JEB™

ISSN: 0254-8704 (Print)
ISSN: 2394-0379 (Online)
CODEN: JEBIDP

Genome wide identification and characterization of Calcium Dependent Protein Kinase gene family in *Cajanus cajan*

Authors Info

Dhammaprakash Pandhari Wankhede^{1*}, Meenu Kumari², Tambi Richa³, J. Aravind¹ and Subramani Rajkumar¹

¹ICAR-National Bureau of Plant Genetic Resources, New Delhi-110 012, India

²ICAR-IIHR, Central Horticultural Experiment Station, Aiginia, Bhubaneswar-751 019, India

³Kuroda Lab, Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan

*Corresponding Author Email : d.wankhede@icar.gov.in

Key words

CDPK family, Expression analysis, Hidden Markov model, Phylogenetic analysis, Pigeonpea

Publication Info

Paper received : 22.09.2016
Revised received : 21.10.2016
Accepted : 05.11.2016

Abstract

Aim: Calcium-dependent protein kinases (CDPKs) are Ca^{2+} binding proteins which play a crucial role in plant growth and development, various biotic and abiotic stress and hormone signal transduction pathways through Ca^{2+} signal transduction mechanism. The main objective of the present study was to identify and characterize CDPK gene family from Pigeonpea genome and assign scientific nomenclature.

Methodology: A Hidden Markov model (HMM) was built from known CDPKs of other plants and searched against Pigeonpea genome. Phylogenetic analysis of CDPKs was done using Molecular Evolutionary Genetics Analysis (MEGA) 6 program. Modeller 9.15 was used to design 3-D models of CDPK proteins following homology modeling approach. Expression of identified CDPKs was studied using CLC genomics workbench making use of transcriptome Sequence Read Archive (SRA) data.

Results: Total 23 members of CDPK gene family were identified from Pigeonpea genome and characterized. The identified CDPK family was divided into four groups on the basis of sequence similarity and phylogeny. The Pigeonpea CDPKs were named following orthologous nomenclature on the basis of *Arabidopsis* CDPKs. A detailed *in-silico* characterization of Pigeonpea CDPK was performed. The transcript level evidences of Pigeonpea CDPKs was seen in the form of Expressed Sequence Tags (ESTs). Expression analysis revealed that most of the Pigeonpea CDPKs genes were expressed in two Pigeonpea cultivars, Asha and Upas 120.

Interpretation: The present study gives an insight into various structural and functional aspects of CDPK family in Pigeonpea. Further, this work is expected to facilitate more elaborative and comprehensive research on CDPKs in Pigeonpea.

Introduction

Calcium as a second messenger plays a crucial role in signal transductions mechanism in eukaryotes. In plants, several cues of internal (hormones) and external natures (such as heat, cold, drought, light, wounding, pathogen elicitors etc.) are known to regulate intracellular concentrations of Ca^{2+} (Cheng *et al.*, 2002; Bhattacharjee, 2009; Hamel *et al.*, 2014). Ca^{2+} , thus plays a pivotal role not only in growth and development but also in generating responses to several biotic and abiotic stress conditions. Influx of Ca^{2+} occurs from various subcellular locations such as vacuole, apoplastic space, mitochondrion, chloroplast and nuclear envelop. Specificity of complex Ca^{2+} signalling appears to be regulated by stimuli specific Ca^{2+} signature in terms of magnitude, kinetics and cellular sources, which in turn differentially activates Ca^{2+} channels, $\text{H}^+/\text{Ca}^{2+}$ antiporters and H^+ and Ca^{2+} -ATPases (Bush, 1995; Cheng *et al.*, 2002). Several Ca^{2+} sensor proteins sense these altered calcium signature and convert these specific upstream signal to appropriate downstream cellular responses (Hashimoto and Kudla, 2011). Among different classes of Ca^{2+} sensors, CaMs (calmodulins) and CaM-like proteins, CBLs (calcineurin B-like proteins), and CDPKs/CPKs (calcium-dependent protein kinases) play a major role (Sanders *et al.*, 2002). The CDPKs particularly generate interest as it possesses both calmodulin-like domains, as well as protein kinase domain (conserved serine threonine kinase domain). Owing to both Ca^{2+} sensing and kinase activity, CDPK proteins can directly bind calcium and independent of calmodulin proteins can translate Ca^{2+} signal to downstream responses by phosphorylation of specific target proteins. The binding of Ca^{2+} to EF hands (Calmodulin domain) of CDPK results in conformational changes removing autoinhibitory domain from the kinase domain which then activates CDPK (Wernimont *et al.*, 2011).

Role of CDPK has been shown in growth and development of plants, as well as in biotic and abiotic stress signal transductions. In *Arabidopsis*, AtCPK1 has been shown to play role in pathogen resistance by phosphorylating Phenylalanine ammonia-lyase (PAL), and thereby accumulation of salicylic acid (SA) (Coca and Segundo, 2010).

