AJAB

Asian J Agric & Biol. 2019;7(1): 27-38.

Original Article

Protein modeling and evolutionary analysis of Calmodulin Binding Transcription Activator (CAMTA) gene family in *Sorghum bicolor*

Rashid Mehmood Rana[1], Saima Saeed[1], Fahad Masoud Wattoo[1], Muhammad Waqas Amjid[1], Muhammad Azam Khan[2]

[1]Department of Plant Breeding & Genetics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan

[2]Department of Horticulture, PMAS-Arid Agriculture University, Rawalpindi, Pakistan

Received:
March 09, 2018

Accepted:
July 31, 2018

Published:
March 30, 2019

**Corresponding author email:*
rashid.pbg@uaar.edu.pk

Abstract

Calmodulin Binding Transcription Activator (CAMTA) family is present in almost all plants and in many animals. CAMTA are named so due to the presence of specific calmodulin binding domain which is an important Ca2+ transducer. Multiple sequence alignment and phylogenetic analysis of CAMTA proteins in *Sorghum bicolor*, *Oryza sativa*, *Zea mays*, *Glycine max* and *Arabidopsis thaliana* showed highly conserved sequence and evolutionary similarity. In *Sorghum bicolor* six CAMTA proteins were identified to be located in nucleus. These proteins were named on the basis of their location on the chromosomes. Alignment and phylogenetic tree clearly indicates close similarity in monocot and dicot ancestry which appears on the same clade but diverged from each other with time. Almost all CAMTA proteins share same domain organizations. A highly conserved motif sequence in these species was identified which might play some important functional roles. In order to understand structural and DNA binding patterns of SbCAMTA proteins, 3-D models of proteins structure and their domains revealed many important DNA binding residues playing their role in protein-protein interaction and structural modification.. A further detailed study of the CAMTA protein members in sorghum may explore their mode of interaction and exact function in signaling mechanism under abiotic stresses.

Keywords: Binding residues, Calmodulin, Dicots, Evolution, Monocots, Proteins

How to cite this:

Rana RM, Saeed S, Wattoo FM, Amjid MW and Khan MA, 2019. Protein modeling and evolutionary analysis of Calmodulin Binding Transcription Activator (CAMTA) gene family in *Sorghum bicolor*. Asian J. Agric. Biol. 7(1): 27-38

Introduction

A set of transcription factors binds to the specific cis acting element of genes to regulate their expression in response to the various signals. Environmental stresses especially abiotic stresses are the most common problems to be faced by a plant. Plant adapt to these worse conditions by producing various signals (Ca2+signaling) (Mizoi et al., 2012). These signals are sensed by many EF-hand containing proteins. These proteins interact with different transcription factors like calmodulin binding transcription activator (CAMTA) to regulate stress responsive proteins by binding to their highly conserved cis acting element (G/A/C)CGCG (C/G/T) (Du et al., 2011; Kudla et al., 2010; Sanders et al., 2002; Yang and Poovaiah,

2002a). All CAMTA members do have a highly conserved DNA binding domain CG-1 (Bouché et al., 2002), IQ motif or calmodulin binding region (Bähler and Rhoads, 2002), and ankyrin repeat (Rubtsov and Lopina, 2000). Some members also have non-specific DNA binding sites or TIG domain (Aravind and Koonin, 1999). Calmodulin or CAM like proteins bind four calcium ions to their domains and target many kinases, phosphatases, metabolic enzymes, ion channels, ion exchangers etc. by undergoing different conformational changes to produce regulatory effects (Bouché et al., 2005; Cohen, 1982; Poovaiah et al., 2013; Yang and Poovaiah, 2003).

Calmodulin binding domain of CAMTA transcription factors play essential role in calcium signaling. (Choi et al., 2005). This evolutionary conserved transcription factor family is present in all organisms from humans to most of the plants like *Sorghum bicolor, Oryza sativa, Zea mays, Glycine max* and *Arabidopsis thaliana* etc. (Finkler et al., 2007). These proteins regulate various plants specific physiological responses against pathogen attack, senescence and environmental stresses. This vast involvement of CAMTA gene family from calcium dependent regulation gene expression under drought and cold (Doherty et al., 2009) to auxin signaling (Reddy et al. 2011) compels to investigate about evolutionary diversification of this gene family. Calmodulin, a most common Ca2+ sensor, interacts in different ways to regulate expression of many transcription factors like CAMTA in a calcium dependent manner. However it is yet to be identified that how CAMTA proteins interact with DNA to respond drought stress in sorghum. Therefore, the current study was planned to elucidate the evolution, structural modeling and DNA-Protein interaction of CAMTA proteins in plants. The CAMTA proteins from *Sorghum bicolor, Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana* were used to investigate evolution of CAMTA transcription family, while protein sequences from sorghum CAMTA were used for structural modeling ligand identification. Our results suggested the role of purifying selection during evolution. Amino acid substitutions suggested that relaxed constraints might have played significant role in functional divergence among CAMTA members. Presence of conserved motifs in CAMTA members is a very important aspect to further explore their functions. Analysis of motif sequence in these organisms revealed a highly conserved motif in all CAMTA members that might have some functional activities. 3-D structures of CAMTA domains explored many critical residues involved in DNA binding. Overall, this study reports genomic organization, comparative phylogenic analysis, and structural characterization of CAMTA proteins to facilitate understanding of their interaction mechanism.

Material and Methods

Retrieval and identification of CAMTA sequences

The CAMTA protein, CDS (CoDing Sequence) and genomic DNA sequences of sorghum bicolor, Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana were retrieved from phytozome v10.3using sequences fromplantfdbv3.0 as query. The retrieved protein sequences were further screened for the presence of functional CAMTA domains like CG1, TIG, ankyrin repeats and IQ or calmodulin binding domain using SMART (smart.emble.heidelberg.de).Duplicate or redundant protein sequences were also discarded. Structure of SbCAMTA genes was predicted from GSDS (http://gsds.cbi.pku.edu.cn/) along with their locations on chromosomes (http://blast.ncbi.nlm.nih.gov/Blast.cgi). In silico sub cellular localization of SbCAMTA family protein was predicted by a ProtComp v 9.0 (www.softberry.com).

