



# Genome-Wide Identification of Calcium Dependent Protein Kinase Gene Family in Plant Lineage Shows Presence of Novel D-x-D and D-E-L Motifs in EF-Hand Domain

Tapan K. Mohanta<sup>1\*</sup>, Nibedita Mohanta<sup>2</sup>, Yugal K. Mohanta<sup>3</sup> and Hanhong Bae<sup>1\*</sup>

<sup>1</sup> School of Biotechnology, Yeungnam University, Gyeongsan, South Korea, <sup>2</sup> Department Of Biotechnology, North Orissa University, Baripada, India, <sup>3</sup> Department Of Botany, North Orissa University, Baripada, India

## OPEN ACCESS

### Edited by:

Rajeev K. Varshney,  
International Crops Research Institute  
for the Semi-Arid Tropics, India

### Reviewed by:

Swarup Kumar Parida,  
National Institute of Plant Genome  
Research, India  
Caiguo Zhang,  
University of Colorado, Denver, USA

### \*Correspondence:

Tapan K. Mohanta  
nostoc.tapan@gmail.com;  
Hanhong Bae  
hanhongbae@ynu.ac.kr

### Specialty section:

This article was submitted to  
Plant Genetics and Genomics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 25 August 2015

**Accepted:** 02 December 2015

**Published:** 24 December 2015

### Citation:

Mohanta TK, Mohanta N, Mohanta YK  
and Bae H (2015) Genome-Wide  
Identification of Calcium Dependent  
Protein Kinase Gene Family in Plant  
Lineage Shows Presence of Novel  
D-x-D and D-E-L Motifs in EF-Hand  
Domain. *Front. Plant Sci.* 6:1146.  
doi: 10.3389/fpls.2015.01146

Calcium ions are considered ubiquitous second messengers in eukaryotic signal transduction pathways. Intracellular Ca<sup>2+</sup> concentration are modulated by various signals such as hormones and biotic and abiotic stresses. Modulation of Ca<sup>2+</sup> ion leads to stimulation of calcium dependent protein kinase genes (CPKs), which results in regulation of gene expression and therefore mediates plant growth and development as well as biotic and abiotic stresses. Here, we reported the CPK gene family of 40 different plant species (950 CPK genes) and provided a unified nomenclature system for all of them. In addition, we analyzed their genomic, biochemical and structural conserved features. Multiple sequence alignment revealed that the kinase domain, auto-inhibitory domain and EF-hands regions of regulatory domains are highly conserved in nature. Additionally, the EF-hand domains of higher plants were found to contain four D-x-D and two D-E-L motifs, while lower eukaryotic plants had two D-x-D and one D-x-E motifs in their EF-hands. Phylogenetic analysis showed that CPK genes are clustered into four different groups. By studying the CPK gene family across the plant lineage, we provide the first evidence of the presence of D-x-D motif in the calcium binding EF-hand domain of CPK proteins.

**Keywords:** CPK, CaM, CBL, calcium signaling, evolution, EF-hands

## INTRODUCTION

Plants are constantly challenged by different environmental factors that guide regulation of various physiological and developmental programs in a systematic manner (Choi et al., 2014). Plants sense and process environmental stimuli such as light, temperature, water and nutrients to enable normal growth and development (Krouk et al., 2011; Davidson and Gu, 2012; Lawson et al., 2014; Matos et al., 2014). Plant cells are able to respond to external stimuli in a time dependent manner, after which they process the information received. This is followed by signal transduction and finally appropriate physiological responses (Asano et al., 2005; Clouse, 2011; Tena et al., 2011; Huang et al., 2012; Boudsocq and Sheen, 2013; Wasternack and Hause, 2013). The propagation and amplification of signals involve rapid changes in cellular and sub-cellular levels of signals by second messengers that relay information to downstream effector molecules such as enzymes, transcription

factors, and cytoskeletal proteins, which are responsible for changes in gene expression, metabolic activity and developmental patterns (Mohanta et al., 2012; Kanchiswamy et al., 2013). Among these messengers, calcium dependent protein kinases (CPKs) are the most important enzymes responsible for a diverse array of plant growth and development, as well as biotic and abiotic stress responses (Luan, 2009; Kanchiswamy et al., 2013; Mohanta et al., 2015b).

Calcium ion is considered an ubiquitous second messenger in eukaryotic signal transduction pathways (Hamel et al., 2014). Intracellular  $\text{Ca}^{2+}$  concentrations are modulated by various signals (e.g., hormones, biotic and abiotic stresses) and play significant roles in plant growth and development and stress responses (Dodd et al., 2010; Reddy et al., 2011; Mohanta and Bae, 2015). CPKs are considered novel calcium sensors that act as crucial mediators in response to diverse stress conditions (Cheng et al., 2002; Schulz et al., 2013). The genes encoding CPKs are found in plants as multi-gene family members (Harmon et al., 2001). The CPKs are classified as serine/threonine protein kinases that contain five domains: (i) N-terminal variable domain, (ii) kinase domain, (iii) auto-inhibitory domain, (iv) regulatory domain, and (v) C-terminal variable domain (Harmon et al., 2001). The regulatory domain contains four calcium binding EF-hands, and the conserved glutamic acid (E) and aspartate (D) amino acids present in the EF-hands act as an important calcium sensor (Day et al., 2002; Hrabak et al., 2003). The EF-hands, which act in pairs, are divided into N- and C-terminal EF-hand pairs. The EF-hand motif is highly conserved and contains a 12 amino acids helix-loop-helix structure that coordinates the calcium ion. The EF-hand module binds  $\text{Ca}^{2+}$  ion in the loop and undergoes conformational changes that result in the exposure of hydrophobic pockets, which facilitates protein-protein interactions with a variety of downstream interacting partners (Harmon et al., 2000; Sulmann et al., 2014). The N- and C-terminal domains of CPKs are highly variable in nature and contain myristoylation and palmitoylation sites for sub-cellular targeting, whereas the kinase domain is catalytic in nature and contains an ATP binding site (Ito et al., 2010). While CPKs are regulated by diverse mechanisms, sensitivity to calcium ion can be influenced by the type of substrate, and each isoform of CPKs responds to a specific set of calcium signals.

There are several gene families that act as potential calcium sensors in plants including calcium dependent protein kinase (CPK), calmodulin (*CaM*), calmodulin like (*CML*), calcineurin B-like (*CBL*), and CPK related kinase (*CRK*) (Hrabak et al., 2003). Among these calcium sensor genes, CPKs are unique in that they contain both a kinase catalytic domain and an EF-hand containing calcium binding domain, which leads to signaling and sensor activity in a single gene product (Harmon et al., 2000). The fusion of the kinase domain with an EF-hand containing regulatory domain probably arose via early fusion of an upstream protein kinase with downstream *CaM* to conduct immediate and efficient downstream phosphorylation of target proteins.

To date, there has been little information available regarding the unified nomenclature, genomics, biochemistry and conserved structures of CPK genes in plants. Therefore, in this study, we conducted large scale identification of CPK gene family members

using 40 different plant species (950 CPK genes). Additionally, we provided a unified nomenclature system and analyzed the conserved structural and functional aspects of CPKs (Ouyang et al., 2007; Goodstein et al., 2012; Lamesch et al., 2012; Nystedt et al., 2013).

## MATERIALS AND METHODS

### Identification of CPK Genes

To understand the signature motif of the CPK gene family in plants, genome-wide identification of CPK gene family members of 40 plant species was conducted. The CPK genes from *Arabidopsis thaliana* were downloaded from “The Arabidopsis Information Resources” (TAIR) database (Lamesch et al., 2012), while the CPK genes from rice were downloaded from the “rice Genome Annotation Resources” (<http://rice.plantbiology.msu.edu/>) database (Ouyang et al., 2007). The CPK protein sequences from *Arabidopsis thaliana* and *Oryza sativa* were used as search queries in the publicly available phytozome database (<http://www.phytozome.net/>) (Goodstein et al., 2012) and *P. abies* genome database (<http://congenie.org/start>) (Nystedt et al., 2013) to identify the CPK genes from other plant species. In total, 40 species were included in this study (Table 1). The BLASTP search was conducted to identify the CPK gene family members of unknown species by taking the protein sequences of AtCPKs and OsCPKs. All CPK proteins of the 40 species were analyzed using the SCAN PROSITE software with default parameters to confirm the presence of all the domains of CPKs (de Castro et al., 2006). Only CPK proteins that contained the kinase domain, auto-inhibitory domain, and regulatory domain with four calcium binding EF-hands was considered for further investigation. All the CPK proteins identified from these approaches were again subjected to BLASTP searches against the TAIR and rice genome annotation database to confirm the BLAST hits with CPK genes of *Arabidopsis thaliana* and *Oryza sativa*. The palmitoylation sites of all CPK proteins were identified by CSS Palm software version 2.0 (Ren et al., 2008). The molecular masses of all CPK proteins were predicted using protein calculator version 3.3 (<http://www.scripps.edu/~cdputnam/protcalc.html>).

