Evolution Analyses of CAMTA Transcription Factor in Plants and Its Enhancing Effect on Cold-tolerance

Peixuan Xiao, Jia-Wu Feng, Xi-Tong Zhu and Junxiang Gao*

Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, China

The calmodulin binding transcription activator (CAMTA) is a transcription factor that is widely present in eukaryotes with conserved structure. It contributes to the response to biotic and abiotic stresses and promotes the growth and development of plants. Although previous studies have investigated the number and function of CAMTAs in some species, there is still a lack of comprehensive understanding of the evolutionary process, phylogenetic relationship, expression patterns, and functions of CAMTAs in plants. Here we identified 465 CMATA genes from 112 plants and systematically studied the origin of CAMTA family, gene expansion, functional differentiation, gene structure, and conservative motif distribution. Based on these analyses, we presented the evidence that CAMTA family was originated from chlorophyta, and we speculated that CAMTA might experience obvious structure variation during its early evolution, and that the number of CAMTA genes might gradually increase in higher plants. To reveal potential functions of CAMTA genes, we analyzed the expression patterns of 12 representative species and found significant species specificity, tissue specificity, and developmental stage specificity of CAMTAs. The results also indicated that the CAMTA genes might promote the maturation and senescence. The expression levels and regulatory networks of CAMTAs revealed that CAMTAs could enhance cold tolerance of rice by regulating carbohydrate metabolism-related genes to accumulate carbohydrates or by modulating target genes together with other transcription factors. Our study provides an insight into the molecular evolution of CAMTA family and lays a foundation for further study of related biological functions.

Keywords: CAMTA transcription factors, phylogenetics analysis, gene expression, cold tolerance, gene regulatory network

INTRODUCTION

The divalent ion of calcium (Ca^{2+}) is an ubiquitous second messenger in eukaryotes (Galon et al., 2010b), and it plays an important role in the growth and development of plants under biotic and abiotic stress (Kudla et al., 2010; Reddy et al., 2011). Calmodulins (CaM) is the most important type of Ca^{2+} receptor, and the Ca^{2+}/CaM complex can bind to a large population of target proteins, including protein kinases, phosphatases, transcription factors, metabolic enzymes, ion channels,

transporters, and molecular motors (Snedden and Fromm, 2001; Yang and Poovaiah, 2003; Bouché et al., 2005; Reddy et al., 2011; Poovaiah and Du, 2018). Currently, at least 90 transcription factors have been identified as CaM-binding proteins such as CAMTA, MYB, WRKY, NAC, bZIP, and MADS-box proteins (Reddy et al., 2002; Popescu et al., 2007; Kim et al., 2009; Galon et al., 2010b). Among these transcription factors, the calmodulin-binding transcription activator (CAMTA) is a conserved family and the most characteristic transcription factor related to calmodulin (Bouché et al., 2005; Finkler et al., 2007).

Tobacco ethylene response gene (*NtER1*) is the earliest CAMTA gene identified in plants. *NtER1* is highly expressed in senescent tobacco leaves and petals, suggesting it plays a regulatory role in tobacco development, and it is a trigger for senescence and death (Yang and Poovaiah, 2000). Six CAMTA genes (*AtCAMTA1-AtCAMTA6*) have been identified in *Arabidopsis*. They are highly responsive to environmental signals and stresses such as extreme temperature, ultraviolet radiation, salt, and injury, as well as hormones (Reddy et al., 2000; Galon et al., 2010a). For example, *AtCAMTA1*, *AtCAMTA2*, and *AtCAMTA3* work together to directly bind to the promoter of C-repeat binding factor (CBF2), thereby inducing CBF2 expression and improving the cold tolerance of plants (Doherty et al., 2009; Kim et al., 2013). *AtCAMTA1* also actively modulates drought response by regulating several stress response genes, including RD26, ERD7, RAB18, LTPs, COR78, CBF1, and HSPs (Pandey et al., 2013). As a plant immune negative regulator, AtCAMTA3 can regulate salicylic acid (SA)-mediated immune response by interacting with the *EDS1* gene promoter and inhibiting its expression (Du et al., 2009). Moreover, CAMTA homologues have been identified in some plants, such as the *LeER66* in tomato, the *BnCAMTA* in rape, and the *OsCBT* gene in rice (Zegzouti et al., 1999). A typical CAMTA protein has multiple characteristic domains, including a CG-1 domain, a TIG domain, an ankyrin (ANK) repeat domain, an IQ motif, and a CaM-binding (CaMB) domain. The CG-1 domain is a DNA-binding domain, while the TIG domain is involved in non-specific DNA binding. The ANK domain is involved in protein-protein interaction. The CaMB domain and a varying number of IQ motifs (IQXXXRGXXXR) bind to CaM in a calcium-dependent or calcium-independent manner, respectively (da Costa e Silva, 1994; Aravind and Koonin, 1999; Bork et al., 1999; Song et al., 2006).

In recent years, CAMTA genes have been idendified in various plants, such as *Z. mays* (Yue et al., 2015), *M. truncatula* (Yang et al., 2015), *V. vinifera* (Shangguan et al., 2014), *and G. max* (Wang G. et al., 2014). In addition, A previous study analyzed the evolution of CAMTA genes based on 35 plant genomes representing four major lineages and the 6 chlorophyta genomes available at that time (Rahman et al., 2016b). With more and more plant genomes deciphered, it is necessary to systematically analyze the CAMTA family from global perspectives. In this study, we analyzed the CAMTA genes from 112 plants, examined their gene structure and protein domain composition to reveal the origin and evolution of CAMTA, investigated the expression pattern in different tissues and different developmental stages, explored the role of CAMTAs in rice under low temperature

stress. The study will provide a comprehensive insight into the evolution, expression, and function of CAMTA family, which might be conducive to further functional studies.

MATERIALS AND METHODS

Data Collection

In this study, a total of 112 plant genomes were collected for subsequent analysis (**Supplementary Table 1**). Among them, a total of 64 plants were derived from the Phytozome database (Goodstein et al., 2012), 16 species from the Ensemble database (Hubbard et al., 2002), and 26 species from the NCBI database.[1] Besides, *N. colorata* was collected from the National Genomics Data Center (Xue et al., 2021), *G. montanum* from the Dryad database,[2] and *A. filiculoides* and *S. cucullata* from the Fernbase database.[3]

Identification of Calmodulin Binding Transcription Activators From Plant Genomes

CAMTA protein sequences were integrated to establish a local protein database for identifying the CAMTA homologous sequences. The sequences of six *Arabidopsis* CAMTA protein were downloaded from the *Arabidopsis* Information Resource database[4] (Lamesch et al., 2012), including AT1G67310.1, AT2G22300.1, AT3G16940.1, AT4G16150.1, AT5G09410.2, and AT5G64220.1. BLASTP (v2.10.0) was used to perform sequence alignment, *e*-value was set as 1e-10. The *Arabidopsis* CAMTA protein sequences were used as queries to align with the constructed protein database. Only the longest transcript from genes containing multiple transcripts was retained, and the shorter transcripts were removed. Then, four conserved domains or motifs of proteins including CG-1 domain, TIG domain, ANK repeats, and IQ motifs were identified by using three databases, namely, CDD, SMART, and PFAM, and the CaMB domain was identified using the Calmodulin target database. Only the sequences with these identified domains were defined as CAMTA protein.

Subcellular Localization and Gene Ontology Annotation

The subcellular localization of all CAMTA protein sequences was performed using WoLF PSORT[5] (Horton et al., 2007). Omics Box (v1.4.12)[6] was employed for gene ontology annotation of the CAMTA family. First, protein sequences of the identified CAMTA genes were imported into Omics Box. Then, the gene functional annotation was conducted with default parameters. Finally, the results were visualized by using matplotlib (v3.3.3) (Hunter, 2007).

---

[1]https://www.ncbi.nlm.nih.gov/
[2]https://datadryad.org/stash/
[3]https://www.fernbase.org/
[4]https://www.arabidopsis.org/
[5]https://wolfpsort.hgc.jp/
[6]https://www.biobam.com/omicsbox/

Phylogenetic Analysis of Calmodulin Binding Transcription Activator Proteins

CAMTA protein sequences were aligned using MAFFT (v7.453) (Katoh and Standley, 2013), which is an efficient multi-sequence alignment software. ProtTest (v3.4.2) (Darriba et al., 2011) was used to select the best fitting model of protein evolution according to the AIC and BIC indicators, and the optimal amino acid substitution model was JTT+I+G. The phylogenetic tree based on the maximum likelihood method (bootstrap value = 1,000) was constructed by IQtree2 (v2.0.6) (Minh et al., 2020) and RaxML-HPC BlackBox of CIPRES (v3.3) (Miller and Schwartz, 2011), respectively. Finally, the tree was trimmed by using Interactive Tree Of Life (iTOL v5)[7] (Letunic and Bork, 2019).

Structure Analysis of Calmodulin Binding Transcription Activator Genes

The information of untranslated region (UTR), coding sequence (CDS), and intron phase was extracted and transformed into bed format with the GFF3 files of 112 plants as the input. The structure of CAMTA genes was visualized using Gene Structure Display Server (GSDS v2.0)[8] (Hu et al., 2015). The phylogenetic tree of the CAMTA protein and the structure of CAMTA genes were displayed in the same figure. Since the lengths of the introns of some genes were too long to be displayed, the lengths of the introns were set as the same size. The phase and the number of the introns were also analyzed and displayed in the gene structure.

