Mol Biol Rep (2013) 40:2645–2662
DOI 10.1007/s11033-012-2351-z

Genome-wide identification, classification, and expression analysis of CDPK and its closely related gene families in poplar (*Populus trichocarpa*)

Ran Zuo · Ruibo Hu · Guohua Chai · Meiling Xu · Guang Qi · Yingzhen Kong · Gongke Zhou

Received: 24 August 2012 / Accepted: 9 December 2012 / Published online: 15 December 2012

Abstract Calcium-dependent protein kinases (CDPKs) are Ca^{2+}-binding proteins known to play crucial roles in Ca^{2+} signal transduction pathways which have been identified throughout plant kingdom and in certain types of protists. Genome-wide analysis of CDPKs have been carried out in *Arabidopsis*, rice and wheat, and quite a few of CDPKs were proved to play crucial roles in plant stress responsive signature pathways. In this study, a comprehensive analysis of *Populus* CDPK and its closely related gene families was performed, including phylogeny, chromosome locations, gene structures, and expression profiles. Thirty *Populus CDPK* genes and twenty closely related kinase genes were identified, which were phylogenetically clustered into eight distinct subfamilies and predominately distributed across fifteen linkage groups (LG). Genomic organization analyses indicated that purifying selection has played a pivotal role in the retention and maintenance of *Populus* CDPK gene family. Furthermore, microarray analysis showed that a number of *Populus CDPK* and its closely related genes differentially expressed across disparate tissues and under various stresses. The expression profiles of paralogous pairs were also investigated to reveal their evolution fates. In addition, quantitative real-time RT-PCR was performed on nine selected *CDPK* genes to confirm their responses to drought stress treatment. These observations may lay the foundation for future functional analysis of *Populus* CDPK and its closely related gene families to unravel their biological roles.

Keywords *Populus trichocarpa* · CDPK · Phylogenetic analysis · Gene structure · Expression analysis

Introduction

Plants have developed a sophisticated network of signaling pathway to survive the changeable environment. As a universal second messenger, Ca^{2+} plays an important role in the signal transduction pathways [1]. In plants, various stimuli cues, such as light, hormones, biotic and abiotic stresses, could elicit transition changes of intercellular Ca^{2+} concentration [2–6], and the Ca^{2+} signatures are sensed and decoded by different Ca^{2+} sensors which subsequently transduct them into a series of downstream effects such as phosphorylation of target proteins [7, 8].

Electronic supplementary material The online version of this article (doi:10.1007/s11033-012-2351-z) contains supplementary material, which is available to authorized users.

R. Zuo · R. Hu · G. Chai · M. Xu · G. Qi · G. Zhou (✉)
CAS Key Laboratory of Biofuels, Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of BioEnergy and BioProcess Technology, Chinese Academy of Sciences, Qingdao, Shandong, People's Republic of China
e-mail: zhougk@qibebt.ac.cn

R. Zuo
e-mail: zuoran@qibebt.ac.cn

R. Hu
e-mail: hurb@qibebt.ac.cn

G. Chai
e-mail: chaigh@qibebt.ac.cn

M. Xu
e-mail: meilingxu1@163.com

G. Qi
e-mail: qiguang@qibebt.ac.cn

Y. Kong (✉)
Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA
e-mail: yingzhenkong@hotmail.com

There are four classes of Ca^{2+} sensors/Ca^{2+}-binding proteins identified in plants, including calmodulins (CaM), calmodulin-like proteins (CaML), calcineurin B-like proteins (CBL) and calcium-dependent protein kinases (CDPK) [9–12]. CaM, CaML and CBL contain Ca^{2+}-binding domains but lack effector domains. As a result, they act only as Ca^{2+} sensors and transmit the Ca^{2+} signal by interacting with target proteins and regulating their activities [3]. In contrast, CDPK proteins contain a variable N-terminal domain and three functional domains, including a catalytic Ser/Thr protein kinase domain, an autoinhibitory domain which functions as a pseudo substrate and a CaM domain containing EF-hand motifs for Ca^{2+}-binding capacity [9, 13]. This unique structure enables the CDPK proteins to function as both Ca^{2+} sensors and effectors.