### Nomenclature of CPKs

In the traditional nomenclature system, numbers have been assigned to CPK genes according to the serial number in which they are cloned and identified. However, this nomenclature system makes it difficult to determine the exact function of the orthologous counterpart. Plant genomic data are currently increasing daily; therefore, it is necessary to provide a unique and systemic nomenclature to all the CPK genes. The development of a new nomenclature system can provide brief functional information regarding the orthologous CPK genes in other organisms. Therefore, an orthologous based nomenclature system was used to assign the name to each CPK gene of the studied organisms as proposed by different research groups (Supplementary Table 1; Hamel et al., 2006; Kanchiswamy et al., 2013; Mohanta and Mohanta, 2013; Mohanta et al., 2014, 2015a). The *Arabidopsis thaliana* and *Oryza sativa* CPK genes were taken as the orthologous genes for this purpose (Hamel et al.,

**TABLE 1 | Genome size of different plant species used in this study and number of CPK genes present in their respective genomes.**

Sl. No	Name of species	Type of organism	Genome size (Mbs)	Total no. of loci/genome	Total no. of CPK genes/genome
1	<i>Aquilegia coerulea</i>	Dicot	302	24823	16
2	<i>Arabidopsis thaliana</i>	Dicot	135	27416	34
3	<i>Brachipodium distachyon</i>	Monocot	272	31694	27
4	<i>Brassica rapa</i>	Dicot	283.8	40492	49
5	<i>Capsella rubella</i>	Dicot	134.8	26521	32
6	<i>Carica papaya</i>	Dicot	135	27332	15
7	<i>Chlamydomonas reinhardtii</i>	Lower eukaryote	111.1	17741	14
8	<i>Citrus clemantina</i>	Dicot	301.4	24533	26
9	<i>Citrus sinensis</i>	Dicot	319	25376	24
10	<i>Coccomyxa subellipsoidea</i>	Lower eukaryote	49	9629	2
11	<i>Cucumis sativus</i>	Dicot	203	21494	18
12	<i>Eucalyptus grandis</i>	Dicot	691	36376	22
13	<i>Fragaria vesca</i>	Dicot	240	32831	14
14	<i>Glycine max</i>	Dicot	975	54175	41
15	<i>Gossypium raimondi</i>	Dicot	761.4	37505	40
16	<i>Linum usitatissimum</i>	Dicot	318.3	43471	47
17	<i>Malus domestica</i>	Dicot	881.3	63514	28
18	<i>Manihot esculenta</i>	Dicot	532.5	34085	26
19	<i>Medicago truncatula</i>	Dicot	241	50894	11
20	<i>Micromonas pusila CCMP1545</i>	Lower eukaryote	22	10660	2
21	<i>Mimulus guttatus</i>	Dicot	321.7	26718	25
22	<i>Oryza sativa</i>	Monocot	372	39049	30
23	<i>Ostreococcus lucimarinus</i>	Lower eukaryote	13.2	7796	3
24	<i>Panicum virgatum</i>	Monocot	1358	98007	53
25	<i>Phaseolus vulgaris</i>	Dicot	521.1	27197	25
26	<i>Physcomitrella patens V.13</i>	Bryophyte	480	33362	25
27	<i>Picea abies</i>	Gymnosperm	1960	28354	11
28	<i>Populus trichocarpa</i>	Dicot	422.9	41335	28
29	<i>Prunus persica</i>	Dicot	227.3	26873	17
30	<i>Ricinus communis</i>	Dicot	400	31221	15
31	<i>Selaginella moellendorffii</i>	Pteridophyte	212.5	22273	9
32	<i>Setaria italica</i>	Monocot	405.7	35471	27
33	<i>Solanum lycopersicum</i>	Dicot	900	34727	28
34	<i>Solanum tuberosum</i>	Dicot	800	35119	21
35	<i>Sorghum bicolor</i>	Monocot	697.5	33032	28
36	<i>Thellungiella halophila</i>	Dicot	238.5	26351	31
37	<i>Theobroma cacao</i>	Dicot	346	29452	17
38	<i>Vitis vinifera</i>	Dicot	487	26346	17
39	<i>Volvox carteri</i>	Lower eukaryote	125.4	14971	6
40	<i>Zea mays</i>	Monocot	2500	63540	47

2006; Schlicker et al., 2006). In this study, the nomenclature system adopted for naming *Arabidopsis*, poplar and rice mitogen-activated protein kinases (MPKs) was extended to other genes (Hamel et al., 2006; Mohanta et al., 2015a). For each CPK gene, a name was assigned by taking the first letter of the genus and the first letter of the species, after which CPK and the orthology based number of *Arabidopsis thaliana* or *Oryza sativa* were used. In the nomenclature system, the first letter of the genus was kept upper case and the first letter of the species was kept lower case. When the first letter of the genus and species coincided

with one another, only the first letter of the genus and the first, second, third or later letter of the species name was considered in conjunction with the suitability of pronunciation. For example, *C. rubella* and *C. reinhardtii* share “C” as first letter of the genus and “r” as the first letter of the species. In this case, *C. rubella* CPKs were named CrCPK, while *C. reinhardtii* CPKs were named CreinCPK. The monocot specific CPKs were named according to *Oryza sativa*, while the rest of CPKs were named according to *Arabidopsis thaliana*. This nomenclature system can be useful to provide a unique identity to each CPK gene throughout the

plant kingdom. The sequence/structural similarities resemble the functional similarities of genes (Aravind et al., 2002; Schlicker et al., 2006; Li et al., 2008). Specifically, the unique orthologous gene of one species may resemble the other gene and undergo similar cellular function. A similar approach is being used to predict potential function of newly sequenced genes and their gene products. It is nearly not possible to study the role of each *CPK* with different functional aspects. Therefore, an orthology based nomenclature system of *CPK* genes will provide basic information regarding its ortholog counterpart (Wright and Bruford, 2006).

### Multiple Sequence Alignment

The *CPK* protein sequences of plants were subjected to multiple sequence alignment to identify the conserved domains and motifs. Multiple sequence alignment was carried out using the Multalin software (<http://multalin.toulouse.inra.fr/multalin/>). The statistical parameters used to carry out multiple sequence alignment were as follows: protein weight matrix, BLOSUM62-12-2; gap penalty at opening, default; gap penalty at extension, default; gap penalty at extremities, none; one iteration only, no; high consensus value, 90% (default); and low consensus value 50% (default).

### Construction of Phylogenetic Tree

All the protein sequences of *CPKs* were subjected to multiple sequence alignment in clustal omega to generate a clustal file. The generated clustal file of *CPKs* was converted to MEGA file format using the MEGA6 software and then used to construct the phylogenetic tree using the same software (Tamura et al., 2013). Different statistical parameters used to construct the phylogenetic tree were; statistical method, maximum likelihood; test of phylogeny, bootstrap method; number of bootstrap replication, 5000; model/method, Jones-Taylor-Thornton; rates among sites, uniform; gaps/missing data treatment, partial deletion; site coverage cutoff, 95%; and branch swap filter, very strong.

### Statistical Analysis

Tajima's relative rate test was conducted to understand the rate of evolution of plant *CPKs*. To accomplish this, MEGA6 software was used to conduct Tajima's relative rate test and Tajima's test of neutrality of the *CPKs* in MEGA file that used to generate the phylogenetic tree. The statistical parameters used for Tajima's relative rate test were; analysis, Tajima's relative rate test; scope, for three chosen sequences; substitution type, amino acids; gaps/missing data treatment, complete deletion. Tajima's test of neutrality was carried out to understand the polymorphism of *CPK* genes in plants. The statistical parameters used to carry out Tajima's test of neutrality were: analysis, Tajima's neutrality test; scope, all selected taxa; substitution type, amino acids; and gaps/missing data treatment, complete deletion.

Correlation regression analysis (<http://www.mathportal.org/calculators/statistics-calculator/correlation-and-regression-calculator.php>) and paired *t*-test (<http://www.mathportal.org/calculators/statistics-calculator/t-test-calculator.php>) was conducted using online tool Math Portal. In both the analysis,

genome size was provided in X-axis and *CPK* gene family size was provided in Y-axis. The paired *t*-test was conducted at significance level of  $p < 0.001$ .

### RNA ISOLATION AND qRT-PCR

Wild type *Brassica rapa* L. sub species *oleifera* were grown in soil in greenhouse under 16 h light and 8 h dark cycle at 22–25°C for 20 days. Cold treatment was conducted by subjecting the plants to 4°C and drought/heat treatment was applied at 37°C. Treated plant leaves were immediately transferred to liquid nitrogen for further analysis. All experiments were carried out in triplicates. Total RNA was isolated from treated and control samples using Trizol reagent. RNA was treated with RNase free DNAase to ensure that there is no DNA contamination in the RNA sample. First strand cDNA synthesis was carried out using a Fermentas RevertAid first strand cDNA synthesis kit according to the manufacturer's instruction. The primers for the *B. rapa CPK* genes were designed using primer3 software targeting either the extreme 5' end (extreme 5' ends are not conserved) or 3' UTR region which produced an amplicon of 120–200 bp (primer length between 20 and 24 bp) with a melting temperature of 58–60°C (Supplementary Table 2). The primer lengths were kept around 18–22 nucleotides. Quantitative real-time PCR was conducted using a Mx3000P real-time PCR system with SYBR green master mix (2x) (Fermentas) and ROX as the passive reference standard to normalize the SYBR fluorescent signal. PCR amplification was conducted in a 25 µl reaction mixture containing 1 µl cDNA as template, 12.5 µl SYBR green master mix (2x), 1 µl of each forward and reverse primer and nuclease free water up to 25 µl. The thermal profile for quantitative real time PCR reaction was; initial activation at 95°C for 10 min, followed by 40 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s. Each quantitative real time PCR was carried out in triplicate on three biological replicates. Primers showing efficiency between 90–105% were considered significant. Relative expression was calculated using *ACT2* as the reference gene. The relative expression of *BrCPK* genes was calculated using the  $2^{-\Delta\Delta CT}$  method (Schmittgen and Livak, 2008).

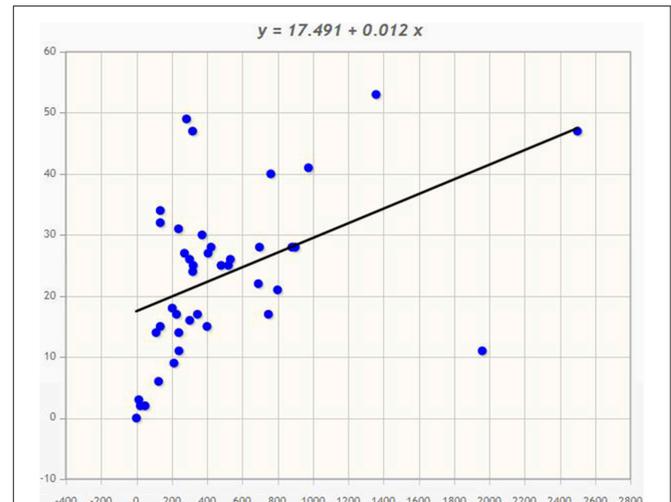
## RESULTS AND DISCUSSION

### Larger the Genome Size is Not Directly Proportional to Bigger the Gene Family

Angiosperms are considered as the most advanced groups of land plants. Although the *CPKs* of several angiosperms have been widely investigated, such studies have been limited to specific species (Hrabak et al., 2003; Asano et al., 2005; Kanchiswamy et al., 2013; Kong et al., 2013; Zuo et al., 2013; Liu et al., 2014). Therefore, very little information is known about their different genomic aspects. We found that genome size and presence of the *CPK* gene number vary from species to species (Table 1). Among the investigated plants, *Ostreococcus lucimarinus* has the smallest genome (13.2 Mbs), while *Zea mays* has the largest (2500 Mbs). The genome of *Coccomyxa subellipsoidea* encodes the lowest