Multiple sequence alignment, phylogenetic analysis and motif discovery

Alignment of sorghum, rice, maize, soybean, and Arabidopsis CAMTA proteins along with an unrelated E. coli protein (gi1498749) as outgroup was performed with clustalX (Thompson et al. 1997). The phylogenetic tree was constructed using MEGA6using neighbor joining method. Bootstrap test of phylogeny was performed with 1000 replicates using pair wise deletions and p-distance model. DNA binding motif sequences were identified using meme suit (meme-suit.org/tools/meme).

Selection tests

MEGA6 selection models were employed to detect positive selection pressure on coding sites during the evolution of CAMTA genes. Nucleotide substitution rates (non-synonymous to synonymous (dN/dS)) were applied to estimate maximum likelihood ratio of selection pressure (ω) on codon sites in a specific period of time by using best models of likelihood (ML).

Functional divergence analysis
Functional divergence analysis was done by using DIVERGE v2.0 to detect amino acids residues responsible for functional divergence. The coefficient of functional divergence (θ), likelihood ratio testing (LRT), and posterior probabilities were estimated between 5 phylogenetic groups.

Prediction of 3D structural model and DNA binding domain
Structural model of six sorghum CAMTA proteins were developed using RaptorX (www.raptorX.uchicago.edu). RaptorX (http://raptorx.uchicago.edu/StructurePrediction/predict/) predicts 3D Structure of input sequence by aligning it with experimentally determined templates. Models quality was assessed on the bases of alignment score, e-value and global distance test (GDT) scores. Evaluation of the models was performed by comparing it with ramachandran plot using RAMPAGE. DNA binding domain protein models were determined by using Raptor X (http://raptorx.uchicago.edu/BindingSite/). Ligands binding to the many critical residues were determined and analyzed for identification of any critical residue important for DNA binding to CAMTA domains.

Results

Sequence retrieval and domain analysis
Raw sequence of CAMTA proteins, CDS, gDNA were screened for the presence of functional domains. After screening, a total of 6, 7, 6, 15, and 6 CAMTA proteins were found in *Sorghum bicolor, Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana* respectively. Subsequent analysis of six SbCAMTA proteins revealed the presence of CG-1 and Ankyrin repeat domains and IQ motifs in all six proteins with an additional TIG domain in SbCAMTA1, SbCAMTA2, SbCAMTA3, and SbCAMTA5 (Fig. 1). All three of these domains and IQ motif were also found conserved in *Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana.* The structure of SbCAMTA genes on the basis of information obtained from genomic DNA and CDS revealed 13 exons in each gene except SbCAMTA6 which contained 10 exons (Fig 1). Location of SbCAMTA proteins on chromosomes, and their molecular weights and isoelectric potentials are shown in Table 1. All six SbCAMTA proteins were located in nucleus when analyzed through ProtCom v 9.0 (Table 1).

Alignment, phylogenetic analysis and motif discovery
Multiple sequence alignment of CAMTA protein from *Sorghum bicolor, Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana* was performed using clustalX. Alignment output highlighted about 44 amino acids showing 100% similarity indicating high conservation of these amino acids in all the five organisms (Supplementary Fig. 1). Presence of highly conserved sequences clearly indicates that this protein might show some functional homology in these organisms. Multiple sequence alignment of proteins was used to construct phylogenetic tree. In phylogenetic tree, outgroup (gi1498749) clearly defined the basic evolutionary node and clustered proteins in three major groups. Evolutionarily, SbCAMTA is much more similar to ZmCAMTA (sister lines) as compared to other organisms except SbCAMTA5 more close to OsCAMTA4 (Fig. 2). Alignment analysis also showed high sequence similarity with a little difference (only at a few bases) in protein sequence of both organisms. Conserved motifs in protein sequences were discovered by using all CAMTA sequence of five organisms as query at online web server MEME suit. Highly conserved motifs were present in CAMTA proteins of all five organisms as presented in Figure 3.

Selection tests in CAMTA proteins
MEGA6 applies best model to select for positively substituting codons in CAMTA members. Each codon has specific dN/dS which provide maximum likelihood of selection pressure on coding site. Negative selection constraints were found because ω or d_N/d_S ratios were not significant (less than 1) (Table 2).Best models of selection pressure were selected based on lowest AIC (Akaika Information Criterion) and BIC (Bayesian Information Criterion) values. Substitution models showed restricted variation of bases (base frequencies >1). Non uniform variation (Gamma) and non-varying sites (invariants) were calculated in all models where applicable. Gamma shape estimates were ≥1 showing that most of the sites evolved slowly but some at faster rates. Proportion of invariable sites decreased (<1) as the heterogeneity rate increased.

Functional divergence analysis
Functional divergence analysis of aligned sorghum, rice, maize, soybean, and Arabidopsis sequences was performed through DIVERGE v2.0 to identify amino

acids involved in functional divergence. CAMTA proteins were grouped in three clusters and comparison of these groups like c1/c2, c1/c3, and c2/c3 identified4, 8, and 8 amino acid residues respectively, contributing to divergence. Likelihood ratio test (LRT) and standard error (SE) showed significant functional constraints (Table 3; Fig. 4). High values of posterior probabilities of amino acid residues like in C1/C2, C1/C3 and C2/C3 suggest their role in functional divergence.

Structural protein modeling

Structural modeling of SbCAMTA proteins was done by using template based modeling to develop tertiary structure through RaptorX (Figure 6). SbCAMTA protein sequences showed maximum homology to the crystal structure of the TIG domain of human calmodulin binding transcription activator (CAMTA1; PDB ID: 2cxka). Query sequence of all six SbCAMTA proteins showed more than 65% similarity to the template based on the alignment score.

Probability value (e-value) and global distance test (GDT) score assess models quality predicting their similarity to experimentally determined structures. GDT evaluates quality of 3D structure based on the number of residues in protein structure by setting a specific cutoff value (>50)All models generated through RaptorX gave high quality structures with e-value less than 10^{-4} and GDT of SbCAMTA1, SbCAMTA2, SbCAMTA3, SbCAMTA4, SbCAMTA5, and SbCAMTA6 were GDT= 70.8, 70.7, 73.7, 76.8, 74.8, and 70.9 respectively (Table 4). Analysis of secondary structure of SbCAMTA proteins revealed frequent occurrence of loops and alpha helical regions as compared to beta sheets. Presence of high percentage of loops or turns makes protein structure more stable as it allows better interaction of beta sheets and alpha helices. Backbone confirmation and structure evaluation was performed through ramachandran plot analysis using RAMPAGE web server. 3D structures predicted through raptorX were of good quality as most (90%) of the identified residues were in allowed and favored region. A very few percentage of residues were found in disallowed or generous regions (Table 4).