Calcium-dependent protein kinases are activated by a relieving autoinhibition mechanism [14]. As a result of Ca^{2+} binding to the CaM domain, the CDPK protein undergoes a conformational change that displaces the autoinhibitory domain from the kinase domain, which subsequently activates the enzyme. According to the recent research, all C-terminal regions to the catalytic domain work together for activation [15, 16].

Calcium-dependent protein kinases are the best characterized Ca^{2+} sensors in plants, which have been identified throughout the plant kingdom and in certain types of protists [17]. A number of studies from various plant species indicated that CDPKs participate in plant response to diverse stimulus, including light, hormones, mechanical wounding, abiotic stress and pathogen elicitors. In rice (*Oryza sativa*), transgenic plants overexpressing *OsCDPK7* showed enhanced tolerance to cold, salt and drought stresses [18]. *OsCDPK13*, *OsCPK12* and *OsCPK21* were also involved in responses to cold, low nitrogen and salt stress, respectively [19–21]. In *Arabidopsis*, CPK10 was reported to participate in ABA and Ca^{2+}-medicated stomatal regulation in response to drought stress [22]. Two homologs, AtCPK4 and AtCPK11, acted as positive regulators in ABA signaling pathways involved in seed germination, seedling growth, stomatal movement and salt stress tolerance [23], while AtCPK12 had been characterized as a negative ABA-signaling pathway regulator recently [24]. In addition, AtCPK6 and AtCPK23 were demonstrated to play crucial roles in responses to drought and salt stresses [25, 26]. Although all these studies indicate the involvement of CDPKs in plant responses to environmental stresses, biological functions of the majority CDPKs remain uncharacterized until now.

A genome-wide analysis of CDPKs has identified 34 *CDPK* genes in *Arabidopsis* [9, 13]. Similarly, 29 *CDPK* genes were initially revealed in rice genome [27] and two more members were annotated in a later release of rice genomic sequence [28]. In wheat (*Triticum aestivum* L.), 20 *CDPK* genes including 14 full-length cDNA sequences were comprehensively studied [29].

There are also four types of Ser/Thr protein kinases which are closely related to the CDPKs, namely, CDPK-related kinases (CRK), calcium and phosphoenolpyruvate carboxylase kinases (PPCKs), phosphoenolpyruvate carboxylase kinase-related kinases (PEPRKs) and calcium and calmodulin-dependent protein kinases (CCaMKs) [9]. Genome-wide analysis has identified eight CRKs, two PEPRKs and two PPCKs in *Arabidopsis* [9], and five CRKs, two PEPRKs and one CCaMKs in rice [27], respectively. Function analysis has been performed on a few CDPK closely related kinases. DMI3, a Medicago truncatula CCaMK, was demonstrated to be required for bacterial and fungal symbioses [30], and LeCRK1 was associated with the tomato fruit ripening process [31].

Compared to the extensive studies of *CDPK* genes in many other plant species, few research has been conducted in model tree species *Populus* so far. Considering the importance of *Populus* in wood production and environment protection as well as the emergence of *CDPK* genes as promising candidates for plant stress tolerance modification, it was of interest for us to characterize the CDPK gene family in *Populus*.

In this study, we performed a genome-wide analysis of CDPK gene family in *Populus*, which identified 30 *CDPK* genes and 20 CDPK closely related kinase genes. Furthermore, we examined nine *Populus CDPK* genes to confirm their inducible expression patterns under water-deficient conditions. Our results could provide a subset of potential candidate *CDPK* genes for future engineering modification of stress tolerance characteristics in *Populus*.

Materials and methods

Database search and sequence retrieval

Sequences of *Arabidopsis*, rice and wheat *CDPK* and its closely related genes were obtained from the *Arabidopsis* Information Resource (TAIR, http://www.Arabidopsis.org/, release 10.0), rice genome annotation database (http://rice.plantbiology.msu.edu/, release 5.0) or GenBank (http://www.ncbi.nlm.nih.gov/genbank/) respectively. Sequences of *Populus* were downloaded from Phytozome (http://www.phytozome.net/). Local blast was performed using *Arabidopsis* CDPK and its closely related kinase proteins as queries for the identification of genes from *Populus*. Manual reannotation was also performed using online web server FGENESH (http://linux1.softberry.com/berry.phtml). All putative candidates were manually verified with the InterProScan program (http://www.ebi.ac.uk/Tools/pfa/iprscan/) to confirm the presence of the protein kinase domain and the CaM domain.