number of *CPK* genes (2), while the genome of *Panicum virgatum* encodes the highest (53) among the studied species. In this study we found that, the larger genome size is not directly proportional to a larger gene family. The *CreinCPK17-5* (*Chlamydomonas reinhardtii*) is the largest *CPK*, containing an ORF of 5940 nucleotides (1979 amino acids), while *CpCPK2* (*Carica papaya*) is the smallest *CPK*, with an ORF length of 693 nucleotides (231 amino acids; Supplementary Table 1). Correlation regression analysis of *CPK* genes compared to the genome size didn't result any linear correlation (Figure 1). Eleven species were found to be clustered within size of 10–20 *CPK* members and 14 species were found to be clustered with size of 20–30 *CPK* members (Figure 1). The regression coefficient analysis was found to be  $r = 0.4487$ . The regression coefficient should be  $+1/-1$  to infer a direct correlation between the genome size and the size of gene family. The paired *t*-test between the genome size and gene family size differs significantly at  $p < 0.001$ . The *t*-value was found to be 5.9796 with degree of freedom 40 and critical value 3.551. The average size of *CPK* genes of green algae and *Physcomitrella patens* is larger than those of advanced higher plants. The extra sequences of lower eukaryotic plants are probably deleted during evolution from lower eukaryotes to higher eukaryotes owing to the need to adapt to terrestrial habitats (Rensing et al., 2008). These findings indicate that evolution of *CPK* genes occurred via loss of gene size; hence, higher plants do not contain extra sequences. Further genomic analysis revealed that, many *CPK* genes contain either 6, 7, or 8 introns with a maximum of 11 introns in at least one gene, except for *Carica papaya*, *Ostreococcus lucimarinus*, and *Ricinus communis* (Supplementary Table 1). No introns were detected in some of the *CPK* genes of *O. lucimarinus* and *P. patens*. The *O. lucimarinus* is a lower aquatic algae and *P. patens* is the first land plant, explaining their conserved and ancient origin (Palenik et al., 2007; Jain et al., 2008; Rensing et al., 2008). The molecular masses of *CPKs* were reportedly vary from 40 kD to 90 kD (Martín and Busconi, 2001; Tuteja and Mahajan, 2007); however, we found that, they ranges from 26.134 kD (*CpCPK2*) to 198.916 kD (*CreinCPK17-5*). The isoelectric point of *CPKs* ranges from 4.61 (*PaCPK33*, *OICPK3*) to 9.33 (*AtCPK16*) (Supplementary Table 3). The analysis of amino acid composition of *CPKs* showed that, leucine (L) was present most frequently (8.46 times per *CPK*) (Supplementary Table 4), while tryptophan (W) was present least frequently (0.91 times per *CPK*). Aspartic acid (D) and glutamic acid (E), which were responsible for calcium sensing, were found to be 7.14 and 7.83 times per *CPK*, respectively (Supplementary Table 4).

These finding indicates that, when more energy is required to synthesize a particular amino acid, there is less chance of that amino acid occurring more frequently (Akashi and Gojobori, 2002). The average abundance of T in *CPK* is only 0.91 times per *CPK*, and the energy required to synthesize W is 74.3. High energy input is required to synthesize W; therefore, the *CPK* protein can only incorporate 0.91 W per *CPK*. The energy required to synthesize D and E is 12.7 and 15.3 unit, respectively (Akashi and Gojobori, 2002); accordingly, the plants can able to accommodate 7.14 D and 7.83 E per *CPK*.



**FIGURE 1 | Correlation regression analysis of Genome size and *CPK* gene family size.** The correlation regression analysis shows, the genome size of an organism is not directly proportional to the size of the gene family. The bigger the genome size isn't directly proportional to the bigger the gene family. In the majority of cases, plant encodes 10–30 *CPKs* per genome. The X-axis denotes the genome size and Y-axis denotes the *CPK* gene family size.

## CPKs Contains Four Conserved D-x-D Motifs in EF-Hand Domain

Even though many studies have been carried out, no specific conserved domain/motif has been reported for kinase domains, auto-inhibitory domain and EF-hand domains of the *CPK* protein. We believe, this is the first report showing the presence of conserved domains/motifs in *CPKs*. To identify the conserved domains/motifs, multiple sequence alignments of monocot, dicot and lower eukaryotic plants were carried out separately. We found that, monocot and dicot plants shared some common conserved sequences in the kinase domains, C-x-G-G-E-L-x-D-R-I, H-R-D-L-K-P-E-N-F-L, D-x-V-G-S-x-Y-Y, A-P-E-V-L, D-V/I-W-S, G-V-I-x-Y-I-L-L, G-x-P-P-F-W, P-W-P-x-I-S, A-K-D-L-V, and H-P-W (Figure 2, Table 2, Supplementary Table 5, Supplementary Figures 1–3), while the kinase domain of lower eukaryotic plants such as *C. reinhardtii*, *V. carteri*, *M. pusila*, *O. lucimarinus*, and *P. patens* contains little bit different conserved sequences, M-E-L-C-x-G-G-E-L-F, H-R-D-L-K-P-E-N-F-L, D-F-G-L-S-V/x, A-P-E-V-L/x, D-I-W-S-x-G-V, and P-F-W (Supplementary Table 5, Supplementary Figure 4).

The conserved sequences D-x-V-G-S-x-Y-Y, G-V-I-x-Y-I-L-L, G-x-P-P-F-W, P-W-P-x-I-S, A-K-D-L-V, and H-P-W present in higher plants are absent from the kinase domain of lower eukaryotic plants (Supplementary Table 5). Earlier it was thought that, auto-inhibitory domain does not have any conserved sequence. But, we found that, the auto-inhibitory domains of monocot, dicot and lower eukaryotic plants are also highly conserved in nature. The common conserved domains present in the auto-inhibitory domains of monocot and dicot plants are K-P-L-D, F-S-A-M-N-K-L, and A-L-x-x-I-A (Figure 3A, Supplementary Table 6). The auto-inhibitory domain of lower eukaryotic plants doesn't contain K-P-L-D and



**FIGURE 2 | Presence of conserved domains in the kinase domain of plant calcium-dependent protein kinases (CPKs) (red).** The major conserved domains are C-S-G-G-E-L-F-D-R-I, H-R-D-L-K-P-E-N-F-L-F, D-V-V-G-S-P-Y-Y, A-P-E-V-L, D-V-W-S, G-V-M-V-Y-I-L-L, G-V-P-P-F-W, P-W-P-H-V-S, A-K-E-L-V, and H-P-W. The complete details describing the conserved domains of CPKs are provided in the Supplementary Figures.

A-L-x-x-I-A domains and only contains the A-M-N-K-L domain (Supplementary Table 6).

The EF-hands of CPK bind the calcium ions and play a significant role in gene regulation (Valmonte et al., 2014). The Ca<sup>2+</sup> ion binds to either D or E amino acid of the EF-hand which leads the auto-inhibitory domain being released from the kinase domain, and subsequent gene regulation (Hwang et al., 2000; Weljie et al., 2003; Tuteja and Mahajan, 2007; Dadacz-Narloch et al., 2011; Boudsocq and Sheen, 2013). Therefore, the presence of D or E amino acid in the EF-hand is very crucial. The EF-hand containing the regulatory domains of CPKs is also highly conserved in nature. The monocot and dicot plants share some common motifs in their EF-hand domain which are; E-E-I/x, D-x-D, D/E-E-L, D-Y-x-E-F, F-D-x-D, E-E-L, D-G-x-I, and Y-x-E-F-x-x-M-M (Figure 3A, Table 2, Supplementary Table 5). Additionally, one extra domain E-D-x<sub>(4)</sub>-A-F is conserved in monocot plants which is absent in dicot plants. The lower eukaryotic plants possess only one D-x-D and D-x-x-E motif in their EF-hand domain. The monocot and dicot (higher eukaryotes) plants possess at least four D-x-D (one D-x-D motif in each EF-hand) and two D/E-E-L motifs in their EF-hands, while lower eukaryotes contain only two D-x-D and one D-x-x-E motifs in their EF-hands (Supplementary Table 5). The two D-x-D motifs of lower eukaryotic plants are present only in the third and fourth EF-hands. Instead of presence of D-x-D and D/E-E-L motifs in CPKs, they also found to contain some substitute motifs like x-D/E-L, Q-E-L (BdCPK30, PvCPK30-1, PvCPK30-2,

SiCPK30, SbCPK30, and OsCPK9) and E-E-F/M (BdCPK7-2, PvCPK8-2, SiCPK7-1, SbCPK8-1, ZmCPK8-1, ZmCPK32-1, OsCPK8, PvCPK7-1, BdCPK16-2, OsCPK3, OsCPK4, SiCPK16, SbCPK28, ZmCPK16-1/2, and PvCPK16-2) (Supplementary Table 7). Earlier, Cheng et al. (2002) reported that EF-hands are most conserved at position 1 and 2, while they are least conserved at position 4 (Cheng et al., 2002). In this study, we found that, in the first EF-hand of higher eukaryotic plants (monocot and dicot), the E-E-I/x motif is conserved at position 1, 2, and 3 (Figure 3A). However, all the four EF-hands contain the D-x-D motif at position 14, 15, and 16 and the D/E-E-L motif is at position 24, 25, and 26 respectively (Figures 3A,B, 4). The first EF-hand contains an E-M-F motif at position 8, 9, and 10. In the second EF-hand, there is an E-x-E motif at position 3, 4, and 5, while in the third EF-hand; the x-E-D motif is present at position 3 and 4. In the fourth EF-hand, there is no conserved amino acid at position 1, 2, or 3 but there is presence of a conserved D/E residue at position 11 and 12. According to our findings, the conserved domains/motifs at position 14, 15, 16, 24, 25, and 26 are crucial for all EF-hands that are responsible for Ca<sup>2+</sup> binding. The specific conserved motif E-E-I/x at position 1, 2, and 3 and E-M-F at position 8, 9, and 10 are first EF-hand specific (Figure 3A). The E-x-E motif at position 3, 4, and 5 is second EF-hand specific, while the x-E-D motif at position 3 and 4 is third EF-hand specific (Figure 3A). The presence of conserved D/E amino acids at position 11 and 12 is fourth EF-hand specific. The presence of the

