Overexpression of GmCAMTA12 enhanced drought tolerance in Arabidopsis and Soybean

CURRENT STATUS: POSTED

Muhammad Noman
Jilin Agricultural University

ORCiD: https://orcid.org/0000-0002-5580-671X

Aysha Jameel
Jilin Agricultural University

Weidong Qiang
Jilin Agricultural University

Naveed Ahmad
Jilin Agricultural University

Liu Weican
Jilin Agricultural University

Fawei Wang
Jilin Agricultural University

Haiyan Li

2944649591@qq.com Corresponding Author

DOI: 10.21203/rs.2.14149/v1

SUBJECT AREAS
Plant Physiology and Morphology     Plant Molecular Biology and Genetics

KEYWORDS
Arabidopsis, CaM (Calmodulin), Calmodulin-Binding Transcription Activators (CAMTA), cis-elements, drought, qPCR, soybean hairy roots
Abstract
Background: Earlier, fifteen transcription factors in the CAMTA family of soybean were reported to
differentially express in multiple stresses however, their functional analyses had not yet been
attempted. To characterize their role in drought stress, we comprehensively analyzed GmCAMTA
family in silico and determined their expression pattern after which we cloned and overexpressed the
2769 bp CDS of GmCAMTA12 in Arabidopsis and soybean and carried out drought assays. Results:
The bioinformatics analysis revealed multiple stress-related cis motifs including ABRE, SARE, G-box
and W-box, 10 unique miRNA targets in GmCAMTA transcripts and 48 proteins in GmCAMTAs’
interaction network. The stably transformed homozygous lines of Arabidopsis overexpressing
GmCAMTA12 exhibited enhanced tolerance to drought in soil as well as on MS media containing
mannitol. In their drought assay, the average survival rate of transgenic Arabidopsis lines OE5 and
OE12 (Overexpression Line 5 and Line 12) was 83.66% and 87.87% respectively, which was ~30%
higher than that of wt. In addition, the germination and root length assays as well as physiological
indexes such as proline and MDA contents, CAT activity and leakage of electrolytes affirmed the
better performance of OE lines. Likewise, GmCAMTA12 overexpression in soybean promoted more
developed hairy roots in OE chimeric plants as compare to that of VC (Vector control). In parallel, the
improved growth performance of OE in Hoagland-PEG and on MS-mannitol was revealed by their
phenotypic, physiological and molecular measures. Furthermore, with the overexpression of
GmCAMTA12, the downstream genes including AtAnnexin5, AtCaMHSP, At2G433110 and AtWRKY14
were upregulated in Arabidopsis. Likewise, in soybean hairy roots, GmELO, GmNAB and GmPLA1-IIId
were significantly upregulated as a result of GmCAMTA12 overexpression and majority of these
upregulated genes in both plants possess CAMTA binding CGCG/CGTG motif in their promoters.
Conclusions: Taken together, we report that GmCAMTA12 plays substantial role in tolerance of
soybean against drought stress and could prove as a novel candidate for engineering of soybean and
other plants against drought stress. Some research gaps were also identified for future studies to
extend our comprehension of Ca-CaM-CAMTA-mediated stress regulatory mechanisms.

Background
A successful sustainable agriculture should ensure food security and must be eco-friendly as well as safe to human (1). With the existing population growth rate, the current food production rate needs to be increased at least up to 70% by 2050 (2,3). Despite advanced farming practices, the abiotic stresses due to drought, salinity, water and temperature fluctuations are causing 50-80% losses in crop yield and therefore should be as effectively managed as possible (4). Soon, the warmer earth will cause more humid atmosphere but less humid soil leading to more frequent drought which would negatively affect the rate of photosynthesis, uptake of CO$_2$, accumulation of biomass and yield (5,6).

Thus, developing stress-resistant crops with stable yields under adverse conditions is an important strategy to ensure future food security (2,7). Deep insights into the mechanisms underlying signaling crosstalk mediating stress tolerance would aid plant researchers to upgrade the plant indigenous machinery through biotechnology and genetic engineering.

In plant, the ubiquitous secondary messenger Calcium is the key to maintain the harmonious and homeostatic conditions via signaling (8,9). Calcium ions perceive and encode the environmental, developmental or hormonal signals into a definite frequency which is decoded and relayed by the protein molecules next to them including calmodulin (CaM). By interacting with calcium, calmodulin (calcium-modulated) proteins sense and convey the signals to calmodulin-binding proteins (10). Here, CaM through signal frequency readjustment channelizes them before transducing to CAMTA TFs, thus CaM acts like a prism (as prism dissects white light into its components). A transcription factor (TF) family reported and initially referred to as Ethylene-induced calmodulin binding protein (11) or signal responsive (SR) protein while now called as Calmodulin-binding transcription activator (CAMTA) is present in almost all eukaryotes. CAMTAs were first detected in Nicotiana tabaccum while studying calmodulin binding proteins (11-13). After their emergence, all multicellular eukaryotes studied to date have been reported to be equipped with variable number of CAMTA genes such as Arabidopsis thaliana (6) (13), Lycopersicum esculantum (7) (14), Medicago truncatula (7) (15), Citrus (9) (16), Populus trichocarpa (7) (17), Nicotiana tabacum (13) (18), Musa acuminata (5) (19) and phaseolus vulgaris (8) (20). Glycine max possesses 15 CAMTA genes all of which differentially express under various stress conditions (21).
CAMTA TFs are an integral element of Calcium-mediated biotic/abiotic stress and hormonal signaling pathway (22–25). Stress signals are conveyed and modulated through Ca-CaM-CAMTA pathway and a rapid and calculated response is observed by carrying out the transcription of stimulus specific genes (26). Calmodulin Binding Transcription Activators through their CG-1 domain recognize and specifically bind the ((A/C)CGCG(C/G/T), (A/C)CGTGT) sequence of the target genes thereby directly interacting and regulating their transcription (12). The CG-1 motif is a pivotal member of a rapid stress response element (RSRE) found in the promoters of many genes which are rapidly activated in response to stress (27). Previous experiments report the negative role of Arabidopsis CAMTA3 in regulating plant immunity as demonstrated in the AtMCAMTA3 (loss-of-function) mutant Arabidopsis (28–30). Similarly, AtCAMTA1 and AtCAMTA3 have been shown to function in drought and regulate auxin (22,31). Moreover, AtCAMTA1 and AtCAMTA3 regulated cold response by inducing CBF pathway genes as the double ATCAMTA1 and ATCAMTA2 mutant exhibited impaired freezing tolerance. (32,33). AtCAMTA3 role in regulating SA pathway genes working in freezing tolerance was recently determined (34). Two recent studies highlighted the role of AtCAMTA3 (35) and AtCAMTA6 (36) in salt tolerance. TaCAMTA4 in wheat was demonstrated to negatively regulate defense response against Puccinia triticina (25).

Soybean is an important crop cultivated globally for food, feed (37), pharmaceutical and soil nitrogen improving purposes (38). However environmental adversities including drought affect soybean growth and yield (39). Earlier, the in silico analysis of GmCAMTAs was conducted, however the potential targets of miRNA in GmCAMTA transcripts and the protein-protein interaction network were not reported (21). In addition, the previous study also did not analyze the expression pattern of GmCAMTAs in soybean leaves in response to drought (21). In an extension to that study and in order to decipher the role of soybean CAMTA family in drought, we first comprehensively analyzed (in silico) the GmCAMTAs including their physicochemical properties, chromosomal distribution, cis-motifs, miRNA targets and protein-protein interaction network. Next, we determined the spatiotemporal expression pattern of GmCAMTAs in roots as well leaves of soybean under PEG stress and selected an efficient member of the GmCAMTA family to functionally characterize. We constructed the
overexpression construct by cloning the 2769 bp CDS of \textit{GmCAMTA12} and transformed into Arabidopsis and soybean hairy roots. Through various drought assays, we demonstrated that the transgenic Arabidopsis and chimeric soybean (OE- Overexpressing \textit{GmCAMTA12}) plants exhibited enhanced tolerance and performed better under drought stress than their non-transgenic counterpart. qPCR of the downstream genes in Arabidopsis and soybean also displayed altered expression as a result of \textit{GmCAMTA12} overexpression. This study proves that \textit{GmCAMTA12} plays role in drought stress and as a transcription factor it regulates the downstream genes involved in drought tolerance. \textit{GmCAMTA12} could be exploited in developing drought tolerant crops.

\textbf{Results}

\textbf{Physico-chemical properties of GmCAMTA proteins}

Various physico-chemical properties like number of amino acids, protein molecular weight (MW), pI (Isoelectric point), number of atoms, instability and aliphatic indexes and GRAVY (grand average of hydropathy) determined online with the ProtParam tool are given in table 1. GmCAMTA11 and GmCAMTA9 are the shortest polypeptides comprising of 910 and 911 aa, while GmCAMTA1 is the longest possessing 1122 aa. Overall, their average length is \(\sim 1004\)aa with a range of \(\sim 200\) aa mutual difference. The molecular weight of GmCAMTAs ranges between 126989.43 and 102394.56 kDa with an average MW of 112848.3 and the number of atoms is proportional to the molecular weight of each protein. Similarly, GmCAMTA8 has the highest pI value of 7.64 showing that GmCAMTAs have relatively lower pI. Moreover, they are also hydrophilic in nature as the GRAVY ranges between -0.625 (GmCAMTA1) and -0.394 (GmCAMTA12). Almost all of the GmCAMTAs are thermally stable as their aliphatic indexes match with that of the other globular proteins (highest in GmCAMTA12). While none of them is stable in test tube as the instability index of all GmCAMTAs is higher than 40. Except GmCAMTA8, the number of Asp+Glu (negatively charged aa) is higher than Arg+Lys (positively charged aa).