Finally, all obtained protein sequences were further examined by the Hidden Markov Model of Pfam (http://pfam.sanger.ac.uk/search)/SMART (http://smart.embl-heidelberg.de/) tools. Predicted molecular masses were calculated using DNAMAN software. The EF-hand and N-myristoylation motifs were predicted by PROSITE (http://prosite.expasy.org/scanprosite/) [32] and the palmitoylation sites were predicted by CSS-Palm program [33]. Data of putative alternative splicing sites of *Populus* genes were acquired from Phytozome.

Phylogenetic analysis

Multiple alignments of amino acid sequences were performed by Clustal X (version 1.83) program [34]. The unrooted phylogenetic trees were constructed with MEGA5.0 [35] using the Neighbor-Joining (NJ) method and the bootstrap test carried out with 1,000 replicates.

Chromosomal location and gene duplication

Genes were mapped on chromosomes by identifying their chromosomal position provided in the Phytozome database. Identification of segmental duplications resulting from salicoid genome-wide duplications was accomplished based on duplication coordinates from the *Populus* genome assembly v2.1. Blocks in the same colors represent the homeologous chromosomal segments.

Amino acid sequences from segmentally duplicated pairs were aligned first by Clustal X v1.83 and the aligned sequences were subsequently transferred into original cDNA sequences using the PAL2NAL program (http://www.bork.embl.de/pal2nal/) [36],which uses the CODEML program of PAML [37] to estimate synonymous (Ks) and nonsynonymous (Ka) substitution rates. Divergence time (T) was calculated using a synonymous mutation rate of λ substitutions per synonymous site per year as $T = Ks/2\lambda$ ($\lambda = 9.1 \times 10^{-9}$ for *Populus*) [38].

Gene structure analysis

The exon/intron organization for individual gene was illustrated with Gene structure display server (GSDS) program (http://gsds.cbi.pku.edu.cn/) [39] by alignment of the cDNAs with their corresponding genomic DNA sequences from Phytozome (http://www.phytozome.net/poplar, release 2.1).

Microarray analysis

The microarray data for various tissues/organs and developmental stages available at NCBI Gene Expression Omnibus (GEO) database [40] under the series accession number GSE13990 were used for the tissue-specific expression analysis. The series GSE13990 includes Affymetrix microarray data from nine different tissue samples representing three biological replicates [41], The Affymetrix CEL files representing nine tissues/organs as well as photoperiodic treatments were downloaded from GEO database at NCBI and imported into GeneSpring GX (V11.5) software (Agilent Technologies) for further analysis. The data was normalized by the Gene Chip Robust Multiarray Analysis (GCRMA) algorithm followed by log transformation and average calculation. After normalization and log transformation of data for all the *Populus* genes present on the chip, the log signal intensity values for *Populus* probe IDs corresponding to CDPK and its closely related gene model (v1.1) were extracted as a subset for further analyses. The tab-delimited files for the average log signal intensity values were imported into Genesis program (v1.75) to generate heatmaps [42]. Hierarchical clustering was performed based on Pearson coefficients with average linkage rule.

For abiotic and hormone treatments, Affymetrix microarray data available at NCBI GEO database under the series accession numbers GSE17230 (drought stress on leaves), GSE17223 (drought stress on root tips), GSE9673 (fungal infection) and GSE17686 were analyzed [43, 44]. GSE17686 is composed of the following five subset series: GSE14893 (nitrogen limitation, genotype 1979), GSE14515 (nitrogen limitation, genotype 3200), GSE16783 (1 week after leaf wounding), GSE16785 (90 h after leaf wounding) and GSE16773 (methyl jasmonate-elicited suspension cell cultures). The Affymetrix CEL files representing different abiotic and hormone treatments were downloaded from GEO database at NCBI and preprocessed by using GeneSpring GX (V11.5) software (Agilent Technologies). The data was normalized by GCRMA algorithm followed by log transformation and average calculation. After normalization and log transformation of data for all the *Populus* genes present on the chip, the log signal intensity values for *Populus* probe IDs corresponding to *CDPK* and its closely related gene models (v1.1) were extracted as a subset for further analyses. Expression was indicated as fold change of experimental treatments relative to control samples. The tab-delimited files for the average log signal intensity values were imported into Genesis program (v1.75) to generate heatmaps [42]. Hierarchical clustering was performed based on Pearson coefficients with average linkage rule.

