalgondar, M; Bevan, DR; Winkel, BSJ; Mohamed, A; Cheng, C; Shih, M; Poulton, JE; Esen, A: **Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1.** *Plant Mol Biol* 2004, **55**:343–367.

24. Opassiri R, Pomthong B, Onkoksoong T, Akiyama T, Esen A, Ketudat Cairns JR: **Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 beta-glucosidase.** *BMC Plant Biol* 2006, **6**:33.
25. Gomez-Anduro G, Cenicerros-Ojeda EA, Casados-Vazquez LE, Bencivenni C, Sierra-Beltran A, Murillo-Amador B, Tiessen A: **Genome-wide analysis of the beta-glucosidase gene family in maize (*Zea mays* L. var B73).** *Plant Mol Biol* 2011, **77**(1-2):159-183.
26. Young ND, Debelle F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H *et al.* **The *Medicago* genome provides insight into the evolution of rhizobial symbioses.** *Nature* 2011, **480**(7378):520-524.
27. Naoumkina M FM, Sumner LW, Tang YH, Liu CJ, Dixon RA: **Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(46):17909–17915.
28. Nakano RT, Yamada K, Bednarek P, Nishimura M, Hara-Nishimura I: **ER bodies in plants of the *Brassicales* order: biogenesis and association with innate immunity.** *Frontiers in Plant Science* 2014, **5**:73.
29. Dharmawardhana DP, Ellis BE, Carlson JE: **cDNA cloning and heterologous expression of coniferin beta-glucosidase.** *Plant Mol Biol* 1999, **40**(2):365-372.
30. Sue M, Yamazaki K, Yajima S, Nomura T, Matsukawa T, Iwamura H, Miyamoto T: **Molecular and structural characterization of hexameric beta-D-glucosidases in wheat and rye.** *Plant Physiol* 2006, **141**(4):1237-1247.
31. Morant AV, Bjarnholt N, Kragh ME, Kjaergaard CH, Jorgensen K, Paquette SM, Piotrowski M, Imberty A, Olsen CE, Moller BL *et al.* **The beta-glucosidases responsible for bioactivation of hydroxynitrile glucosides in *Lotus japonicus*.** *Plant Physiol* 2008, **147**(3):1072-1091.
32. Ahn YO, Shimizu B, Sakata K, Gantulga D, Zhou ZC, Bevan DR, Esen A: **Scopolin-hydrolyzing beta-glucosidases in roots of *Arabidopsis*.** *Plant Cell Physiology* 2010, **51**(1):132-143.
33. Thayer SS, Conn EE: **Subcellular localization of dhurrin beta-glucosidase and hydroxynitrile lyase in the mesophyll cells of sorghum leaf blades.** *Plant Physiol* 1981, **67**(4):617-622.
34. Miyahara T, Sakiyama R, Ozeki Y, Sasaki N: **Acyl-glucose-dependent glucosyltransferase catalyzes the final step of anthocyanin formation in *Arabidopsis*.** *J Plant Physiol* 2013, **170**(6):619-624.
35. Roepke J, Bozzo GG: ***Arabidopsis thaliana* beta-glucosidase BGLU15 attacks flavonol 3-O-beta-glucoside-7-O-alpha-rhamnosides.** *Phytochemistry* 2015, **109**:14-24.
36. Veitch NC: **Isoflavonoids of the leguminosae.** *Natural Product Reports* 2013, **30**(7):988-1027.
37. Zamioudis C, Hanson J, Pieterse CMJ: **beta-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in *Arabidopsis* roots.** *New Phytol* 2014, **204**(2):368-379.
38. Seshadri S, Akiyama T, Opassiri R, Kuaprasert B, Cairns JK: **Structural and enzymatic characterization of Os3BGLu6, a rice beta-glucosidase hydrolyzing hydrophobic glycosides and (1->3)- and (1->2)-linked disaccharides.** *Plant Physiol* 2009, **151**(1):47-58.