Analysis of Calmodulin Binding Transcription Activator Genes Expression

The transcription data were from the NCBI SRA database[9] and TraVA database[10] (**Supplementary Table 2**). To compare the expression of the CAMTAs in multiple plants, we collected RNA-Seq data in *B. vulgaris*, *B. distachyon*, *C. quinoa*, *E. grandis*, *G. raimondii*, *H. annuus*, *M. domestica*, *O. sativa*, *P. trichocarpa*, *S. lycopersicum*, *V. vinifera,* and *Z. mays,* in which most samples derived from the leaves. In order to compare the expression of the CAMTAs in different tissues of the same plant, RNA-Seq data were sampled from different tissues in *O. sativa*, *Z. Mays*, and *A. thaliana*. The RNA-Seq data under low temperature stress were collected to investigate the expression of the CAMTA genes. The RNA-Seq data were converted into fastq format using SRA Toolkit (v2.10.4) (Leinonen et al., 2011) and filtered using Trimmomatic (v0.38) (Bolger et al., 2014), respectively. Then Hisat2 (v2.1.0) (Kim et al., 2019) was used to construct the genome index and perform read mapping, and RSEM (v1.3.3) (Li and Dewey, 2011) was used to quantify the CAMTAs. The expression heatmap was drawn using pheatmap in R package (v1.0.12) (Kolde, 2012).

---

[7]https://itol.embl.de/
[8]http://gsds.gao-lab.org/
[9]https://www.ncbi.nlm.nih.gov/sra/
[10]http://travadb.org/

Genome-Wide Identification of Calmodulin Binding Transcription Activator Target Genes and Differentially Expressed Genes

The frequency matrix of *Arabidopsis* CAMTA motifs was downloaded from JASPAR database[11] (Fornes et al., 2019), and MEME FIMO (v5.1.1) (Grant et al., 2011) was used to predict CAMTA target genes in the whole genome. Under cold stress, differentially expressed genes (DEGs) were identified by the tool DESeq2 (v1.26.0) (Love et al., 2014) and edgeR (v3.32.1) (Robinson et al., 2009) with fold change set as 2, and false discovery rate (FDR) set as 0.01. The intersection between DEGs and potential target genes detected by MEME were designated as target genes regulated by CAMTA. GENIE3 (v1.8.0) (Huynh-Thu et al., 2010) was used to analyze the regulatory effect of CAMTA genes on target genes, the weight threshold was set as 0.25. Gephi (v0.9.2) (Bastian et al., 2009) was used to visualize the regulatory network.
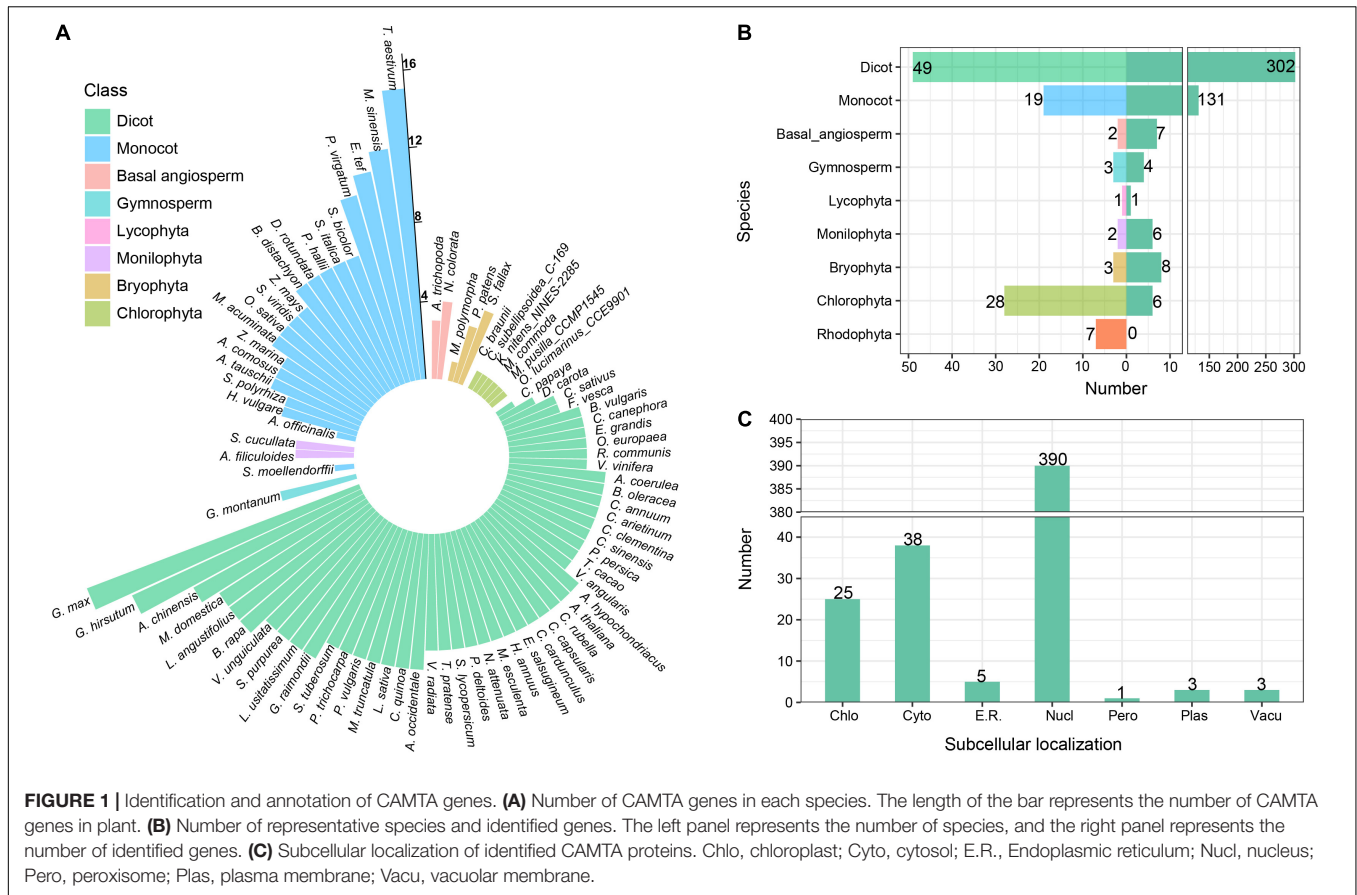
RESULTS

Identification of Calmodulin Binding Transcription Activator Genes

The homologous sequences of CAMTA genes were retrieved from the constructed local database using BLASTP with protein sequences of six *Arabidopsis* CAMTA genes as queries. The longest transcript from the genes containing multiple transcripts was retained, and the redundant sequences were removed. Then, the conserved domains were manually validated against CDD, SMART, PFAM, and Calmodulin Target Database. After the strict screening, a total of 465 CAMTA protein sequences were obtained from 112 plants (**Figures 1A,B** and **Supplementary Table 3**). These representative plants included rhodophyta, chlorophyta, bryophyta, monilophyta, lycophyta, gymnosperms, and angiosperms, among which dicots exhibited the most abundant, followed by chlorophyta. None of the 7 rhodophyta species was found to contain CAMTA protein sequence. CAMTA protein sequences were identified from only 6 out of 28 chlorophyta plants, and they were single copies, implying that CAMTA might be originated from chlorophyta.

A total of 131 CAMTA genes were identified from 19 monocots with 4 or more CAMTAs from each monocot except *A. officinalis*. Among them, the number of CAMTA genes from *T. aestivum*, *M. sinensis*, *E. tef*, and *P. virgatum* was 15, 12, 11, and 10, respectively. In total, 302 CAMTAs were identified from 49 dicots with 4 or more from each dicot except 1 from *C. papaya*, 2 from *D. carota*, 3 from *C. sativus*, and *F. vesca*. In addition, 15, 13, and 10 CAMTAs were obtained, respectively, from *G. max*, *G. hirsutum*, and *A. chinensis*. For basal angiosperm, 4 CAMTAs were identified from *N. colorata* and 3 from *A. trichopoda*. For gymnosperm, 4 CAMTAs were obtained with all of them from *G. montanum*. However, no complete CAMTA genes were identified from *P. taeda* and *P. abies*. For Lycophyta,

---

[11]http://jaspar.genereg.net/

FIGURE 1 | Identification and annotation of CAMTA genes. **(A)** Number of CAMTA genes in each species. The length of the bar represents the number of CAMTA genes in plant. **(B)** Number of representative species and identified genes. The left panel represents the number of species, and the right panel represents the number of identified genes. **(C)** Subcellular localization of identified CAMTA proteins. Chlo, chloroplast; Cyto, cytosol; E.R., Endoplasmic reticulum; Nucl, nucleus; Pero, peroxisome; Plas, plasma membrane; Vacu, vacuolar membrane.

only 1 CAMTA gene was identified from *S. moellendorffii*. For Monilophyta, 3 CAMTA genes obtained from *A. filiculoides* and *S. cucullata*, respectively. For Bryophyta, 1 CAMTA gene was identified from *M. polymorpha*, 3 from *P. patens*, and 4 from *S. fallax*.

Based on the above results, we found that the CAMTA genes did not exist in rhodophyta, but they appeared in some chlorophyta plants and evolutionarily late land plants. Considering that most chlorophyta plants live in fresh water, and their living environments are greatly different from those of rhodophyta in seawater, the appearance of new genes might contribute to chlorophyta's adaptation to the new environment. Therefore, we speculated that the CAMTA genes might be originated from some chlorophyta plants, which is consistent with the previous research results (Zhang et al., 2020). In our study, the number of CAMTA genes varied greatly in different phyla and exhibited a trend of gradual increase during evolution process, indicating many gene duplication events in the long-term evolution of CAMTA genes. In higher angiosperm, the copy number of CAMTAs varied greatly across different species. *G. max* had the maximum copies (up to 15), followed by *G. hirsutum* with 13 copies, while *A. officinalis*, *C. papaya,* and *C. sativus* had only 1~3 copies. Moreover, no obvious pattern of CAMTA number distribution was found in the species under the same class. This showed that the evolution of CATMAs was a very complex process under the influence of multiple

factors, thus the distribution of CAMTA number cannot be explained merely from the perspective of a single factor. For example, plants suffered various biotic and abiotic stresses from aquatic to terrestrial environment. As a transcription factor playing important roles in stress response, CAMTA was bound to be affected. In addition, genome-wide duplication events and chromosome deletion might also lead to the change in CAMTA gene number.