\textbf{Table 1. Physico-chemical properties of all GmCAMTA proteins}
| Proteins              | No. of aa | MW (kDa)   | pl  | Asp+Glu | Arg+Lys | No. of atoms | II    | Al    | GRAVY |
|----------------------|-----------|------------|-----|---------|----------|--------------|-------|-------|-------|
| Glyma.05g178200      | 1122      | 126989.4   | 5.79| 159     | 130      | 17668        | 47.68 | 75.97 | -0.625|
| (GmCAMTA1)           |           | 3          |     |         |          |              |       |       |       |
| Glyma.08g135200      | 1102      | 124429.5   | 5.66| 156     | 124      | 17322        | 43.50 | 77.79 | -0.596|
| (GmCAMTA2)           |           | 6          |     |         |          |              |       |       |       |
| Glyma.15g053600      | 1088      | 122478.3   | 5.69| 151     | 118      | 16999        | 49.97 | 76.17 | -0.552|
| (GmCAMTA3)           |           | 6          |     |         |          |              |       |       |       |
| Glyma.08g072100      | 1079      | 121261.4   | 5.72| 148     | 122      | 16874        | 45.99 | 77.27 | -0.522|
| (GmCAMTA4)           |           | 3          |     |         |          |              |       |       |       |
| Glyma.05g117000      | 1088      | 121941.0   | 5.58| 150     | 123      | 16966        | 45.00 | 75.90 | -0.514|
| (GmCAMTA5)           |           | 2          |     |         |          |              |       |       |       |
| Glyma.08g178900      | 1081      | 121609.2   | 5.58| 150     | 117      | 16890        | 50.60 | 77.12 | -0.541|
| (GmCAMTA6)           |           | 2          |     |         |          |              |       |       |       |
| Glyma.17g038800      | 999       | 112283.9   | 6.81| 131     | 126      | 15695        | 40.79 | 80.39 | -0.490|
| (GmCAMTA7)           |           | 0          |     |         |          |              |       |       |       |
| Glyma.15g143400      | 911       | 103084.7   | 7.64| 105     | 106      | 14419        | 42.07 | 80.72 | -0.428|
| (GmCAMTA8)           |           | 4          |     |         |          |              |       |       |       |
| Glyma.09g038300      | 911       | 102969.3   | 6.55| 109     | 103      | 14385        | 43.48 | 80.42 | -0.420|
| (GmCAMTA9)           |           | 7          |     |         |          |              |       |       |       |
| Glyma.05g148300      | 983       | 109314.7   | 5.43| 132     | 103      | 15159        | 49.09 | 73.92 | -0.547|
| (GmCAMTA10)          |           | 4          |     |         |          |              |       |       |       |
| Glyma.18g005100      | 962       | 108036.7   | 5.39| 137     | 109      | 14981        | 46.47 | 76.13 | -0.520|
| (GmCAMTA11)          |           | 7          |     |         |          |              |       |       |       |
| Glyma.17g031900      | 922       | 103869.7   | 7.39| 104     | 104      | 14503        | 41.29 | 81.3  | -0.394|
| (GmCAMTA12)          |           | 0          |     |         |          |              |       |       |       |
| Glyma.07g242000      | 921       | 104293.9   | 7.03| 107     | 105      | 14545        | 41.60 | 79.60 | -0.455|
| (GmCAMTA13)          |           | 9          |     |         |          |              |       |       |       |
| Glyma.11g251900      | 910       | 102394.5   | 5.54| 130     | 106      | 14222        | 45.31 | 78.58 | -0.531|
| (GmCAMTA14)          |           | 6          |     |         |          |              |       |       |       |
| Glyma.08g105200      | 965       | 107767.1   | 5.81| 124     | 104      | 14945        | 50.69 | 73.30 | -0.552|
| (GmCAMTA15)          |           | 1          |     |         |          |              |       |       |       |

(aa- amino acids, MW- molecular weight, pl- isoelectric point, Asp- Aspartate, Glu- Glutamate, Arg- Arginine, Lys- Lysine, II- instability index, Al- aliphatic index)

**Phylogenetics and structure of GmCAMTAs**

As mentioned earlier, CAMTA TFs are multiple-stress responsive. Enrichment of *cis* motifs involved in
signal response in the promoters of Medicago CAMTA genes hints that they are likely to respond variedly to various signals like other CAMTAs (15). We dissected the regulatory region of GmCAMTAs (~2 kb upstream) online with PLANTCare which detected stimulus-specific cis motifs in their promoters. Overall, there are light (G-box, MRE and AE-box), drought (MBS), salt (MYB), pathogen (TC-rich repeats), wound (WUN, WRE), low temperature (LTR), gibberellin (GARE, P-box), auxin (AUXRE) and abscisic acid (ABRE) responsive cis elements as shown in figure 1A. The presence of multiple cis motifs in GmCAMTA genes represent their responsivity to multiple stimuli. Moreover, the cis element (CG-box) is the binding site of CAMTA TFs, thus CG-box presence within GmCAMTA promoters indicate the interaction of one GmCAMTA TF with another. The promoter analysis of GmCAMTA12 revealed MYC, MYB, MBS and G-box indicating its potential role in drought stress.

Using the online GSDS tool, the gene structure of all the GmCAMTAs was visualized in order to mutually compare their structural diversity. The length of GmCAMTA genes lie in the range between 3196 bp (GmCAMTA14) to 3947 bp (GmCAMTA6) with an average length of 3607 bp. The exons (yellow), introns (black lines) as well as 5` and 3` UTR regions (blue) of each gene are shown in figure 2B. A close observation of the number of exon-intron reveals a similar pattern such as except 3 genes (GmCAMTA7, GCAMTA10 and GmCAMTA15) which are comprised of 12 exons, the rest have 13 exons and 12 introns of variable length, showing their close mutual evolutionary relationship. This fixed numbering of intron and exon is a conserved characteristic of CAMTA which is descended from ancestors and is also demonstrated in the CAMTA family of other species (23). Similarly, except the last one/two exons, an ascending order in the exon size from 5` to 3` UTR can be observed across all the genes.

The four domains (CG-1, ANK, IQ and CaMBD) is the common conserved characteristic of all CAMTA TFs (23). Scanning the amino acid sequences of GmCAMTAs with online protein domains illustrator tool showed the same 4 domains in all 15 members (figure 1C). GmCAMTA TFs through IQ (calcium-independent)/CaMBD (calcium-dependent) interact with Calmodulin while through CG-1 domain, they bind the target DNA in a sequence-specific manner (CGCG/CGTG) at their promoter region. “ANK repeats” mediate protein-protein interactions. All these conserved domains along with other
properties make GmCAMTA proteins the “transcription factors”. The high sequence specificity is common in the Calmodulin binding domain of Arabidopsis and soybean CAMTAs as shown in figure 1D.

An ML (Maximum Likelihood) tree was constructed which traced the evolutionary relationship among the CAMTA gene families of soybean, Arabidopsis, maize and tomato (figure 1E). Using MEGAX, the evolutionary tree was constructed from the full length aligned CAMTA protein sequences of the four species. A total of 37 CAMTAs including 6 from Arabidopsis, 15 from soybean, 7 from tomato and 9 from maize clustered into 4 distinct groups, A, B, C and D. GmCAMTA1-6, AtCAMTA1-3, SICAMTA1 and 2 and ZmCAMTA3, 6, 7a and 7b might have co-evolved and thus clustered together in group A, representing the largest clade. Similarly, GmCAMTA10, 11, 14, 15, AtCAMTA4, SICAMTA3, 4 and ZmCAMTA1 clustered in group C showing their mutual high homology. GmCAMTA8, 9, 12 and 13 grouped with AtCAMTA5, 6, SICAMTA5, 6 and ZmCAMTA5 making clade D. Two orthologs (GmCAMTA7 and SICAMTA7) comprised group B, which is the smallest clade. Clustering in the phylogenetic reconstruction indicates more mutual similarity and probably weak homology to the members of the other three bigger clusters. It is noticeable that except B, in the rest of clades, CAMTAs of the same species are at the tips of the same branches and vice versa retaining their intraspecific homology.

**miRNA targets in GmCAMTA transcripts**

miRNA target prediction is important in finding their role in plant growth, development in normal as well as stress conditions (18). Keeping the Expectation score ≤ 5, the 639 miRNAs were scanned and top of the list miRNA (for each GmCAMTA transcript) with the minimum E. value was taken using the online psRNA TARGET tool (40). A total of 10 unique miRNAs were predicted which potentially target the GmCAMTA transcripts by inhibiting translation or through cleavage (Table 2). Their length ranges between 19 bp (gma-miR1533) to 24 bp (gma-miR343b). The accessibility of target site (UPE) which is associated with identification of target site and energy required to cleave transcript, varies from 11.8 (gma-miR1533) to 21.6 (gma-miR5780c). The translation of GmCAMTA1 and GmCAMTA2 transcripts is potentially inhibited by the common gma-miR5780c while gma-miR6299 cleaves GmCAMTA8, GmCAMTA9, GmCAMTA12 and GmCAMTA13. Similarly, GmCAMTA5 and GmCAMTA6 have potential
targets for gma-miR1533 while \textit{GmCAMTA11} and \textit{GmCAMTA14} are predicted to be cleaved by gma-miR2111b. gma-miR9726 is predicted to cleave \textit{GmCAMTA3} and gma-miR1522 \textit{GmCAMTA15}. These \textit{in silico} predictions require experimental validation which would extend our understanding of the mechanisms of \textit{Ca-CaM-CAMTA}-mediated stress tolerance in plants.

Table 2. \textbf{miRNA potential targets in \textit{GmCAMTA} transcripts}

| miRNA       | Transcript | Alignment                                      | E.V. | UPE Target Access. | Inhibition |
|-------------|------------|------------------------------------------------|------|--------------------|------------|
| gma-miR5780c| \textit{GMCAMTA1,2} | AGGUUCAGGGACUAUAUAUUU GAAAGGCCCCUAUAUAUAGA | 3.0  | 21.668             | Translation |
| gma-miR9726 | \textit{GMCAMTA3}  | CUUCUUUUUUAUUACGGAUAU AGAGGAAAGUACUGCUUAUA  | 4.0  | 18.124             | Cleavage   |
| gma-miR343b | \textit{GMCAMTA4}  | AGGCUGUUGGACUAGACAAUCU AACCAUCA--UUUGGUCUGUGGA | 4.0  | 19.255             | Cleavage   |
| gma-miR1533 | \textit{GMCAMTA5}  | AGUAUUAAUUAAAAAUAUA ACAUGAUUUUUUUUUUAUA  | 4.5  | 12.197             | Cleavage   |
| gma-miR5369 | \textit{GMCAMTA6}  | ACUGUAGGAAGAAAAGAGU UCACAUUUUUUUUUUCUUG | 4.0  | 11.815             | Cleavage   |
| gma-miR5041-5p| \textit{GMCAMTA7} | AACUCGUUCAACUCUCUACUU CAGAACAGUUGAAUAUGGAG | 3.5  | 13.098             | Cleavage   |
| gma-miR6299 | \textit{GMCAMTA8,9,12,13} | AGUGUUGGUUUAUUAAAAUUA AGCUAGUUGUUAAAUUGGAGU | 4.0  | 15.803             | Cleavage   |
| gma-miR5030 | \textit{GMCAMTA10} | GGCCAUUUUGUGUUAACACAGA UUGUAAUAUUGAAUGUGGACU | 4.5  | 18.387             | Cleavage   |
| gma-miR2111b| \textit{GMCAMTA11,14} | AUUUGGAGUCCUCUACGCUAAU AGACUCCUUGGGUGUAGAUC | 3.5  | 13.344             | Cleavage   |
| gma-miR1522 | \textit{GMCAMTA15} | UAAAGUAAAUUUCGUUAUU UAUUCUGGUACAAUAUU  | 5.0  | 14.966             | Cleavage   |