**TABLE 2 | Presence of conserved motifs in EF-hands of CPKs.**

Conserved motifs in EF-hands			
Monocot	Dicot	Lower eukaryotes	Altogether
E-E-I/x	E-E-I/x		E-E-I/x
E-M-F	E-M-F		
D-x-D	D-x-D		D-x-D
	G-x-I		G-x-I
D-E-L	D/E-E-L		D-E-L
	D/E-x-E		D-E-x-E
D-x-D	D-x-D	D-x-D	D-x-D
D-x-x-E-F	D-Y-x-E-F	D-x-x-E	E-F-I-x
E-D-X <sub>(4)</sub> -A-F			
F-D-x-D	F-D-x-D		D-x-D
			G-x-I
	G-I-x-Y		
E-E-L	D-E-L		D-E-L
D-x-D	D-x-D	D-x-D	E-D-x-D
D-G-x-I	D-G-R/x-I		D-x-D/G-x-I
Y-x-E-F-x-x-M-M	Y-x-E-F-x-x-M-M		Y-x-E-F-x-x-M-M

Monocot and dicot plants share common E-E-I/x, E-M-F, and D-x-D motifs in EF-hands. Higher plants contain 4 D-x-D motifs in their EF-hands, while lower eukaryotes have only two D-x-D motifs.

conserved of CPK motifs from lower to higher plants explains the homologous nature of gene evolution (Thornton and DeSalle, 2000). All EF-hands shows the differences in their first, second, third and fourth amino acid positions. When compared to higher eukaryotes, lower eukaryotes possess only two D-x-D motifs at the second and third EF-hand (Supplementary Table 5). These D-x-D motifs are also present in position 14, 15, and 16 of the second and third EF-hands in lower eukaryotic plants. Some of the lower eukaryotes contain an E-E-I/x motif at the first EF-hand. In the first and fourth EF-hands of lower eukaryotes, D is conserved at position 14, while in the second EF-hand, D and E are conserved at position 22 and 25, respectively. These differences in numbers and positions of EF-hands are likely cause variations in the allosteric properties of the Ca<sup>2+</sup> binding and activation threshold (Hrabak et al., 2003). The number of EF-hand is very important to determine the Ca<sup>2+</sup> regulation in CPK activity. Site directed mutagenesis of conserved E residues in each EF-hand showed that, a closer proximity of the EF-hand to the auto-inhibitory domain shows an effect on Ca<sup>2+</sup> regulation (Zhao et al., 1994).

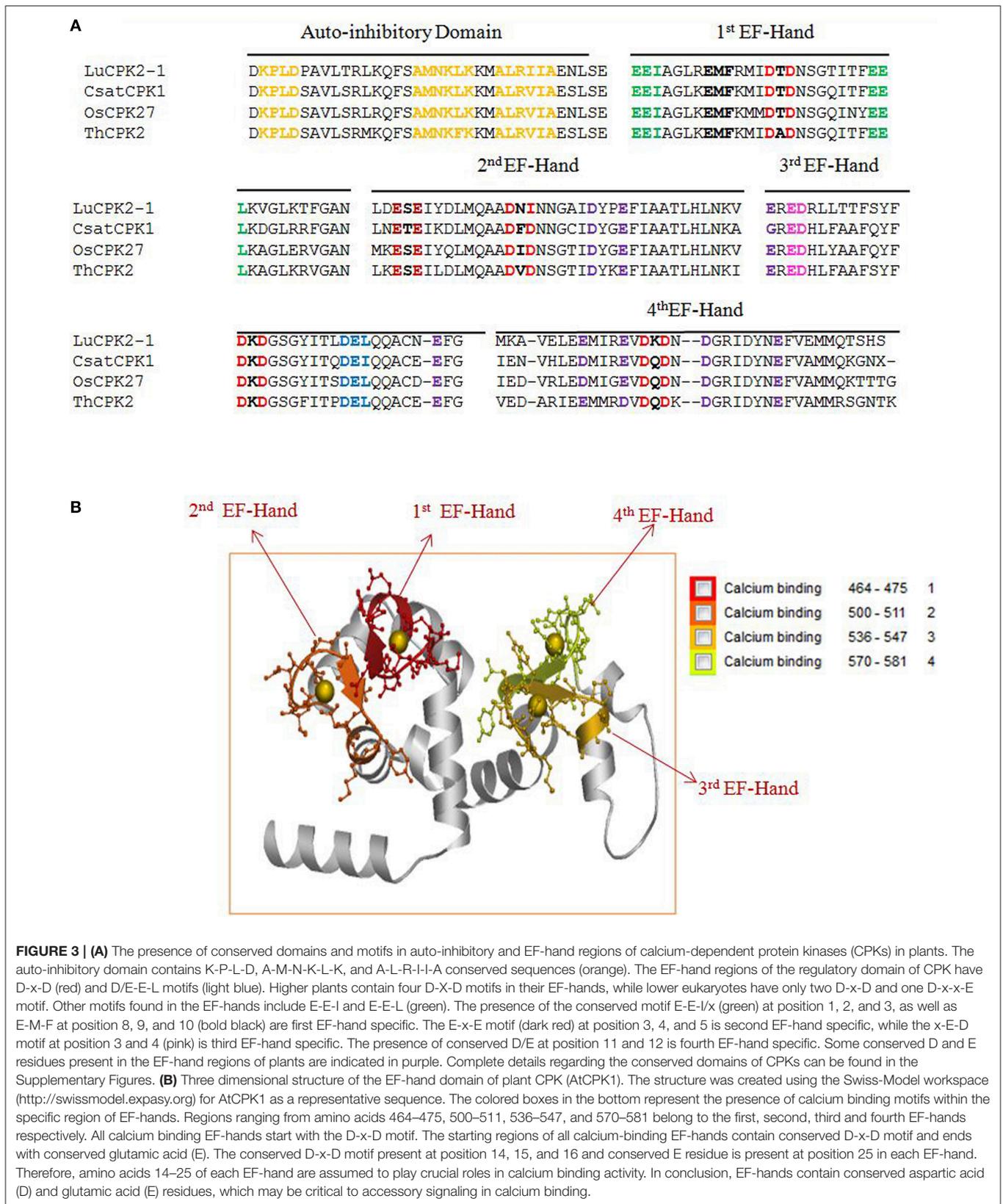
## CPKs Contains Membrane Targeting Palmitoylation and Myristoylation Sites

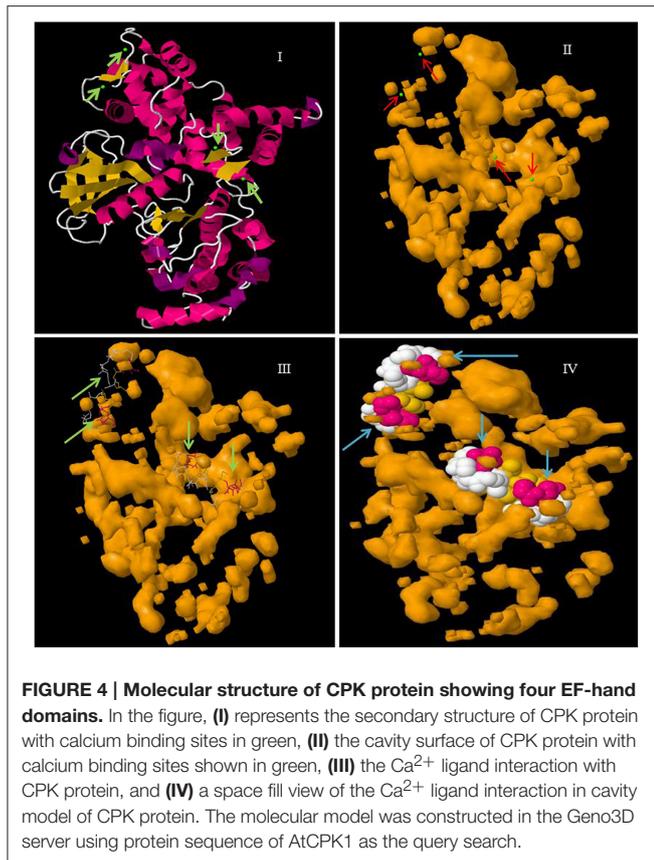
Very little information is available regarding the N-terminal variable domain of CPK. It has been reported that, the N-terminal variable region contains sub-cellular targeting information (Harmon et al., 2000; Curran et al., 2011). However, none of the 34 *Arabidopsis* CPKs are predicted to be integral membrane proteins (Cheng et al., 2002). Despite the variability at the N-terminal domain, most CPKs have a G residue at the second

position and a C residue at 3, 4, 5, and 6 positions. This N-terminal G residue undergoes modification by myristic acid which is known as myristoylation (Smotrys and Linder, 2004). The N-myristoylation has been reported to promote protein-membrane attachment and protein-protein interactions (Cheng et al., 2002). The mutation of the N-terminal G residue abolishes lipid modification and prevents membrane association (Martín and Busconi, 2001). Among the 950 CPKs, 844 have the G residue at the second position (Table 3). In *Arabidopsis* it has reported that, the N-terminal G residue of AtCPK2 is myristoylated, and the first ten amino acids of CPKs are known to be critical for localization to the endoplasmic reticulum (ER) membrane (Lu and Hrabak, 2002). The second lipid modification leads to addition of palmitate to the C residue at position 3, 4, or 5 of the N-terminal region of CPKs (Cheng et al., 2002). Some major conserved palmitoylation motifs of CPKs found during this study were M-G-C, M-G-N-C, M-G-N-C-C, M-G-N-T-C-V, and M-G-N-C-C-R (Table 3). Both myristoylation at the N-terminal G residue and palmitoylation at the N-terminal C residues at position 4 and 5 have been validated experimentally in membrane bound OsCPK2 (Martín and Busconi, 2001). When myristoylation of OsCPK2 was abolished by mutating N-terminal G amino acid, the protein could no longer be palmitoylated, indicating that, myristoylation might be the prerequisite to palmitoylation (Martín and Busconi, 2001).

## Phylogenetic Analysis Shows CPKs are Grouped into Four Groups

In this study, we constructed an unrooted phylogenetic tree of CPKs from monocot, dicot and lower eukaryotic plants (Figure 5). The result revealed that CPK genes fell into four different groups; group A (red), B (blue), C (fuschia), and D (purple) (Figure 5). Previous studies by different researchers also reported the presence of four groups in CPKs (Lu and Hrabak, 2002; Hamel et al., 2014). The phylogenetic tree showed that, the basal architecture of the CPK gene family of three classes (monocot, dicot and lower eukaryotic plants) is conserved and emerged from a common ancestor. Significant sequence similarities exist among the CPKs, indicating that they are probably arose very recently via gene duplication with similar or overlapping functions. The orthologous genes are considered as evolutionary counter parts that must be derived from a common ancestor, while paralogs are homologous genes that evolved through duplication within the same genome (Chen et al., 2005). Paralogs tend to evolve due to new functions for better adaptation (Makarova et al., 2005; Wu et al., 2006). When duplication precedes speciation, each of the paralogs gives rise to a distinct line of orthologous descent (Vleeshouwers et al., 2001; Ting et al., 2004; Chauve et al., 2008). Conversely, when duplication occurs after a particular speciation event in one or both lineages independently, one-to-one orthologous relationships cannot be delineated in principle, resulting in co-orthologs (Li et al., 2010; Mahmood et al., 2012; Yan et al., 2012). The lineage specific expansion of paralogous gene families, which in some cases accounts for a sizable fraction of the genome, is considered as one of the major mechanisms of adaptation (Jordan et al., 2001; Lespinet et al., 2002).