Probe sets corresponding to genes were identified using an online Probe Match tool available at POParray (http://aspendb.uga.edu/poparray). For probe sets matching several genes models, only those exhibited the highest hybridization signals consistently across multiple samples were considered. The list of probe sets corresponding to *Populus* CDPK and its closely related genes was provided in Table S1.

Plant material and growth conditions

Plant material was collected from clonally propagated one-year-old *Populus deltoides* grown in the growth camber under long-day conditions (16 h light/8 h dark) at 25–28 °C. Drought stress treatment was conducted following the previous method with minor modification [45]. Briefly, the intact root systems of plants were removed from the pots, washed gently with water to remove soil and then laid down on filter paper with 70–80 % humility at 25 °C under dime light. Two biological replicates were performed for each stress treatment. After exposure to stresses after 0, 1, 3, 6, 12, 24, 36 and 48 h, root tips from three different plants were harvested at various time points, flash frozen in liquid nitrogen, and stored at −80 °C for further analysis.

RNA isolation and qRT-PCR

Total RNA from root tips was isolated by CTAB method with minor modifications [46]. RNA integrity was verified by 2 % agar gel electrophoresis. Before cDNA synthesis, RNA was treated with RQ1 RNase-free DNase (Promega, Madison, WI, USA) according to the manufacturer's instructions to ensure no DNA contamination, and then the first-strand cDNA synthesis was carried out with