The potential miRNA target sites predicted with the online psRNATarget tool. Keeping a Max. Expectation value = 5, Max UPE= 25. E.V – Expectation Value, M – Multiplicity

\textbf{Chromosomal distribution and regulatory network of \textit{GmCAMTAs}}
The genome browser tool in NCBI mapped *GmCAMTA* genes to their respective chromosomes. The 15 *GmCAMTA* genes are unequally distributed over 8 out of 20 chromosomes of soybean as shown in figure 2. The figure depicts the complete size of each chromosome with the exact position of genes. Chromosome 8 has the highest number of *GmCAMTA* genes, i.e.; 4, while chromosome 7, 9, 11 and 18 have got one only in each. In prokaryotes, due to the absence of nucleus, transcription and translation occur simultaneously in coupling phase. On the other hand, translation of mRNA is always executed outside the nucleus in eukaryotic cytoplasm and the proteins that work in nucleus have a nuclear localization sequence (signal). Transcription factors also work in nucleus thus after their translation in cytoplasm, they are directed to the nucleus which is mediated through their NLS. To find NLS, protein sequences of each GmCAMTA were submitted to the online cNLS mapper tool at http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi. Keeping a cut off score of 5, at least one NLS in all GmCAMTAs and even more than one in some proteins were detected. Likewise, all of the 6 Arabidopsis CAMTAs possess only one NLS in the CG-1 domain of each protein (12). In contrast, rice CAMTAs have one NLS in the C-terminal and another in N of CG-1 domain (41). Experimental evidence shows that these domains perform diverse functions in the regulation of gene expression (42). The nuclear localization sequence in each of the 15 GmCAMTAs and predict that all of these transcription factors are localized in the nucleus.

In order to find the interaction network of GmCAMTA proteins to relate them with other pathways, the protein sequences were individually put in STRING database, which predicted a number of interactors (43). Thus they can aid in linking proteins of interest to other pathways and could lead to the discovery of novel pathways as well. The String database displayed a network of ~10 interactors for each GmCAMTA protein among which some sets were redundant. Thus a total of 48 unique proteins were predicted for 15 GmCAMTAs (Figure S1, Additional File 4). This vast interaction network (experimentally determined and software predicted) indicates the complex regulatory upstream/downstream pathways of CAMTAs. However, these in silico predicted interactions require experimental validation. Besides these predicted interactions, their orthologues in other species such as Arabidopsis or other legumes could also be exploited to search other potential interactors in
soybean. Information of the 48 proteins is mentioned in table S1, Additional File 3.

**GmCAMTAs as early drought stress-responsive TFs**

The spatiotemporal expression in roots and leaves under 0, 1, 3, 6, 9 and 12 hours of simulated drought stress is shown graphically in figure 3. In roots, *GmCAMTA2* was highly expressed during 3 hours of drought followed by *GmCAMTA7* and *GmCAMTA10*. In contrast, *GmCAMTA14* was downregulated during all the 5 time points followed by *GmCAMTA8*, *GmCAMTA9* and *GmCAMTA11*. Overall, these transcription factors upregulated abruptly during 1 and 3 hours of stress and downregulated afterwards (figure 3a). The expression profile of *GmCAMTA*s in leaves at different stress durations is different as compared to roots (figure 3b). In leaves, they look uniform except *GmCAMTA4* which is the highest upregulated gene followed by *GmCAMTA5*, *GmCAMTA11* and *GmCAMTA12*. Interestingly, like roots, majority of these 15 genes retained the 3 hours trend in the leaves as well. It is obvious that the expression is relatively the highest at 3 hours after which it decreased until 12 hours.

The differential expression of *GmCAMTA* family is a result of their tightly regulated transcription. We speculate that in stress conditions, though the calcium ions continuously convey the stress signals through calcium signatures to the cytoplasm as well as nucleus however the intensity/amount of these signals is weighed and adjusted by the next signal relaying molecules, such as CaM, before conveying to CAMTA transcription factors. At times, this signal transduction to CAMTA is continuous with the same intensity, however, after certain period (3 h in our case), CAMTAs response is not the same throughout the course and seems to be unconcerned and even downregulated as the stress period continues. From the control samples (0 h) in leaves and roots, it is also obvious that the expression of all CAMTAs is active at all times. In brief, the spatiotemporal expression pattern revealed that *GmCAMTA*s are upregulated in the early phase of drought thus are early drought stress-responsive transcription factors.

**Arabidopsis overexpressing GmCAMTA12 exhibited enhanced drought tolerance**

To evaluate the contribution of GmCAMTA12 protein in drought stress, we engineered Arabidopsis plants (*35s:GmCAMTA12*) to constitutively express *GmCAMTA12* gene under 35s promoter. Prior to
Arabidopsis transformation, the *Agrobacterium tumefaciens* strain EHA105 transformants harboring overexpression cassette were verified (through PCR). (Figure S2, Additional File 4). After floral dip, several overexpressing lines were obtained of which we selected two independent stable homozygous lines (OE5-Overexpressing *GmCAMTA12*-Line5 and OE12-Overexpressing *GmCAMTA12*-Line12) for functional analysis. The expression of *GmCAMTA12* in two transgenic Arabidopsis lines was validated through qPCR.

For drought assay, the two lines of OE *GmCAMTA12* (OE5 and OE12) and the wild type plants were subjected to drought stress by withholding water for two weeks and then re-watered as shown in figure 5A. Initially, all the plants were growing normally until water was withheld. However, upon encountering drought, nearly all the wild type and transgenic Arabidopsis stopped growth, wilted and started turning yellow afterwards. After 14 days of continuous drought treatment, most of the wild type plants were completely dried as obvious from their phenotype (dried leaves). Unlike wt, most of the OE lines retained life processes which was evident from chlorophyll they retained in their leaves. After re-watering, majority of transgenic Arabidopsis rejuvenated but most of the wild type did not. The plants were then allowed to grow under normal conditions until seed harvest. As expected, the seed yield of wt and transgenic lines was unequal and OE lines developed more seeds than wild type Arabidopsis. Under well-watered conditions, the survival rate in soil under drought was ~100%, however withholding the water for two weeks and then re-watering, less than 60% wt plants survived while OE5 and OE12 showed ~83% and ~87% survival rate as shown in figure 4D. Obviously, the constitutive overexpression of *GmCAMTA12* had enhanced the drought survival efficiency of transgenic Arabidopsis leading to better growth and development.

In their root length assay on ½ MS-mannitol medium as shown in figure 4B, wt plants grew longer roots than OE plants at 0 mM concentration of mannitol, however, on MS-mannitol, OE plants specifically OE12 developed longer roots than wt at all three concentrations (50 mM, 100 mM and 200 mM) of mannitol. Interestingly, the roots of OE12 plants were the longest at 200 mM mannitol. Root length (cm) is shown graphically in figure 4E.

Between the two overexpression lines, OE12 performed better than OE5 in both drought assays
(Figure A, B), thus we subsequently inoculated seeds of wt and OE12 on ½ MS with various concentrations of mannitol to evaluate the germination rate (Figure 4C). As expected, we observed higher germination rate of transgenic line OE12 than that of wt at all four concentrations of mannitol. The germination rate was ~30% higher than wt at 50 mM mannitol. At 100 mM, the germination rate decreased in both types at a similar pace (~75 % in OE and ~45 % in wt). As the mannitol concentration increased to 150 and 200 mM, we saw a dramatic decline in the germination efficiency of wt (30% and 25%) as compared to OE (>65% and >60%) as shown in figure 4F. We can say that the constitutive expression of GmCAMTA12 has enhanced the germination efficiency of transgenic seeds under drought.

In order to check the performance of wt and OE lines at physiological level, the physiological indexes like proline and MDA contents, CAT activity and relative electrolyte leakage were determined in all plants subjected to stress. Under well-watered conditions, we observed no significant difference in the level of proline contents which was quite low in wt and OE lines. In contrast, proline contents calculated in drought-stressed plants had significantly elevated. In wt, the average value of proline was ~400 µg/g, while in OE, it was recorded in the range of 850 to 900 µg/g (Figure 4G).

Malodinaldehyde (MDA) is a well-known biomarker for sorting out stress-induced membrane damage due to oxidative stress. MDA contents in wt and OE plants during normal conditions matched mutually but its level was doubled in the drought-stressed wt plants compared with drought treated OE lines (figure 4H). Catalase (CAT) is a major antioxidant enzyme which accumulates during abiotic stresses and scavenges \( \text{H}_2\text{O}_2 \). The CAT activity in wt and OE plants under usual conditions was nearly equal (figure 4I) however, drought treatment enhanced CAT activity in transgenic plants as compare to wt. During stress, electrolytes, specifically \( \text{K}^+ \) ions leak out of the cells through various channels and thus damage due to stress could be monitored by comparing the electrolyte leakage in wt and transgenic lines. We determined relative electrolyte leakage (REL) in wt and OE lines during normal as well as drought conditions. REL % was nearly levelled under well-watered conditions but the leakage was more pronounced in wt as compared to OE lines during drought (figure 4J). Noticeably, the amino acid
sequence analysis revealed high level of sequence similarity between GmCAMTA12 and AtCAMTA5. We can speculate that GmCAMTA12 being a transcription factor interacted with the downstream target genes (including AtCAMTA5 interactors) and modulated their expression in transgenic Arabidopsis which contributed to their better performance under drought (determined at molecular level in later section).

**GmCAMTA12 overexpression regenerated more developed and drought-efficient hairy roots in soybean**

For further functional validation of GmCAMTA12 in response to drought, the hairy roots system was exploited to overexpress the target gene in soybean. Prior to generating transgenic hairy roots, the *Agrobacterium rhizogenes* strain K599 cells harboring control vector (VC) and OEGmCAMTA12 were verified through gene-specific PCR (Figure S3, Additional File 4). After 1-2 weeks of infection of soybean seedlings with K599 (VC and OE), hairy roots had started regenerating with various frequency. Prior to subsequent experiments, we made sure that the hairy roots were transgenic by using gene-specific PCR from the genomic DNA of a small portion of hairy roots. Using qPCR, we validated the overexpression of the target gene in OE roots which was over 3 times higher than the VC hairy roots. After hairy roots had become long and strong enough by growing for two more weeks, the primary roots were excised and the chimeric plants (with transgenic roots and non-transgenic shoots) were shifted to fresh vermiculite.