Critical residues

Analysis of SbCAMTA 3D structure using raptorX revealed critical residues (ligand binding) responsible for interaction with receptor molecules. Best model containing critical residues was selected on the bases of e-value, template quality along with their GDT score from a total of three predicted models by raptorX (Figure 7).

Predicted model ofSbCAMTA1 consist residual binding sites at position Asp^{471}, Asp^{496}, Ser^{372}, and Ala^{495} for DA (2'-Deoxyadenosine-5'-monophosphate). Asp^{471}, Ser^{472} and Asp^{496}were present in coils forming a turn indicating their role in high stability and conservation but Ala^{495} is located in helical region. Asp^{496} and Asp^{471} showed 52.6% and 46.3% solvent accessibility and are quite exposed to binding as compared to the other residues. SbCAMTA2 domain binds to residues in coil region with Asp^{532} that is highly exposed (67.7%) to binding. SbCAMTA3 has most of the buried residues although present in helical turns of protein. Similarly SbCAMTA4also contributed DA binding to Asp^{436} in helical turns with more than 50%solvent accessibility. Three critical residues Pro^{334}, Pro^{335}, and Gln^{355} were also identified in SbCAMTA5 having affinity to bind with a DNA linking ligand DG (2'-deoxyguanosine-5'-monophosphate). This region contains highly conserved amino acid based on their high degree of presence in coils. Pro^{335} was highly exposed (about 90.1%) to the ligand binding, whereas Pro^{334} was neither highly exposed nor too buried sharing a medium binding capacity.SbCAMTA6 has one important residue Lys^{452}located in coils with more than 50% exposure to the binding of DA.

Discussion

Calmodulin binding transcription activator (CAMTA) is a key transcription factor family in triggering Ca^{2+}regulated stress responses (Finkler et al., 2007; Lenzoni et al., 2018). In CAMTA proteins, IQ or calmodulin (CaM) domain is the most important domain that senses interacellular changes in calcium levels under stress (Yang and Poovaiah, 2002a) and modulates signal transduction pathway (Chung et al., 2000). Calmodulin maintains homeostasis of ROS (Reactive oxygen species) in cells as it binds and activates ROS degrading enzymes (Yang and Poovaiah, 2002b). CAM proteins control many aspects of cellular respiration as a classical calcium binding protein (Bouché et al., 2005; DeFalco et al., 2010; Du et al., 2009; Reddy et al., 2011). In stresses, free cytoplasmic Ca^{2+} is sensed by CaM proteins which manage nuclear transport of proteins in GTP independent manners by stimulating effector proteins (still unknown) (Bachs et al., 1992; Carafoli, 1987;

Pruschy et al., 1994; Vogel, 1994). Keeping in view the importance of CaM dependent gene activation under abiotic stresses for plants, we conduct analysis of CAMTA proteins including their identification, characterization, evolutionary mechanism and protein structural analysis to explore their binding residues and motif sequences in *Sorghum bicolor, Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana.*

The query sequence of *Sorghum bicolor, Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana* from plantfdb v3.0 were used to retrieve protein, CDS, and genomic DNA sequence from phytozome v10.3. After screening, a total of 6, 6, 7, 15, and 6 CAMTA proteins were found in *Sorghum bicolor, Oryza sativa, Zea mays, Glycine max, Arabidopsis thaliana* respectively. CAMTA proteins of Sorghum bicolor were named as SbCAMTA1, SbCAMTA2, SbCAMTA3, SbCAMTA4 SbCAMTA5, and SbCAMTA6. These genes were named according to their location on chromosomes in ascending order (Table 1). Domain analysis through "SMART" showed that all CAMTA members have calmodulin binding domains and DNA binding motif like CG-1, IQ, and ankyrin repeats which are used to identify CAMTA proteins in many multicellular organisms (Bouché et al., 2002) but showed variation for the presence of TIG. Multiple sequence alignment of SbCAMTA proteins in comparison with rice, maize, soybean, and Arabidopsis showed high degree of homology in sequence as observed through "clustalX". Phylogenetic tree divided all species (sorghum, rice, maize, soybean, and Arabidopsis) in three major clusters. Dicots and monocots appeared in the same major cluster while divided clearly in sub-clusters which indicates common ancestry and diversion of these groups (Chaw et al., 2004; Zanis et al., 2002). Phylogenetic tree clearly indicates the close similarity between monocots, like SbCAMTAs showed higher similarity to ZmCAMTA proteins except SbCAMTA6 which is quite similar to OsCAMTA7. Homology of CAMTA genes of maize and rice has also been reported previously (Yue et al. 2015). Our results are further supported by the previous studies which claim that all monocots of order poales have originated from common ancestor sharing a number of morphological traits (Chase et al., 1995; Donoghue and Doyle, 1988; Doyle and Donoghue, 1992; Duvall et al., 1993; Hamby and Zimmer, 1992; Tucker and Douglas, 1996). Whereas dicot (Soybean and *Arabidopsis thaliana*) proteins like GmCAMTAs shows similarity to the AtCAMTAs only GmCAMTA17 that appears apart away from the rest of members which might be due to the presence of IPT or non-specific DNA binding domain. Previous results also showed closer homology of AtCAMTA to the dicots (Bowers et al., 2003). Selection analysis detects restricted selection pressure on coding sites by applying many best fit evolutionary models which were selected on the basis of lowest AIC and BIC values (Akaike, 1974). Omega (ω) estimates of best models were pointing towards negative selection constraints based on dN/dS ratios inferring high stability of these species however posterior probability analysis depicted many critical residues to be involved in functional divergence (Type I). Interestingly, all the critical residues except '154' were present in CG-1 domain on N-terminal of CAMTA proteins suggesting evolution of some specie specific DNA binding sites (Huang et al., 2008). Highly conserved motif sequence was identified in protein sequence of all five organisms probably having some role in protein folding (structural motif). Advanced analysis of motif sequence in SbCAMTA protein can be applied to further explore their exact functional role. All of the SbCAMTA proteins were nuclear localized as revealed by in silico sub cellular localization depicting their possible; role in modulating expression of different proteins to respond to abiotic stresses (Bouché et al., 2002).