Table 1 CDPK gene family in *Populus*

Gene symbol	Gene model (V2.1)	Gene model (V1.1)	*Arabidopsis* orthologue locus	Score	*E*-value	No. of EF hands	N-Myrist	N-Palmit	M.W. (kDa)
PtCDPK1	POPTR_0006s21490.1	eugene3.00061131	At5g04870.1	831	0.0	4	Y	N	67.1
PtCDPK2	POPTR_0008s01530.1	estExt_Genewise1_v1.C_LG_VIII1176	At5g04870.1	954	0.0	4	N	Y	64.7
PtCDPK3	POPTR_0010s25090.1	estExt_fgenesh4_pg.C_LG_X2219	At5g04870.1	956	0.0	4	N	Y	65.1
PtCDPK4	POPTR_0016s06700.1	gw1.XVI.2117.1	At5g04870.1	821	0.0	4	Y	N	68.5
PtCDPK5	POPTR_0001s10070.1	estExt_fgenesh4_pm.C_LG_I0337	At4g23650.1	809	0.0	4	Y	Y	58.2
PtCDPK6	POPTR_0003s13380.1	N.A	At4g23650.1	806	0.0	4	Y	Y	59.0
PtCDPK7	POPTR_0019s11290.1	N.A	At4g09570.1	825	0.0	4	N	N	58.3
PtCDPK8	POPTR_0004s21710.1	grail3.0066015802	At2g17290.1	967	0.0	4	N	Y	62.5
PtCDPK9	POPTR_0009s16970.1	fgenesh4_pm.C_LG_IX000013	At2g17290.1	899	0.0	3	N	Y	60.0
PtCDPK10	POPTR_0001s26430.1	fgenesh4_pg.C_LG_I001840	At5g12480.1	903	0.0	4	N	Y	59.8
PtCDPK11	POPTR_0009s05740.1	estExt_Genewise1_v1.C_LG_IX3554	At5g12480.1	902	0.0	4	N	Y	60.0
PtCDPK12	POPTR_0013s11690.1	gw1.41.252.1	At1g35670.1	824	0.0	4	N	N	58.7
PtCDPK12	POPTR_0013s11690.2	gw1.41.252.1	At1g35670.1	758	0.0	3	N	N	53.2
PtCDPK13	POPTR_0019s00630.1	estExt_fgenesh4_pg.C_LG_XIX0080	At1g35670.1	777	0.0	4	N	N	56.6
PtCDPK14	POPTR_0016s12460.1, POPTR_0016s12450	fgenesh4_pg.C_LG_XVI001087	At3g51850.1	756	0.0	4	N	Y	59.7
PtCDPK15	POPTR_0006s10230.1	N.A	At3g51850.1	757	0.0	1	N	Y	48.4
PtCDPK16	POPTR_0006s21390.1	gw1.VI.1822.1	At2g38910.1	883	0.0	4	N	Y	66.5
PtCDPK17	POPTR_0016s06590.1 POPTR_0016s06580	gw1.XVI.2082.1	At2g38910.1	520	e-147	4	N	Y	66.4
PtCDPK18	POPTR_0004s01530.1	N.A	At4g04720.1	799	0.0	4	Y	Y	60.1
PtCDPK19	POPTR_0021s00750.1	N.A	At4g04720.1	817	0.0	4	Y	Y	59.9
PtCDPK20	POPTR_0007s02120.1	N.A	At2g31500.1	739	0.0	4	Y	Y	60.7
PtCDPK21	POPTR_0005s11560.1	fgenesh4_pg.C_scaffold_57000083	At5g66210.2	875	0.0	4	Y	Y	63.1
PtCDPK22	POPTR_0007s09580.1	estExt_fgenesh4_pg.C_LG_III0688	At5g66210.2	875	0.0	4	Y	Y	63.2
PtCDPK23	POPTR_0002s01850.1	N.A	At1g76040.2	754	0.0	4	Y	Y	60.8
PtCDPK24	POPTR_0005s26640.1	N.A	At1g76040.2	723	0.0	4	N	Y	57.6
PtCDPK25	POPTR_0012s07360.1	eugene3.00120658	At1g74740.1	931	0.0	4	Y	Y	63.1
PtCDPK26	POPTR_0015s07740.1	fgenesh4_pg.C_LG_XV000415	At1g74740.1	907	0.0	4	Y	Y	64.0
PtCDPK27	POPTR_0006s05140.1	estExt_fgenesh4_pm.C_LG_VI0180	At3g57530.1	863	0.0	4	N	Y	59.8
PtCDPK28	POPTR_0016s05490.1	estExt_fgenesh4_pm.C_LG_XVI0164	At3g57530.1	887	0.0	4	N	Y	60.4
PtCDPK29	POPTR_0001s28150.1	fgenesh4_pm.C_LG_I000794	At5g19360.1	894	0.0	4	Y	Y	56.5
PtCDPK30	POPTR_0009s07330.1	fgenesh4_pg.C_LG_IX000914	At5g19360.1	908	0.0	4	Y	Y	58.6

approximately 2 mg RNA using the Revert Aid First Strand cDNA Synthesis Kit (MBI, Fermentas) and oligo-dT primers according to the manufacturer's procedure. Primers were designed using Beacon Designer v7.0 (Premier Biosoft International, California, USA) with melting temperatures 58–60 °C, primer lengths 20–24 bp and amplicon lengths 50–200 bp. All the primer sequences were listed in Table S2. qRT-PCR was conducted on LightCycler® 480 Detection System (Roche, Penzberg, Germany) using SYBR Premix Ex Taq (TaKaRa, Toyoto, Japan). Reactions were prepared in a total volume of 20 µl containing: 10 µl of 2xSYBR Premix, 2 µl of cDNA template, 0.4 µl of each specific primer to a final concentration of 200 nM. The reactions were performed as the following conditions: initial denaturation step of 95 °C for 10 s followed by two-step thermal cycling profile of denaturation at 95 °C for 5 s, and combined primer annealing/extension at 60 °C for 1 min for 40 cycles. No-template controls were included for each primer pair and each PCR reaction was performed in triplicate. To verify the specificity of the amplicon for each primer pair, a melting curve analysis was performed ranging from 60 to 95 °C with temperature increasing steps of 0.06 °C/s (five acquisitions per °C) at the end of each run. Baseline and threshold cycles (Ct) were automatically determined using the LightCycler 480 Software (release 1.5.0). Relative expression was calculated as described previously [47] using *UBQ10* as reference gene.