When the transgenic roots were ~ 10 cm long, the VC and OE chimeric soybean plants were extirpated from vermiculite and transferred to Hoagland solution as shown in figure 5A. After a 3 days acclimation, both VC and OE chimeric plants were treated with 6% PEG6000. The plants started to wilt with leaves curling and shoot apexes drooping on encountering drought. However, the wilting was more apparent in VC than in OE chimeric plants as VC leaves and apexes were more wilted.

*GmCAMTA12 overexpression induced profuse hairy roots* (figure 5B) due to which the aerial non transgenic part of the OE chimeric soybean plant was also more developed than the VC chimeric plants. The roots were analyzed using the root scanning (figure 4C). The OE hairy roots showed higher values with total root length (figure 5J), surface area (figure 5K), root volume (figure 5L), number of
branches (figure 5M) and projected area (figure 5N).

The proline and MDA contents, CAT activity and relative electrolyte leakage were determined to check the impact of GmCAMTA12 overexpression at physiological level. In control samples (Hoagland), proline contents had nearly equal amount in VC and OE hairy roots, however under drought, proline content level was recorded significantly higher in OE hairy roots (figure 5F). MDA shows the level of membrane damage as it is the final product of lipid peroxidation. In the absence of stress, MDA contents in OE type were slightly less than VC hairy roots, however with PEG treatment, VC had substantial increase in MDA contents as compare to OE roots (figure 5G). In contrast, CAT activity was significantly higher in OE hairy roots than VC under drought stress (figure 5H). For REL, VC hairy roots had higher electrolyte leakage (%) than OE in response to drought (figure 5I). Comparative physiology of OE and VC hairy roots shows that the overexpression of GmCAMTA improved the drought tolerance of OE.

To analyze their growth efficiency under drought stress, 0.1 g of VC and OE hairy roots was weighed under sterile conditions and inoculated on GM media containing 4 various mannitol concentrations (50 mM, 100 mM, 150 mM and 200 mM) including a control (0 mM mannitol) with three replicates of each type (Figure S4, Additional File 4). All the plates were kept in dark in growth room with 28 °C for 10 days of culturing after which the fresh and dry weights were determined. The control samples of both types (VC and OE) hairy roots had nearly same fresh and dry weights however on stress media, OE hairy roots showed better performance as the weight of OE roots was more than VC roots at all 4 concentrations of mannitol (figure 5D, 5E). It further flaunted that the overexpression of GmCAMTA12 has improved the drought survival efficiency of OE roots.

Expression analysis of GmCAMTA12 orthologues’ regulatory network in Arabidopsis

In Arabidopsis, AtCAMTA5 is the orthologue of GmCAMTA12, thus the regulatory network of AtCAMTA5 was predicted with STRING database. To test our hypothesis whether the transcript abundance of GmCAMTA12 TF in transgenic Arabidopsis would interact with genes in the regulatory network of AtCAMTA5, we analyzed the expression of 10 interactors predicted with STRING database (figure 6A).
Total RNA from drought treated WT and OE plants (figure 6B) was isolated and reverse transcribed into cDNA. Using cDNA and gene-specific primers, we conducted a qPCR which deciphered the differential expression pattern of the 10 genes in wt and OE plants (figure 6C). Among them, NIP30 (NEFA-interacting nuclear protein - AT3G62140) which in humans is involved in negative regulation of proteasomal protein catabolic process, is slightly upregulated in response to drought. AtWRKY14 (AT1G30650) encoding WRKY transcription factor14 possesses a DNA binding domain and specifically interacts with W box (a common elicitor-responsive cis element). WRKY14 is also nuclear localized transcription factor and regulates many important processes through gene regulation. GmCAMTA12 overexpression upregulated AtWRKY14 which indicates that along with other transcription factors including WRKY TFs, the spectrum of CAMTA TFs regulatory networks is much wider than we think. Thus the mutual interaction of CAMTA and WRKY should be further investigated. Peptdyl-prolyl cis-trans isomerase AtCYP59 (AT1G53720) mediates posttranslational modifications specifically protein folding. With GmCAMTA12 overexpression, the upregulation of AtCYP59 is almost equal to that of WRKY14. AtANN5 (AT1G68090) is a calcium binding protein, plays role in pollen development and is induced by cold, heat, drought and salt stresses. As expected, its expression is relatively the highest among the 10 interactors. SR (Serine racemase- AT4G11640) is involved in serine biosynthetic pathway. Its expression also seems to elevate with GmCAMTA12 overexpression in OE Arabidopsis. ELO3 (Elongator complex3- AT5G50320) is a part of Elongator multiprotein complex and regulates initiation and elongation of transcription. No apparent change in the expression level of ELO3 in wt and OE was observed. Like ANN5, CaMHSP (Calmodulin Binding Heat Shock Protein - AT3G49050) also called as Alpha/beta-Hydrolases superfamily protein exhibited higher transcript level. It is involved in lipid metabolic pathway. B120 (G-type lectin S-receptor-like serine/threonine-protein kinase - AT4G21390) is involved in protein kinase activity and recognition of pollen. Its expression level was downregulated in transgenic Arabidopsis. AT2G43110 (U3 containing 90S pre-ribosomal complex subunit) was also upregulated with the overexpression of GmCAMTA12. AT3G19850 (BTB/POZ domain-containing) mediates protein degradation by facilitating ubiquitination. Its expression level elevated with GmCAMTA12 upregulation. All of these genes possess CGCG/CGTG motif in their
promoter sequences which also validates the sequence-specific binding of CAMTA TFs. Most of these interactions are based on text-mining and should be determined experimentally. The genomic, CDS and proteomic sequences of these 10 interactors can be found in (Table S4. Additional File 5)

**GmCAMTA12 overexpression orchestrated downstream genes in transgenic hairy roots**

In order to find whether the constitutive overexpression of *GmCAMTA12* modulates the genes with which *GmCAMTA12* TF interacts, we selected 8 genes in the *GmCAMTA12* regulatory network in soybean predicted with STRING (figure 6D). In chimeric soybean plants possessing VC and OE (figure 6E) hairy roots treated with 6% PEG6000, we analyzed the expression pattern of the eight of the predicted interactors of *GmCAMTA12* to know their *GmCAMTA12*-mediated regulation. Using gene-specific primers, qPCR of the selected interacting proteins in transgenic hairy roots as well as non-transgenic leaves was carried out. The genes displayed a differential expression profile in the control and drought treated chimeric soybean plants. In VC roots (figure 6H), GmNIP30 (NEFA interacting protein - *GLYMA19G40910*) was slightly upregulated in response to drought and its expression was indifferent to *GmCAMTA12* overexpression in OE roots (figure 6I). GmPLA1-IId (Phospholipase A1-IId - *GLYMA12G15430*) involved in lipid metabolic process was upregulated in VC roots under drought stress while in OE roots, its upregulation was two-fold higher than that in VC. GmNAB (Nucleic Acids Binding - *GLYMA18G40360*) was downregulated in VC roots while upregulated with *GmCAMTA12* overexpression in OE roots in response to drought. The expression of GmELO (Catalytic histone acetyltransferase subunit of the RNA polymerase II elongator complex - *GLYMA06G18150*) was the highest and it equally expressed in both VC and OE roots. GmSR1 (Serine racemase-1 involved in D-Serine biosynthetic process - *GLYMA05G37930*) was upregulated in VC roots however, with *GmCAMTA12* overexpression, GmSRI was downregulated in OE roots under drought. Similarly, GmSR2 (Serine racemase-2 - *GLYMA08G01670*) was repressed in the absence of *GmCAMTA12* overexpression, however in OE roots under drought, it was slightly upregulated. Simultaneously, GmUC1 (Uncharacterized but possessing a Myb-DNA binding domain - *GLYMA20G32540*) transcript was slightly upregulated in VC while it was significantly induced in OE hairy roots in response to drought. In contrast, GmUC2 (Uncharacterized2 - *GLYMA17G07200*) was positively regulated in VC
and negatively regulated in OE roots. Unlike roots, the expression profile of these 8 genes was nearly similar in chimeric soybean leaves possessing VC (figure 6hF) and OE (figure 6G) hairy roots in response to drought. Five out of eight of these genes possess CGCG/CGTTG motif. The genomic, CDS and proteomic sequences of the 8 genes can be found in Table S5, Additional File 5.

Discussion
Calcium as a ubiquitous secondary messenger orchestrates almost every cellular process in response to environmental stimulus. Plants employ the divalent cation, calcium (Ca$^{2+}$) in relaying these endogenous (developmental) and exogenous (environmental) signals to appropriate cellular responses. Calcium alone specifically encodes a myriad of distinct signals by using spatial and temporal Ca$^{2+}$ spikes as well as the frequency and amplitude of Ca$^{2+}$ oscillations (10). Next to the secondary messenger lies the signal relaying molecules including Calmodulin, which further tune the calcium signatures and pass them to CAMTAs. Calmodulin Binding Transcription Activators have a short history of two decades, which after first reported have been genome-wide identified in numerous plant species (23). CAMTAs are important in the sense that they are the transcription factors and an intermediate in the calcium-mediated (Ca-CaM-CAMTA) stress signaling.