## CPKs are Undergone Vivid Gene Duplication During the Course of Evolution

The CPK gene family is characterized by the presence of several paralogs that share a significant level of homology. These closely related CPK genes most likely emerged following a recent gene duplication event, which would explain why CPK genes have not yet diversified much. In some cases, duplication may be due to diversification events or lineage specific. Gene duplication may increase due to auto-polyploidization, which is the most common mechanism of gene duplication in plants.

The genomes of *C. subellipsoidea* and *M. pusila* CCMP1545 contain only two CPK genes, whereas *O. lucimarinus* contains only three CPK genes, indicating no significant duplication of CPK genes in these species. Another lower eukaryotic alga, *C. reinhardtii*, contains 14 CPK genes with several paralogs; while *V. carteri* has six CPK with no paralogous genes, indicating that, gene duplication may have started from *Chlamydomonas*. The first land plant, *P. patens*, contains 25 CPK genes in its genome, including several paralogous genes; however, the pteridophyte *S. moellendoffii* contains only nine CPK genes (Banks et al., 2011; Baker et al., 2013). Although *Picea abies* has undergone vivid duplication (Morse et al., 2009; Mackay et al., 2012; Nystedt et al., 2013), we found that the model gymnosperm *P. abies* contains only 11 CPK with no paralogous genes. Only the genome of *P. patens* in bryophytes has been sequenced and confirmed to have undergone recent large scale duplication (Bisova et al.,

**TABLE 3 | Protein myristoylation and palmitoylation are important post-translational modification events required for sub-cellular targeting of CPKs.**

Sl No.	Consensus myristoylation and palmitoylation site sequence	No. of CPK genes
1	MGNC	268
2	MGNCC	201
3	MGC	60
4	MGNTCV	57
5	QFGTTYLC	45
6	MGNCCR	40
7	MGLC	36
8	MGGC	35
9	MGNNC	26
10	MGSC	26
11	MGNSC	21
12	QFGTFLC	20
13	MGIC	17
14	MGNCNAC	16
15	MGQC	15
16	MGNAC	14
17	QFGTTYQC	13
18	MGNVC	12
19	MGVV	12
20	MGNQC	8
21	QFGVTYLC	5
22	QFGITYLC	5
23	VHLVMELC	4
24	MGNCNTC	4
25	MELC	4

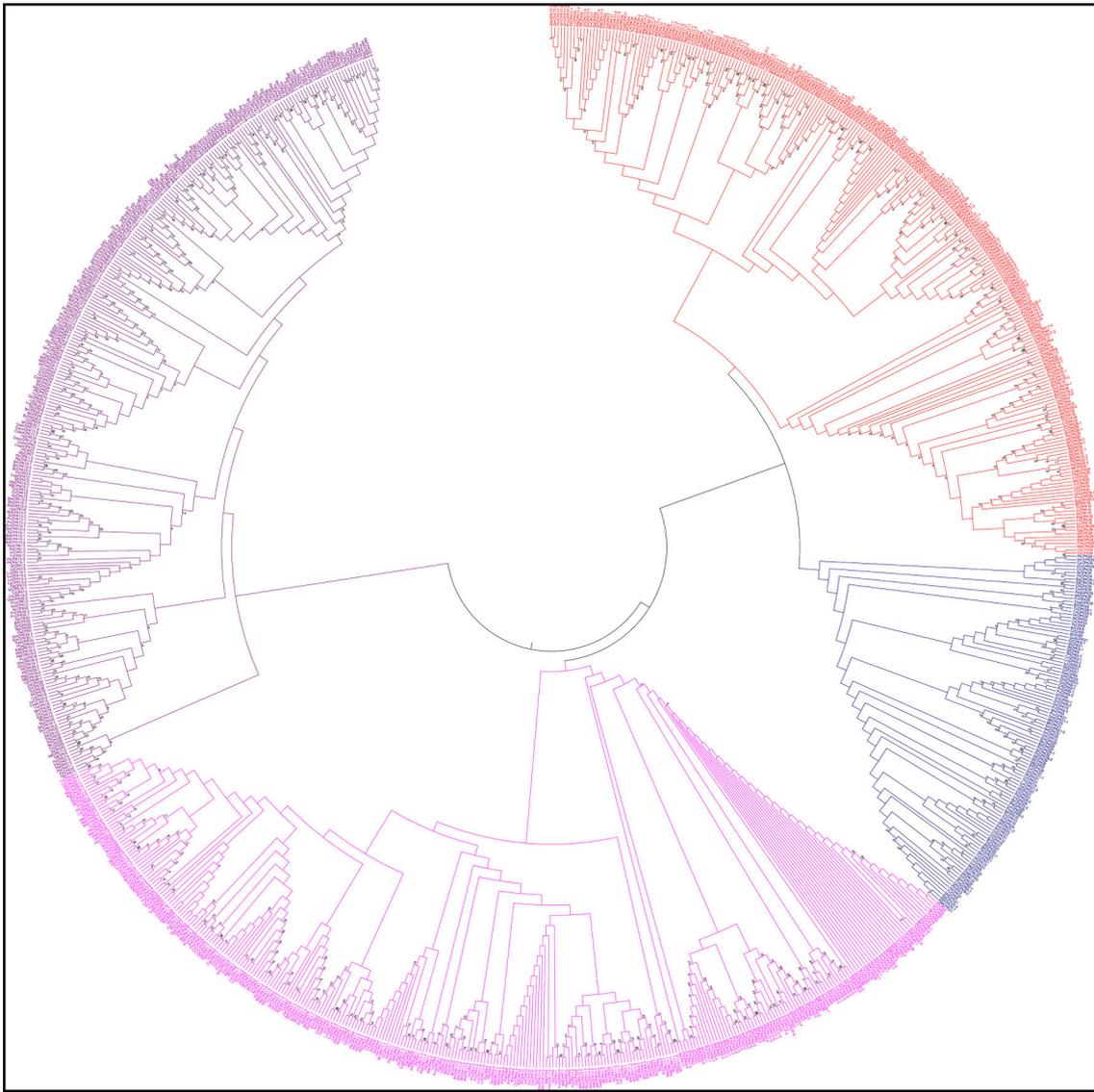
Protein myristoylation is usually carried out in glycine (G) residues and is conserved at the second position (green). Similarly, protein palmitoylation is carried out in cysteine (C) residues (red). The 25 likely conserved consensus sequences of CPK palmitoylation sites found in this study are listed. The presence of C amino acid at position 3, 4, and 5 plays a crucial role in CPK palmitoylation.

2005; Zimmer et al., 2008). Nevertheless, it is difficult to predict whether recent duplication events contributed to expansion of the CPK family of other non-vascular plants.

## CPKs are Evolved by Balancing Selection

Statistical analysis is very important to understand the significance of the study. Therefore, we used Tajima's relative rate test to analyze three samples (sequences) selected at random (Tajima, 1998). We repeated this analysis three times using different sequences. For all three cases, the *p*-value and X<sup>2</sup>- (Chi-square) test were found to be significant (Table 4).

Tajima's neutrality test (*D*-test) was conducted to understand the polymorphism of CPKs (Tajima, 1989). We found that, Tajima's *D*-value for all CPKs is 5.269218 (Table 5). A negative Tajima's *D* signifies very low frequency of polymorphism relative to the expectation indicating expansion of population size by purifying selection. Similarly, a positive Tajima's *D* indicates a high level of polymorphism (Tajima, 1989; Fu and Li, 1993;



**FIGURE 5 | Phylogenetic tree of calcium-dependent protein kinases (CPKs).** The CPKs of monocots, dicots, green algae, bryophytes, pteridophytes, and gymnosperms were searched from different genome databases using *CPK* genes of *Arabidopsis thaliana* as queries. Identified sequences were only accepted if the corresponding proteins had serine/threonine protein kinase domain and four calcium binding EF-hand domains. A phylogenetic tree was constructed using Molecular Evolutionary Genetics Analysis 6 (MEGA 6) by neighbor joining method with 1000 bootstrap replicates. The constructed phylogenetic tree showed that, CPKs are clustered into four different groups: A (red), B (blue), C (fuschia), and D (purple).

Simonsen et al., 1995). The high level of polymorphism indicates decrease in population size by balancing selection. In the present study, the Tajima's  $D$ -value for all CPKs was 5.269218 (Table 5) which indicates that, plant CPK genes are undergoing balancing selection and hence decreasing in population size. A Tajima's  $D$  of zero ( $D = 0$ ) indicates that, the average heterozygosity is equal to the number of segregating sites. This can be inferred as an observed variation similar to the expected variation, indicating that the population evolved in mutation-drift equilibrium. A Tajima's ( $D < 0$ ) indicates greater lower average heterozygosity than the number of segregating sites and that there will be very rare alleles at low frequency. A Tajima's  $D$ -value ( $D > 0$ ) indicates

greater average heterozygosity than the number of segregating sites and multiple alleles presents at high frequency. This leads to balancing selection by sudden population contraction. A  $D$ -value greater than +2 or less than -2 is considered as significant; accordingly, the  $D$ -value in the present study is +5.269218 which is significant (Tajima, 1989; Fu and Li, 1993; Simonsen et al., 1995).

### CPK Genes are Differentially Expressed During Stress Conditions

The expression analysis of some *B. rapa* CPK genes treated with cold and heat/drought stress were analyzed. In cold

**TABLE 4 | Results of Tajima's test for 3 sequences in three replicate analyses.**

Configuration	AtCPK18 PpCPK16-3 SmCPK16	ThCPK12 OsCPK3 CreinCPK17-6	GrCPK11-2 OsCPK3 CreinCPK17-6
Identical sites in all three sequences	307	121	122
Divergent sites in all three sequences	49	164	169
Unique differences in sequence A	66	23	22
Unique differences in sequence B	33	57	54
Unique differences in sequence C	29	75	73
P-value	0.00091	0.00014	0.00024
X <sup>2</sup> -test	11.00	14.45	13.47
Degree of freedom	1	1	1

A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages. The analysis involved three amino acid sequences. All positions containing gaps and missing data were eliminated.