Template based protein models of all SbCAMTA were developed using 2cxka (TIG domain of HsCAMTA1) as template and verified through ramachandran analysis. 3D structures of SbCAMTA proteins were used to find DNA binding residues. Most of the binding residues were located in helix loop helix region. We identified that CAMTA proteins in *Sorghum bicolor* like SbCAMTA1 and SbCAMTA2 contains negatively charged amino acids like Asp^{471}, Asp^{496} and Asp^{532} respectively are present in EF hand motif. These negatively charged amino acids are known to be involved in binding of $Ca^{2+}$by interacting with positive charges of calcium (Lewit-Bentley and Réty 2000). SbCAMTA5 bind to the ligand through proline rich sites (Pro^{334} and Pro^{335}) that are very important for binding domains and protein-protein interactions (Sedgwick and Smerdon 1999). Residues of SbCAMTA1 (Ser^{472}, Asp^{471}, Asp^{496}, Ala^{495}), SbCAMTA5 (Pro^{334}, Pro^{335}, Gln^{355}) and SbCAMTA6 (Val^{455}, Lys^{452}) are present in DNA binding domain i.e. TIG probably involved in binding of CAMTA proteins by interacting with different bases of DNA (Aravind and Koonin, 1999; Ghosh et al., 1995; Müller et al., 1995). The identification of these residues gave us

quite deeper insights on how CAMTA domains interact with DNA to perform their functions. However, a more detailed analysis is required to clearly elucidate CAMTA-DNA interaction.

Taken together, in this study six CAMTA proteins with functional domains (CG-1, and IQ) in sorghum bicolor localized in nucleus and characterized them for evolution, and functional divergence in comparison with rice, maize, soybean, and Arabidopsis. Template based protein modeling of SbCAMTAs to determine their 3-D structures and DNA binding residues was performed to elucidate important binding residues that help in binding of SbCAMTA proteins to the DNA. Further investigation of SbCAMTA proteins is required to explore their role in signaling mechanism that has not been identified yet.

Table 1: Characterization of SbCAMTA proteins

Protein name	Accession number	reference sequence ID	Gene ID	No. of exons	Chromosome number	Location on chromosome	Size of protein	Molecular weight	Isoelectric points	Sub cellular localization	Score
SbCAMTA1	XM_002489167.1	XP_002489212	8155742	12	unknown	unknown	1024	113750.82	5.63	Nucleus	8.5
SbCAMTA2	XP_002467764	XM_002467719.1	8077118	13	Chr 1	56,966,508 - 56,975,715	1027	114657.45	5.86	Nucleus	8.5
SbCAMTA3	XP_002465719	XM_002465674.1	8081906	12	Chr 1	67576601-67583786	997	111335.15	5.82	Nucleus	8.3
SbCAMTA4	XP_002462876	XM_002462831.1	8077453	13	Chr 2	68189250-68198270	949	106292.6	8.58	Nucleus	8.5
SbCAMTA5	XP_002456865	XM_002456820.1	8079812	10	Chr 3	71520747-71527627	848	95477.97	6.97	Nucleus	8.5
SbCAMTA6	XP_002463205	XM_002463160.1	8061564	13	Chr 2	73763662-73770810	1015	113453.54	5.55	Nucleus	8.5

Table 2: Maximum likelihood and their gamma distributions and invariables under model of ω variables CAMTA proteins

Model	InL	Gamma (G)	Invariant (I)	dN/dS or ω
T92+G+I	-23627.83795	1.579432312	0.102447417	ω<1
GTR+G+I	-23610.62064	1.591799302	0.099741127	ω<1
T92+G	-23653.39960	0.962352104	n/a	ω<1
GTR+G	-23634.39287	0.983865348	n/a	ω<1
TN93+G+I	-23648.95627	1.595505218	0.102009223	ω<1
HKY+G+I	-13658.03958	1.584511287	0.101740254	ω<1
TN93+G	-23674.02317	0.973235972	n/a	ω<1
HKY+G	-23682.86602	0.969385832	n/a	ω<1
K2+G+I	-23744.11312	1.671751687	0.105065793	ω<1
K2+G	-23772.34829	0.994989742	n/a	ω<1
T92+I	-24133.44713	n/a	0.133678497	ω<1
GTR+I	-24105.69142	n/a	0.13357486	ω<1
TN93+I	-24144.53521	n/a	0.133650434	ω<1
HKY+I	-24160.65904	n/a	0.133635034	ω<1
K2+I	-24213.11331	n/a	0.133730352	ω<1
JC+G+I	-24215.83658	1.85685698	0.106743017	ω<1
JC+G	-24247.91901	1.075108901	n/a	ω<1
JC+I	-24640.24036	n/a	0.133668904	ω<1
T92	-24632.64291	n/a	n/a	ω<1
GTR	-24693.10255	n/a	n/a	ω<1
TN93	-24690.89564	n/a	n/a	ω<1
HKY	-24710.43611	n/a	n/a	ω<1
K2	-24768.29674	n/a	n/a	ω<1
JC	-25179.09447	n/a	n/a	ω<1

Inl, Maximum likelihood value; +G, Gamma distribution; +I, Invariants; dn/ds, Non synonymous to synonymous ratio.
GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor. ω< 1= less significant (posing towards negative selection constraints)

Table 3: Amino acid residues responsible for functional divergence with their LRT and SE for significant constraints

Clusters	Amino acids	Posterior probabilities	SE	LRT
C1/C2	52, 58, 108, 135	0.90, 0.95, 0.94, 0.92	0.118189	19.47689
C1/C3	52, 53, 58, 65, 75, 115, 135, 154	0.98, 0.95, 0.99, 0.97, 0.94, 0.97, 0.96, 0.90	0.085877	51.0528
C2/C3	39, 54, 68, 87, 208, 146, 148, 154	0.94, 0.91, 0.92, 0.97, 0.94, 0.91, 0.90	0.081082	38.39325

SE, Standard error; LRT, likelihood ratio test (show significant functional constraints