Earlier, the identification and expression analysis of Soybean CAMTAs was conducted (21) however, their functional characterization remained unexplored. Thus, in order to better comprehend the role of GmCAMTAs in drought, and take a holistic snapshot, we first attempted to fill the research gaps through comprehensive in silico analysis of GmCAMTA family. Soybean possesses 15 GmCAMTAs (21), the second highest number after Brassica napus, which possesses 18 genes (24). To our curiosity, such a large stress responsive transcription factor family must have substantial contribution to the drought tolerance of soybean. CAMTA has important role in an array of biotic/abiotic stresses as reported in earlier studies in Arabidopsis (22,31-34,44), tomato (14,28), tobacco (18), Brassica napus (24), Arabidopsis (36) and wheat (25). Dissecting GmCAMTAs with various bioinformatics tools, their gene structure depicted 13 exon pattern which is consistent with its orthologues in Arabidopsis, maize, tomato and others (23). Promoter enrichment analysis revealed the cis-elements including ABRE, DRE, G-box, W-box, WRKY, ARE and MYB, all of which respond to various stresses (21).
$SI_\text{CAMTAs}$ are differentially expressed during the development and ripening of tomato and the presence of $ERE$ cis-element in all $GmCAMTAs$ provides the basis for $CAMTA$ role in fruit development and ripening (28). Light responsive elements ($G$-box) are common in all $GmCAMTAs$ and we suggest that their role in light stress should be investigated. Also, their specific stress-responsive cis-motifs should be tested individually or collectively in designing stress-inducible synthetic promoters. Protein structure is important to know for understanding their action mechanism. The major basic domains CG-1, ANK, IQ and CaMB are common in all CAMTAs and a close look into the motif sequence of CaMB domain shows residues which are highly conserved across the species. Phylogeny of CAMTAs of 4 species traced the evolutionary relationship among the homologues as well as orthologues which is consistent with the previous results (15,17-19,21,23,24). Some homologues show more similarity and might have co-evolved (figure 1). Later the same set of proteins were found to interact with these homologues. A protein interaction network analysis found a few experimentally determined as well as predicted interactions together of 48 unique proteins of various pathways (figure 2). Thus we recommend to experimentally determine those interacting proteins to unveil $GmCAMTA$ TFs link with the pathways of the interacting proteins. In Arabidopsis, 32 proteins were predicted with String in the interaction network of $AtCAMTAs$ (23). An experimental-based detailed map of all the $GmCAMTA$ interactors would deepen our understanding of intricate mechanisms of $GmCAMTA$-mediated development and immunity against biotic/abiotic factors. Likewise, miRNAs are important transcriptional regulators by directly cleaving their target transcripts thus their potential target sites in $GmCAMTA$ was important to determine to understand $CAMTA$-mediated regulation. In silico analysis of $PtCAMTAs$ revealed potential targets for 4 distinct miRNAs (18). A set of 10 unique miRNAs was detected bringing more complexity to $CAMTA$-mediated stress response mechanisms in soybean (table 2). We recommend to experimentally determine the miRNAs targeting $GmCAMTAs$ which might also lead to engineer soybean and other crops against drought more accurately.

The expression analysis of $GmCAMTAs$ in soybean leaf and root reveals that although all of them express constitutively but $GmCAMTA2$, $GmCAMTA4$, $GmCAMTA5$, $GmCAMTA11$ and $GmCAMTA12$ are highly upregulated against drought. Secondly, almost all $GmCAMTAs$ expressed and peaked in 3 h
after which they were repressed indicating their early responsivity to drought stress (figure 3). As expected, overexpressing GmCAMTA12 in Arabidopsis and hairy roots and drought assays thereof validated the previous and current qPCR-based results. Interestingly, GmCAMTA12 enhanced the drought survivability and growth performance of transgenic Arabidopsis (figure 4) as well as hairy roots (figure 5) which was validated at phenotypic, physiological and molecular level. Like GmCAMTA12, another transcription factor from soybean GmNAC85 is also stress-responsive and its constitutive overexpression significantly enhanced the drought tolerance in transgenic Arabidopsis (45). Similarly, RNAseq of soybean overexpressing GmWRKY54 revealed that GmWRKY54 conferred drought tolerance by enhancing ABA/Ca$^{2+}$ signaling to close stomata as well as regulating numerous stress responsive transcription factors (6). In contrast, AtCAMTA1-mutant Arabidopsis exhibited enhanced drought sensitivity while also affecting the expression of other drought responsive genes (31). A recent global transcriptome analysis using RNA-seq revealed that a large number of genes involved in diverse stress responses are regulated either directly or indirectly by AtCAMTA3 as about 3000 genes were misregulated in the AtCAMTA3-mutant Arabidopsis (34). We recommend to experimentally verify the downstream targets of GmCAMTAs through protein-protein interaction experiments including two-hybrid/pull down assays, which might link/lead to novel pathways. GmCAMTA12 overexpression altered the expression of the genes in the regulatory network of AtCAMT5, which is the orthologue of GmCAMTA12 in Arabidopsis. All of the 10 interacting genes possess CGCG/CGTG motif in their 2000 bp upstream region validating the sequence-specific interaction of GmCAMTAs (12). Likewise, the expression pattern of the interactors of GmCAMTA12 in chimeric soybean plants showed no change in leaves but in roots the three genes, GmELO, GmNAB and GmPLA1-IId displayed comparatively higher expression as a result of GmCAMTA12 overexpression. It means that the overexpression in hairy roots is limited to only transgenic roots and did not affect its interactors in aerial non-transgenic part however, the more developed hairy roots did promoted healthier shoot.

Conclusion
In conclusion, the better performance of OE Arabidopsis and chimeric soybean plants on MS-mannitol
and Hoagland-PEG was validated at physiologic and molecular level. The comparison of Proline and MDA contents, CAT activity and relative electrolyte leakage between VC and OE hairy roots shown graphically, flaunts the improved development of chimeric plants with hairy roots overexpressing GmCAMTA12. It was further verified at molecular level by determining the transcript level of the genes in the regulatory network of GmCAMTA12. The GmCAMTA overexpression should also be analyzed in other tissues including soybean flower and seeds as well as in various developmental stages to find its integrated role in processes besides stress tolerance.

However, GmCAMTA12 nuclear localization, promoter-GUS analysis, CRISPR/cas9-mediated gene-knockout and global transcriptomics of stably transformed transgenic soybean and Arabidopsis are under investigation. Taken together, a more systematic approach should be adopted to decipher the integrated role of GmCAMTA TFs with the Calcium-Calmodulin signaling crosstalk in drought.

Methods

In silico analysis

By keyword search (Gene ID/locus/accession no.) in three databases (NCBI https://www.ncbi.nlm.nih.gov/, Phytozome https://phytozome.jgi.doe.gov/pz/portal.html, Plantgrn.noble http://plantgrn.noble.org/), we retrieved the genomic, transcriptomic and proteomic sequences of the fifteen members of Glycine max CAMTA gene family. Similarly, proteomic sequences of the corresponding orthologues in Arabidopsis thaliana, Zea mays and Solanum lycopersicum were also retrieved and a dataset was created for bioinformatics analyses. See Additional File 1 for the datasets created.

The physico-chemical properties including molecular weights, theoretical isoelectric points, Aspartate + Glutamate, Arginine + Lysine, number of atoms, instability index, aliphatic index and GRAVY (Grand average of hydropathicity) of all GmCAMTA proteins were calculated using the ProtParam tool in the ExPASy (https://web.expasy.org/protparam/).

To trace the evolutionary relationship, an ML tree was constructed with MEGAX using default parameters after multiply aligning the protein sequences of Glycine max, Arabidopsis thaliana, Zea mays and Solanum lycopersicum. The aligned genomic and proteomic sequences were also screened
for conserved sites/domain using BioEdit/ClustalX (Additional File 1).

The exons and introns along the length of GmCAMTA genomic sequences were monitored using the online tool, GSDS (Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/index.php) [13]. To check the cis-motifs within the regulatory regions of GmCAMTAs, (2 kb upstream) 5` UTR of each gene was resolved at online database PLACE (https://sogo.dna.affrc.go.jp/). The promoter sequences are in Additional file 1. All the 15 GmCAMTA protein sequences were aligned with ClustalX and the conserved motifs were identified using the online SMART tool (Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de/). The CaMBD in Arabidopsis and Soybean CAMTA family proteins was particularly analyzed for conserved sequences. The alignment is shown in Additional file 2.

The potential miRNA targets along the length of GmCAMTA transcripts were predicted online at psRNATarget server (http://plantgrn.noble.org/psRNATarget/). To predict the proteins interacting with GmCAMTAs, each protein was separately submitted to STRING database (https://string-db.org/). The online prediction software displayed the proteins (experimentally proved/textmining-based) interacting with GmCAMTAs. All the unique interactors of whole GmCAMTA family predicted with STRING are listed in Additional File 3, Table S1.

The loci of CAMTA family over 20 chromosomes were determined using NCBI Genome Data Viewer at https://www.ncbi.nlm.nih.gov/genome/ggv/. Similarly, the subcellular localization of GmCAMTA transcription factors was determined by screening NLS in the protein sequence with online tool cNLS mapper at http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi. NLS (nuclear localization signal/sequence) is an amino acid sequence which if present in a protein characterizes it to be working in nucleus.

Primers for qPCR analysis were designed with Primer premier 5 software as well as NCBI online primer tool https://www.ncbi.nlm.nih.gov/tools/primer-blast/. The multiple alignment of GmCAMTAs CDS showed frequent conserved sequences so to minimize ambiguity, the primers were designed with special care against the unique sites in each GmCAMTA CDS. Moreover, is was ensured that the primers amplify all of alternative transcripts of respective gene. We attempted to design the pair
targeting two exons amplifying a stretch of 200-300bp cDNA. Those primers which failed to give amplification, created more than one peak or did not appear on agarose gel under UV were redesigned until corrected. Moreover, the primers specificity was determined with online nBLAST tool in NCBI [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome]. Suzhou GENEWIZ Biological Technology Services Company, China, synthesized all primers for this experiment listed in Additional File 3, Table S2.

Expression analysis

The seeds of soybean variety “Willliams 82” (kindly provided by the Engineering Research Center of the Chinese Ministry of Education for Bioreactor and Pharmaceutical Development, Changchun) were washed with 75% ethanol and then sterile water. They were hydroponically cultured in Hoagland nutrient solution in a growth room with ~28°C room temperature, 16-8 h (light-dark) photoperiod and a relative humidity of 50%. When the plants have opened their unifoliate leaf pair completely, they were subjected to 6% (PEG 6000) stress simulated drought. The plants were treated with drought for 1, 3, 6, 9 and 12 h in triplicate along with control (Hoagland). At the exact time points, we collected leaf and root samples, froze in liquid nitrogen and stored at -80C for further processing.

Total RNA from the root and leaf samples (ground in liquid nitrogen) of soybean plants (treated with PEG for various time periods) was extracted using RNAiso Plus (Takara, Japan) according to the manufacturer’s protocol. RNA concentration was determined with NanoDrop2000 Spectrophotometer (Thermo Scientific). The quality of RNA samples was checked over 1% agarose gel using 0.5X T.B.E. buffer. Bright and clear bands of 25S/18S RNA showed the intact RNA. The samples which showed degraded RNA or relatively unclear bands were re-extracted until intact RNA was detected. The concentration of all good quality RNA samples was adjusted to ~500ng/ul and it was reverse transcribed into cDNA using the PrimeScript RT reagent kit with gDNA Eraser (Takara) according to the company’s protocol. After the removal of genomic DNA with gDNA eraser, 1 ug total RNA was used as template to synthesize cDNA and stored at -20C.