Role of CPK1/2 has also been shown in cell death in response to pathogen effectors. Further, other *Arabidopsis* CDPKs such as AtCPK4/AtCPK5/AtCPK6/AtCPK11 phosphorylates a group of WRKY transcription factors which in turn are responsible for generation of responses restricting pathogen growth through Nucleotide-binding domain leucine-rich repeat (NLR) protein (Gao *et al.*, 2013). CDPK (CPK10) also plays an important role in stomatal regulation in drought stress (Zhu *et al.* 2007). Further, *Arabidopsis* CDPKs have been shown to act as positive regulator (CPK4 and CPK11) in ABA signaling pathways (Zhu *et al.*, 2007).

Analysis of *Arabidopsis* genome sequence revealed presence of total 34 CDPK genes (Cheng *et al.*, 2002). Genome wide analysis of rice and poplar genome has shown presence of 31 and 30 CDPKs, respectively (Ray *et al.*, 2007; Zuo *et al.*, 2013). Pigeon pea once considered as an orphan crop has now received importance due to its commercial value. It finds its origin in India. But production of this commercially important crop has been affected due to biotic stress. Recently, the draft genome of Pigeonpea has been published (Varshney *et al.*, 2011; Singh *et al.*, 2012). However, there has been no detailed study related to CDPK gene family in Pigeonpea and their contribution towards resilience against stresses. Hence, the present work aimed at identifying of CDPK gene family in Pigeonpea and its characterization.

Materials and Methods

Sequence and database search for CDPKs : The CDPK sequences from *Arabidopsis* and rice were obtained from National Center for Biotechnology Information (NCBI) and The Arabidopsis Information Resource (TAIR). In order to retrieve CDPK sequences from diverse plant species, standard Protein BLAST was performed using *Arabidopsis* and rice CDPK sequences. Resultant hits with more than 50% identity and annotated CDPK function were downloaded. Pigeonpea genome files were downloaded from NCBI genome database. The total *C. cajan* protein sequence file having 48,331 protein sequences was used to construct a local protein database. A standalone BLAST search was performed using total 164 CDPK protein sequences from diverse plants species used as query against total Pigeonpea protein sequences with e-values 0.001. Sequences with more than 50% identity were retrieved and used for subsequent analysis. Further, using CLUSTAL Omega, 164 CDPKs from diverse plant species were aligned to build HMM profile using HMMER3.0 and used to search Pigeonpea local protein database.

Phylogenetic tree construction and nomenclature : Full-length *Arabidopsis* and Pigeonpea CDPK amino acid sequences were aligned using ClustalW program. The resulting alignment was used in Molecular Evolutionary Genetics Analysis 6 (MEGA6) program (Tamura *et al.*, 2013) for generation of unrooted phylogenetic tree using neighbor-joining method with 1000 bootstrap replicates. Pigeonpea CDPKs were named as per the name of respective putative orthologs in *Arabidopsis*.

Multiple sequence alignment, conserved domain/signature detection and localization : MultAlin, a multiple alignment program (Corpet, 1988) was used for multiple sequence alignments of Pigeonpea CDPK amino acid sequences. Conserved domains and signature sequences were determined using conserved domain search program and PROSITE database of protein domains, families and functional sites. Subcellular localization was predicted using PSORT Prediction

tool (Nakai and Horton, 1999).

Gene structure analysis : The gene features (composition and position of exon, intron) for CDPK genes were illustrated using Gene Structure Display Server (GSDS) program (Hu *et al.*, 2015). The CDPK genes structural features were retrieved from genome annotation file (GFF) from NCBI genome database and used as input for visualization.

ESTs and transcriptome analysis : Pigeonpea ESTs available at dbEST database, NCBI were used. Total ESTs of 25578 were used to construct local database. A local blast search was performed using CcCDPKs coding sequences (CDS) as a query against total Pigeonpea ESTs as database. Transcriptome sequence data (454 GS-FLX sequence reads) of pigeonpea cultivar UPAS120 (SRP002557) and Asha (SRP002556) (Dutta *et al.*, 2011) available in Sequence read Archives (SRA), NCBI was used for expression study. CLC genomics workbench was used for expression analysis as per 'RNAseq Protocol'. For alignment, two mismatches at most were allowed. Unique read counts were normalized by calculating RPKM (Read Per Kilobase of transcript per Million mapped reads).

Homology modeling : 3D models of CDPKs were generated computationally following homology modeling approach as mentioned in Rao *et al.* (2011) using MODELLER 9.15 (Fiser *et al.*, 2000). The template structure used for CcCDPK6 were 4MVF (identity-39%, e value-7e-108), 3IGO (identity-40%, e value-4e-102) and 4M97 (identity-38%, e value-1e-96). For CcCDPK3 also three templates used were, 2AAO (identity-61%, e value-5e-63), 2WEI (identity-43%, e value-8e-66), and 4M97 (identity-39%, e value-5e-96).