Results and discussion

Identification of CDPK gene family in *Populus*

To identify *CDPK* genes in *Populus*, *Populus* genome database (release 2.1, http://www.phytozome.net/poplar.php) was searched using known CDPK proteins as query sequence. Initially, a total of 46 non-redundant putative *CDPK* genes were isolated. After manual reannotation and confirmation of the protein kinase domain and CaM domain, 30 *Populus CDPK* genes were identified and designated as *PtCDPK1-PtCDPK30* following the nomenclature proposed in the previous study [48].

As comparative genomic study revealed a ratio of 1.4–1.6 putative poplar homologs for each *Arabidopsis* gene [49], it was hypothesized that *CDPK* genes in *Populus* would be a large multi-gene family as 34 *CDPK* genes were identified in *Arabidopsis* [9, 13]. However, according to the present study, the number of *Populus CDPK* genes was even smaller than that of *Arabidopsis*, which was also the case for rice and wheat CDPK gene families. What's more, only less than half of the *Arabidopsis CDPK* genes

Table 2 CDPK closely related gene families in *Populus*

Gene symbol	Gene model (V2.1)	Gene model (V1.1)	*Arabidopsis* orthologue locus	Score	*E*-value	No. of EF hands	N-Myrist	N-Palmit	M.W. (kDa)
PtCRK1	POPTR_0006s03890.1	estExt_fgenesh4_pg.C_LG_VI0324	At2g41140.1	872	0.0		Y	Y	63.2
PtCRK2	POPTR_0016s03470.1	estExt_Genewise1_v1.C_LG_XVI1225	At2g41140.1	934	0.0		Y	Y	64.7
PtCRK3	POPTR_0009s10880.1	fgenesh4_pm.C_LG_IX000297	At3g19100.1	804	0.0		Y	Y	66.5
PtCRK4	POPTR_0002s17720.1	fgenesh4_pm.C_LG_II000819	At2g46700.1	693	0.0		Y	Y	67.6
PtCRK5	POPTR_0014s09920.1	eugene3.00140489	At2g46700.1	689	0.0		Y	Y	67.6
PtCRK6	POPTR_0012s03040.1	fgenesh4_pg.C_LG_XII000275	At5g24430.1	765	0.0		N	N	59.7
PtCRK7	POPTR_0005s14250.1	fgenesh4_pg.C_LG_V000496	At3g50530.2	898	0.0		Y	Y	67.6
PtCRK8	POPTR_0007s11320.1	fgenesh4_pm.C_LG_VII000349	At3g50530.2	899	0.0		Y	Y	67.5
PtCRK9	POPTR_0004s15110.1	eugene3.01300040	At1g49580.1	863	0.0		Y	Y	67.3
PtPPCK1	POPTR_0019s03480.1	eugene3.00190185	At1g08650.1	353	9e-098		N	Y	31.1
PtPPCK2	POPTR_0013s04320.1	fgenesh4_pg.C_LG_XIII000226	At1g08650.1	338	4e-093		N	Y	30.9
PtPPCK3	POPTR_0010s08220.1	gw1.273.33.1	At1g08650.1	305	4e-083		N	N	30.7
PtPEPRK1	POPTR_0003s12080.1	N.A	At1g12580.1	600	e-171		N	N	57.3
PtPEPRK1	POPTR_0003s12080.2	N.A	At1g12580.1	600	e-171		N	N	52.6
PtPEPRK2	POPTR_0012s13820.1	estExt_Genewise1_v1.C_LG_XII0412	At1g12580.1	583	e-166		N	N	57.6
PtPEPRK3	POPTR_0015s13790.1	N.A	At1g12580.1	597	e-171		N	N	57.8
PtPEPRK3	POPTR_0015s13790.2	N.A	At1g12580.1	597	e-171		N	N	57.8
PtPEPRK4	POPTR_0005s27290.1	N.A	At1g12680.1	653	0.0		N	N	51.2
PtPEPRK5	POPTR_0007s03570.1	gw1.VII.1272.1	At1g12680.1	489	e-138		N	N	51.9
PtPEPRK6	POPTR_0017s07570.1	gw1.XVII.263.1	At1g12680.1	489	e-138		N	N	51.7
PtCCaMK1	POPTR_0008s01210.1	N.A		528	e-150	2	N	N	43.7
PtCCaMK2	POPTR_0010s25360.1	N.A		422	e-118	1	N	N	44.6