After the reverse transcription of RNA into cDNA and qPCR primer synthesis, qPCR was run according
to the experimental design. Prior to the qPCR of all the samples, a standard curve was created with a reaction using a serially diluted cDNA in duplicate. Actin 11 was used as reference with leaf cDNA while Elongation factor α with that of root. Before determining the expression pattern of soybean CAMTA gene family in leaves and roots, the personal error was minimized by the standardization of experiment. An initial reaction was run on Stratagene Mx3000P thermocycler and primers validity was determined first by confirming the single peak (amplification plot), the ct value for each gene, and then by running the reaction product on agarose gel. Each product gave a single band parallel to right marker band. A 20 µL reaction mixture contained 10µL SYBR Premix Ex Taq, 0.4µL ROX Reference Dye II, 0.4µL of each gene-specific Primer, 2µL of cDNA, and 6.8µL ddH2O. High PCR efficiency is indispensable for robust and more precise RT-qPCR. The formula $E = 10^{(-1/Slope)-1}$ is used to calculate amplification efficiency and the slope of the calibration curve. The primers having ~100% PCR amplification efficiencies were used. The RT-qPCR profile for our samples consisted of an initial denaturation of 95°C for 2 min followed by 40 cycles of 94°C for 5s and 62°C for 20s. The fold change in relative expression level was evaluated using $2^{-\Delta\Delta Ct}$ formula.

**Gene transformation and Drought assays**

From the bioinformatics as well as qPCR analyses, we selected the gene GmCAMTA12 (Glyma.17G031900) to clone and transform for functional validation. The complete CDS (2769 bp) of GmCAMTA12 was PCR-amplified using cDNA as template, forward primer ACTAGTATGGCGAATAACTTAGCGG and reverse primer CCCGGGCTATGTCTTCAGTTGCATGTCAA. For amplification, LA Taq kit (Takara, China) was used according to the manufacturer’s protocol. The PCR profile was; an initial denaturation at 94 °C for 1 min followed 35 cycles of 94 °C for 30 sec, 55 °C for 30 sec and 72 °C for 2 min 30sec, and a final extension at 72 °C for 10 minutes. Using CT101 cloning kit (Transgene, China), the gene was cloned into pEASY®-T1 cloning vector and transformed into Trans1-T1 chemically competent cells (Transgene, China) following heat shock method according to the manufacturer’s instructions. The positive clones were obtained on a selective LB-Kan⁺ agar plate and screened through vector and gene specific PCR as well as restriction. After double check, three
separate clones were sequenced using M13 forward and reverse primers. When the sequences of all three clones were analyzed without any point mutation, the gene was restricted from pEASY®-T1 and sub-cloned by restricting from cloning vector using SpeI and Smal (Takara) and ligated into the corresponding sites of expression vector pBASTA (with Kanamycin and Glyphosate resistance genes) using T4 Ligase (Promega) constituting the recombinant overexpression construct, pBASTA-GmCAMTA12. The pBASTA-GmCAMTA12 construct was transformed into DH5α strain of E.coli (Transgene) and positive clones were obtained on a selective LB-Kan+ agar plate. After a double check (PCR and restriction) of single colonies harboring the overexpression construct, the colonies were preserved in sterile 80% glycerol. The recombinant vector was transformed into chemically competent Agrobacterium tumefaciens (EHA105) and Agrobacterium rhizogenes (K599) by a 5 min liquid nitrogen treatment followed by a heat shock of 5 min at 37 °C. The positive clones of EHA105 were identified on YEP Kan+Rif plates while K599 were selected using YEP Kan+Strep plates. After PCR verification, the cells were preserved at -80 °C for future use.

Arabidopsis Col-0 seeds were kindly provided by Engineering Research Center of the Ministry of Education of Bioreactor and Pharmaceutical Development, Changchun. To synchronize the germination of seeds, wt seeds were soaked in water and kept at 4 °C for 48 hours and then sowed on a mixture of humus + vermiculite (3:7) in a growth room in dark at 23 °C. The germinated seedlings were then allowed to grow in 16 h photoperiod until 40 days with regular watering. We transformed the inflorescence (unopened flowers) of mature healthy Arabidopsis plants through Floral Dip method (46) and harvested T₁ (Transgenic 1) seeds. The T₁ seeds were germinated in excess under the same conditions as for wild type seeds. A primary screening of T₁ plants (at six leaf stage) was carried out by spraying with 1% Basta (glyphosate) and the basta-survived plants were grown in separate pots in fresh soil (humus + vermiculite) under the same conditions as for wild type. Each plant represented a separate transformation event. The non-transformants died while from the green plants, we extracted genomic DNA from leaves by 2X CTAB and a PCR with gene specific primers using the gDNA as template was performed for secondary screening. T₁ plants positive with PCR were grown to harvest
T₂ seeds. Similarly, T₃ seeds from homozygous lines were harvested following the primary screening by 1% Basta and secondary with PCR to ensure stable transformation. The seeds of Arabidopsis were surface sterilized by keeping in 70 % ethanol for 1 min followed by 50 % bleach for 10 min after which the seeds were washed 6 times by sterile water. The wild-type and two T₃ transgenic Arabidopsis lines were germinated under sterile conditions on half MS plates added with various mannitol concentration. For germination, we vernalized wt and OE seeds by keeping in dark for 3 days at 4 °C. Each petri plate was marked and divided in to three parts, one of which was allotted to wt while the other two for Line-5 (OE5) and Line-12 (OE12) seeds. The seeded plates were kept in a growth chamber with 16 h photoperiod and 23 °C temperature. The germination of each line was observed and recorded for one week and the germination rate was determined by the number of seeds germinated divided by the total number of seeds. For root length analysis, wt and T₃ lines were first germinated on MS media under sterile conditions and one week old seedlings of same length were transferred to square plates containing MS with various concentrations of mannitol. The plates were placed in a rack vertically so that the roots grow downwards to ground. The root growth was observed and the plants photographed. The seeds of wild-type and two T₃ transgenic Arabidopsis lines were germinated and grown for one month in soil at 23 °C, watered regularly and then subjected to drought stress by withholding water for 14 days and photographed before, during and on drought recovery after re-watering. Various physiological parameters like MDA (47) and proline contents (48) as well as CAT activity (49) and relative electrolyte leakage were determined. Cell membrane integrity is lost as K⁺ ions (a chief electrolyte) leaks out of the cell leading to cell death in stress. This measurement is an indicator of the cell damage caused by stress (50). To determine REL, leaf sample (~2 g) was thoroughly rinsed with deionized water and then subjected to vibration for 2 h at 25 °C in a test tube containing 10 mL deionized H₂O. Using a conductivity meter, we determined the conductivity of solution (C1). A second measurement of conductivity (C2) was taken after boiling the solution for 25 min and then cooled to RT. REL % was calculated using the formula (C1÷C2) 100.
Soybean hairy root is an established platform for the functional analysis of a gene by overexpression. For hairy roots, seeds of soybean variety JU 72 (provided by the Engineering Research Center of the Chinese Ministry of Education for Bioreactor and Pharmaceutical Development, Changchun) were washed with 75% ethanol and rinsed in sterile water a few times. The seeds were sown in pots containing autoclaved vermiculite and kept in a growth room with 16 h photoperiod and 28 °C temperature and watered regularly. In parallel, K599 from glycerol was streaked on YEP Kan + Strep plates and incubated at 28 °C for 48 h. A single colony was picked and inoculated in YEP containing Kanamycin and Streptomycin. The culture was incubated in a shaking incubator at 200 rpm and 28 °C for 48 hours. A 100 µl of inoculum rom the culture was spread on YEP Kan + Strep plates and incubated at 28 °C for 48 h. When the seedlings had just sprouted, the cotyledons unfolded and the first unifoliate leaves were not appeared yet, they were ready to infect with K599 harboring the overexpression construct. The *Agrobacterium rhizogenes* culture from the plate was picked with a needle and injected into the hypocotyl right at the base of cotyledon (51). K599 harboring empty pBASTA were used to regeneration of VC (Vector Control) hairy roots. After infection, the seedlings were covered with a transparent plastic to ensure high humidity. Within two weeks, hairy roots started to sprout from the site of infection at variable frequencies with at least one root per seedling. The soybean plants with hairy roots were allowed to grow for two weeks after which excess of autoclaved vermiculite was added to cover the hairy roots and watered regularly. We extracted the genomic DNA from both type of the transgenic roots to confirm VC and OE through gene specific primers. The roots were scanned with a scanner and analyzed with the software. After the hairy roots were ~10 cm long, the plants were uprooted, primary roots were cut and the chimeric plants (transgenic root and non-transgenic shoot) were transferred to fresh autoclaved vermiculite and watered regularly. It is important to notice that the shoot of each plant was cut after the second trifoliate leaves. Thus the newly grown leaves were compared between the chimeric plants VC and OE transgenic roots. To check the performance of transgenic hairy roots under drought, plants with 10 cm long transgenic hairy roots were transferred to Hoagland solution and after acclimation, they were treated with 6% PEG6000. Proline and MDA contents, CAT activity and REL
were determined in VC and OE hairy roots. Moreover, the 0.1 g of the hairy roots (VC and OE) were weighed in sterile conditions and grown on MS media at various concentrations of mannitol including 0 mM, 50 mM, 100 mM, 150 mM and 200 mM. After 10 days, fresh and dry weights of the hairy roots were determined. The dry weight was measured keeping hairy roots at 60 °C overnight.

Expression analysis of the GmCAMTA12 orthologue’s regulatory network in wt and OE Arabidopsis

The regulatory network of the orthologue of GmCAMTA12 in Arabidopsis was predicted online with STRING database. The primers for qPCR were designed against the 10 genes with Primer-BLAST tool (Table S4). In parallel, RNA from wt and OE lines was extracted, quantified and reverse transcribed into cDNA with Takara kit following kit’s instructions. To determine the expression pattern of the 10 genes, qPCR was run using the primers of each gene with three biological replicates. AtActin11 was used an internal reference. The data was analyzed with 2^{-ddCt} method.

Analysis of GmCAMTA12 regulatory network in chimeric soybean plants

Using STRING database, the interactors of GmCAMTA12 in soybean were predicted. The primers for qPCR of 8 genes designed with Primer-BLAST are given in table S5. To determine the effect of overexpression of GmCAMTA12 on its interacting proteins, RNA from all samples was extracted using RNAiso plus (Takara) and reverse transcribed into cDNA with Takara rt kit and qPCR was run for all samples in triplicate. Actin 11 was used as internal control. Relative expression level was determined using 2^{-ddCt} method.