Subcellular localization information provides an important clue for exploring protein function. WoLF PSORT was used to analyze the subcellular localization of the identified CAMTA proteins (**Figure 1C**). The results showed that 390 CAMTA proteins were located in the nucleus, accounting for 83.87% of all the sequences; 38 CAMTAs were located in the cytosol, accounting for 8.17%; 25 CAMTAs were located in chloroplast, accounting for 5.37%. The remaining CAMTAs were located in the endoplasmic reticulum, peroxisome, plasma membrane, and vacuolar membrane, accounting for 2.59% in total. CAMTA proteins were mostly located in the nucleus, which was in accordance with the transcriptional activation function of CAMTA as a transcription factor. Some CAMTAs contain nuclear localization signal (NLS) sequence, thus they can enter the nucleus to perform their function. For example, in *M. truncatula*, all of the MtCAMTA proteins were predicted to contain an NLS at the N-terminus of the protein, which was consistent with their function in the nucleus as transcription

factor (Yang et al., 2015). In *B. napus L.*, all the BnCAMTA proteins were also predicted to contain an NLS at the N-terminus of the protein (Rahman et al., 2016a).

Phylogenetic Analysis of Calmodulin Binding Transcription Activator Protein Sequences

In order to explore the phylogenetic relationships of CAMTA families, a phylogenetic tree of CAMTA proteins was constructed based on the maximum likelihood (**Figure 2**). The six CAMTA proteins of chlorophyta were used as the outgroup and root node. The 459 protein sequences can be grouped into three clades, namely, Group I, Group II, and Group III. The base nodes of each group represented ancient plants such as bryophyta or monilophyta, followed by basal angiosperm, monocot, and dicot, which was in line with the evolutionary process of plants. In addition, the plants were regularly clustered together by class in each group. Therefore, from the perspective of evolution and clustering, our phylogenetic tree was reasonable. Each group was originated from lower plants, and it contained bryophyta or monilophyta, gymnosperm, and angiosperm, indicating the differentiation of CAMTA protein in early evolution stage of land plants. Considering the great differences among sea water, fresh water, and land environment, we speculated that the adaptation to the new living environment might contribute to this differentiation. These differentiations were maintained during the long evolutionary process, thus a large number of variations were accumulated in the sequence. The largest clade Group II consisted of two subgroups, which were termed as II-A and II-B. As shown in **Figure 2**, both II-A and II-B contained basal angiosperm, monocot, and dicot, while there were more ancient species in II-A, including bryophyta, monilophyta, lycophyta, and gymnosperm. These results suggested that subgroup II-B might diverge from subgroup II-A in basal angiosperm. Although II-B was originated later than II-A, it contained the largest number of species, which exceeded Group I and Group III, indicating that the subgroup experienced rapid gene expansion. Although there were no obvious subgroups in Group I and Group III, these two groups had large number of monocot and dicot species, indicating that their protein sequences constantly evolved and their number expanded overall. The variation of sequence composition in each group also promoted the change of protein domain. Thus we further analyzed protein domain in the following section.

To sum up, we inferred that CAMTA was originated from chlorophyta, and then differentiated into three groups, and CAMTAs expanded rapidly to be adapted to the great changes from aquatic to land environment. In particular, the largest clade Group II underwent great variation and differentiated into two subgroups. The available data support that differentiation events occurred in basal angiosperm.

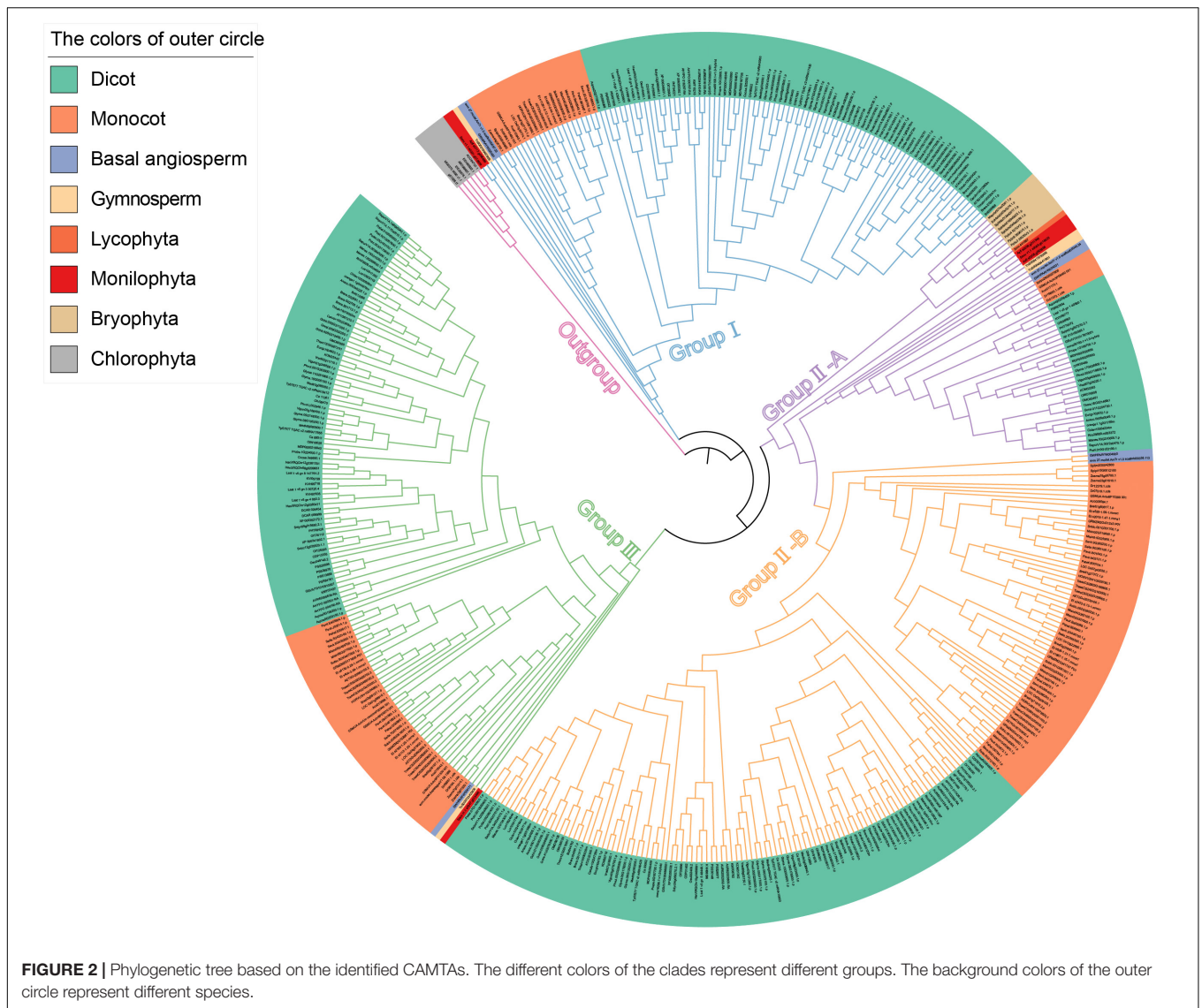## Structure Analysis of Calmodulin Binding Transcription Activator Genes

Knowledge of the intron phases helps to understand the evolution of the gene structure and alternative splicing of CAMTA genes.

Introns can be divided into three types based on phases: phase 0, phase 1, and phase 2. Phase 0 intron does not disrupt a codon, and phase 1 intron disrupts a codon between the first and second bases, whereas phase 2 intron disrupts a codon between the second and third bases (Choudhuri, 2014). The intron phase can affect exon translation. In this study, we counted the number of introns of the CAMTA genes and found that the number of introns in CAMTA genes varied greatly across species, ranging from 0 to 34 (**Figure 3A** and **Supplementary Table 4**). Among our identified CAMTA genes, *O. lucimarinus CCE9901*, *P. patens*, and *S. fallax* contained no introns, which was consistent with a previous report (Rahman et al., 2016b). This indicated that CAMTA experienced significant structure variation during its early evolution. The number of introns in other plants was mostly between 10 and 13. In 255 genes, each gene contained 12 introns, accounting for 54.83%, and in other 90 genes, each gene contained 11 introns, accounting for 19.35%. In most cases, the same species and clade had similar exon-intron structure and intron number. Only a small number of CAMTA genes showed various exon-intron structure and intron number with more than 20 introns contained in each gene, which was much higher than the average of 10∼13. For example, Dr04389.1.cds had 34 introns, and MDP0000265800 had 21 introns. Overall, we discovered that the number of introns in most of the same species varied slightly, indicating that the gene structure was relatively conserved in the same species. The conservation of CAMTA structure implied that CAMTA was related to important functions of life activities.

The analysis of the intron phases indicated that intron phases exhibited certain distribution patterns (**Figure 3B**). Pattern 0200-1102-0110 and pattern 1102-0110-0200 accounted for 20.86 and 18.06%, respectively, far more than other patterns. Pattern 1102-0101-0200, 1102-0101-0200, 0201-0102-0110, and 0200-1020-110 accounted for 6.88, 4.94, 4.30, and 4.30%, respectively. The exon-intron structure of CAMTA was visualized by GSDS (**Figure 3C** and **Supplementary Figure 1**). The representative species were selected to basically cover the evolution process of plants, including chlorophyta, bryophyta, monilophyta, gymnosperm, basal angiosperm, monocot, and dicot plants (**Supplementary Table 5**). Interestingly, we noticed that the length of exons and the phase of introns showed a correlation to a certain extent. As shown in **Figure 3C**, the intron phase pattern 0200, 0201, and 1102 were abundant in CAMTA genes. The lengths of the exons corresponding to intron phase pattern 0200 or 0201 (pink) were significantly longer than those corresponding to 1102 or 0110 (dark red).