39. Elmaghrabi AM, Rogers HJ, Francis D, Ochatt S: **Toward unravelling the genetic determinism of the acquisition of salt and osmotic stress tolerance through in vitro selection in *Medicago truncatula*.** *Methods in Molecular Biology* 2018, **1822**:291-314.
40. Gou JQ, Debnath S, Sun L, Flanagan A, Tang YH, Jiang QZ, Wen J, Wang ZY: **From model to crop: functional characterization of *SPL8* in *M. truncatula* led to genetic improvement of biomass yield and abiotic stress tolerance in alfalfa.** *Plant Biotechnol J.* 2018, **16**(4):951-962.
41. Roepke J, Gordon HOW, Neil KJA, Gidda S, Mullen RT, Freixas Coutin JA, Bray-Stone D, Bozzo GG: **An apoplastic beta-glucosidase is essential for the degradation of flavonol 3-O-beta-glucoside-7-O-alpha-rhamnosides in *Arabidopsis*.** *Plant Cell Physiology* 2017, **58**(6):1030-1047.
42. Opassiri R, Maneesan J, Akiyama T, Pomthong B, Cairns JRK: **Rice Os4BGLu12 is a wound-induced β -glucosidase that hydrolyzes cell wall- β -glucan-derived oligosaccharides and glycosides.** *Plant Science* 2010, **179**(3):273-280.
43. Ishihara H, Tohge T, Viehover P, Fernie AR, Weisshaar B, Stracke R: **Natural variation in flavonol accumulation in *Arabidopsis* is determined by the flavonol glucosyltransferase BGLU6.** *J Exp Bot* 2016, **67**(5):1505-1517.
44. Jin SX, Kanagaraj A, Verma D, Lange T, Daniell H: **Release of hormones from conjugates: chloroplast expression of beta-glucosidase results in elevated phytohormone levels associated with significant increase in biomass and protection from aphids or whiteflies conferred by sucrose esters.** *Plant Physiol* 2011, **155**(1):222-235.
45. Yamaguchi-Shinozaki K, Shinozaki K: **Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters.** *Trends Plant Sci* 2005, **10**(2):88-94.
46. Bendtsen JD, Nielsen H, Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *Journal of Molecular Biology* 2004, **340**(4):783-795.
47. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching.** *Nucleic Acids Res* 2009, **37**:W202-208.
48. Chen C, Xia R, Chen H, He Y: **TBtools, a Toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface.** *BioRxiv* 2018:289660.
49. Wang YP, Tang HB, Debarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin H, Marler B, Guo H *et al*: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity.** *Nucleic Acids Res* 2012, **40**(7):e49.
50. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van-de-Peer Y, Rouze P, Rombauts S: **PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.** *Nucleic Acids Res* 2002, **30**(1):325-327.

Table

Table 1. Properties and locations of predicted BGLU proteins in *M. truncatula*.