Results and Discussion

Availability of Pigeonpea genome sequence has facilitated the genome wide analysis study in once considered the 'orphan crop'. To identify CDPK gene family in Pigeonpea, HMM as well as standalone BLAST approach was used. Total 164 CDPK amino acid sequences from diverse plant species including *Arabidopsis* and rice CDPK sequences were searched against total protein sequences (48,331) of Pigeonpea. Simultaneously, a HMM profile was built using 164 CDPK sequences from diverse plant species and profile search was carried out against total Pigeonpea protein sequences. Results of the local BLAST search and HMM were compared, repeated sequences were removed and the sequences were parsed manually. Unique sequences obtained after comparing both BLAST and HMMER approach were further used to screen for the presence of protein kinase domain and Calmodulin domain (EF-hands) using NCBI conserved domain search and PROSITE database of protein domains, families and functional sites. The protein is considered genuine CDPK only if it had consensus S/T protein kinase domain, conserved D and K amino acid residues

within active site (D[L/I/V]K) motif, and a calmodulin-like domain having at least one EF-hand motif (Hamel *et al.*, 2014). The above screening resulted in identification of 23 CDPK proteins from Pigeonpea genome. As per the genome annotation of Pigeonpea, there are total 28 calcium dependent protein kinases in Pigeonpea. However, screenings showed that five of the 28 sequence were either lacking the conserved Ser/Thr protein kinase domain (KYP37630.1) or the Ca²⁺ binding EF hand (KYP73635.1, KYP74920.1, KYP38683.1 and KYP71837.1), so they were excluded from the study. In comparison to *Arabidopsis*, rice and soybean and poplar, CDPK gene family of Pigeonpea appears to be smaller as compared to other plant species (Cheng *et al.*, 2002, Ray *et al.*, 2007; Liu *et al.*, 2016). The small gene family size could either be due to non-availability of completed genome and it has also been speculated that a few CDPKs in plant genomes might have lost due to functional redundancy (Zuo *et al.*, 2013).

The identified CDPKs were 462 to 580 amino acids in length with predicted molecular weight ranging from 55.85 kDa to 65.22 kDa and isoelectric point (pI) ranging from 5.27 to 9.15 (Table 1). Possible subcellular localization of the Pigeonpea CDPKs have also been predicted (Table 1).

In order to avoid ambiguity in CDPK names, in the present study a unified orthologous nomenclature was proposed to the identified Pigeonpea CDPKs with reference to *Arabidopsis* CDPK as reported for poplar CDPK (Hamel *et al.*, 2014). The unified nomenclature is expected to facilitate species-to-species comparisons and might help functional characterization of homologs having clear orthologous relationships model plant *Arabidopsis* (Hamel *et al.*, 2014). Further, there exists redundancy in the nomenclature given in genome annotation to Pigeonpea CDPKs. For example, three CDPKs (KYP37692.1, KYP74724.1 and KYP74725.1) were named as CDPK4; two CDPKs (KYP50696.1, KYP56058.1) were named as CDPK30, etc. This prevalent redundancy also necessitated the robust and unified nomenclature for Pigeonpea CDPK. Hence, the identified Pigeonpea CDPKs were named following orthologous nomenclature with reference to *Arabidopsis*. The Pigeonpea CDPK closest to its putative ortholog in *Arabidopsis* CDPK were named accordingly, viz. CcCDPK3 was closest or similar to AtCPK3 (Table 2, Fig. 1). Additional numberings were followed to differentiate paralogs such as CcCDPK 9-1 and CcCDPK 9-2 which were similar to AtCPK9. Such orthologous nomenclature was followed for naming other gene family, where reliable orthologous relationship could be drawn (Hamel *et al.*, 2006). Such nomenclature shows only evolutionary relationships, and conservation of biological functions among orthologs needs to be established.

In order to further characterize the identified CDPKs and to study the phylogenetic relationship, the full length CDPK sequences (57 CDPKs) from Pigeonpea and *Arabidopsis* were