Domain Analysis of Calmodulin Binding Transcription Activator Proteins

We identified a total of 465 CAMTA protein sequences and calculated their length, molecular weight, and isoelectric point (**Supplementary Table 6**). The average length of CAMTA proteins was 1004.5 aa. Protein PHT66961 (from *C. annuum*) was the shortest with a length of 745 aa and a molecular weight of 84.2 kDa. Protein Dr04389.1.cds (from *D. rotundata*) was the longest with a length of 2309 aa and a molecular weight of 254.4 kDa. The maximum, minimum, and average of isoelectric points were

FIGURE 2 | Phylogenetic tree based on the identified CAMTAs. The different colors of the clades represent different groups. The background colors of the outer circle represent different species.

9.34, 5.02, and 6.55, respectively. Using CDD, SMART, PFAM, and Calmodulin Target Database, we identified CG-1 domain, TIG domain, CaMB domain, ANK repeats, and IQ motifs.

Although the CAMTA proteins contained very conserved domains, the distribution patterns of these domains varied among the species and the groups (**Figure 4** and **Supplementary Figure 2**). As described above, the selection criterion of the representative proteins was similar to those of the representative species (**Figure 4** and **Supplementary Table 5**). Generally, CAMTA proteins included CG-1 DNA-binding domain, TIG domain involved in non-specific DNA binding, ankyrin repeats, IQ motifs of Ca^{2+}-independent calmodulin-binding domain, and Ca^{2+} dependent calmodulin-binding domain. All the CAMTA proteins contained ANK repeats except MDP0000255517. The number of ANK repeats was quite different, leading to their significantly different length. The number of IQ motifs contained in CAMTA proteins varied from 0 to 3, and this variation was slightly weaker than that of ANK

repeats. Three IQ motifs were contained in each of 15 CAMTA proteins, two in each of 356 CAMTA proteins, and one in each of 89 CAMTA proteins, and no IQ motif in 3 CAMTA proteins. In addition, two CAMTA proteins contained IQ-like motif.

Some CAMTA proteins contained TIG domain, while others did not. According to the presence and absence of TIG domain, the CAMTA protein was divided into TIG-CAMTA and non-TIG CAMTA (Rahman et al., 2016b), and their distribution patterns were different across phylogenetic clusters and species. Considering this, we analyzed the TIG domain of each species, and found a total of 401 TIG CAMTAs and 64 non-TIG CAMTAs. In all the chlorophyta species except *C. braunii*, the CAMTAs fell into TIG ones. In bryophyta except for two CAMTA proteins from *S. fallax*, all the other CAMTAs were TIG CAMTAs. In monilophyta, lycophyta, and gymnosperm, all the CAMTA proteins were TIG CAMTAs. These results indicated that non-TIG CAMTAs might originate from the late-stage chlorophyta, and they persisted in early land plants from
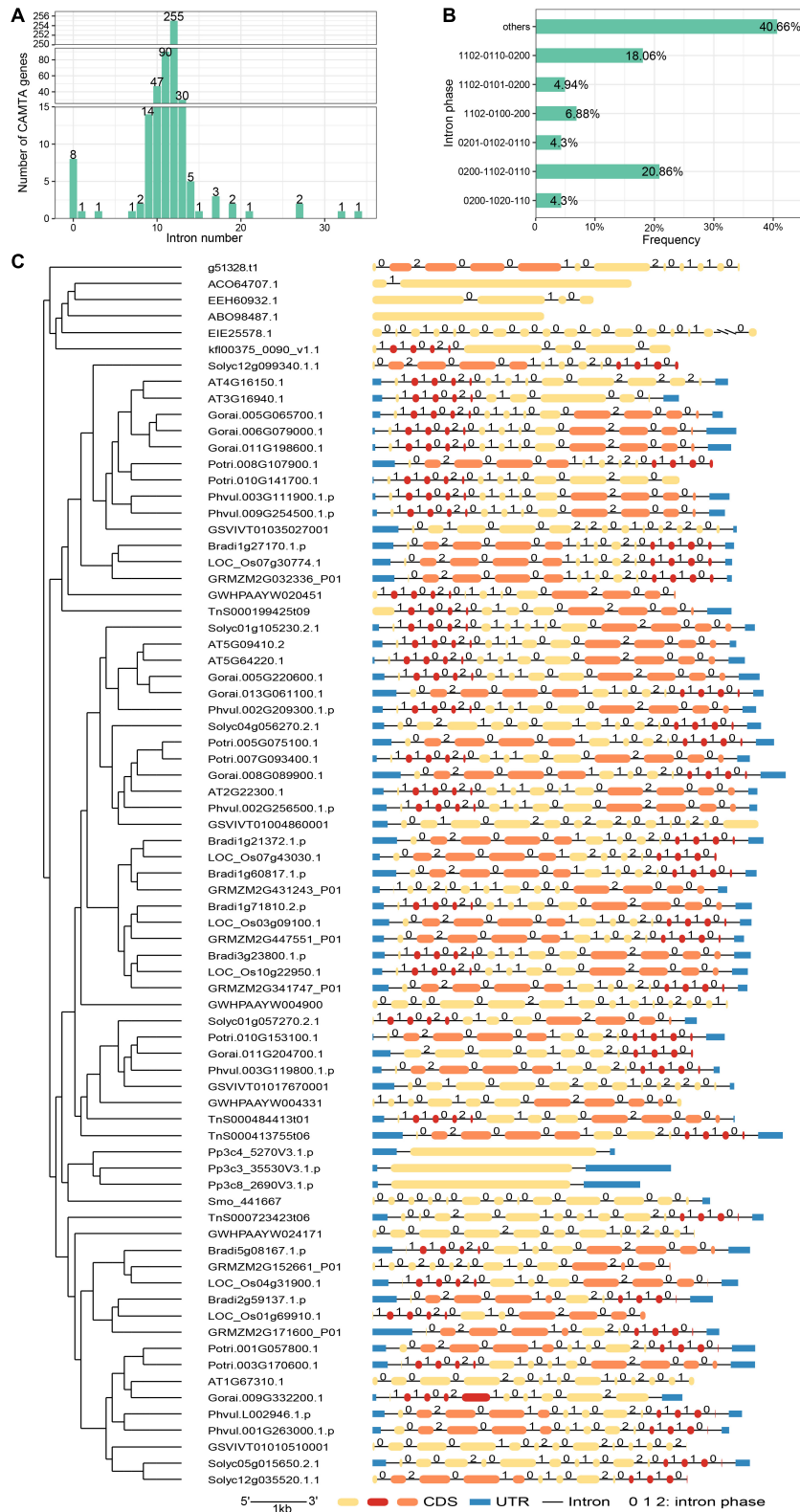
FIGURE 3 | Gene structure analysis of CAMTAs. **(A)** Distribution of intron number. **(B)** Percentage of intron phase. **(C)** CAMTA structures of some representative species. Pink indicates the exons corresponding to intron phase pattern 0200 and 0201, and dark red denotes those corresponding to intron phase pattern 1102 or 0110.
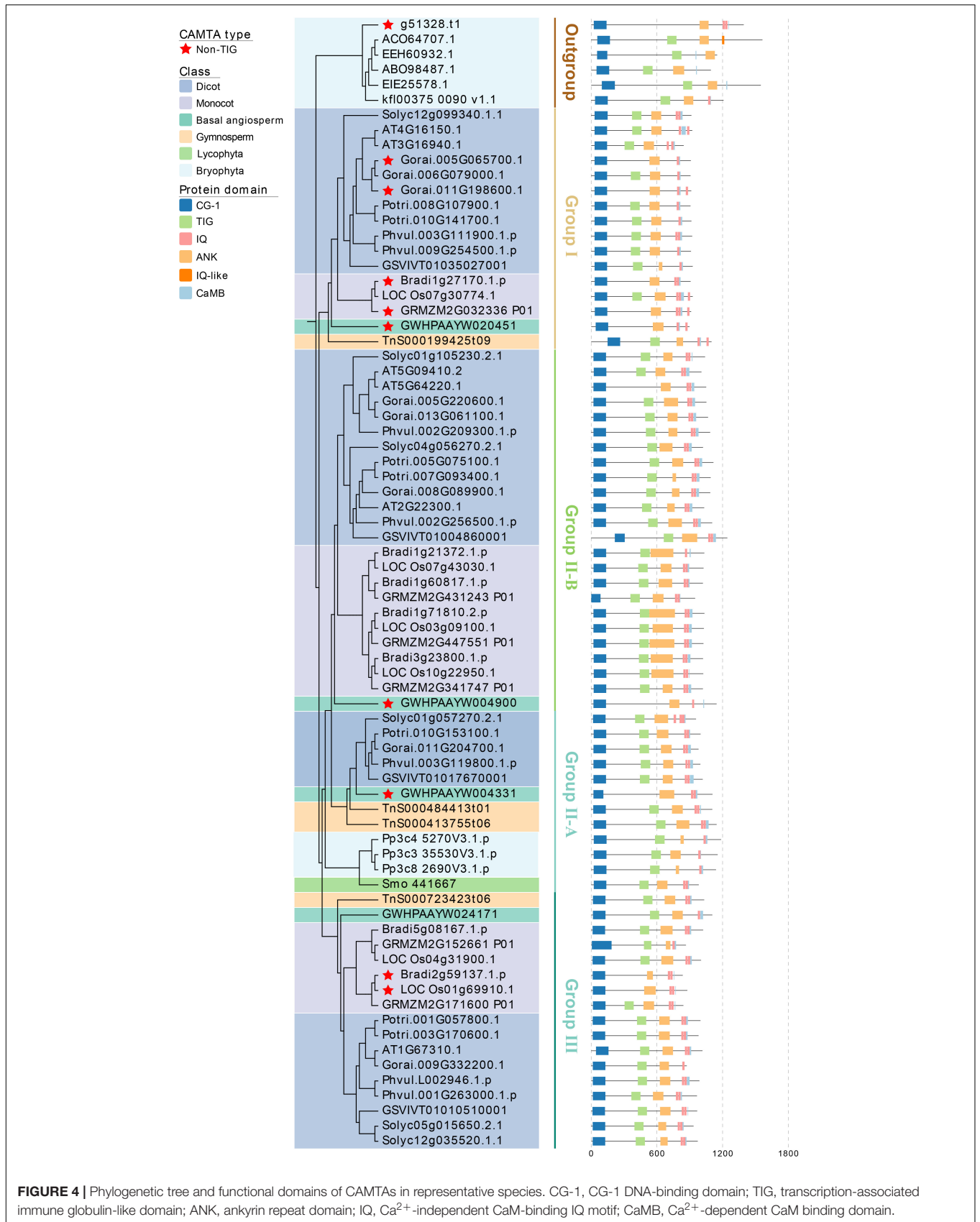
FIGURE 4 | Phylogenetic tree and functional domains of CAMTAs in representative species. CG-1, CG-1 DNA-binding domain; TIG, transcription-associated immune globulin-like domain; ANK, ankyrin repeat domain; IQ, Ca^{2+}-independent CaM-binding IQ motif; CaMB, Ca^{2+}-dependent CaM binding domain.