Abbreviations

| Abbreviation | Full Form                        |
|--------------|---------------------------------|
| CAMTA        | Calmodulin Binding Transcription Activator |
| CaM          | Calmodulin                      |
| OE           | Overexpression                   |
| PEG          | Poly Ethylene Glycol            |
| TF           | Transcription Factor            |
| VC           | Vector control                  |
| MDA          | Malondialdehyde                 |
| CAT          | Catalase                        |

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: All authors read and approved the final manuscript.

Availability of data and materials: The dataset supporting the conclusions of this article are
included within the article and its additional files.

**Competing interests:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**Funding:** This research was supported by the Special Program for Research of Transgenic Plants (2016ZX08010-002), the National Key Research and Development Program (2016YFD0101005), the National Natural Science Foundation of China (31601323) and the Natural Science Foundation of Science Technology Department of Jilin Province (20170101015JC, 20180101028JC, 20190201259JC).

**Author Contributions:** Conceptualization, H.-Y.L.; methodology, W.-C.L. and M.N.; software, M.N.; validation, A.J.; investigation, M.N.; resources, H.-Y.L.; data curation, N.A; writing—original draft preparation, M.N.; writing—review and editing, M.N. and F.-W.W.; visualization, W.-D.Q.; supervision, F.-W.W.; project administration, H.-Y.L.; funding acquisition, H.-Y.L.. All authors have approved the manuscript.

**Acknowledgments:** We are grateful to Prof. Li Haiyan who kindly provided all the requirements for this research.

**References**

1. Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M, et al. Solutions for a cultivated planet. Nature [Internet]. 2011 Oct 12;478:337. Available from: https://doi.org/10.1038/nature10452

2. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al. Food Security: The Challenge of Feeding 9 Billion People. Science (80-) [Internet]. 2010 Feb 12;327(5967):812 LP – 818. Available from: http://science.sciencemag.org/content/327/5967/812.abstract

3. Hickey LT, N. Hafeez A, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, et al. Breeding crops to feed 10 billion. Nat Biotechnol [Internet]. 2019;37(7):744-54. Available from: https://doi.org/10.1038/s41587-019-0152-9
4. Shinozaki K, Uemura M, Bailey-Serres J, Bray E WE. Responses to abiotic stress. In: Buchanan BB, Gruissem W JR, editor. Biochemistry and molecular biology of plants [Internet]. 2nd ed. Wiley, Oxford; 2015. p. 1051-1100. Available from: http://scholar.google.com/scholar_lookup?title=Responses%20to%20abiotic%20stress&author=K.%20Shinozaki&author=M.%20Uemura&author=J.%20Bailey-Serres&author=E.%20Bray&author=E.%20Weretilnyk&pages=1051-1100&publication_year=2015

5. Prudhomme C, Giuntoli I, Robinson EL, Clark DB, Arnell NW, Dankers R, et al. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proc Natl Acad Sci [Internet]. 2014 Mar 4;111(9):3262 LP – 3267. Available from: http://www.pnas.org/content/111/9/3262.abstract

6. Wei W, Liang D-W, Bian X-H, Shen M, Xiao J-H, Zhang W-K, et al. GmWRKY54 improves drought tolerance through activating genes in abscisic acid and Ca2+ signaling pathways in transgenic soybean. Plant J [Internet]. 2019 Jul 4;0(0). Available from: https://doi.org/10.1111/tpj.14449

7. Zhang H, Li Y, Zhu J-K. Developing naturally stress-resistant crops for a sustainable agriculture. Nat Plants [Internet]. 2018;4(12):989-96. Available from: https://doi.org/10.1038/s41477-018-0309-4

8. Wu M, Li Y, Chen D, Liu H, Zhu D, Xiang Y. Genome-wide identification and expression analysis of the IQD gene family in moso bamboo (Phyllostachys edulis). Sci Rep [Internet]. 2016 Apr 20;6:24520. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27094318

9. Khan S-A, Li M-Z, Wang S-M, Yin H-J. Revisiting the Role of Plant Transcription Factors in the Battle against Abiotic Stress. Vol. 19, International Journal of Molecular Sciences . 2018.
10. Kim MC, Chung WS, Yun D-J, Cho MJ. Calcium and calmodulin-mediated regulation of gene expression in plants. Mol Plant [Internet]. 2009/01/06. 2009 Jan;2(1):13–21. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19529824

11. Reddy ASN, Reddy VS, Golovkin M. A Calmodulin Binding Protein from Arabidopsis Is Induced by Ethylene and Contains a DNA-Binding Motif. Biochem Biophys Res Commun [Internet]. 2000;279(3):762–9. Available from: http://www.sciencedirect.com/science/article/pii/S0006291X00940324

12. Yang T, Poovaiah BW. A Calmodulin-binding/CGCG Box DNA-binding Protein Family Involved in Multiple Signaling Pathways in Plants. J Biol Chem [Internet]. 2002 Nov 22;277(47):45049–58. Available from: http://www.jbc.org/content/277/47/45049.abstract

13. Bouché N, Scharlat A, Snedden W, Bouchez D, Fromm H. A Novel Family of Calmodulin-binding Transcription Activators in Multicellular Organisms. J Biol Chem [Internet]. 2002 Jun 14;277(24):21851–61. Available from: http://www.jbc.org/content/277/24/21851.abstract

14. Zegzouti, Jones, Frasse, Marty, Maitre, Latch, et al. Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. Plant J. 1999 Jun;18(6):589-600.

15. Yang Y, Sun T, Xu L, Pi E, Wang S, Wang H, et al. Genome-wide identification of CAMTA gene family members in Medicago truncatula and their expression during root nodule symbiosis and hormone treatments. Front Plant Sci [Internet]. 2015 Jun 19;6:459. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26150823

16. Zhang J, Pan X, Ge T, Yi S, Lv Q, Zheng Y, et al. Genome-wide identification of citrus CAMTA genes and their expression analysis under stress and hormone treatments. J
17. Wei M, Xu X, Li C. Identification and expression of CAMTA genes in Populus trichocarpa under biotic and abiotic stress. Sci Rep [Internet]. 2017;7(1):17910. Available from: https://doi.org/10.1038/s41598-017-18219-8

18. Kakar KU, Nawaz Z, Cui Z, Cao P, Jin J, Shu Q, et al. Evolutionary and expression analysis of CAMTA gene family in Nicotiana tabacum yielded insights into their origin, expansion and stress responses. Sci Rep [Internet]. 2018;8(1):10322. Available from: https://doi.org/10.1038/s41598-018-28148-9

19. Meer L, Mumtaz S, Labbo AM, Khan MJ, Sadiq I. Genome-wide identification and expression analysis of calmodulin-binding transcription activator genes in banana under drought stress. Sci Hortic (Amsterdam) [Internet]. 2019;244:10–4. Available from: http://www.sciencedirect.com/science/article/pii/S0304423818306320

20. Büyük İ, İlhan E, Şener D, Özsoy AU, Aras S. Genome-wide identification of CAMTA gene family members in Phaseolus vulgaris L. and their expression profiling during salt stress. Mol Biol Rep [Internet]. 2019;46(3):2721–32. Available from: https://doi.org/10.1007/s11033-019-04716-8

21. Wang G, Zeng H, Hu X, Zhu Y, Chen Y, Shen C, et al. Identification and expression analyses of calmodulin-binding transcription activator genes in soybean. Plant Soil [Internet]. 2015;386(1):205–21. Available from: https://doi.org/10.1007/s11104-014-2267-6

22. Galon Y, Aloni R, Nachmias D, Snir O, Feldmesser E, Scrase-Field S, et al. Calmodulin-binding transcription activator 1 mediates auxin signaling and responds to stresses in Arabidopsis. Planta [Internet]. 2010;232(1):165–78. Available from: https://doi.org/10.1007/s00425-010-1153-6
23. Rahman H, Yang J, Xu Y-P, Munyampundu J-P, Cai X-Z. Phylogeny of Plant CAMTAs and Role of AtCAMTAs in Nonhost Resistance to Xanthomonas oryzae pv. oryzae. Front Plant Sci [Internet]. 2016 Feb 29;7:177. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26973658

24. Rahman H, Xu Y-P, Zhang X-R, Cai X-Z. Brassica napus Genome Possesses Extraordinary High Number of CAMTA Genes and CAMTA3 Contributes to PAMP Triggered Immunity and Resistance to Sclerotinia sclerotiorum [Internet]. Vol. 7, Frontiers in Plant Science . 2016. p. 581. Available from: https://www.frontiersin.org/article/10.3389/fpls.2016.00581

25. Wang Y, Wei F, Zhou H, Liu N, Niu X, Yan C, et al. TaCAMTA4, a Calmodulin-Interacting Protein, Involved in Defense Response of Wheat to Puccinia triticina. Sci Rep [Internet]. 2019;9(1):641. Available from: https://doi.org/10.1038/s41598-018-36385-1

26. Yang T, Poovaiah BW. An early ethylene up-regulated gene encoding a calmodulin-binding protein involved in plant senescence and death. J Biol Chem. 2000 Dec;275(49):38467–73.

27. Benn G, Wang C-Q, Hicks DR, Stein J, Guthrie C, Dehesh K. A key general stress response motif is regulated non-uniformly by CAMTA transcription factors. Plant J [Internet]. 2014/08/19. 2014 Oct;80(1):82–92. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25039701

28. Yang T, Peng H, Whitaker BD, Jurick WM. Differential expression of calcium/calmodulin-regulated SISRs in response to abiotic and biotic stresses in tomato fruit. Physiol Plant [Internet]. 2013 Jul 1;148(3):445–55. Available from: https://doi.org/10.1111/ppl.12027

29. Du L, Ali GS, Simons KA, Hou J, Yang T, Reddy ASN, et al. Ca2+/calmodulin regulates
salicylic-acid-mediated plant immunity. Nature [Internet]. 2009 Jan 4;457:1154. Available from: https://doi.org/10.1038/nature07612

30. Nie H, Zhao C, Wu G, Wu Y, Chen Y, Tang D. SR1, a calmodulin-binding transcription factor, modulates plant defense and ethylene-induced senescence by directly regulating NDR1 and EIN3. Plant Physiol [Internet]. 2012/02/16. 2012 Apr;158(4):1847-59. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22345509

31. Pandey N, Ranjan A, Pant P, Tripathi RK, Ateek F, Pandey HP, et al. CAMTA 1 regulates drought responses in Arabidopsis thaliana. BMC Genomics [Internet]. 2013;14(1):216. Available from: https://doi.org/10.1186/1471-2164-14-216

32. Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF. Roles for Arabidopsis CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. Plant Cell [Internet]. 2009 Mar;21(3):972-84. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19270186

33. Kim Y, Park S, Gilmour SJ, Thomashow MF. Roles of CAMTA transcription factors and salicylic acid in configuring the low-temperature transcriptome and freezing tolerance of Arabidopsis. Plant J. 2013 Aug;75(3):364-76.