Table 4: Domain and Structure analysis of SbCAMTA proteins

Protein number	Domain analysis					Structural analysis		Ramachandran analysis	
	Ligand	Residues	e-value	GDT	Pocket similarity	No. of residues	e -value	Allowed region	Favored regions
SbCAMTA1	DA	S472, D471, D496, A495	7.53e-07	70.8	101	821	7.56e-07	756 (90.1%)	61 (7.3%) Outlier: 22 (2.6%)
SbCAMTA2	MPD	D532, S531, A643, Q542, R540	4.88e-07	70.7	10	730	4.36e-07	670 (89.6%)	59 (7.9%) Outlier: 19 (2.5%)
SbCAMTA3	DG	F404, L403, Q402	3.28e-06	73.7	10	834	3.65e-06	768 (90.1%)	48 (5.6%) Outlier: 36 (4.2%)
SbCAMTA4	DA	Y437, N436, Q460	4.54e-06	76.8	52	691	5.53e-06	656 (92.5%)	32 (4.5%) Outlier: 21 (3.0%)
SbCAMTA5	DG	P334, P335, Q355	7.49e-06	79.3	49	678	7.49e-06	616 (88.8%)	48 (6.9%) Outlier: 30 (4.8%)
SbCAMTA6	DA	V455, K452	1.28e-06	26.5	48	736	1.36e-06	661 (87.9%)	59 (7.8%) Outlier: 32 (4.3%)

e-value, probability value; GDT, Global Distance Test

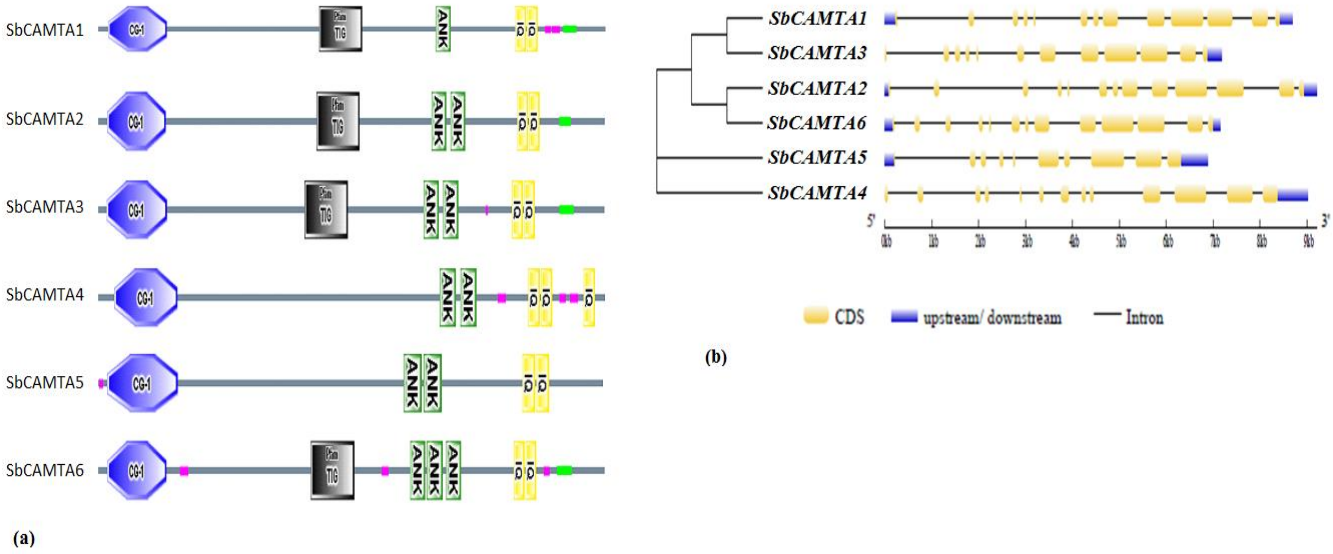

Fig. 1: Domain organization (a) and gene structure (b) of SbCAMTA proteins. a)

Presentation of CAMTA proteins in Sorghum bicolor was obtained from SMART.emble.heidelberg.de. CG-1 domain found in N-terminal region with an ankyrin repeat separated by TIG domain inferring its role in protein-protein interaction. IQ motifs present on C- terminus are known to be associated with the CAM binding proteins. b) Gene structure of SbCAMTA proteins were predicted at (http://gsds.cbi.pku.edu.cn/). A number of exons interfered with the introns are shown in above figure along with their upstream and downstream regions.

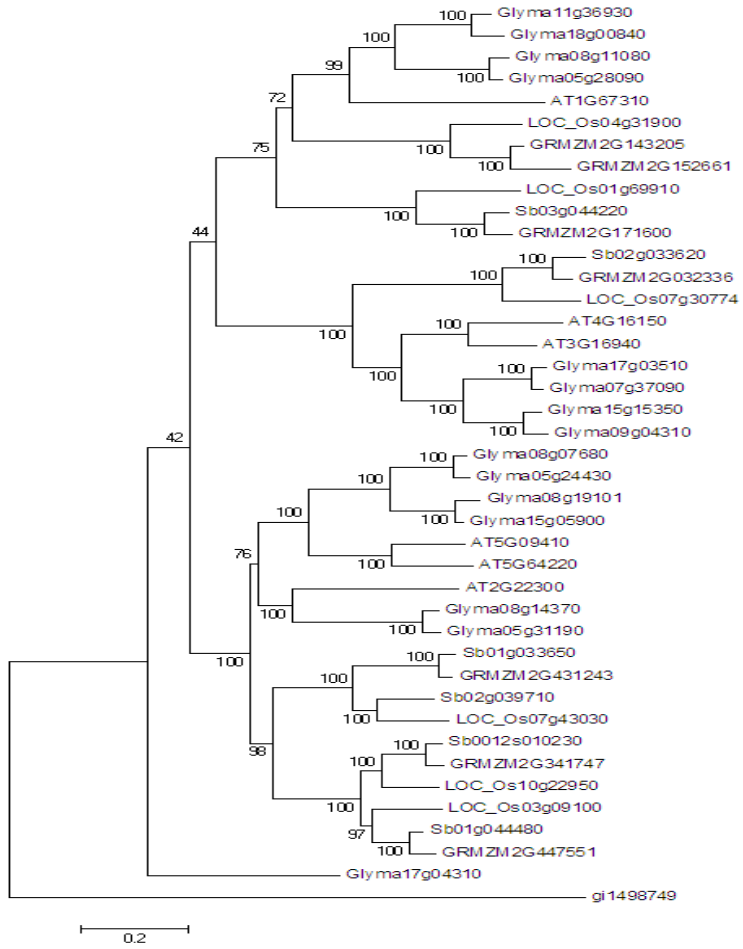

Fig. 2 Phylogenetic relationship of CAMTA proteins

The tree was generated at MEGA6 by neighbor joining method and bootstrap value of 1000 at each branch. The tree of CAMTA proteins from *Sorghum bicolor* (SbCAMTA1= Sb0012s010230; SbCAMTA2= Sb01g033650; SbCAMTA3= Sb01g044480; SbCAMTA4= Sb02g033620; SbCAMTA5= Sb03g044220; SbCAMTA6= Sb02g039710), *Oryza sativa*, *Zea mays*, *Glycine max*, *Arabidopsis thaliana* was developed by using an unknown protein (gi1498749) as outgroup.