bryophytes to angiosperms. Moreover, the distribution of TIG CAMTAs and non-TIG CAMTAs between monocots and dicots, as well as among group I, II, and III were uneven. As shown in **Supplementary Figure 2**, among the 302 CAMTAs of the monocots, the proportions of TIG, and non-TIG CAMTAs were 87.09 and 12.91%, respectively; among the 131 CAMTAs of dicots, the proportion of TIG and non-TIG were 77.86 and 22.14%, respectively. Group III had only 3 non-TIG CAMTAs, which were far less than those from group I and II. In group I, there were 21 CAMTAs from monocots, of which 17 were non-TIG, and 84 CAMTAs from dicots, of which 21 were non-TIG. Only a few non-TIG CAMTAs were identified from the two subgroups of group II. Most non-TIG CAMTAs in group II-B were from dicots, whereas non-TIG CAMTA in group II-A were not found in the dicots. The difference in protein domain distribution of CAMTAs suggested the differentiation of their function.

Expression Analysis of Calmodulin Binding Transcription Activator Genes

Gene expression profiles were closely related to their functions. To better understand the functions of CAMTA genes, we analyzed the transcription patterns of multiple representative species including *B. vulgaris*, *B. distachyon*, *C. quinoa*, *E. grandis*, *G. raimondii*, *H. annuus*, *M. domestica*, *O. sativa*, *P. trichocarpa*, *S. lycopersicum*, *V. vinifera*, and *Z. mays* (**Supplementary Table 2**). In order to compare gene expression between different species, we collected RNA-seq data from the same tissue. Most data were from leaves except that *C. quinoa* data derived from whole seeds and *P. trichocarpa* data from stems. We calculated the transcripts per million (TPM) of each CAMTA gene for every species, and then computed the average of TPM to represent the expression level according to the number of CAMTA genes in the species. The results showed that the expression level had significant species specificity (**Figure 5A**). The expression levels of the CAMTA genes in *O. sativa* and *V. vinifera* were significantly higher than the average, while those in *P. trichocarpa*, *H. annuus,* and *M. domestica* were significantly lower than the average. No obvious expression pattern was observed in dicot and monocot. In addition, Group II-B CAMTAs were identified from all the 12 plants, and their expression levels were generally high. However, group II-A genes were only identified from 5 plants, and their expression levels were generally low except in *V. vinifera* (**Supplementary Figure 3**).

In order to explore the expression differences of CAMTA genes in different tissues, we collected transcription data from multiple tissues for further analysis, including 25 tissues of *A. thaliana*, 8 tissues of *O. sativa*, and 10 tissues of *Z. mays* (**Supplementary Table 2**).

The expression levels of CAMTA family showed significant tissue specificity, that is, the same tissue generally clustered together, but different tissues had significant differences (**Figure 5B**). For example, the expression level of AT3G16940 in mature leaves and siliques was much higher than that in other tissues, and that of AT5G64220 was significantly lower in seeds
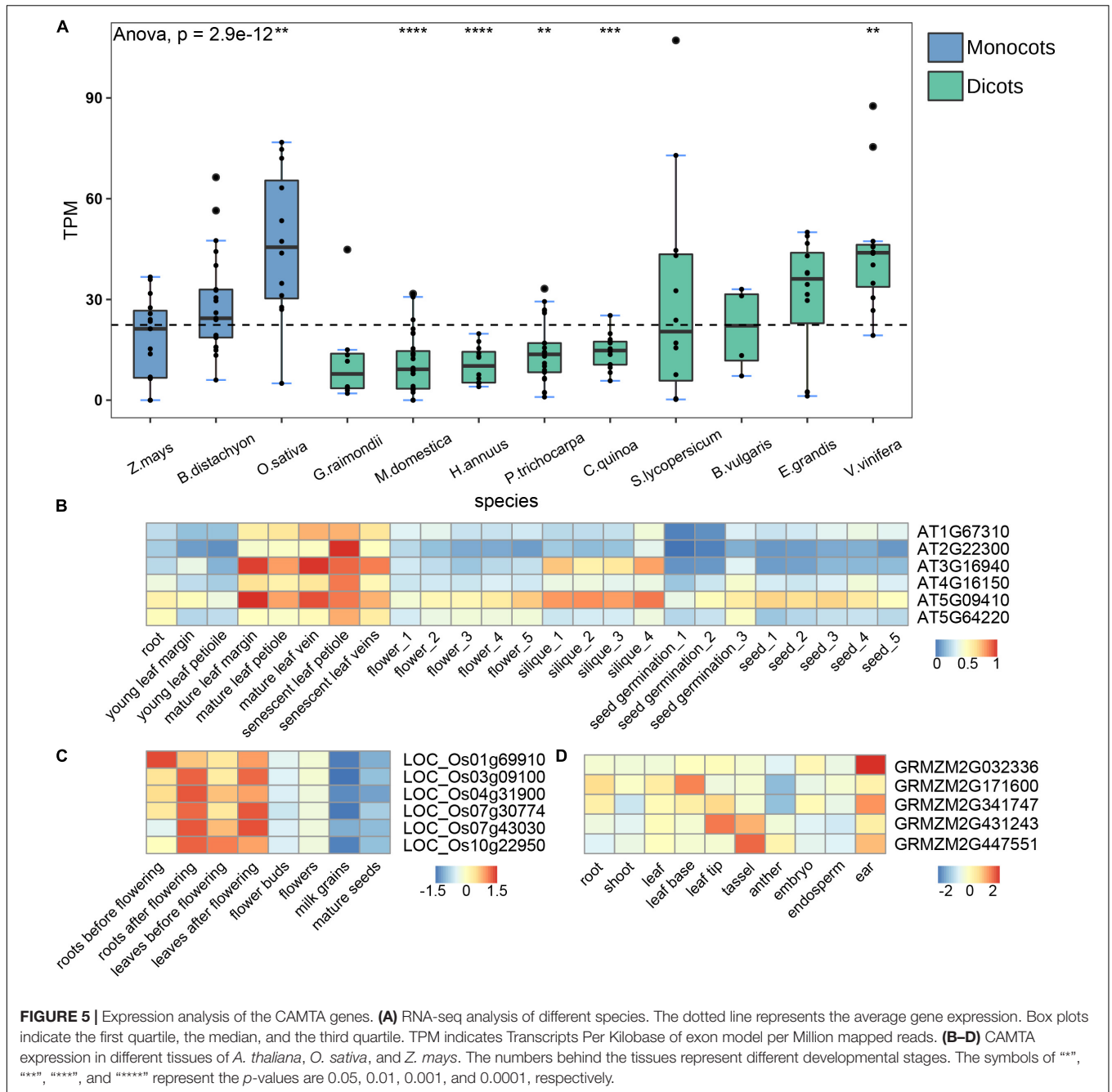
and young leaves. The expression level of CAMTAs also showed the specificity of developmental stage. They were significantly higher in mature and senescent leaves than in young leaves. Previous studies have shown that PtCAMTA1 is highly expressed in mature leaves of *P. tomentose*, and that some CAMTA genes in tomato show high expression in fruits, indicating that CAMTA genes are closely related to fruit development and maturity (Yang T. et al., 2012; Wei et al., 2017), which are basically consistent with ours. Therefore, we speculated that the CAMTA genes might promote the maturation and senescence of *A. thaliana*.

To examine whether the results of CAMTA genes from *A. thaliana* are applicable to other model plants, we compared the expressions of CAMTAs in roots, leaves, flowers, and seeds of *O. sativa* at different developmental stages and found significant specificity. As shown in **Figure 5C**, the expression levels of almost all CAMTA genes in leaves after flowering were substantially higher than those before flowering. Similar results were observed between roots after flowering and before flowering, between flowers and flower buds, between mature seeds and milk grains. Overall, the expression levels of CATMA genes were higher in late developmental stage than in early developmental stage. These results further support the role of CAMTA in promoting maturation. We also analyzed RNA-seq data collected from the root, shoot, leaf, leaf base, leaf tip, tassel, anther, embryo, endosperm, ear of *Z. mays*, and obtained the results of tissue specificity, which were similar with our findings in *A. thaliana* (**Figure 5D**). For example, the expression level of GRMZM2G171600 was the highest in leaf base, but the lowest in anther. In brief, the expression level of CAMTA genes exhibited significant species specificity, tissue specificity, and developmental stage specificity. The CAMTA genes might promote the maturation and the senescence of plants.