**TABLE 5 | Results of Tajima's Neutrality Test.**

m	S	p <sub>s</sub>	Θ	π	D
950	24	1.000000	0.134532	0.427892	5.269218

The analysis involved 950 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 24 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

m, number of sequences; n, total number of sites; S, number of segregating sites; p<sub>s</sub>, S/n; Θ, p<sub>s</sub>/a<sub>1</sub>; π, nucleotide diversity, and D is the Tajima test statistic.

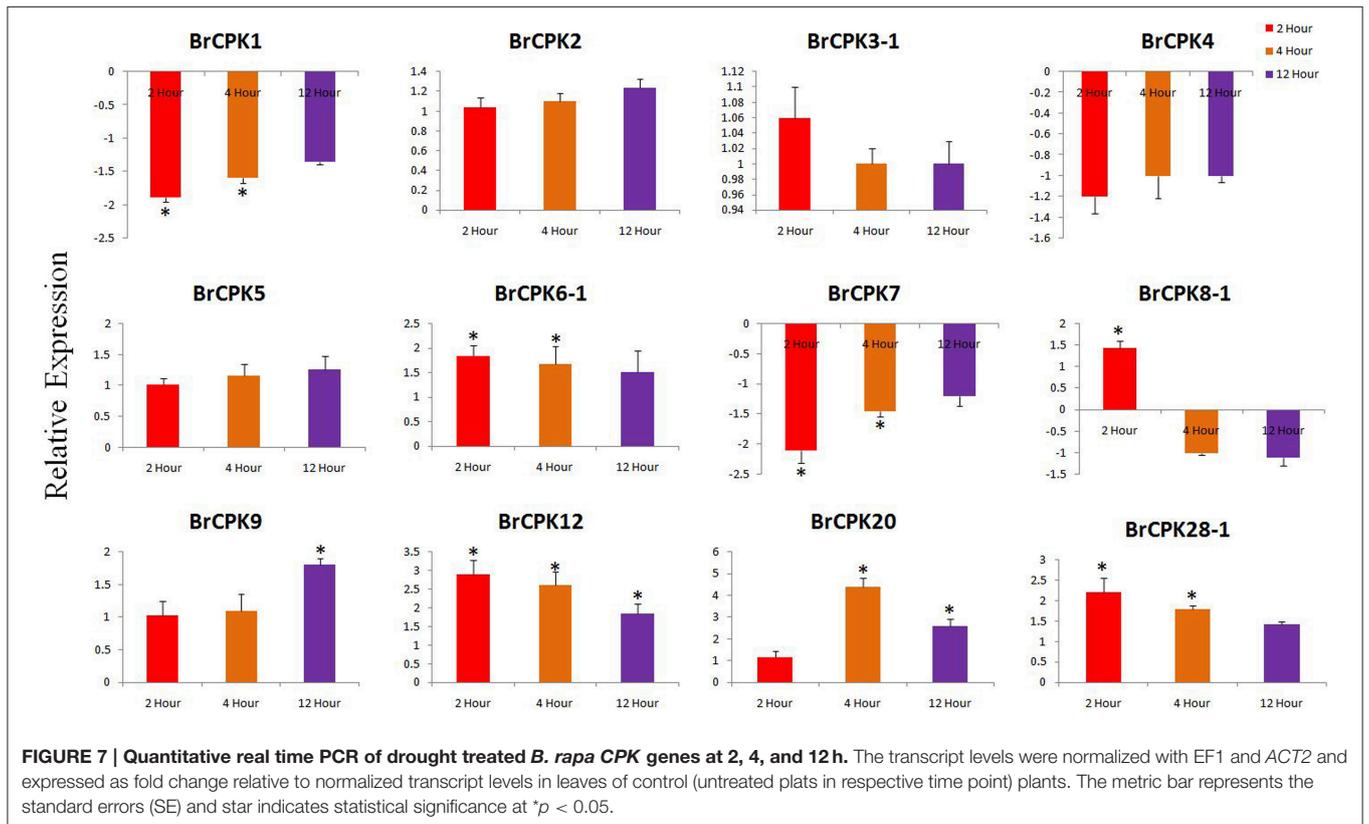
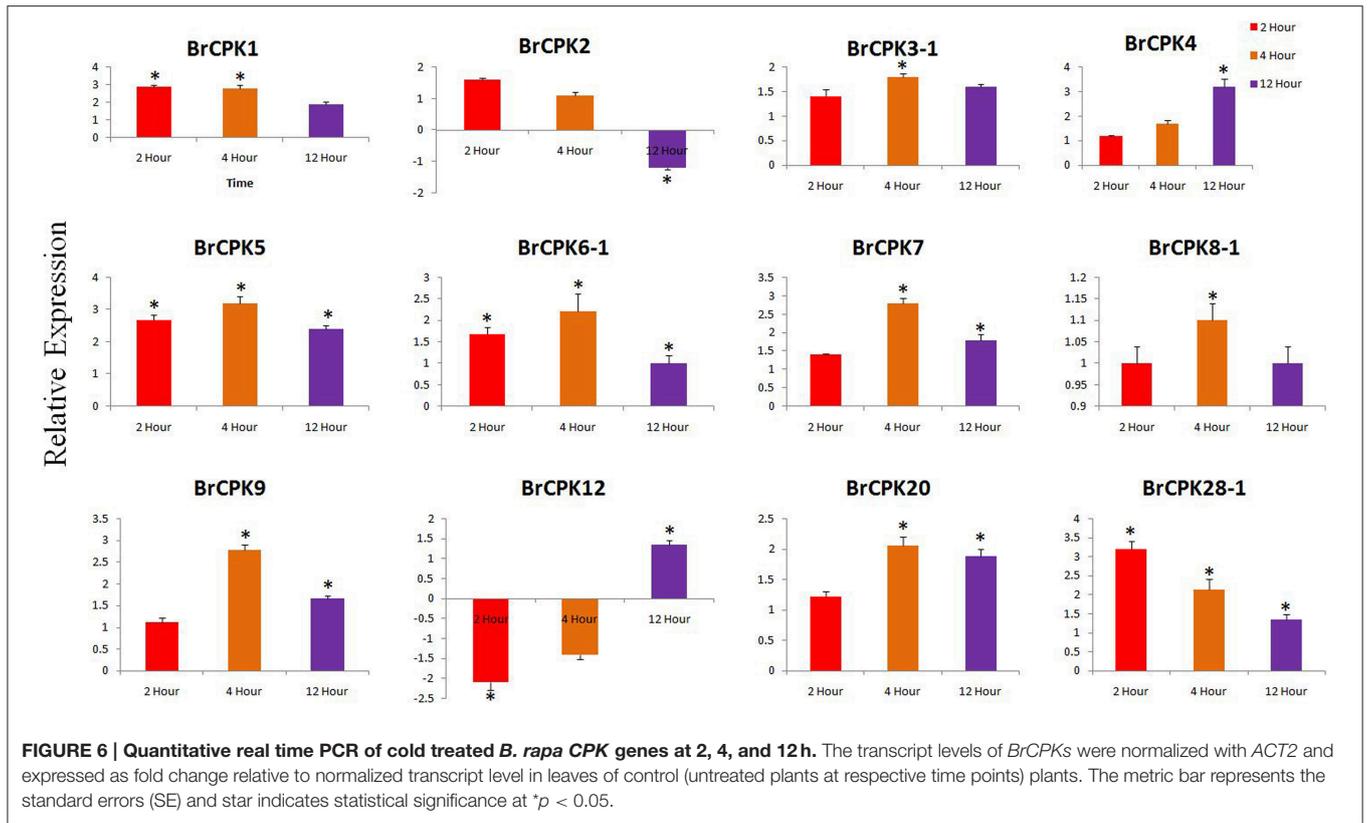
treated *B. rapa*, expression of *BrCPK1* was up-regulated at 2, 4, and 12 h (Figure 6). Additionally, *BrCPK2* was up-regulated at 2 and 4 h and got down regulated at 12 h. The *BrCPK4* was gradually up-regulated from 2 to 12 h (Figure 6). The expression of *BrCPK5* was up-regulated by more than 2-fold in all three time periods. This indicates that, *BrCPK1*, *BrCPK2*, *BrCPK4*, and *BrCPK5* are cold stress responsive genes. The up-regulation at 2 and 4 h time point indicates the early response of calcium signaling genes to cold stress. Down regulation of *BrCPK2* at 12 h indicates, this gene plays significant role in early stage of calcium signaling event and gradually its expression decreases at late stage. The *BrCPK12* was down regulated at 2 and 4 h and up-regulated at 12 h; while *BrCPK28-1* up-regulated at 2 h and then gradually down regulated at 12 h (Figure 6). We found that, the maximum number of CPK genes were up-regulated in response to cold stress indicating that these genes play significant roles in cold stress response in *B. rapa*.

Expression analysis of *B. rapa* CPK genes subjected to drought stress revealed dynamic responses (Figure 7). The *BrCPK1* down regulated in all three time periods similar to *BrCPK4* and *BrCPK7* (Figure 7). However, *BrCPK8-1* was up-regulated at 2 h, then

gradually down regulated. *BrCPK12* was up-regulated at all the three time points. Similarly, the maximum expression of *BrCPK20* was observed at 4 h, when it was up-regulated by more than 4-fold relative to control. The *BrCPK1* and *BrCPK4* was up-regulated in cold stress and down regulated in drought stress. These result indicates, *BrCPKs* responds to cold and drought stress differently. Early studies show that, CPKs are significantly modulated by cold, drought and salt stresses. The *OsCPK4* was found to be induce under cold and drought stresses (Ray et al., 2007). The *BrCPK4* significantly up-regulated at 4 and 12 h time point similar to *OsCPK4*. Rice seedlings grown under cold stress for 3 h shows induction of *OsCPK4*, *OsCPK12*, *OsCPK15*, and *OsCPK21* (Ray et al., 2007). The *BrCPK12* found to be down regulated at 2 and 4 h time point and later get up-regulated at 12 h time point. This result shows that, *BrCPK4* and *BrCPK12* are responsive to cold stress in time dependent manner. *BrCPK7* shows up-regulation due to cold stress. The over expression line of *OsCPK7* shows enhanced tolerance to cold stress (Saijo et al., 2000). This shows that, *BrCPK7* also responds to cold stress and it's over expression might responsible for cold tolerance. Unlike cold stresses, CPKs are also highly modulated due to drought stresses too (Figure 7). The *AtCPK3* and *AtCPK6* regulate ABA mediated stomatal closure (Mori et al., 2006). Thus, up-regulation of *BrCPK3-1* and *BrCPK6-1* in drought stress might be directly linked to stomatal closure in ABA mediated manner. *AtCPK6* over expression plant shows enhanced tolerant to drought stress (Xu et al., 2010). So, up-regulation of *BrCPK6* in drought stress directly reflects its role in drought stress response.