Fig. 3 Conserved motifs in Sorghum bicolor, Oryza sativa, Zea mays, Glycine max, and Arabidopsis Thaliana discovered at MEME suit

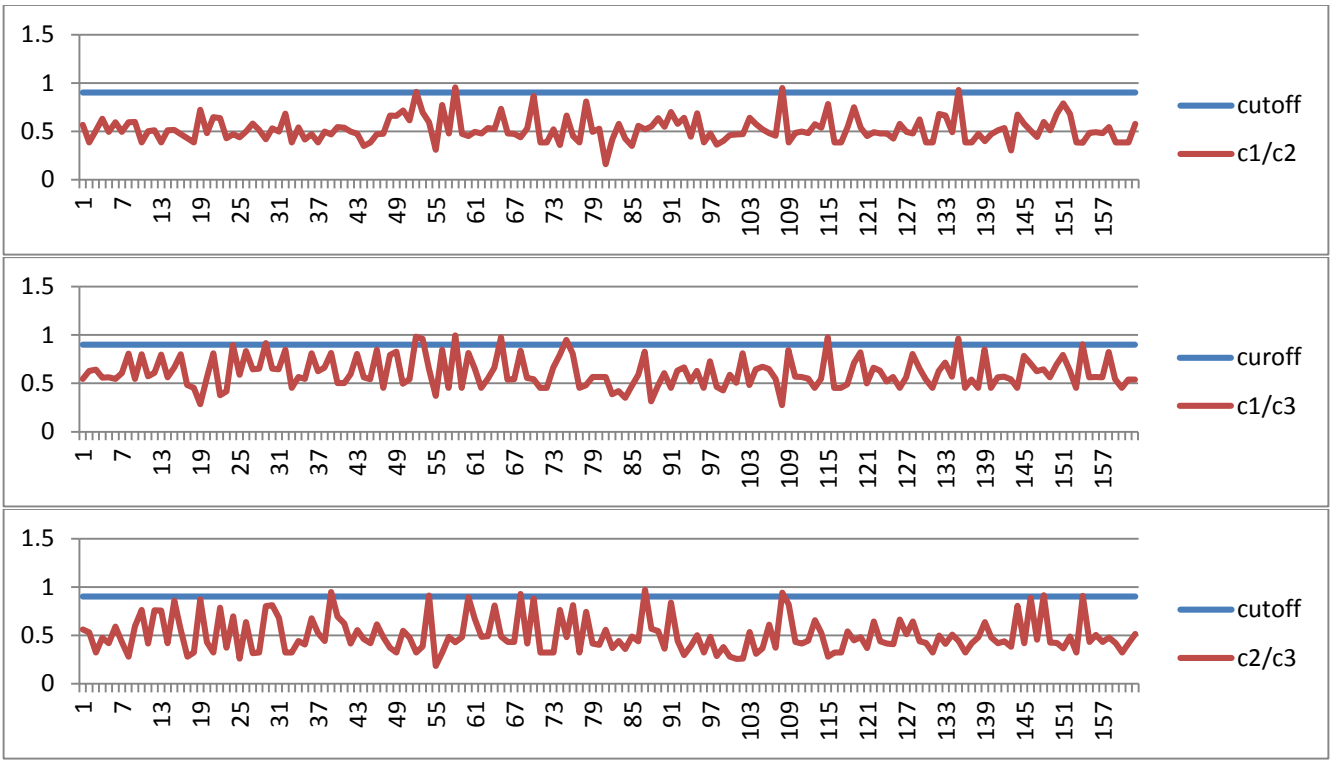

Fig. 4 Type I functional divergence analysis CAMTA proteins among sorghum, rice, maize, soybean, and Arabidopsis (cutoff=0.90)