Function Analysis of Calmodulin Binding Transcription Activator Genes in Rice Under Cold Stress

The Gene Ontology (GO) database was used to analyze molecular function (MF), cellular component (CC), and biological process (BP) of identified CAMTA genes (**Figure 6A**). The annotation assigned these CAMTA genes into six biological process categories, and most of these genes were involved in regulation of biological process, cellular process, and metabolic process. In the molecular function, CAMTA genes were enriched in molecular function regulator (31%), transcription regulator activity (31%), and binding (38%). In the cellular component, we found that all the CAMTA genes were enriched in cellular anatomical entity (**Figure 6A**).
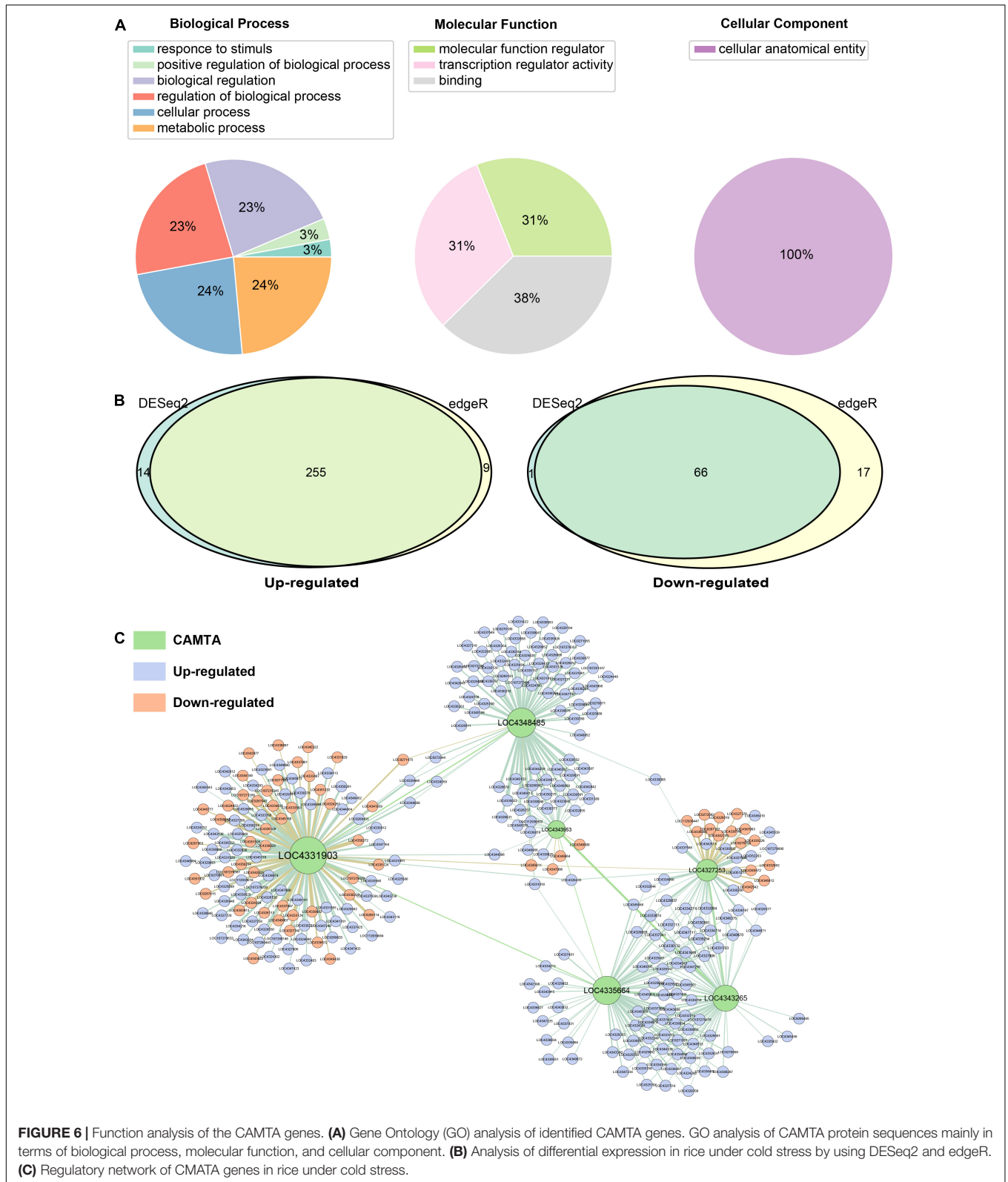
Low temperature is one of the main environmental stresses that plants frequently experience under natural conditions. As a transcription factor, CAMTA has a variety of functions, and its role in promoting maturation was described in the previous sub-section. We further investigated its role in resisting low temperature stress. We used two sets of RNA-Seq data including PRJNA430015 and PRJNA607661 (**Supplementary Table 2**), with or without cold stress to explore the function of rice CAMTAs and their target genes. DESeq2 and edgeR were

FIGURE 5 | Expression analysis of the CAMTA genes. **(A)** RNA-seq analysis of different species. The dotted line represents the average gene expression. Box plots indicate the first quartile, the median, and the third quartile. TPM indicates Transcripts Per Kilobase of exon model per Million mapped reads. **(B–D)** CAMTA expression in different tissues of *A. thaliana*, *O. sativa*, and *Z. mays*. The numbers behind the tissues represent different developmental stages. The symbols of "*", "**", "***", and "****" represent the *p*-values are 0.05, 0.01, 0.001, and 0.0001, respectively.

separately used for differential expression analysis. By integrating two results, a total of 255 CAMTA target genes were found to be up-regulated, and 66 down-regulated (**Figure 6B**).

DAVID was used to annotate the differentially expressed target genes (**Supplementary Table 7**). The results showed that some down-regulated genes were involved in carbohydrate metabolic process, such as beta-glucosidase 20/26 (LOC9270758/LOC4344146) and glucan endo-1, 3-beta-glucosidase GII (LOC4326519). Some other down-regulated genes were involved in ATP binding, such as serine/threonine-protein kinase Nek3/SAPK3 (LOC4349411/LOC4342542) and

UMP-CMP kinase 1 (LOC4339882). Some up-regulated genes were involved in transcriptional regulation, protein catabolism, and defense response regulation. In addition, we found that the AP2/ERF transcription factor gene family was up-regulated, including dehydration responsive element-binding protein 1C (LOC4339974) and ethylene-responsive transcription factor 5/7/11 (LOC4336947, LOC4349395, LOC4337341). Among them, dehydration responsive element-binding protein 1C is also called transcription factor CBF1, which can bind to GCC-box elements to induce the expression of cold stress-related genes. In addition, there were some other cold resistance-related

FIGURE 6 | Function analysis of the CAMTA genes. **(A)** Gene Ontology (GO) analysis of identified CAMTA genes. GO analysis of CAMTA protein sequences mainly in terms of biological process, molecular function, and cellular component. **(B)** Analysis of differential expression in rice under cold stress by using DESeq2 and edgeR. **(C)** Regulatory network of CMATA genes in rice under cold stress.

transcription factors, such as NAC, MYB, bZIP, and WRKY (Liu et al., 2012; Yang A. et al., 2012; Wang L. et al., 2014; Diao et al., 2020). We also constructed regulatory network of six CAMTA

genes in rice using GENIE3 and found various regulatory patterns (**Figure 6C**). First, LOC4331903 had the most targets (123 genes), while LOC43953 had the least targets (only 33

genes). Second, LOC4331903, LOC4327253, and LOC4343953 had both up-regulated and down-regulated target genes, but the target genes of LOC4335664 and LOC4343265 were all up-regulated. These results suggested that the regulation roles and patterns of the CAMTAs under cold stress have differentiated, and that LOC4331903 might have more important regulatory functions and more complex regulatory patterns than other CAMTAs. In addition, we also found that some target genes were regulated by both LOC4343953 and LOC4348485, and some target genes by three CAMTAs, namely, LOC4327253, LOC4335664, and LOC4343265. However, LOC4331903 shared very few target genes with other CAMTAs, namely, this CAMTA independently regulated target genes. This implied that LOC4327253, LOC4335664, and LOC4343265 might jointly regulate their target genes, LOC4343953 and LOC4348485 tended to cooperate, but LOC4331903 was more likely to function independently.

On the one hand, CAMTA may increase the cold resistance of plants by regulating the expression of carbohydrate-related genes to accumulate more carbohydrate. On the other hand, under cold stress, CAMTA transcription factors can regulate the transcription of target genes or work together with other transcription factors to increase rice resistance to low temperature stress.

DISCUSSION

Origin of Calmodulin Binding Transcription Activator Genes

Although genome-wide identification of CAMTA family has been completed in *A. thaliana*, *O. sativa*, *Z. mays*, *G. max*, *M. truncatula,* and other plants, the origin of CAMTA is still a controversial issue. A previous study identified the CAMTA genes from 6 chlorophyta genomes available at that time and 35 plant genomes, and found that no gene containing both a CG-1 domain and an IQ domain or a CaMB domain in any chlorophyta genome (Rahman et al., 2016b). Based on these findings, CAMTA genes were assumed to emerge from the embryophyta lineage ancestor. As mentioned in the Introduction section, the CAMTA proteins consist of multiple functional domains or motifs, including CG-1, TIG, ANK, IQ, and CaMB. Among them, the CG-1 is a unique DNA-binding domain, which is necessary for CAMTA to bind to DNA and perform its function. The IQ motifs can bind to CaM and CaM-like proteins in a Ca^{2+}-independent manner. Similar to the IQ motifs, CaMB domain is also a calmodulin-binding domain, but it binds to CaM and CaM-like proteins in a Ca^{2+}-dependent manner. In the previous study, only six chlorophyta genomes were used, in which no gene contained both CG-1 and IQ motif or CaMB domain simultaneously. In our study, CAMTA genes were undetectable in rhodophyta, but they were identified from 6 out of 28 chlorophyta plants. All these identified CAMTA genes harbored the CG-1 domain, the ANK domain, and at least one of CaMB domain, IQ, or IQ-like motif (**Figure 4**). The results showed that except that CaMB domains of our identified chlorophyta CAMTAs were slightly smaller than those of land plant CAMTAs, there was no significant difference in CAMTA