34. Prasad KVSK, Abdel-Hameed AAE, Xing D, Reddy ASN. Global gene expression analysis using RNA-seq uncovered a new role for SR1/CAMTA3 transcription factor in salt stress. Sci Rep [Internet]. 2016 Jun 2;6:27021. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27251464

35. Kim YS, An C, Park S, Gilmour SJ, Wang L, Renna L, et al. CAMTA-Mediated Regulation of Salicylic Acid Immunity Pathway Genes in Arabidopsis Exposed to Low Temperature and Pathogen Infection. Plant Cell [Internet]. 2017 Oct 1;29(10):2465 LP - 2477. Available from: http://www.plantcell.org/content/29/10/2465.abstract

36. Shkolnik D, Finkler A, Pasmanik-Chor M, Fromm H. CALMODULIN-BINDING
TRANSCRIPTION ACTIVATOR 6: A Key Regulator of Na<sup>+</sup> Homeostasis during Germination. Plant Physiol [Internet]. 2019 Jun 1;180(2):1101 LP – 1118. Available from: http://www.plantphysiol.org/content/180/2/1101.abstract

37. Soystats 2019. A publication of the American Soybean Association. 2019.

38. Li Y, Chen Q, Nan H, Li X, Lu S, Zhao X, et al. Overexpression of GmFDL19 enhances tolerance to drought and salt stresses in soybean. PLoS One. 2017;12(6):e0179554.

39. Du Y-T, Zhao M-J, Wang C-T, Gao Y, Wang Y-X, Liu Y-W, et al. Identification and characterization of GmMYB118 responses to drought and salt stress. BMC Plant Biol [Internet]. 2018;18(1):320. Available from: https://doi.org/10.1186/s12870-018-1551-7

40. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W155-9.

41. Choi MS, Kim MC, Yoo JH, Moon BC, Koo SC, Park BO, et al. Isolation of a Calmodulin-binding Transcription Factor from Rice (Oryza sativa L.). J Biol Chem [Internet]. 2005 Dec 9;280(49):40820–31. Available from: http://www.jbc.org/content/280/49/40820.abstract

42. Han J, Gong P, Reddig K, Mitra M, Guo P, Li H-S. The fly CAMTA transcription factor potentiates deactivation of rhodopsin, a G protein-coupled light receptor. Cell. 2006 Nov;127(4):847–58.

43. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan;47(D1):D607–13.

44. Galon Y, Nave R, Boyce JM, Nachmias D, Knight MR, Fromm H. Calmodulin-binding transcription activator (CAMTA) 3 mediates biotic defense responses in Arabidopsis.
45. Nguyen KH, Mostofa MG, Li W, Van Ha C, Watanabe Y, Le DT, et al. The soybean transcription factor GmNAC085 enhances drought tolerance in Arabidopsis. Environ Exp Bot [Internet]. 2018;151:12–20. Available from: http://www.sciencedirect.com/science/article/pii/S0098847218300832

46. Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H. Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc. 2006;1(2):641–6.

47. Xia Z, Xu Z, Wei Y, Wang M. Overexpression of the Maize Sulfite Oxidase Increases Sulfate and GSH Levels and Enhances Drought Tolerance in Transgenic Tobacco. Front Plant Sci. 2018;9:298.

48. Bates LS, Waldren RP, Teare ID. Rapid determination of free proline for water-stress studies. Plant Soil [Internet]. 1973;39(1):205–7. Available from: https://doi.org/10.1007/BF00018060

49. Tang Y, Liu K, Zhang J, Li X, Xu K, Zhang Y, et al. JcDREB2, a Physic Nut AP2/ERF Gene, Alters Plant Growth and Salinity Stress Responses in Transgenic Rice. Front Plant Sci. 2017;8:306.

50. Demidchik V, Straltsova D, Medvedev SS, Pozhvanov GA, Sokolik A, Yurin V. Stress-induced electrolyte leakage: the role of K+-permeable channels and involvement in programmed cell death and metabolic adjustment. J Exp Bot. 2014 Mar;65(5):1259–70.

51. Kereszt A, Li D, Indrasumunar A, Nguyen CDT, Nontachaiyapoom S, Kinkema M, et al. Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. Nat Protoc. 2007;2(4):948–52.
Additional Files Legends

**Additional File 1**- Genomic, CDS and Proteomic Sequences of GmCAMTAs, Sequence dataset of phylogenetic tree

**Additional File 2**- Alignment of protein sequence ATCAMTAs and GmCAMTA

**Additional File 3**- Table S1: Proteins in the interaction network of GmCAMTAs, Table S2: Nuclear localization signal in GmCAMTAs, Table S3- Primers for qPCR of GmCAMTAs.

**Additional File 4**- Figure S1: Protein-protein interaction network of *GmCAMTAs* predicted with STRING database. Figure S2: Gene-specific PCR of the EHA105 Transformants harboring OE construct. Figure S3: Gene-specific PCR of the K599 Transformants harboring OE construct. Figure S4- Hairy roots culture for fresh and dry weight assay on MS with various concentrations of Mannitol.

**Additional File 5**- Sequence dataset of AtCAMTA5 interactors. Table S4: Primers for the qPCR of the AtCAMTA5 regulatory network. Sequence dataset of the GmCATMA12 regulatory network. Table S5: Primers for the qPCR of GmCAMTA12 regulatory network

**Figures**
Figure 1

Comprehensive in silico analysis of GmCAMTA family. (A) Cis elements in the GmCAMTA promoter region. Each type of cis motif present in GmCAMTA promoters is shown in unique color/color pattern. (B) Exon-intron assembly of GmCAMTA genes. (C) Domain organization of GmCAMTA proteins. Four different domains are represented in different colored shapes.
(D) Alignment showing the conserved motif sequence of the Calmodulin Binding Domain of Arabidopsis and Soybean CAMTA TFs. Each conserved residue at the definite position along the row (throughout the orthologues) is shaded in unique color. The functional residues in CaMB domain of these CAMTAs are indicated in the motif below the alignment. In the square brackets “[]” are the amino acids allowed in this position of the motif; “X” represents any amino acid and the round brackets “()” indicate the number of amino acids. (E) ML phylogenetic tree.

Figure 2

Chromosomal distribution of GmCAMTA genes. GmCAMTA genes are unequally distributed among chromosome 5, 7, 8, 9, 11, 15, 17 and 18.
Spatiotemporal expression analysis of GmCAMTA in drought. (A) Relative fold expression of GmCAMTAs in Roots. Soybean plants were treated with 6% PEG6000 in Hoagland’s solution for 5 different durations (1h, 3h, 6h, 9h and 12h) along with a control (0h). (B) Relative fold expression of GmCAMTAs in Leaves.
Figure 4

Phenotypic and physiological assays of wt and OE under drought stress. (A) Drought assay of wild type (WT) and transgenic (OE5 and OE12) Arabidopsis grown on soilrite subjected to 14 days of drought stress. (B) Root length assay on MS-mannitol. (C) Germination assay on MS-mannitol. (D) Column chart showing the survival rate (%) of WT and OE in soil. (E) Root length (cm) of WT, OE5 and OE12 lines. (E) Germination rate (%) of WT and OE plants. (G) Proline contents, (H) MDA contents, (I) CAT activity and (J) Relative electrolyte leakage (REL %) of wt and OE plants in control and drought treatment.
Figure 5

A: Comparison of plant growth under control and drought conditions. VC (Vicia faba) and OE (Oryza sativa) plants are shown.

B: Detailed root systems of VC and OE plants under control conditions.

C: Comparative root systems of VC and OE under drought conditions.

D: Bar chart showing fresh weight (mg) of VC and OE plants at different concentrations (0 mM, 50 mM, 100 mM, 150 mM, 200 mM).

E: Bar chart showing dry weight (mg) of VC and OE plants at different concentrations.

F: Bar chart showing proline content (μg/g) of VC and OE plants under control and drought conditions.

G: Bar chart showing MDA content (pmol/g) of VC and OE plants under control and drought conditions.

H: Bar chart showing CAT activity (U/min/g) of VC and OE plants under control and drought conditions.

I: Bar chart showing REL (%) of VC and OE plants under control and drought conditions.

J: Bar chart showing total length (cm) of VC and OE plants.

K: Bar chart showing surface area (cm²) of VC and OE plants.

L: Bar chart showing root volume (cm³) of VC and OE plants.

M: Bar chart showing number of branches of VC and OE plants.

N: Bar chart showing projected area (cm²) of VC and OE plants.
Comparative phenotype and physiology of chimeric soybean plants. (A) Chimeric soybean plants bearing VC (Vector control) roots and OE (overexpressing GmCAMTA12) roots in control and drought conditions. (B) Comparison of chimeric soybean plants having VC and OE roots. (C) VC and OE hairy roots observed with a root scanner. (D) Fresh and (E) dry weight of VC and OE hairy roots after culturing on MS-mannitol for 10 days. Comparison of VC and OE hairy roots in terms of (F) Proline contents, (G) MDA contents, (H) CAT activity and (I) relative electrolyte leakage between VC and OE hairy roots. (J) Analysis of the total length, (K) surface area, (L) root volume, (M) number of branches and (N) projected area.
Expression analysis of genes in the regulatory network of GmCAMTA12 and AtCAMTA5. (A) STRING-predicted regulatory network of AtCAMTA5. (B) Drought treated WT and OE Arabidopsis. (C) Expression profile of the 10 interactors of AtCAMTA5 in wt and OE Arabidopsis. (D) STRING-predicted regulatory network of GmCAMTA12 in soybean. (E) Chimeric soybean plants with VC and OE hairy roots. (F) Expression analysis of the 8 interactors of GmCAMTA12 in VC and (G) OE leaves. (H) Expression analysis of the 8 interactors in VC and (I) OE roots.

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

Additional File 2-Alignment.pdf
Additional File 3, Table S1, Table S2 and Table S3.pdf
Additional File 5-PPI sequence dataset, Table S4, Table S5.pdf
Additional File 1-Sequence dataset .pdf
Additional File 4- Figure S1, S2, S3 and S4.pdf