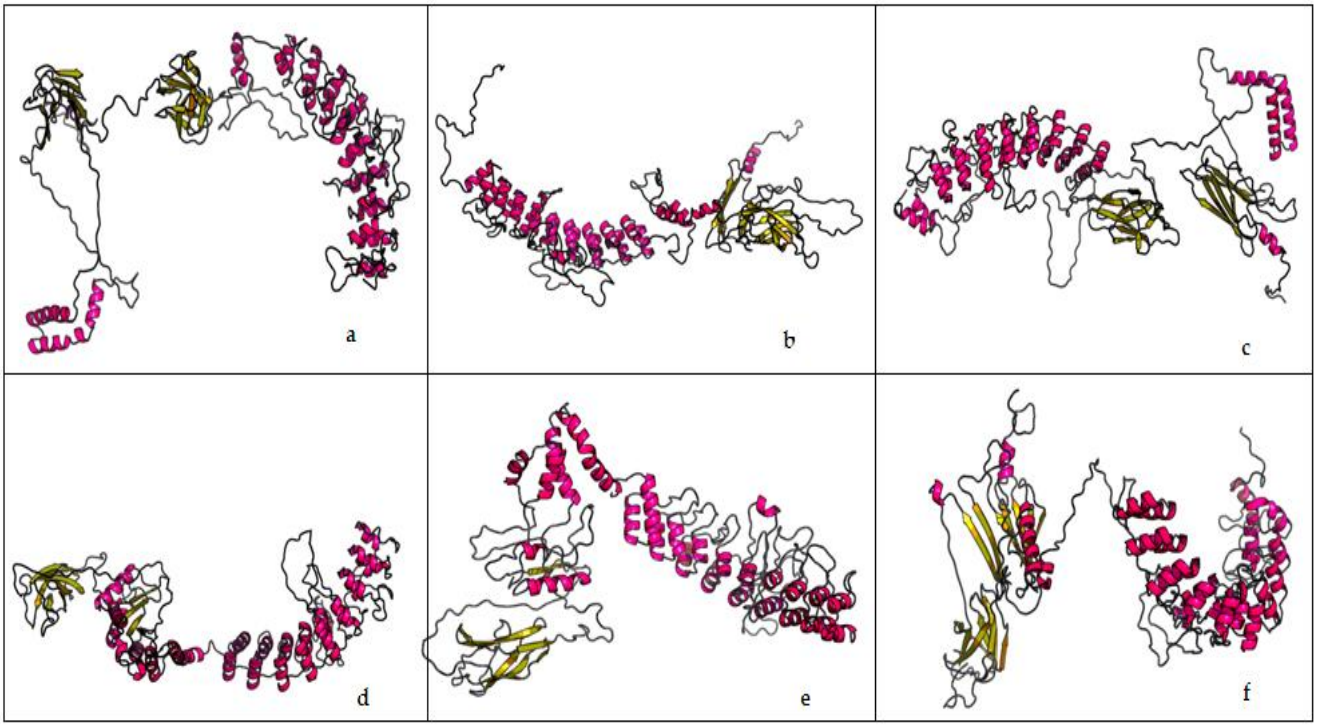

Fig. 5 3-D structure analysis of SbCAMTA proteins. Template based modeling of SbCAMTA1 (a), SbCAMTA2 (b), SbCAMTA3 (c), SbCAMTA4 (d), SbCAMTA5 (e), SbCAMTA6 (f) proteins to compute their 3-D structures using TIG domain of human calmodulin binding transcription activator (CAMTA1; PDB ID: 2cxka) as a template through RaptorX Structure Prediction.

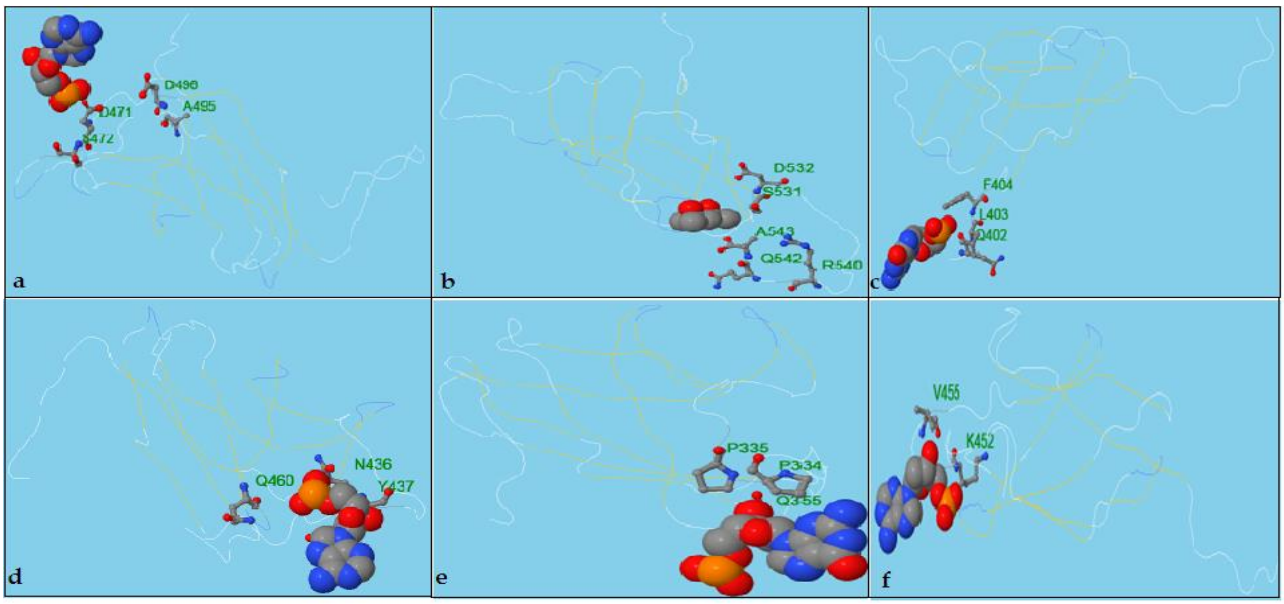

Fig. 6. 3-D structures of SbCAMTA domains.

Domain structure of SbCAMTA1 (a), SbCAMTA2 (b), SbCAMTA3 (c), SbCAMTA4 (d), SbCAMTA5 (e), and SbCAMTA6 (f) were also generated through RaptorX Binding Prediction to explore their binding residues. (a) SbCAMTA1 binds through S472, D471, D496, A495, (b) SbCAMTA2 D532, S531, A643, Q542, R540, (c) SbCAMTA3 contains F404, L403, Q402, (d) SbCAMTA4 contains Y437, N436, Q460, (e) SbCAMTA5 contains P334, P335, Q355, and (f) SbCAMTA6 contains V455, K452

Contribution of Authors

Rana RM: Designed study, Data Analysis, Proof read manuscript
Saeed S: Executed study, Data Analysis, Result write-up
Wattoo FM: Discussion write-up, Proof read manuscript
Amjid MW: Data analysis, Proof read manuscript
Khan MA: Discussion write-up, Proof read manuscript

Disclaimer: None.
Conflict of Interest: None.
Source of Funding: None.

References

Akaike H, 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19: 716-723.

Aravind L and Koonin EV, 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J. Mol. Biol. 287: 1023-1040.

Bachs O, Agell N and Carafoli E, 1992. Calcium and calmodulin function in the cell nucleus. BBA-Biomembranes. 1113: 259-270.

Bähler M and Rhoads A, 2002. Calmodulin signaling via the IQ motif. FEBS Lett. 513: 107-113

Bouché N, Scharlat A, Snedden W, Bouchez D and Fromm H, 2002. A novel family of calmodulin-binding transcription activators in multicellular organisms. J. Biol. Chem. 277: 21851-21861.

Bouché N, Yellin A, Snedden WA, Fromm H, 2005. Plant-specific calmodulin-binding proteins. Annu. Rev. Plant Biol. 56: 435-466.

Bowers JE, Chapman BA, Rong J and Paterson AH, 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 422: 433-438.

Carafoli E, 1987. Intracellular calcium homeostasis. Annu. Rev. Biochem. 56: 395-433.

Chase MW, Duvall MR, Hills HG, Conran JG, Cox AV, Eguiarte LE, Hartwell J, Fay MF, Caddick LR, Cameron KM and Hoot S, 1995. Molecular phylogenetics of Lilianae, pp 109-137. In: P. J. Rudall, P. J. Cribb, D. F. Cutler and C. J. Humphries (eds.), Monocotyledons: systematics

and evolution, 1st edn. Royal Botanic Garden, Kew, UK.