genes between chlorophyta and land plants. However, the size of CaMB domains identified from chlorophyta were similar to that in some CAMTAs from basal angiosperms and lycopodiopsida species, such as GWHPAAYW020451, GWHPAAYW004900, and Smo 441667. The CaMB domains with similar size could also be found in the some CAMTAs of even a few monocots and dicots, such as Gorai.006G079000.1, Solyc01g105230.2.1, LOC Os10g22950.1, and LOC Os01g69910.1. These genes have the functional domains of CAMTA, especially characteristic CG-1, CaMB, or IQ, thus we speculated that CAMTAs might be originated from chlorophyta. CAMTAs were detected from a small number of chlorophyta and all the evolutionarily young land species, but not from the evolutionarily old rhodophyta, which confirmed our speculation. More importantly, a recent study has reported the evolution and diversity of transcription factors in *A. angustus* and 18 green plants, and has identified CAMTAs from *K. nitens* and *C. braunii* (Zhang et al., 2020). *K. nitens* belongs to the nematophyta order of chlorophyta, and usually grows in filamentous form in temperate streams (John, 2003). *C. braunii* is an evolutionarily advanced chlorophyta species, and it resembles a land plant. These evidences also support our speculation that CAMTA is originated from chlorophyta. In addition, some studies have indicated that many gene families are originated from chlorophyta species. For example, WRKY, a star transcription factor, used to be regarded as unique to land plants. However, in recent years, it has gradually become a mainstream view that WRKY is originated from chlorophyta (Chen et al., 2017).

Gene Structure and Protein Domain

To obtain an insight into the structural diversity of CAMTA genes, the intron-exon organization was analyzed. The results showed that the number of introns in the CAMTA genes drastically varied, ranging from 0 to 34, but mainly concentrated between 10 and 13. In general, there were fewer introns in lower plants. These results were consistent with those of one previous study on a small number of plants (Rahman et al., 2016b). We also had some new findings when analyzing the phases of the introns. Phase pattern 0200-1102-0110 and 1102-0110-0200 were far more than other patterns in CAMTAs. Besides, the length of exons and the phase of introns exhibited a certain extent of correlation. For instance, the exons corresponding to 0200 or 0201 were generally significantly longer than those corresponding to 1102 or 0110. These results suggested that intron phase might be related to gene structure, but the possible reason for this phenomenon needs to be further studied.

Although CAMTAs contained a very conserved domain, we still observed the following gene structure variations. First, most CAMTAs contained TIG domain, but a small part did not. The distribution pattern of TIG domain in different phylogenetic clusters and species varied. Second, the length of CAMTA proteins varied greatly (745 ~ 2309 aa), which might be mainly due to the different repetition numbers of ANK domains. Third, most CAMTA proteins had two IQ motifs (>70%), but some of them had one IQ motif (~20%). In addition, very few CAMTA proteins contained three IQ motifs or no IQ motif. The variation of CAMTA domains indicated their functional differentiation.

Further research and more evidence are needed to reveal the relationship between the protein structures and their functions.

Calmodulin Binding Transcription Activator Expression and Its Function of Cold Resistance

The studies of CAMTA genes in the model plant *A. thaliana*, *O. sativa*, and *Z. mays* have generated abundant functional characterization data, which are valuable for functional prediction of their orthologous genes in other plants. CAMTAs have been reported to be related to maturity and development of plant organs in *P. tomentose* and *S. lycopersicum* (Yang T. et al., 2012; Wei et al., 2017). Based on these findings, we collected transcription data of a total of 43 tissues from *A. thaliana*, *O. sativa*, and *Z. mays* for further analysis. The results showed that the transcription level had significant tissue specificity and development stage specificity. Our analysis also indicated that CAMTA genes had the potential to regulate organ maturation and senescence.

In addition to modulating the maturation and development of organs and tissues, CAMTAs has also been found to play important role in cold stress response in *A. thaliana* (Doherty et al., 2009; Kim et al., 2013). *O. sativa* is very sensitive to low temperature stress, and it makes a series of physiological and metabolic responses under low temperature stress. However, most existing studies of cold stress-related genes in rice are limited to the preliminary mapping of quantitative trait loci (QTLs), and so far only 8 cold resistance-related genes have been cloned and identified (Mao et al., 2019). Especially, the function of CAMTAs in rice under cold stress remains largely unknown. We collected two sets of RNA-Seq data including PRJNA430015 and PRJNA607661 to investigate the roles of CAMTAs under cold stress. The intersection of DEGs detected by DESeq2 and edgeR and potential target genes detected by MEME was defined as target genes regulated by CAMTAs. A total of 255 CAMTA target genes were up-regulated, and 66 were down-regulated. Some down-regulated genes were involved in carbohydrate metabolism. Based on these results, we speculated that CAMTAs might improve the cold resistance of plants by regulating the genes related to carbohydrate metabolism to accumulate more carbohydrates. Some up-regulated target genes might be related to low temperature tolerance in *O. sativa*, of which transcription factor CBF1 is a case in point. The promoter region of CBF1 contains a CG-1 sequence at 1 kb upstream of the start codon. CAMTA has been confirmed to induce the expression of CBF1 under the cold stress in *A. thaliana* (Doherty et al., 2009), which is in line with our results that CAMTA also affected the expression level of CBF1 in rice, thus indicating the above cold-resistance mechanism may also be applicable to rice.

CONCLUSION

In this study, we identified a total of 465 CAMTA genes from 112 plants. CAMTA gene was not detected in rhodophyta, but was identified from 6 out of 28 chlorophyta and all the evolutionarily young species. Therefore, we speculated that CAMTA might be originated from chlorophyta, which was also supported by a recent study that identified CAMTAs in *K. nitens* and *C. braunii*. We constructed a phylogenetic tree of CAMTA proteins based on the maximum likelihood, and inferred that CAMTAs might differentiate into three groups, and all these 3 groups expanded rapidly due to the great changes from aquatic to terrestrial environment. Gene structure and protein domain analyses indicated that the protein domains of CAMTA were basically conservative. However, the number of introns in the CAMTA genes dramatically varied, ranging from 0 to 34, and a certain extent of correlation was observed between the length of exons and the phase of introns. Gene expression profiles showed that the transcription level had significant species specificity, tissue specificity, and developmental stage specificity. The CAMTA genes presented the potential to regulate organ maturation and senescence. Under cold stress, CAMTAs in rice might regulate carbohydrate accumulation and target genes transcription, and these CAMTA genes work together with other transcription factors to resist low temperature stress. Our study provides a comprehensive insight into the origin and molecular evolution of CAMTA family and a valuable reference for further research on its biological functions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JG and PX conceived, designed the research, and wrote the manuscript. PX conducted the identification, phylogenetic, and structure analysis. PX, J-WF, and X-TZ carried out the gene expression and functional analysis. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.758187/full#supplementary-material

Supplementary Figure 1 | Structure analysis of identified CAMTA genes.

Supplementary Figure 2 | Phylogenetic tree and functional domains of identified CAMTAs. CG-1, CG-1 DNA-binding domain; TIG, transcription-associated

immune globulin-like domain; ANK, ankyrin repeat domain; IQ, Ca^{2+}-independent CaM-binding IQ motif; CaMB, Ca^{2+}-dependent CaM binding domain.

Supplementary Figure 3 | Expression level of CAMTA genes in 12 species.

Supplementary Table 1 | A total of 112 plant genomes for subsequent analysis.

Supplementary Table 2 | The transcription data for subsequent analysis.

Supplementary Table 3 | A total of 465 identified CAMTA proteins.

Supplementary Table 4 | Intron number and intron phase of identified CAMTA genes.

Supplementary Table 5 | The selected representative species.

Supplementary Table 6 | Length, molecular weight, and isoelectric point of identified CAMTA proteins.

Supplementary Table 7 | Gene Ontology analysis of identified CAMTA genes.

REFERENCES

Aravind, L., and Koonin, E. V. (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* 287, 1023–1040. doi: 10.1006/jmbi.1999.2653

Bastian, M., Heymann, S., and Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks," in *Proceedings of the International AAAI Conference on Weblogs and Social Media*. San Jose, CA: PKP Publishing Services. Available online at: https://ojs.aaai.org/index.php/ICWSM/article/view/13937

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bork, P., Doerks, T., Springer, T. A., and Snel, B. (1999). Domains in plexins: links to integrins and transcription factors. *Trends Biochem. Sci.* 24, 261–263. doi: 10.1016/S0968-0004(99)01416-4

Bouché, N., Yellin, A., Snedden, W. A., and Fromm, H. (2005). Plant-specific calmodulin-binding proteins. *Annu. Rev. Plant Biol.* 56, 435–466. doi: 10.1146/annurev.arplant.56.032604.144224

Chen, F., Hu, Y., Vannozzi, A., Wu, K., Cai, H., Qin, Y., et al. (2017). The WRKY transcription factor family in model plants and crops. *Crit. Rev. Plant Sci.* 36, 311–335. doi: 10.1080/07352689.2018.1441103

Choudhuri, S. (2014). *Bioinformatics for Beginners*. Amsterdam: Elsevier Press.

da Costa e Silva, O. (1994). CG-1, a parsley light-induced DNA-binding protein. *Plant Mol. Biol.* 25, 921–924. doi: 10.1007/BF00028887

Darriba, D., Taboada, G., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088

Diao, P., Chen, C., Zhang, Y., Meng, Q., Lv, W., and Ma, N. (2020). The role of NAC transcription factor in plant cold response. *Plant Signal. Behav.* 15:1785668. doi: 10.1080/15592324.2020.1785668

Doherty, C. J., Van Buskirk, H. A., Myers, S. J., and Thomashow, M. F. (2009). Roles for *Arabidopsis* CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. *Plant Cell* 21, 972–984. doi: 10.1105/tpc.108.063958