Table 1 : Characteristics of Pigeonpea CDPK proteins

Group	Pigeonpea CDPK Name	Protein ID ¹	Amino acid ²	Mol wt kDa ³	pI ⁴	AA in PK do-main ⁵	PK act site ⁶	EF hand ⁷	N-Term. Amino acid ⁸	Myristoylation ⁹	Subcellular Localization ¹⁰
I	CcCDPK1	KYP49458.1	576	64.33	6.08	257	234	4	MGNTCIGS	No	N, ChSt, ChTM, ChTS
	CcCDPK2	KYP56975.1	576	64.41	5.62	257	235	4(1,2)	MGNTCVGL	No	MP, N
	CcCDPK4-1	KYP74724.1	492	55.28	5.3	258	149	4	MQKHGFAS¥	No	ERM, MP, PM
	CcCDPK4-2	KYP74725.1	498	55.80	5.53	258	156	4	MSEGGNKV¥	No	MP, N
	CcCDPK6	KYP37692.1	487	54.82	5.4	257	148	4	MRRGLDNQ¥	No	Cy, MP
	CcCDPK11-1	KYP69500.1	496	55.85	5.27	258	154	4	MSKPNSGP¥	No	MP, N
	CcCDPK11-2	KYP48975.1	490	55.54	5.16	257	140	4	MGWVLPQR	No	MP, N
	CcCDPK20	KYP44264.1	580	65.22	5.47	258	231	4(4)	MGNNCVGP	No	ERM, PM, N
II	CcCDPK3	KYP58662.1	462	52.44	5.74	258	133	4	MEDVRSMY¥	No	N, MP
	CcCDPK9-1	KYP73408.1	529	59.85	6.52	258	205	4	MGSCLSKR	Yes	ERM, PM, MP
	CcCDPK9-2	KYP52211.1	553	62.09	6.31	258	223	4	MGCFLSKD	Yes	ERM, PM, MP
	CcCDPK17-1	KYP73246.1	522	58.44	5.72	257	191	4	MGNCCSGG	Yes	ERM, ChTM, PM, ChSt
	CcCDPK17-2	KYP39277.1	547	61.00	5.17	257	215	4	MGNCCSQG	yes	ERM, PM, ChTM
	CcCDPK21	KYP58234.1	522	58.10	5.69	257	198	4	MGCHGSKE	Yes	PM, ChTM, MP
	CcCDPK29	KYP44787.1	522	59.63	6.25	257	195	4	MGHCFSKP	Yes	ERM, N, PM, MP
	CcCDPK8	KYP63277.1	524	59.64	6.36	258	170	4(1)	MGNCCARP	No	ERM, PM, N
III	CcCDPK10	KYP50696.1	544	61.38	6.15	258	186	4	MGNCNTCV	Yes	ERM, N, PM, ChTM
	CcCDPK13	KYP37486.1	490	55.57	5.52	258	140	4	MTVLAGVA¥	No	ERM, N, PM, MP
	CcCDPK24	KYP46292.1	548	62.20	6.08	258	189	4	MGSCVSSQ	Yes	ERM, PM
	CcCDPK30	KYP56058.1	535	60.70	6.18	258	177	4	MGNCCNGCM	Yes	N, PM
	CcCDPK32	KYP70536.1	536	60.51	6.14	258	181	4(1)	MGNCCATP	No	ERM, N, PM, MP
	CcCDPK28-1	KYP74047.1	535	60.41	9.15	260	201	4(2)	MGICHSAT	Yes	ChSt, N, ChTM, ChTS
IV	CcCDPK28-2	KYP60395.1	554	62.73	9.06	260	218	4	MGMCFSAT	Yes	ChSt, N, ChTM, ChTS

N-Nucleus, ERM-endoplasmic reticulum (membrane), MP- microbody (peroxisome), PM- plasma membrane, ChSt-chloroplast stroma, ChTM-chloroplast thylakoid membrane, ChTS- chloroplast thylakoid space, Cy- cytoplasm Column 7: Number of EF hand in CcCDPKs, Number in bracket indicates position of incomplete EF hand Column 8 ¥ indicate absence of N terminal Glycine required for Myristoylation

used to make unrooted phylogenetic tree using MEGA6. As shown in Fig. 1, the phylogenetic tree clearly formed typical four subgroups from twenty three CDPKs, group I-IV which are congruent with earlier studies from *Arabidopsis* (Cheng *et al.*, 2002). Among the four groups, group I was the largest group with eight CDPKs followed by group II (seven CDPKs), group III (six CDPKs), and group IV (two CDPKs) (Table 3). In *Arabidopsis*, the largest CDPK group was group II, followed by group I, III and IV (Cheng *et al.*, 2002). Clear orthologs of fifteen *Arabidopsis* CDPKs (AtCDPK5, 7, 14, 15, 16, 18, 19, 22, 23, 25, 26, 27, 31, 33 and 34) were not found in Pigeonpea. However, in four cases two Pigeonpea paralogs were identified for a single *Arabidopsis* CDPK (CcCDPK4-1, CcCDPK4-2 versus AtCPK4; CcCDPK9-1, CcCDPK9-2 versus AtCPK9; CcCDPK11-1, CcCDPK11-2 versus AtCPK11 and CcCDPK17-1, CcCDPK17-2 versus AtCPK17). As in case of *Arabidopsis* Mitogen Activated Protein Kinase (MAPK) family, paralogous relationship have not been reflected well in CPK nomenclature (Hamel *et al.*, 2006; Cheng *et al.*, 2002), however, the same could be captured in the present nomenclature of Pigeonpea CDPK.

CDPK proteins characteristically possessed four distinct domains namely; a variable amino-terminal domain, a protein

kinase domain, an autoinhibitory domain, and a calmodulin-like domain (Cheng *et al.*, 2002). The amino terminal domain of Pigeonpea CDPK was highly variable and varied from 7 (CcCDPK3) to 109 (CcCDPK1) amino acids long (Fig. 2). Sequence alignment of all twenty three CcCDPK amino terminal region (before kinase domain) showed least similarity compared to other stretch of proteins. The variable amino terminal domains shown to confer substrate specificity *in vivo* and direct subcellular localization of CDPK proteins (Asai *et al.*, 2013). Amino terminal domains of CDPKs also harbor potential sites for myristoylation (addition of a myristic acid) and palmitoylation (addition of palmitate). These lipid modifications are implicated to play a key role in binding of CDPK to membrane and stabilizing the interaction (Cheng *et al.*, 2002). Eleven of the CcCDPKs showed potential myristoylation site (Table 1), and also invariably had potential palmitoylation site, characterized by at least one cysteine residue at 3rd, 4th, or 5th position (Table 1) (Milligan *et al.*, 1995). Mutations in amino-terminal myristoylation and palmitoylation sites in the variable amino terminal domain of a potato CDPK, StCDPK5 compromised its localization to plasma membrane and activity towards StRBOHB (Respiratory Burst Oxidase Homolog) *in vivo* (Asai *et al.*, 2013).