## CONCLUSION AND FUTURE PERSPECTIVES

The CPKs constitute a large multi-gene family in various plant species that play important roles in various physiological processes, including plant growth and development, as well as abiotic and biotic stress responses. *ThCPKs* also plays considerable roles in stress tolerance. Extensive efforts using integrative approaches have provided conclusive evidence that CPKs are versatile and evolutionarily conserved genes that transduce Ca<sup>2+</sup> signals as sensors. The CPKs are involved in a sophisticated Ca<sup>2+</sup> signaling network via protein phosphorylation and coordinate dynamic cellular processes for plant growth and development. Further studies will lead to a better understanding of specific and redundant roles of CPKs and their conserved domains in different signaling networks. To fully elucidate the Ca<sup>2+</sup>-mediated signaling network, it is essential to establish cellular, molecular and genetic linkage with Ca<sup>2+</sup> channels, transporters and pumps for spatio-temporal analysis of CPK gene activation and translocation. *In vivo* activation of CPK genes has been limited because of difficulties in maintaining required Ca<sup>2+</sup> levels, which reflect the physiological state in cell extracts. Raising specific anti-phosphopeptide antibodies against CPK protein or their substrates would facilitate monitoring of CPK activation, while identification



of downstream substrates and unique regulatory features of each isoform will provide insight into CPK signaling, enabling understanding of their integrated roles in diverse biological process.

## AUTHOR CONTRIBUTIONS

TM conceived the idea, design the experiment analyzed data and drafted the manuscript. NM was analyzed the data. YM and HB revised the manuscript.

## REFERENCES

- Akashi, H., and Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3695–3700. doi: 10.1073/pnas.062526999
- Aravind, L., Mazumder, R., Vasudevan, S., and Koonin, E., V. (2002). Trends in protein evolution inferred from sequence and structure analysis. *Curr. Opin. Struct. Biol.* 12, 392–399. doi: 10.1016/S0959-440X(02)00334-2
- Asano, T., Tanaka, N., Yang, G., Hayashi, N., and Komatsu, S. (2005). Genome-wide identification of the rice calcium-dependent protein kinase and its closely related kinase gene families: comprehensive analysis of the CDPKs gene family in rice. *Plant Cell Physiol.* 46, 356–366. doi: 10.1093/pcp/pci035
- Baker, C. R., Hanson-Smith, V., and Johnson, A. D. (2013). Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* 342, 104–108. doi: 10.1126/science.1240810
- Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., dePamphilis, C., et al. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963. doi: 10.1126/science.1203810
- Bisova, K., Krylov, D. M., and Umen, J. G. (2005). Genome-wide annotation and expression profiling of cell cycle regulatory genes in *Chlamydomonas reinhardtii*. *Plant Physiol.* 137, 475–491. doi: 10.1104/pp.104.054155
- Boudsocq, M., and Sheen, J. (2013). CDPKs in immune and stress signaling. *Trends Plant Sci.* 18, 30–40. doi: 10.1016/j.tplants.2012.08.008
- Chauve, C., Doyon, J.-P., and El-Mabrouk, N. (2008). Gene family evolution by duplication, speciation, and loss. *J. Comput. Biol.* 15, 1043–1062. doi: 10.1089/cmb.2008.0054
- Chen, X., Zheng, J., Fu, Z., Nan, P., Zhong, Y., Lonardi, S., et al. (2005). Assignment of orthologous genes via genome rearrangement. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2, 302–315. doi: 10.1109/TCBB.2005.48
- Cheng, S., Willmann, M. R., Chen, H., and Sheen, J. (2002). Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiol.* 129, 469–485. doi: 10.1104/pp.005645
- Choi, W.-G., Toyota, M., Kim, S.-H., Hilleary, R., and Gilroy, and, S. (2014). From the cover: salt stress-induced Ca<sup>2+</sup> waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6497–6502.
- Clouse, S. D. (2011). Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development. *Plant Cell* 23, 1219–1230. doi: 10.1105/tpc.111.084475
- Curran, A., Chang, I.-F., Chang, C.-L., Garg, S., Miguel, R. M., Barron, Y. D., et al. (2011). Calcium-dependent protein kinases from *Arabidopsis* show substrate specificity differences in an analysis of 103 substrates. *Front. Plant Sci.* 2:36. doi: 10.3389/fpls.2011.00036
- Dadacz-Narloch, B., Beyhl, D., Larisch, C., López-Sanjurjo, E. J., Reski, R., Kuchitsu, K., et al. (2011). A novel calcium binding site in the slow vacuolar cation channel TPC1 senses luminal calcium levels. *Plant Cell* 23, 2696–2707. doi: 10.1105/tpc.111.086751
- Davidson, D., and Gu, F. X. (2012). Materials for sustained and controlled release of nutrients and molecules to support plant growth. *J. Agric. Food Chem.* 60, 870–876. doi: 10.1021/jf204092h
- Day, I. S., Reddy, V. S., Shad Ali, G., and Reddy, A. S. (2002). Analysis of EF-hand-containing proteins in *Arabidopsis*. *Genome Biol.* 3:RESEARCH0056. doi: 10.1186/gb-2002-3-10-research0056
- de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34, W362–W365. doi: 10.1093/nar/gkl124
- Dodd, A. N., Kudla, J., and Sanders, D. (2010). The language of calcium signaling. *Annu. Rev. Plant Biol.* 61, 593–620. doi: 10.1146/annurev-arplant-070109-104628
- Fu, Y.-X., and Li, W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.
- 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
- Hamel, L.-P., Nicole, M.-C., Sritubtim, S., Morency, M.-J., Ellis, M., Ehltling, J., et al. (2006). Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci.* 11, 192–198. doi: 10.1016/j.tplants.2006.02.007
- Hamel, L.-P., Sheen, J., and Séguin, A. (2014). Ancient signals: comparative genomics of green plant CDPKs. *Trends Plant Sci.* 19, 79–89. doi: 10.1016/j.tplants.2013.10.009
- Harmon, A. C., Gribskov, M., Gubrium, E., and Harper, J. F. (2001). The CDPK superfamily of protein kinases. *New Phytol.* 151, 175–183. doi: 10.1046/j.1469-8137.2001.00171.x
- Harmon, A. C., Gribskov, M., and Harper, J. F. (2000). CDPKs - a kinase for every Ca<sup>2+</sup> signal? *Trends Plant Sci.* 5, 154–159. doi: 10.1016/S1360-1385(00)01577-6
- Hrabak, E., Chan, C., and Gribskov, M. (2003). The *Arabidopsis* CDPK-SnRK superfamily of protein kinases. *Plant Physiol.* 132, 666–680. doi: 10.1104/pp.102.011999
- Huang, G.-T., Ma, S.-L., Bai, L.-P., Zhang, L., Ma, H., Jia, P., et al. (2012). Signal transduction during cold, salt, and drought stresses in plants. *Mol. Biol. Rep.* 39, 969–987. doi: 10.1007/s11033-011-0823-1
- Hwang, I., Sze, H., and Harper, J. F. (2000). A calcium-dependent protein kinase can inhibit a calmodulin-stimulated Ca<sup>2+</sup> pump (ACA2) located in the endoplasmic reticulum of *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6224–6229. doi: 10.1073/pnas.97.11.6224
- Ito, T., Nakata, M., Fukazawa, J., Ishida, S., and Takahashi, Y. (2010). Alteration of substrate specificity: the variable N-terminal domain of tobacco Ca<sup>2+</sup>-dependent protein kinase is important for substrate recognition. *Plant Cell* 22, 1592–1604. doi: 10.1105/tpc.109.073577
- Jain, M., Khurana, P., Tyagi, A. K., and Khurana, J. P. (2008). Genome-wide analysis of intronless genes in rice and *Arabidopsis*. *Funct. Integr. Genomics* 8, 69–78. doi: 10.1007/s10142-007-0052-9
- Jordan, I. K., Makarova, K. S., Spouge, J. L., Wolf, Y. I., and Koonin, E., V (2001). Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Res.* 11, 555–565. doi: 10.1101/gr.1660R
- Kanchiswamy, C. N., Mohanta, T. K., Capuzzo, A., Occhipinti, A., Verrillo, F., Maffei, M. E., et al. (2013). Differential expression of CPKs and cytosolic Ca<sup>2+</sup> variation in resistant and susceptible apple cultivars (*Malus x domestica*) in response to the pathogen *Erwinia amylovora* and mechanical wounding. *BMC Genomics* 14:760. doi: 10.1186/1471-2164-14-760

## ACKNOWLEDGMENTS

This study was carried out with the support of the “Forest Science and Technology Projects (Project no. S121414L030110)” provided by the Korea Forest Service.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.01146>