Du, L., Ali, G. S., Simons, K. A., Hou, J., Yang, T., Reddy, A. S., et al. (2009). Ca2+/calmodulin regulates salicylic-acid-mediated plant immunity. *Nature* 457, 1154–1158. doi: 10.1038/nature07612

Finkler, A., Ashery-Padan, R., and Fromm, H. (2007). CAMTAs: calmodulin-binding transcription activators from plants to human. *FEBS Lett.* 581, 3893–3898. doi: 10.1016/j.febslet.2007.07.051

Fornes, O., Castro-Mondragon, J. A., Khan, A., Van Der Lee, R., Zhang, X., Richmond, P. A., et al. (2019). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 48, 87–92. doi: 10.1093/nar/gkz1001

Galon, Y., Finkler, A., and Fromm, H. (2010b). Calcium-regulated transcription in plants. *Mol. Plant* 3, 653–669. doi: 10.1093/mp/ssq019

Galon, Y., Aloni, R., Nachmias, D., Snir, O., Feldmesser, E., Scrase-Field, S., et al. (2010a). Calmodulin-binding transcription activator 1 mediates auxin signaling and responds to stresses in *Arabidopsis*. *Planta* 232, 165–178. doi: 10.1007/s00425-010-1153-6

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., et al. (2002). *The Ensembl Genome Database Project*. Available online at: http://www.bioperl.org/ (accessed May 17, 2021)

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5:e12776. doi: 10.1371/JOURNAL.PONE.0012776

John, D. M. (2003). "Filamentous and plantlike green algae," in *Freshwater Algae of North America*, eds D. J. Wehr, and R. G. Sheath (Cambridge, MA: Academic Press), 311–352. doi: 10.1016/B978-012741550-5/50009-X

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kim, M. C., Chung, W. S., Yun, D. J., and Cho, M. J. (2009). Calcium and calmodulin-mediated regulation of gene expression in plants. *Mol. Plant* 2, 13–21. doi: 10.1093/mp/ssn091

Kim, Y., Park, S., Gilmour, S. J., and Thomashow, M. F. (2013). Roles of CAMTA transcription factors and salicylic acid in configuring the low-temperature transcriptome and freezing tolerance of *Arabidopsis*. *Plant J.* 75, 364–376. doi: 10.1111/tpj.12205

Kolde, R. (2012). *Pheatmap: Pretty Heatmaps*. Available online at: https://CRAN.R-project.org/package=pheatmap (accessed May 17, 2021).

Kudla, J., Batistič, O., and Hashimoto, K. (2010). Calcium signals: the lead currency of plant information processing. *Plant Cell* 22, 541–563. doi: 10.1105/tpc.109.072686

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090

Leinonen, R., Sugawara, H., and Shumway, M. (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Liu, C., Wu, Y., and Wang, X. (2012). bZIP transcription factor OsbZIP52/RISBZ5: a potential negative regulator of cold and drought stress response in rice. *Planta* 235, 1157–1169. doi: 10.1007/s00425-011-1564-z

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Mao, D., Xin, Y., Tan, Y., Hu, X., Bai, J., Liu, Z. Y., et al. (2019). Natural variation in the HAN1 gene confers chilling tolerance in rice and allowed adaptation to a temperate climate. *Proc. Natl. Acad. Sci. U.S.A.* 116:201819769. doi: 10.1073/pnas.1819769116

Miller, M. A., and Schwartz, T. (2011). "The CIPRES science gateway: a community resource for phylogenetic analyses sustainable software for science view project CIPRES view project," in *Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery* (New York, NY: Association for Computing Machinery). doi: 10.1145/2016741.2016785

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Pandey, N., Ranjan, A., Pant, P., Tripathi, R. K., Ateek, F., Pandey, H. P., et al. (2013). CAMTA 1 regulates drought responses in *Arabidopsis thaliana*. *BMC Genomics* 14:216. doi: 10.1186/1471-2164-14-216

Poovaiah, B. W., and Du, L. (2018). Calcium signaling: decoding mechanism of calcium signatures. *New Phytol.* 217, 1394–1396. doi: 10.1111/nph.15003

Popescu, S. C., Popescu, G. V., Bachan, S., Zhang, Z., Seay, M., Gerstein, M., et al. (2007). Differential binding of calmodulin-related proteins to their targets revealed through high-density *Arabidopsis* protein microarrays. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4730–4735. doi: 10.1073/pnas.0611615104

Rahman, H., Yang, J., Xu, Y. P., Munyampundu, J. P., and Cai, X. Z. (2016b). Phylogeny of plant CAMTAs and role of AtCAMTAs in nonhost resistance to xanthomonas oryzae pv. oryzae. *Front. Plant Sci.* 7:177. doi: 10.3389/fpls.2016.00177

Rahman, H., Xu, Y. P., Zhang, X. R., and Cai, X. Z. (2016a). Brassica napus genome possesses extraordinary high number of CAMTA genes and CAMTA3 contributes to PAMP triggered immunity and resistance to Sclerotinia sclerotiorum. *Front. Plant Sci.* 7:581. doi: 10.3389/fpls.2016.00581

Reddy, A. S. N., Ali, G. S., Celesnik, H., and Day, I. S. (2011). Coping with stresses: roles of calcium- and calcium/calmodulin-regulated gene expression. *Plant Cell* 23, 2010–2032. doi: 10.1105/tpc.111.084988

Reddy, A. S. N., Reddy, V. S., and Golovkin, M. (2000). A calmodulin binding protein from *Arabidopsis* is induced by ethylene and contains a DNA-binding motif. *Biochem. Biophys. Res. Commun.* 279, 762–769. doi: 10.1006/bbrc.2000.4032

Reddy, V. S., Ali, G. S., and Reddy, A. S. N. (2002). Genes encoding calmodulin-binding proteins in the *Arabidopsis* genome. *J. Biol. Chem.* 277, 9840–9852. doi: 10.1074/jbc.M111626200

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Shangguan, L., Wang, X., Leng, X., Liu, D., Ren, G., Tao, R., et al. (2014). Identification and bioinformatic analysis of signal responsive/calmodulin-binding transcription activators gene models in *Vitis vinifera*. *Mol. Biol. Rep.* 41, 2937–2949. doi: 10.1007/s11033-014-3150-5

Snedden, W. A., and Fromm, H. (2001). Calmodulin as a versatile calcium signal transducer in plants. *New Phytol.* 151, 35–66. doi: 10.1046/j.1469-8137.2001.00154.x

Song, K., Backs, J., McAnally, J., Qi, X., Gerard, R. D., Richardson, J. A., et al. (2006). The transcriptional coactivator CAMTA2 stimulates cardiac growth by opposing class II histone deacetylases. *Cell* 125, 453–466. doi: 10.1016/j.cell.2006.02.048

Wang, G., Zeng, H., Hu, X., Zhu, Y., Chen, Y., Shen, C., et al. (2014). Identification and expression analyses of calmodulin-binding transcription activator genes in soybean. *Plant Soil* 386, 205–221. doi: 10.1007/s11104-014-2267-6

Wang, L., Zhu, W., Fang, L., Sun, X., Su, L., Liang, Z., et al. (2014). Genome-wide identification of WRKY family genes and their response to cold stress in Vitis vinifera. *BMC Plant Biol.* 14:103. doi: 10.1186/1471-2229-14-103

Wei, M., Xu, X., and Li, C. (2017). Identification and expression of CAMTA genes in Populus trichocarpa under biotic and abiotic stress. *Sci. Rep.* 7:17910. doi: 10.1038/s41598-017-18219-8

Xue, Y., Bao, Y., Zhang, Z., Zhao, W., Xiao, J., He, S., et al. (2021). Database resources of the national genomics data center, china national center for bioinformation in 2021. *Nucleic Acids Res.* 49, D18–D28. doi: 10.1093/nar/gkaa1022

Yang, A., Dai, X., and Zhang, W. H. (2012). A R2R3-type MYB gene, OsMYB2, is involved in salt, cold, and dehydration tolerance in rice. *J. Exp. Bot.* 63, 2541–2556. doi: 10.1093/jxb/err431

Yang, T., Peng, H., Whitaker, B. D., and Conway, W. S. (2012). Characterization of a calcium/calmodulin-regulated SR/CAMTA gene family during tomato fruit development and ripening. *BMC Plant Biol.* 12:19. doi: 10.1186/1471-2229-12-19

Yang, T., and Poovaiah, B. W. (2000). An early ethylene up-regulated gene encoding a calmodulin-binding protein involved in plant senescence and death. *J. Biol. Chem.* 275, 38467–38473. doi: 10.1074/jbc.M003566200

Yang, T., and Poovaiah, B. W. (2003). Calcium/calmodulin-mediated signal network in plants. *Trends Plant Sci.* 8, 505–512. doi: 10.1016/j.tplants.2003.09.004

Yang, Y., Sun, T., Xu, L., Pi, E., Wang, S., Wang, H., et al. (2015). Genome-wide identification of CAMTA gene family members in medicago truncatula and their expression during root nodule symbiosis and hormone treatments. *Front. Plant Sci.* 6:459. doi: 10.3389/fpls.2015.00459

Yue, R., Lu, C., Sun, T., Peng, T., Han, X., Qi, J., et al. (2015). Identification and expression profiling analysis of calmodulin-binding transcription activator genes in maize (Zea mays L.) under abiotic and biotic stresses. *Front. Plant Sci.* 6:576. doi: 10.3389/fpls.2015.00576

Zegzouti, H., Jones, B., Frasse, P., Marty, C., Maitre, B., Latché, A., et al. (1999). Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. *Plant J.* 18, 589–600. doi: 10.1046/j.1365-313X.1999.00483.x

Zhang, J., Fu, X. X., Li, R. Q., Zhao, X., Liu, Y., Li, M. H., et al. (2020). The hornwort genome and early land plant evolution. *Nat. Plants* 6, 107–118. doi: 10.1038/s41477-019-0588-4