Table 2 : Genetic organization of Pigeonpea CDPK gene family

Group	Name	Scaffold	Locus tag	Chromo-some	Name assigned in genome annotation	GC%	Intron No.	Putative Arabidopsis orthologs	
								Locus ID	AtCDPK names
I	CcCDPK1	KQ483476.1	KK1_028805	NA	CDPK1	43	6	AT5G04870.1	CPK1,ATCPK1
	CcCDPK2	CM003613.1	KK1_003226	11	CDPK2	42	6	AT3G10660.1	CPK2,ATCPK2
	CcCDPK4-1	CM003604.1	KK1_007415	2	CDPK4	44	6	AT4G09570.1	CPK4,ATCPK4
	CcCDPK4-2	CM003604.1	KK1_007416	2	CDPK4	44	6	AT4G09570.1	CPK4,ATCPK4
	CcCDPK6	KQ484078.1	KK1_041096	NA	CDPK4	44	6	AT2G17290.1	CPK6,ATCDPK3,ATCPK6
	CcCDPK11-1	CM003605.1	KK1_008691	3	CDPK SK5	48	6	AT1G35670.1	ATCDPK2, CPK11,ATCPK11
	CcCDPK11-2	KQ483486.1	KK1_029275	NA	CDPK SK5	45	6	AT1G35670.1	ATCDPK2, CPK11,ATCPK11
	CcCDPK20	KQ483633.1	KK1_034228	NA	CDPK20	48	6	AT2G38910.1	CPK20
	CcCDPK3	CM003612.1	KK1_014079	10	CDPK3	48	8	AT4G23650.1	CDPK6,CPK3
	CcCDPK9-1	CM003604.1	KK1_006033	2	CDPK9	44	7	AT3G20410.1	CPK9
II	CcCDPK9-2	KQ483427.1	KK1_025952	NA	CDPK9	41	7	AT3G20410.1	CPK9
	CcCDPK17-1	CM003604.1	KK1_005862	2	CDPK17	45	7	AT5G12180.1	CPK17
	CcCDPK17-2	KQ483921.1	KK1_039415	NA	CDPK17	46	7	AT5G12180.1	CPK17
	CcCDPK21	CM003613.1	KK1_004528	11	CDPK21	42	7	AT4G04720.1	CPK21
	CcCDPK29	KQ483613.1	KK1_033706	NA	CDPK29	43	7	AT1G76040.2	CPK29
	CcCDPK8	CM003609.1	KK1_017846	7	CDPK8	48	7	AT5G19450.2	CDPK19,CPK8
	CcCDPK10	KQ483450.1	KK1_027513	NA	CDPK30	50	6	AT1G18890.1	ATCDPK1,CPK10,CDPK1
III	CcCDPK13	KQ484099.1	KK1_041317	NA	CDPK13	50	6	AT3G51850.1	CPK13
	CcCDPK24	KQ483562.1	KK1_032091	NA	CDPK24	45	7	AT2G31500.1	CPK24
	CcCDPK30	CM003613.1	KK1_002286	11	CDPK30	50	5	AT1G74740.1	CPK30,CDPK1A,ATCPK30
	CcCDPK32	CM003605.1	KK1_009756	3	CDPK32	50	7	AT3G57530.1	CPK32,ATCPK32,CDPK32
	CcCDPK28-1	CM003604.1	KK1_006717	2	CDPK28	44	11	AT5G66210.1	CPK28
IV	CcCDPK28-2	CM003611.1	KK1_022799	9	CDPK28	45	11	AT5G66210.1	CPK28

The conserved catalytic domain of CcCDPKs were 257 to 260 amino acid long and showed all major eleven protein kinase subdomains (Fig. 3). The consensus Gly-X-Gly-X-X-Gly sequences were observed for all but one CcCDPK (CcCDPK4-2) where last Gly was replaced with Ala (Gly-Arg-Gly-His-Lue-Ala) (Fig. 3). In subdomain II, the invariable Lysine which was involved in phospho-transfer reaction was seen in all CcCDPKs. The regions of catalytic domain of CcCDPK from subdomain VI to IX were highly conserved. The invariant or nearly invariant amino acid from subdomain VI (Asp and Asn) and subdomain VII (Asp, Phe and Gly), which are associated with ATP binding (Hanks *et al.*, 1988), were also found conserved in all twenty three Pigeonpea CDPKs. The consensus three amino acid sequences (Ala, Pro, Glu) in subdomain VIII, known to be a crucial indicator of protein kinase domain and Asp-Gly were also found conserved in CcCDPK. In subdomain IX, the invariant Asp and Gly residues were found conserved with consensus Asp-Val/Ile/Leu-Trp-Ser/Thr-Ala/Val/Ile-Gly sequences in all Pigeonpea CDPKs except CcCDPK24 (Fig. 3). The basic amino acids rich stretch which join the catalytic and calmodulin-like domain in CDPKs are autoinhibitory domain and are implicated to have autoinhibitory activity as its acts as pseudo-substrate for catalytic domain (Harmon *et al.*, 1994; Cheng *et al.*,