had their *Populus* homologs. In contrast, generally two or more *Populus* homologs were found for each *Arabidopsis CDPK* gene (Table 1). It can be hypothesized that approximately 30 *CDPK* genes would be sufficient for plants to mediate Ca^{2+} signals [29] and a subset of *CDPK* genes might have lost during the evolutionary process due to the functional redundancy.

The *Populus* CDPKs identified in our study ranged in molecular masses from 48.4 to 68.5 kDa and all of them possessed the typical CDPK structure, including a N-variable domain, a protein kinase domain, an autoinhibitory domain and a CaM-like domain. In addition, all *Populus* CDPKs, except for PtCDPK9, PtCDPK12 and PtCDPK15, were predicted to have four EF-hand motifs in the CaM-like domain. However, both PtCDPK9 and one alternative splicing of PtCDPK12 contained three EF-hand motifs and PtCDPK15 had only one (Table 1). The EF-hand motifs differed in the Ca^{2+}-binding affinities and subsequently in the contribution to Ca^{2+}-regulated kinase activities, among which the N-terminal EF1- and EF2- motifs with lower Ca^{2+}-binding affinities played a more important role in activating the kinases. Besides, the Ca^{2+}-binding affinities could also participate in the CDPK targeting process by affecting the access of substrate proteins [50, 51].

The N-terminus of a subset of CDPK proteins contained myristoylation motif, which was supposed to promote protein-membrane and protein–protein interactions [52]. Studies in *Arabidopsis* and wheat have revealed that proteins possessed myristoylation motifs tend to localized in plasma membrane [29, 53]. In addition, palmitoylation, as a second lipid modification, could assist in stabilizing the membrane interaction [54] In our study, fourteen of the *Populus* CDPKs were predicted to possess myristoylation motifs at the N-terminus, and all of them, except for PtCDPK1 and PtCDPK4, possessed at least one palmitoylation site (Table 1). However, their subcellular localization needed to be characterized experimentally.

Protein kinases closely related to the CDPKs in *Populus*

Previous studies revealed that CRKs, PEPRKs, PPCKs and CCaMKs were protein kinases closely related to CDPKs [9, 27]. Using a similar method applied in the identification