- Kong, X., Lv, W., Jiang, S., Zhang, D., Cai, G., Pan, J., et al. (2013). Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. *BMC Genomics* 14:433. doi: 10.1186/1471-2164-14-433
- Krouk, G., Ruffel, S., Gutiérrez, R. A., Gojon, A., Crawford, N. M., Coruzzi, G. M., et al. (2011). A framework integrating plant growth with hormones and nutrients. *Trends Plant Sci.* 16, 178–182. doi: 10.1016/j.tplants.2011.02.004
- 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
- Lawson, T., Davey, P. A., Yates, S. A., Bechtold, U., Baeshen, M., Baeshen, N., et al. (2014). C3 photosynthesis in the desert plant *Rhazya stricta* is fully functional at high temperatures and light intensities. *New Phytol.* 201, 862–873. doi: 10.1111/nph.12559
- Lespinet, O., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 12, 1048–1059. doi: 10.1101/gr.174302
- Li, Q., Li, L., Yang, X., Warburton, M., Bai, G., Dai, J., et al. (2010). Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. *BMC Plant Biol.* 10:143. doi: 10.1186/1471-2229-10-143
- Li, Y., Hao, P., Zheng, S., Tu, K., Fan, H., Zhu, R., et al. (2008). Gene expression module-based chemical function similarity search. *Nucleic Acids Res.* 36, e137. doi: 10.1093/nar/gkn610
- Liu, W., Li, W., He, Q., Daud, M. K., Chen, J., and Zhu, S. (2014). Genome-wide survey and expression analysis of calcium-dependent protein kinase in *Gossypium raimondii*. *PLoS ONE* 9:e98189. doi: 10.1371/journal.pone.0098189
- Lu, S. X., and Hrabak, E. M. (2002). An arabidopsis calcium-dependent protein kinase is associated with the endoplasmic reticulum 1. *Plant Physiol.* 128, 1008–1021. doi: 10.1104/pp.010770
- Luan, S. (2009). The CBL-CIPK network in plant calcium signaling. *Trends Plant Sci.* 14, 37–42. doi: 10.1016/j.tplants.2008.10.005
- Mackay, J., Dean, J. F. D., Plomion, C., Peterson, D. G., Cánovas, F. M., Pavy, N., et al. (2012). Towards decoding the conifer giga-genome. *Plant Mol. Biol.* 80, 555–569. doi: 10.1007/s11103-012-9961-7
- Mahmood, K., Webb, G., Song, J., Whisstock, J., and Konagurthu, A. (2012). Efficient large-scale protein sequence comparison and gene matching to identify orthologs and co-orthologs. *Nucl. Acids Res.* 40:e44. doi: 10.1093/nar/gkr1261
- Makarova, K. S., Wolf, Y. I., Mekhedov, S. L., Mirkin, B. G., and Koonin, E. V. (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* 33, 4626–4638. doi: 10.1093/nar/gki775
- Martin, M. L., and Busconi, L. (2001). A rice membrane-bound calcium-dependent protein kinase is activated in response to low temperature. *Plant Physiol.* 125, 1442–1449. doi: 10.1104/pp.125.3.1442
- Matos, D. A., Cole, B. J., Whitney, I. P., MacKinnon, K. J.-M., Kay, S. A., and Hazen, S. P. (2014). Daily changes in temperature, not the circadian clock, regulate growth rate in *Brachypodium distachyon*. *PLoS ONE* 9:e100072. doi: 10.1371/journal.pone.0100072
- Mohanta, T. K., Arora, P. K., Mohanta, N., Parida, P., and Bae, H. (2015a). Identification of new members of the MAPK gene family in plants shows diverse conserved domains and novel activation loop variants. *BMC Genomics* 16:58. doi: 10.1186/s12864-015-1244-7
- Mohanta, T., and Bae, H. (2015). Functional genomics and signaling events in mycorrhizal symbiosis. *J. Plant Interact.* 10, 21–40. doi: 10.1080/17429145.2015.1005180
- Mohanta, T. K., and Mohanta, N. (2013). Genome wide identification of auxin efflux carrier gene family in physcomitrella patens. *J. Biotechnol. Sci.* 1, 54–64.
- Mohanta, T. K., Mohanta, N., Mohanta, Y. K., Parida, P., and Bae, H. (2015b). Genome-wide identification of Calcineurin B-Like (CBL) gene family of plants reveals novel conserved motifs and evolutionary aspects in calcium signaling events. *BMC Plant Biol.* 15:189. doi: 10.1186/s12870-015-0543-0
- Mohanta, T. K., Occhipinti, A., Atsbaha Zebelo, S., Foti, M., Fliegmann, J., Bossi, S., et al. (2012). Ginkgo biloba responds to herbivory by activating early signaling and direct defenses. *PLoS ONE* 7:e32822. doi: 10.1371/journal.pone.0032822
- Mohanta, T., Malnoy, M., Mohanta, N., and Kanchiswamy, C. (2014). *In-silico* identification and phylogenetic analysis of auxin efflux carrier gene family in *Setaria italica* L. *Afr. J. Biotechnol.* 13, 211–225. doi: 10.5897/AJB2014.13617
- Mori, I. C., Murata, Y., Yang, Y., Munemasa, S., Wang, Y. F., Andreoli, S., et al. (2006). CDPKs CPK6 and CPK3 function in ABA regulation of guard cell S-type anion- and Ca<sup>2+</sup>- permeable channels and stomatal closure. *PLoS Biol.* 4:e327. doi: 10.1371/journal.pbio.0040327
- Morse, A. M., Peterson, D. G., Islam-Faridi, M. N., Smith, K. E., Magbanua, Z., Garcia, S. A., et al. (2009). Evolution of genome size and complexity in Pinus. *PLoS ONE* 4:e4332. doi: 10.1371/journal.pone.0004332
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., et al. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 35, D883–D887. doi: 10.1093/nar/gk1976
- Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., et al. (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7705–7710. doi: 10.1073/pnas.0611046104
- Ray, S., Agarwal, P., Arora, R., Kapoor, S., and Tyagi, A. K. (2007). Expression analysis of calcium-dependent protein kinase gene family during reproductive development and abiotic stress conditions in rice (*Oryza sativa* L. ssp. indica). *Mol. Genet. Genomics* 278, 493–505. doi: 10.1007/s00438-007-0267-4
- 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
- Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., and Yao, X. (2008). CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Eng. Des. Sel.* 21, 639–644. doi: 10.1093/protein/gzn039
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646
- Saijo, Y., Hata, S., Kyojuka, J., Shimamoto, K., and Izui, K. (2000). Over-expression of a single Ca<sup>2+</sup>-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. *Plant J.* 23, 319–327. doi: 10.1046/j.1365-313x.2000.00787.x
- Schlicker, A., Domingues, F. S., Rahnenführer, J., and Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* 7:302. doi: 10.1186/1471-2105-7-302
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Schulz, P., Herde, M., and Romeis, T. (2013). Calcium-dependent protein kinases: hubs in plant stress signaling and development. *Plant Physiol.* 163, 523–530. doi: 10.1104/pp.113.222539
- Simonsen, K. L., Churchill, G. A., and Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141, 413–429.
- Smotrys, J. E., and Linder, M. E. (2004). Palmitoylation of intracellular signaling proteins: regulation and function. *Annu. Rev. Biochem.* 73, 559–587. doi: 10.1146/annurev.biochem.73.011303.073954
- Sulmann, S., Dell'orco, D., Marino, V., Behnen, P., and Koch, K.-W. (2014). Conformational changes in calcium-sensor proteins under molecular crowding conditions. *Chemistry* 20, 6756–6762. doi: 10.1002/chem.201402146
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 95, 585–595.
- Tajima, F. (1998). Simple methods for testing the molecular evolutionary clock hypothesis. *Genet. Soc. Am.* 135, 599–607.
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tena, G., Boudsocq, M., and Sheen, J. (2011). Protein kinase signaling networks in plant innate immunity. *Curr. Opin. Plant Biol.* 14, 519–529. doi: 10.1016/j.pbi.2011.05.006

- Thornton, J. W., and DeSalle, R. (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genomics Human Genet.* 1, 41–73. doi: 10.1146/annurev.genom.1.1.41
- Ting, C.-T., Tsaur, S.-C., Sun, S., Browne, W. E., Chen, Y.-C., Patel, N. H., et al. (2004). Gene duplication and speciation in *Drosophila*: evidence from the Odysseus locus. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12232–12235. doi: 10.1073/pnas.0401975101
- Tuteja, N., and Mahajan, S. (2007). Calcium signaling network in plants. *Plant Signal. Behav.* 2, 79–85. doi: 10.4161/psb.2.2.4176
- Valmonte, G. R., Arthur, K., Higgins, C. M., and Macdiarmid, R. M. (2014). Calcium-dependent protein kinases in plants: evolution, expression and function. *Plant Cell Physiol.* 55, 551–569. doi: 10.1093/pcp/pct200
- Vleeshouwers, V. G., Martens, A., van Dooijeweert, W., Colon, L. T., Govers, F., and Kamoun, S. (2001). Ancient diversification of the Pto kinase family preceded speciation in Solanum. *Mol. Plant Microbe Interact.* 14, 996–1005. doi: 10.1094/MPMI.2001.14.8.996
- Wasternack, C., and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in *Annals of Botany*. *Ann. Bot.* 111, 1021–1058. doi: 10.1093/aob/mct067
- Weljie, A. M., Robertson, K. M., and Vogel, H. J. (2003). Conformational changes in the Ca<sup>2+</sup>-regulatory region from soybean calcium-dependent protein kinase- $\alpha$ : fluorescence resonance energy transfer studies. *J. Biol. Chem.* 278, 43764–43769. doi: 10.1074/jbc.M306799200
- Wright, M. W., and Bruford, E. A. (2006). Human and orthologous gene nomenclature. *Gene* 369, 1–6. doi: 10.1016/j.gene.2005.10.029
- Wu, F., Mueller, L. A., Crouzillat, D., Pétiard, V., and Tanksley, S. D. (2006). Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* 174, 1407–1420. doi: 10.1534/genetics.106.062455
- Xu, J., Tian, Y. S., Peng, R. H., Xiong, A. S., Zhu, B., Jin, X. F., et al. (2010). AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in *Arabidopsis*. *Planta* 231, 1251–1260. doi: 10.1007/s00425-010-1122-0
- Yan, Y. L., Bhattacharya, P., He, X. J., Ponugoti, B., Marquardt, B., Layman, J., et al. (2012). Duplicated zebrafish co-orthologs of parathyroid hormone-related peptide (PTHrP, Pthlh) play different roles in craniofacial skeletogenesis. *J. Endocrinol.* 214, 421–435. doi: 10.1530/JOE-12-0110
- Zhao, Y., Pokutta, S., Maurer, P., Lindt, M., Franklin, R. M., and Kappes, B. (1994). Calcium-binding properties of a calcium-dependent protein kinase from *Plasmodium falciparum* and the significance of individual calcium-binding sites for kinase activation. *Biochemistry* 33, 3714–3721. doi: 10.1021/bi00178a031
- Zimmer, S. L., Fei, Z., and Stern, D. B. (2008). Genome-based analysis of *Chlamydomonas reinhardtii* exoribonucleases and poly(A) polymerases predicts unexpected organellar and exosomal features. *Genetics* 179, 125–136. doi: 10.1534/genetics.107.086223
- Zuo, R., Hu, R., Chai, G., Xu, M., Qi, G., Kong, Y., et al. (2013). Genome-wide identification, classification, and expression analysis of CDPK and its closely related gene families in poplar (*Populus trichocarpa*). *Mol. Biol. Rep.* 40, 2645–2662. doi: 10.1007/s11033-012-2351-z

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

Copyright © 2015 Mohanta, Mohanta, Mohanta and Bae. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.