2002). The CDPK activity depends largely on the interactions between the kinase, autoinhibitory and calmodulin-like domains. The autoinhibitory domain analysis showed that twelve CcCDPKs were found to have potential auto-phosphorylation site (Arg/Lys/His-X-X-Ser/Thr, where X indicates any amino acid) in the autoinhibitory domain (Fig. 4). Interestingly, only group I and group III CcCDPKs had potential auto-phosphorylation site in the autoinhibitory domain. All CcCDPKs in group I and group III with one exception each had the auto-phosphorylation consensus sequences. CcCDPK1 (group I) had Tyr residue in place of Ser/Thr (Lys- Gln- Phe-Tyr) and CcCDPK8 (group III) possesses Met residue in place of basic amino acid (Met-Gln-Phe-Ser). Whether, these substitutions hinder the potential auto-phosphorylation needs to be validated experimentally. In *Arabidopsis*, sixteen of the thirty four CDPKs show potential auto-phosphorylation site in the autoinhibitory domain (Harmon *et al.*, 1994; Cheng *et al.*, 2002).

The calmodulin-like domains of CcCDPK proteins invariably possessed four EF hands sequences (Table 1). In case of *Arabidopsis*, most of the *Arabidopsis* CDPKs contained four EF hands except a few CDPKs which had one, two or three EF hands (Cheng *et al.*, 2002).

To get the transcript level evidence of CcCDPKs, EST resources available in public domain were used. A local BLAST search was done against all the available Pigeonpea ESTs (25,578) using coding sequences of CcCDPKs as query, which resulted in only six significant EST hits for four CcCDPKs (Table 4). All the six ESTs were from cDNA libraries either from sterility

mosaic disease (SMD) challenged leaf tissue or *Fusarium* wilt challenged root tissue of two different germplasm lines (ICP2376 and ICPL201202) with different time kinetics (Raju *et al.*, 2010). These results indicate possible role of CcCDPKs in biotic stress and also indicate possible involvement of CcCDPK6, CcCDPK9, CDPK28-2 and CcCDPK32 in generating responses against

Fig. 4 : Sequence alignment of autoinhibitory regions of Pigeonpea CDPK. Putative auto-phosphorylation sites in CcCDPKs have been shown in boxes

Fig. 5 : Theoretical 3D models of Pigeonpea CDPK3 and CDPK6 build by homology modeling. Important amino acid residues have been indicated with single letter code and position in yellow or green color font. Four domains of protein kinases are indicated in white font, VD-Variable domain, PKD-Ser/Thr protein kinase domain, AD- Auto-inhibitory domain and CaMLD- Calmodulin like domain

pathogen infection. These findings are in agreement with the earlier reports of role of CDPKs in different biotic stress conditions (Dubiella *et al.*, 2013). A low number of significant EST hits of CcCDPK owes to less ESTs number of Pigeonpea ESTs available in public domain.

For expression analysis, the total reads (SRP002557, SRP002556) were mapped on coding sequences of Pigeonpea following the guidelines of CLC Genomics Workbench for RNA-seq analysis. Unique read counts were normalized by calculating the read per kilo-base per million (RPKM). The RPKM values of CcCDPKs were log₂ transformed and the heat map generated using R software showed expression level of CcCDPKs in Asha and UPAS120 (Fig. 7). As shown in the figure, sixteen and nineteen of 23 CDPKs showed clear expression in Asha and UPAS120, respectively. Four CDPKs, namely CcCDPK17-1, 17-2, 20, and 24 did not show any expression in UPAS120, where as in cultivar Asha, seven CcCDPKs (CcCDPK2, 4-2, 8, 17-1, 17-2,

Fig. 6 : Gene structure of Pigeonpea CDPKs. Information about intron-exons of CDPK genes is retrieved from genome annotation file (GFF file). Gene structure display server program is used to develop graphics for visualisation. Filled boxes indicate exons and lines joining two exons indicate introns

Fig. 7 : Heatmap showing expression patterns of CDPK genes in Pigeonpea variety Asha and Upas120. SRA datasets ((SRP002557, SRP002556) available in public domain have been used for expression study. RNAseq data analysis was done using CLC genomics workbench. The scale represents Log_2 of RPKM value. The maximum value is displayed as dark red and minimum value is displayed as light green. Gene IDs are given on right side

Table 3 : Group wise distribution of members of Pigeonpea CDPK gene family. Arabidopsis CDPKs with no clear ortholog in Pigeonpea have been indicated in bold letter

Group	Pigeonpea CDPKs classification		Arabidopsis CDPKs classification	
	Number of CDPKs	Name of CDPKs	Number of CDPKs	Name of CDPKs
I	8	CcCDPK1, 2, 4-1, 4-2, 6, 11-1, 11-2, 20	10	AtCDPK1, 2, 4, 5 , 6, 11, 12, 20, 25, 26
II	7	CcCDPK3, 9-1, 9-2, 17-1, 17-2, 21, 29	13	AtCDPK3, 9, 15 , 17, 19 , 21, 22, 23, 27 , 29, 31, 33, 34
III	6	CcCDPK8, 10, 13, 24, 30, 32	8	AtCDPK7, 8, 10, 13, 14 , 24, 30, 32
IV	2	CcCDPK28-1, 28-2	3	AtCDPK 16, 18 , 28