Gene name	Pre-protein			Mature protein					
	Aa	MW	CleaVage Site	Aa	MW	Theoretical pI	GraVy	N-gly site	Possible destination
<i>MtBGLU1</i>	640	72.82	23-24	617	70.34	5.50	-0.36	2	Vac,Chl
<i>MtBGLU2</i>	459	51.94	17-18	442	50.12	5.66	-0.35	2	Vac,Chl
<i>MtBGLU3</i>	522	59.67	31-32	491	56.30	5.73	-0.28	2	Vac,Chl
<i>MtBGLU4</i>	522	59.97	33-34	489	56.10	5.95	-0.39	2	Vac,Chl
<i>MtBGLU5</i>	521	59.64	-	-	-	6.13	-0.39	1	Vac,Chl
<i>MtBGLU6</i>	530	60.71	22-23	508	57.99	6.42	-0.27	4	Vac,Chl
<i>MtBGLU7</i>	529	60.78	24-25	505	57.81	6.47	-0.29	3	Vac,Chl
<i>MtBGLU8</i>	512	58.81	24-25	488	56.16	9.02	-0.37	3	Vac,Chl
<i>MtBGLU9</i>	509	58.67	19-20	490	56.44	8.78	-0.38	4	Vac,Chl
<i>MtBGLU10</i>	516	58.97	26-27	490	56.00	8.60	-0.33	4	Vac,Chl
<i>MtBGLU11</i>	494	56.89	21-22	473	54.44	6.20	-0.34	2	Vac,Chl
<i>MtBGLU12</i>	513	58.20	23-24	490	55.91	9.20	-0.30	1	Vac,Chl
<i>MtBGLU13</i>	514	58.32	26-27	488	55.57	9.18	-0.28	2	Vac,Chl
<i>MtBGLU14</i>	531	60.86	23-24	508	58.33	6.73	-0.30	1	Vac,Chl
<i>MtBGLU15</i>	491	56.26	-	-	-	5.41	-0.49	1	Vac,Chl
<i>MtBGLU16</i>	494	56.28	-	-	-	5.30	-0.51	1	Vac,Chl
<i>MtBGLU17</i>	529	59.92	-	-	-	6.03	-0.38	3	Vac,Chl
<i>MtBGLU18</i>	507	57.68	-	-	-	5.92	-0.40	4	Vac,Chl
<i>MtBGLU19</i>	520	58.86	26-27	494	56.20	6.36	-0.27	5	Vac,Chl
<i>MtBGLU20</i>	495	56.13	23-24	472	53.67	8.87	-0.31	1	Vac,Chl
<i>MtBGLU21</i>	528	60.22	28-29	500	57.08	5.75	-0.37	2	Vac,Chl
<i>MtBGLU22</i>	515	59.24	24-25	491	56.57	7.70	-0.43	2	Vac,Chl
<i>MtBGLU23</i>	495	56.59	27-28	468	53.75	6.70	-0.35	2	Vac,Chl
<i>MtBGLU24</i>	495	57.04	27-28	468	53.93	6.50	-0.38	3	Vac,Chl
<i>MtBGLU25</i>	441	50.53	-	-	-	8.95	-0.50	2	Vac,Chl
<i>MtBGLU26</i>	532	61.04	24-25	508	58.34	7.59	-0.31	2	Vac,Chl
<i>MtBGLU27</i>	520	59.55	20-21	500	58.75	5.82	-0.36	1	Vac,Chl
<i>MtBGLU28</i>	525	60.04	25-26	500	57.18	5.80	-0.32	2	Vac,Chl
<i>MtBGLU29</i>	515	58.59	27-28	488	55.76	8.84	-0.33	3	Vac,Chl
<i>MtBGLU30</i>	408	46.13	20-21	388	43.64	5.61	-0.31	1	Vac,Chl
<i>MtBGLU31</i>	516	58.96	21-22	495	56.60	6.97	-0.28	2	Vac,Chl
<i>MtBGLU32</i>	506	57.28	25-26	481	54.55	8.06	-0.24	3	Vac,Chl
<i>MtBGLU33</i>	1030	115.72	513-514	517	59.12	8.46	0.00	3	Ext,Gol
<i>MtBGLU34</i>	522	59.52	26-27	496	56.74	6.04	-0.37	2	Vac,Chl
<i>MtBGLU35</i>	522	59.32	23-24	499	56.77	6.53	-0.26	3	Vac,Chl
<i>MtBGLU36</i>	516	58.35	23-24	493	55.88	5.59	-0.34	2	Vac,Chl
<i>MtBGLU37</i>	526	60.09	21-22	505	57.73	6.10	-0.28	5	Vac,Chl
<i>MtBGLU38</i>	523	59.35	25-26	498	56.54	5.88	-0.34	1	Vac,Chl
<i>MtBGLU39</i>	523	59.56	23-24	500	57.07	6.54	-0.28	2	Vac,Chl
<i>MtBGLU40</i>	519	59.30	23-24	496	56.83	5.63	-0.29	2	Vac,Chl
<i>MtBGLU41</i>	516	58.90	16-17	500	57.08	7.10	-0.40	5	Vac,Chl
<i>MtBGLU42</i>	538	61.25	32-33	506	58.01	9.08	-0.39	4	Vac,Chl
<i>MtBGLU43</i>	538	61.08	32-33	506	57.52	6.82	-0.36	4	Vac,Chl
<i>MtBGLU44</i>	539	61.31	32-33	507	57.75	7.68	-0.34	4	Vac,Chl
<i>MtBGLU45</i>	453	46.86	-	-	-	8.11	-0.46	4	Vac,Chl