Chaw SM, Chang CC, Chen HL and Li WH, 2004. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. J. Mol. Evol. 58: 424-441.

Choi MS, Kim MC, Yoo JH, Moon BC, Koo SC, Park BO, Lee JH, Koo YD, Han HJ, Lee SY and Chung WS, 2005. Isolation of a calmodulin-binding transcription factor from rice (*Oryza sativa* L.). J. Biol. Chem. 280: 40820-40831.

Chung WS, Lee SH, Kim JC, Do Heo W, Kim MC, Park CY, Park HC, Lim CO, Kim WB, Harper JF and Cho MJ, 2000. Identification of a calmodulin-regulated soybean Ca2+-ATPase (SCA1) that is located in the plasma membrane. Plant Cell. 12: 1393-1407.

Cohen P, 1982. The role of protein phosphorylation in neural and hormonal control of cellular activity. Nature. 296: 613-620.

DeFalco TA, Chiasson D, Munro K, Kaiser BN and Snedden WA, 2010. Characterization of GmCaMK1, a member of a soybean calmodulin-binding receptor-like kinase family. FEBS Lett. 584:4717-4724.

Doherty CJ, Van Buskirk HA, Myers SJ and Thomashow MF, 2009. Roles for Arabidopsis CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. Plant Cell. 21: 972-984.

Donoghue M and Doyle J, 1989. Phylogenetic analysis of angiosperms and the relationships of Hamamelidae, pp 17-45. In: P. R. Crane and S. Blackmore (eds.) Evolution, systematics and fossil history of the Hamamelidae, 1st edn. Oxford, NewYork, USA.

Doyle JA and Donoghue MJ, 1992. Fossils and seed plant phylogeny reanalyzed. Brittonia. 44:89-106.

Du L, Ali GS, Simons KA, Hou J, Yang T, Reddy A and Poovaiah B, 2009. Ca2+/calmodulin regulates salicylic-acid-mediated plant immunity. Nature. 457: 1154-1158.

Du L, Yang T, Puthanveettil SV and Poovaiah B, 2011. Decoding of calcium signal through calmodulin: calmodulin-binding proteins in plants. In: Luan S (ed) Coding and Decoding of Calcium Signals in Plants, 1st edn. Springer, USA. pp. 177-233.

Duvall MR, Learn GH, Eguiarte LE and Clegg MT, 1993. Phylogenetic analysis of rbcL sequences identifies Acorus calamus as the primal extant monocotyledon. P. Natl. Acad. Sci. USA. 90: 4641-4644.

Finkler A, Ashery-Padan R and Fromm H, 2007. CAMTAs: calmodulin-binding transcription activators from plants to human. FEBS Lett. 581: 3893-3898.

Ghosh G, Van Duyne G, Ghosh S and Sigler PB, 1995. Structure of NF-κB p50 homodimer bound to a κB site. Nature. 373: 303-310.

Hamby RK and Zimmer EA, 1992. Ribosomal RNA as a phylogenetic tool in plant systematics, pp 50-91. In: P. L. Soltis (ed.), Molecular systematics of plants, 1st (edn). Springer, USA.

Huang J, Wang MM, Bao YM, Sun SJ, Pan LJ and Zhang HS, 2008. SRWD: A novel WD40 protein subfamily regulated by salt stress in rice (*Oryza sativa* L.). Gene. 424: 71-79.

Kudla J, Batistič O and Hashimoto K, 2010. Calcium signals: the lead currency of plant information processing. Plant Cell. 22: 541-563.

Lenzoni G, Liu J, and Knight MR, 2018. Predicting plant immunity gene expression by identifying the decoding mechanism of calcium signatures. New Phytol. 217: 1598-1609.

Lewit-Bentley A and Réty S, 2000. EF-hand calcium-binding proteins. Curr. Opin. Struc. Biol. 10: 637-643.

Mizoi J, Shinozaki K, and Yamaguchi-Shinozaki K, 2012. AP2/ERF family transcription factors in plant abiotic stress responses. Biochem. Biophys. Acta. 1819: 86-96.

Müller CW, Rey FA, Sodeoka M, Verdine GL and Harrison SC, 1995. Structure of the NF-κB p50 homodimer bound to DNA. Nature. 373: 311-317.

Poovaiah B, Du L, Wang H and Yang T, 2013. Recent advances in calcium/calmodulin-mediated signaling with an emphasis on plant-microbe interactions. Plant Physiol. 163: 531-542.

Pruschy M, Ju Y, Spitz L, Carafoli E and Goldfarb D, 1994. Facilitated nuclear transport of calmodulin in tissue culture cells. J. Cell Biol. 127: 1527-1536.

Reddy AS, Ali GS, Celesnik H and Day IS, 2011. Coping with stresses: roles of calcium-and calcium/calmodulin-regulated gene expression. Plant Cell. 23:2010-2032.

Rubtsov AM and Lopina OD, 2000. Ankyrins. FEBS Lett. 482: 1-5.

Sanders D, Pelloux J, Brownlee C and Harper JF, 2002. Calcium at the crossroads of signaling. Plant Cell. 14: S401-S417.

Sedgwick SG and Smerdon SJ, 1999. The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem. Sci. 24: 311-316.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882.

Tucker SC and Douglas AW, 1996. Floral structure, development, and relationships of paleoherbs: Saruma, Cabomba, Lactoris, and selected Piperales, pp 141-175. In: D. W. Taylor and L. J. Hickey (eds.), Flowering Plant Origin, Evolution and Phylogeny, 1st edn. Springer, USA.

Vogel HJ, 1994. Calmodulin: a versatile calcium mediator protein. Biochem. Cell Biol. 72: 357-376.

Yang T and Poovaiah B, 2002a. A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. J. Biol. Chem. 277: 45049-45058.

Yang T and Poovaiah B, 2002b. Hydrogen peroxide homeostasis: activation of plant catalase by calcium/calmodulin. P. Natl. Acad. Sci. 99: 4097-4102.

Yang T and Poovaiah B, 2003. Calcium/calmodulin-mediated signal network in plants. Trends Plant Sci. 8:505-512.

Yue R, Lu C, Sun T, Peng T, Han X, Qi J, Yan S and Tie S, 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.

Zanis MJ, Soltis DE, Soltis PS, Mathews S and Donoghue MJ, 2002. The root of the angiosperms revisited. P. Natl. Acad. Sci. 99: 6848-6853.