Table 4 : Pigeonpea CDPKs with EST hits

CDPK Name	EST Acc No.	Identity	E-value	Cultivar	Tissue	Library information
CcCDPK6	GR471088.1	99.52	0	ICP7035	Leaf	Sterility mosaic disease (SMD) challenged (60 DAS) pigeonpea genotype ICP7035 leaf cDNA library
CcCDPK28-2	GR468306.1	100	2E-150	ICP2376	Root	Fusarium wilt challenged (15 DAI) Pigeonpea genotype ICP2376 root cDNA library
	GR468318.1	100	2E-150	ICP2376	Root	Fusarium wilt challenged (15 DAI) pigeonpea genotype ICP2376 root cDNA library
	GR468294.1	98.13	1E-138	ICP2377	Root	Fusarium wilt challenged (15 DAI) pigeonpea genotype ICP2376 root cDNA library
CcCDPK32	GR464257.1	100	3E-90	ICPL20102	Root	Fusarium wilt challenged (6 DAI) pigeonpea genotype ICPL20102 root cDNA library
CcCDPK9-1	GR465487.1	100	8E-50	ICPL20102	Root	Fusarium wilt challenged (15 DAI) pigeonpea genotype ICPL20102 root cDNA library

20 and 24) did not show any expressions. These results showed expression of most of the identified CDPKs in Pigeonpea validating its biological role in plants. The Pigeonpea CDPKs which showed no expression in two cultivars could be due to limited expression or spatial temporal specificity or stresses. Different stress responsive expression of CDPKs in different tissue types have been demonstrated in other plants (Zuo *et al.*, 2013). The present study reports identification of 23 member CDPK gene family in Pigeonpea, which is considered as an orphan crop. The genome wide identification and characterization of CDPK gene family would help enhance understanding of CDPK mediated signaling in non-model crop plants in general and Pigeonpea in particular. The orthologous nomenclature would facilitate easy comparison between CDPKs of model and non model crop plants and help removing ambiguity in CDPK nomenclature in Pigeonpea. Further, the information generated in the present study is expected to form foundation for further comprehensive investigations to unravel role of CDPKs in various biotic and abiotic stresses, as well as in growth and development. The present study constitutes miniature step forward in unraveling CDPK signaling network in Pigeonpea.

Acknowledgments

Authors acknowledge use of High Performance Computational facility at ICAR-NBPGR. Dr. Soma S. Marla is

acknowledged for critically reading the manuscript. The facility provided and funding from ICAR is gratefully acknowledged

References

- Asai, S., T. Ichikawa, H. Nomura, M. Kobayashi, Y. Kamiyoshihara, H. Mori, Y. Kadota, C. Zipfel, J.D. Jones and H. Yoshioka: The variable domain of a plant calcium-dependent protein kinase (CDPK) confers subcellular localization and substrate recognition for NADPH oxidase. *J. Biol Chem.*, **288**, 14332-40 (2013).
- Bhattacharjee, S.: Involvement of calcium and calmodulin in oxidative and temperature stress of *Amaranthus lividus* L. during early germination. *J. Environ. Biol.*, **30**, 557-562 (2009).
- Bush, D.S.: Calcium regulation in plant cells and its role in signaling. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **46**, 95-122 (1995).
- Cheng, S.H., M.R. Willmann, H.C. Chen and J. Sheen: Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiol.*, **129**, 469-485 (2002).
- Coca, M. and B. Segundo: AtCPK1 calcium-dependent protein kinase mediates pathogen resistance in Arabidopsis. *Plant. J.*, **63**, 526-40 (2010).
- Corpet, F.: Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.*, **16**, 10881-10890 (1988).
- Dubiella, U., H. Seybold, G. Durian, E. Komander, R. Lassig, C.P. Witte, W.X. Schulze and T. Romeis: Calcium-dependent protein kinase/NADPH oxidase activation circuit is required for rapid defense signal propagation. *Proc. Natl. Acad. Sci. USA*, **110**, 8744-9 (2013).