<i>MtBGLU46</i>	488	55.72	-	-	-	6.17	-0.36	4	Vac,Chl
<i>MtBGLU47</i>	530	60.56	30-31	500	58.75	5.32	-0.30	5	Vac,Chl
<i>MtBGLU48</i>	455	52.50	-	-	-	7.79	-0.51	3	Vac,Chl
<i>MtBGLU49</i>	315	35.76	29-30	286	32.34	8.80	-0.08	3	Vac,Chl
<i>MtBGLU50</i>	515	59.21	29-30	286	55.79	9.23	-0.36	5	Vac,Chl
<i>MtBGLU51</i>	515	59.27	29-30	286	55.85	9.22	-0.39	5	Vac,Chl

Supporting Files Legend

Fig. S1 Alignment of multiple MtBGLU amino acid sequences. The sequence conservation is shown as a percentage bar-score below. The sequence logo is shown at the bottom, summarizing the occurrence of given amino acids at specific positions. Two well-conserved active glutamate residues of MtBGLU domain A: catalytic acid/base and B: catalytic nucleophile were presented.

Fig. S2 Phylogenetic relationships and conserved domains of MtBGLU proteins. Left panel: Phylogenetic tree constructed using MEGA-X by the Neighbor-Joining method. The proteins are clustered into five main clades. Right panel: distribution of conserved domain in MtBGLU proteins. The different-colored boxes represent different domains and their positions in each MtBGLU protein sequence.

Fig. S3 Sequence information of the ten conserved motifs of *MtBGLU* gene. Including the sequence logo and given amino acids, as well as amino acid numbers of each motifs.

Fig. S4 Expression profiles of 12 representative *MtBGLUs* in response to different abiotic stress and hormone treatments.

Fig. S5 Phylogenetic relationships and motif compositions of BGLU proteins from *Medicago* and *Arabidopsis*. Left panel: Phylogenetic tree constructed using MEGA-X by the Neighbor-Joining method. The proteins are clustered into five main clades. Right panel: distribution of conserved motifs in BGLU proteins. The different-colored boxes represent different motifs and their positions in each BGLU protein sequence.

Table S1 List of primers used in this research.

Table S2 List of all *MtBGLU* genes identified in the *M. truncatula* genome.

Table S3 *MtBGLU* genes and their functions as identified and characterized in previous reports.

Additional file 1: All tandem and segmental duplicated *MtBGLU* gene pairs in *Medicago*, and the orthologous *BGLU* gene pairs between *Medicago* and *Arabidopsis*/rice.

Additional file 2: List of identified *cis*-acting elements of all *MtBGLU* genes.

Additional file 3: Detailed information on the expression levels of *MtBGLU* genes retrieved from microarray data, including four different tissues and treatments in response to different abiotic stresses and hormones.

Figures

Figure 1

Phylogenetic tree representing relationships among BGLU proteins from Medicago and Arabidopsis. The phylogenetic tree was constructed using MEGA-X based on the Neighbor-Joining (NJ) method; bootstrap was 1,000 replicates. The different-colored arcs indicate different groups of BGLU proteins. The red solid circles and hollow circles represent BGLU proteins from Medicago and Arabidopsis, respectively.

a

Phylogenetic tree of 45 Gypsy_Hydrus_1 sequences. The tree is rooted and shows bootstrap values at the nodes. The sequences are grouped into six clades (I-VI) based on their phylogenetic relationships. The scale bar represents 1000 bp.

b

Genomic organization of the Gypsy_Hydrus_1 retrovirus. The scale bar represents 18000 bp. The genome contains a UTR, a GOR, a Gypsy_Hydrus_1, and a TPT. The Gypsy_Hydrus_1 sequence is highlighted in red.

c

Genomic organization of the Gypsy_Hydrus_1 retrovirus. The scale bar represents 1000 bp. The genome contains a UTR, a GOR, a Gypsy_Hydrus_1, and a TPT. The Gypsy_Hydrus_1 sequence is highlighted in red.