- Dutta, S., G. Kumawat, B.P. Singh, D.K. Gupta, S. Singh, V. Dogra, K. Gaikwad, T.R. Sharma, R.S. Raje, T.K. Bandhopadhyay, S. Datta, M.N. Singh, F. Bashasab, P. Kulwal, K.B. Wanjari, R. Varshney, D.R. Cook and N.K. Singh: Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC. Plant Biol.*, **11**, 17 (2011).
- Fiser, A., R.K. Do and A. Sali.: Modeling of loops in protein structures. *Prot. Sci.*, **9**, 1753-1773 (2000).
- Gao, X., X. Chen, W. Lin, S. Chen and D. Lu: Bifurcation of *Arabidopsis* NLR immune signaling via Ca²⁺-Dependent Protein Kinases. *PLoS Pathol.*, **9**, e1003127 (2013).
- Hamel, L.P., J. Sheen and A. Séguin: Ancient signals: Comparative genomics of green plant CDPKs. *Trends Plant Sci.*, **19**, 79–89 (2014).
- Hamel, L.P., M.C. Nicole, S. Sritubtim, M.J. Morency, M. Ellis, J. Ehltling, N. Beaudoin, B. Barbazuk, D. Klessig, J. Lee, G. Martin, J. Mundy, Y. Ohashi, D. Scheel, J. Sheen, T. Xing, S. Zhang, A. Seguin and B.E. Ellis.: Ancient signals: Comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci.*, **11**, 192–198 (2006).
- Hanks, S.K., A.M. Quinn and T. Hunter: The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Sci.*, **241**, 42-52 (1988).
- Harmon, A.C., B.C. Yoo and C. McCaffery: Pseudosubstrate inhibition of CDPK, a protein kinase with a calmodulin-like domain. *Biochemistry*, **33**, 7278–7287 (1994).
- Hashimoto, K. and J. Kudla: Calcium decoding mechanisms in plants. *Biochimie.*, **93**, 2054-2059 (2011).
- Hu, B., J. Jin, A.Y. Guo, H. Zhang, J. Luo and G. Gao: GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*, **31**, 1296-1297 (2015).
- Liu, H., Z. Che, X. Zeng, X. Zhou, H.M. Siteo, H. Wang and D. Yu: Genome-wide analysis of calcium-dependent protein kinases and their expression patterns in response to herbivore and wounding stresses in soybean. *Funct. Integr. Geno.*, **6**, 481-493 (2016).
- Milligan, G., M. Parenti and A.I. Magee: The dynamic role of palmitoylation in signal transduction. *Trends. Biol. Sci.*, **20**, 181–186 (1995).
- Nakai, K. and P. Horton: PSORT: A program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends. Biochem. Sci.*, **24**, 34-35 (1999).
- Raju, N.L., B.N. Gnanesh, P. Lekha, B. Jayashree, S. Pande, P.J. Hiremath, M. Byregowda, N.K. Singh and R.K. Varshney: The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). *BMC. Plant Biol.*, **10**, 45 (2010).
- Rao, K.P., G. Vani, K. Kumar, D.P. Wankhede, M. Mishra, M. Gupta and A.K. Sinha.: Arsenic stress activates MAP kinase in rice roots and leaves. *Arch. Biochem. Biophys.*, **506**, 73–82 (2011).
- Ray, S., P. Agarwal, R. Arora, S. Kapoor and A.K. Tyagi.: 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. Geno.*, **278**, 493-505 (2007).
- Sanders, D., J. Pelloux, C. Brownlee and J.F. Harper: Calcium at the crossroads of signaling. *Plant Cell*, **14**, 401–17 (2002).
- Singh, N.K., D.K. Gupta, P.K. Jayaswal, A.K. Mahato, S. Dutta, S. Singh, S. Bhutani, V. Dogra, B.P. Singh, G. Kumawat, J.K. Pal, A. Pandit, A. Singh, H. Rawal, A. Kumar, G. Rama Prashat, A. Khare, R. Yadav, R.S. Raje, M.N. Singh, S. Datta, B. Fakrudin, K.B. Wanjari, R. Kansal, P.K. Dash, P.K. Jain, R. Bhattacharya, K. Gaikwad, T. Mohapatra, R. Srinivasan and T.R. Sharma.: The first draft of the pigeonpea genome sequence. *J. Plant Biochem. Biotechnol.*, **21**, 98-112 (2012).
- Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar: MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mole. Biol. Evol.*, **30**, 2725-2729 (2013).
- Varshney, R.K., W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T. Donoghue, S. Azam, G. Fan, A.M. Whaley, A.D. Farmer, J. Sheridan, A. Iwata, R. Tuteja, R.V. Penmetsa, W. Wu, H.D. Upadhyaya, S.P. Yang, T. Shah, K.B. Saxena, T. Michael, W.R. McCombie, B. Yang, G. Zhang, H. Yang, J. Wang, C. Spillane, D.R. Cook, G.D. May, X. Xu and S.A. Jackson.: Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.*, **30**, 83-9 (2011).
- Wankhede, D.P., M. Misra, P. Singh and A.K. Sinha.: Rice mitogen activated protein kinase kinase and mitogen activated protein kinase interaction network revealed by in-silico docking and yeast two-hybrid approaches. *PLoS. One*, **8**, e65011 (2013).
- Wernimont, A.K., M. Amani, W. Qiu, J.C. Pizarro, J.D. Artz, Y.H. Lin, J. Lew, A. Hutchinson and R. Hui.: Structures of parasitic CDPK domains point to a common mechanism of activation. *Proteins*, **79**, 803–820 (2011).
- Zhu, S.Y., X.C. Yu, X.J. Wang, R. Zhao, Y. Li, R.C. Fan, Y. Shang, S.Y. Du, X.F. Wang, F.Q. Wu, Y.H. Xu, X.Y. Zhang and D.P. Zhang.: Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in *Arabidopsis*. *Plant Cell*, **19**, 3019–36 (2007).
- Zuo, R., R.B. Hu, G.H. Chai, M.L. Xu, G. Qi, Y.Z. Kong and G. Zhou.: Genome-wide identification, classification and expression analysis of CDPK and its closely related gene families in poplar (*Populus trichocarpa*). *Mol. Biol. Rep.*, **40**, 2645–62 (2013).