Phylogenetic relationships, gene structure and architecture of conserved motifs in MtBGLU genes. a. The phylogenetic tree was constructed based on the full-length amino acid of MtBGLU proteins using MEGA-X software. Details of groups are shown in different colors. b. Exon-intron structure and conserved domain of MtBGLU genes. Blue boxes indicate untranslated 5'- and 3'-regions; yellow boxes indicate exons; black lines indicate introns. The BGLU domain (glyco_hydro_1) and N-terminal signal peptide are highlighted by green and red boxes, respectively. c. The motif composition of MtBGLU proteins. The motifs, numbers 1-10, are displayed in different colored boxes. The sequence information for each motif is provided in Fig. S3.

Figure 3

Figure 3

Chromosomal location and gene duplication of MtBGLU genes. The tandem duplicated genes are marked by blue arc trajectory and segmentally duplicated genes are connected by red lines.

Figure 4

Synteny analysis of BGLU genes between *Medicago* and two representative plant species (*A. thaliana* and *O. sativa*). Gray lines in the background indicate the collinear blocks within *Medicago*, and *A. thaliana*/*O. sativa*, and the red lines highlight the syntenic MtBGLU gene pairs.

Figure 5

Figure 5

Analysis of the numbers and types of cis-acting elements in MtBGLU genes. a. The phylogenetic tree was constructed based on the amino acid sequences of MtBGLU proteins using MEGA-X software. Details of each group are shown in different colors. b. Colors and numbers of the grid indicated the numbers of different cis-acting elements in these MtBGLU genes. c. Colored block represented different types of cis-acting elements and their locations in each MtBGLU gene.

Figure 6

Figure 6

Expression profiles of MtBGLU genes under different hormone and stress treatments retrieved from microarray data. Each column indicates a sampling time point, and each row indicates an MtBGLU gene, wherein, each treatment was normalized in the same row, except limit N and NAA. The relative expressions levels are log2- transformed and visualized for heatmap. The colors vary from blue to red, and circle from small to large representing the scale of the relative expression levels. Each treatment was normalized in the same row (scale="row"), except limit N and NAA treatments with only two treatment groups. Genes with deeper background color showed that their expression level was significantly increased after stress induction. Information on the corresponding gene chip number and specific treatment are available in Additional file 3.

Figure 7

Expression profiles of MtBGLU genes under different treatments from microarray data. a. The phylogenetic tree was constructed based on the amino acid sequences of MtBGLU proteins using MEGA-X software. Details of each group are shown in different colors. b. Each column indicates a treatment, and each row indicates an MtBGLU gene, wherein, each member was normalized in the same column. The relative expression levels are log2-transformed and visualized for heatmap. The colors vary from blue to red, and circle from small to large representing the scale of relative expression levels. c. Presence of treatments in which each MtBGLU gene was up-regulated by more than two fold. Different colors represent different treatments. There were only two replicates for NaCl and drought treatments, and each of them is calculated as 0.5.

Figure 8

Expression profiles of MtBGLU genes. a. The phylogenetic tree was constructed based on amino acid sequences of MtBGLU proteins using MEGA-X software. Details of clusters are shown in different colors. b. Expression profiles of MtBGLU genes in 9 different tissues from microarray data. c. Expression analysis of 12 MtBGLU genes in six representative samples by qPCR analysis. Data were normalized to actin-related protein 4A gene and vertical bars indicate standard deviation.

Figure 9

Figure 9

Expression profiles of 12 representative MtBGLU genes in response to abiotic stress treatments. Data were normalized to actin-related protein 4A gene and vertical bars indicate standard deviation. Asterisks indicate the corresponding gene significantly up- or down-regulated compared with the untreated control (*P < 0.05, **P < 0.01, Duncan's t-test)

Figure 10

Figure 10

Expression profiles of 12 representative MtBGLU genes in response to different hormonal treatments. Data were normalized to actin-related protein 4A gene and vertical bars indicate standard deviation. Asterisks indicate the corresponding gene significantly up- or down-regulated compared with the untreated control (*P < 0.05, **P < 0.01, Duncan's t-test)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xlsx](#)
- [Supportinginformation.pdf](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile3.xlsx](#)