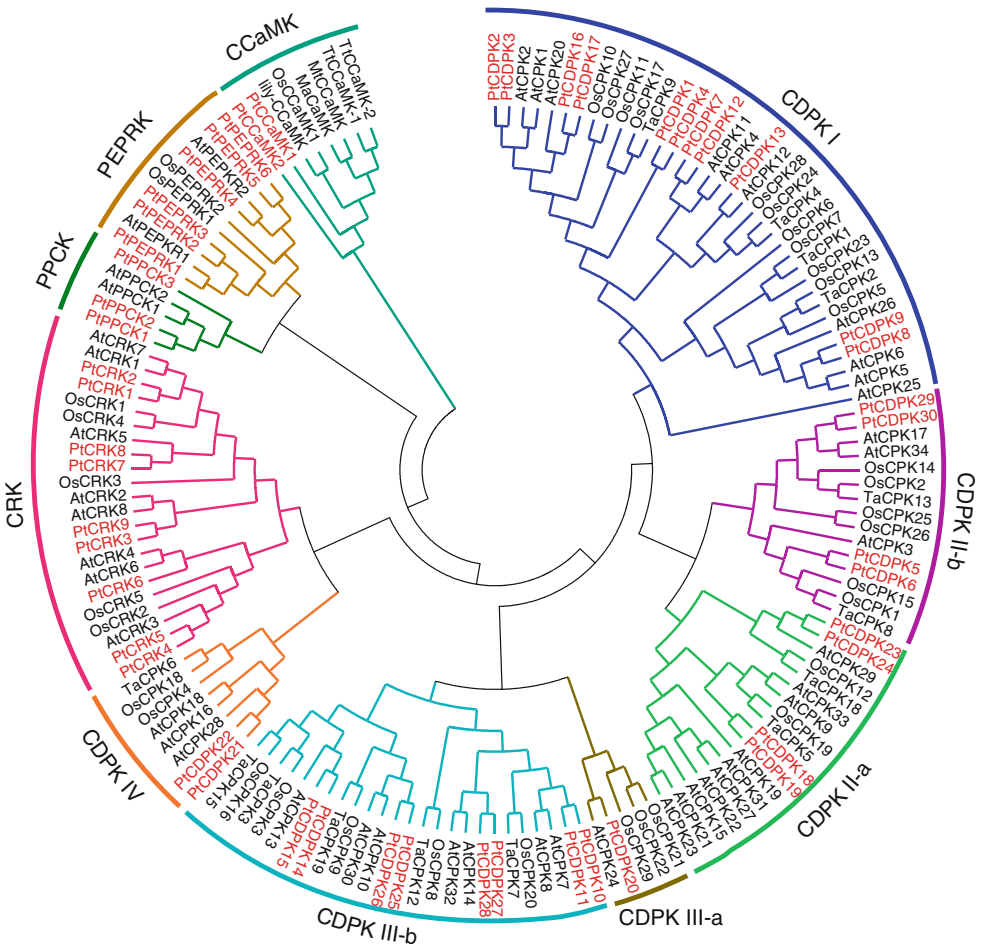


Fig. 1 Phylogeny and distribution of CDPK and its closely related proteins from four plant species. The deduced full-length amino acid sequences were aligned by Clustal X 1.83 and the phylogenetic tree was constructed using MEGA5.0 by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. Each family or subfamily was indicated in a *specific color*. Members from *Populus* were denoted in *red*. (Color figure online)

of *CDPK* genes, we identified nine CRKs, six PEPRKs, three PPCKs and two CCaMKs in *Populus*.

The CRKs shared most similar structures with the CDPKs except for the degenerated calmodulin-like domains [55]. Thus, compared to the CDPKs, CRKs might be consistently active, which was supported by certain biochemical studies [9, 56]. The *Populus* CRKs ranged in molecular masses from 59.7 to 67.6 kDa (Table 2), which were comparable with CRKs from other plant species [9, 27]. Moreover, in consistent with previous findings that most CRKs were N-terminus modified [9, 27], all *Populus* CRKs except PtCRK6 possessed both potential myristoylation and palmitoylation sites at their N-terminus (Table 2). Unlike CRKs, PPCKs are calcium-independent protein kinases containing only catalytic domains, and hence they are among the smallest ATP-dependent proteins and the enzyme activities appear to be consistent [9]. The molecular masses of *Populus* PPCKs ranged from 30.7 to 31.1 kDa (Table 2), which were smaller than the *Arabidopsis* PPCKs' [9]. The catalytic domains of PEPRKs are most closely related to PPCKs and usually possess both N-terminal and C-terminal extensions. With no apparent regulatory domain existed, PEPRKs are also supposed to be consistently active [9]. The PEPRKs identified in our study ranged in molecular masses from 51.2 to 57.8 kDa (Table 2), specifically, alternative splicing sites were found at both C-terminal extension of *PtPEPRK1* and 5′ untranslated region of *PtPEPRK3*. Although not found in *Arabidopsis* [9], CCaMKs have been identified in several plant species including tobacco, lily, legume and rice [27, 30, 57, 58]. The structure of CCaMKs is similar with the CDPKs except for the Ca^{2+}-binding domain, which usually contains three EF-hand motifs and is more similar with visinin rather than calmodulin [9]. Previous research showed that they could be regulated by both Ca^{2+} and calmodulin [59]. In the present study, we identified two putative CCaMKs in *Populus*. Interestingly, both of them possessed truncated C-terminus which contained either one or two EF-hand motifs (Table 2). As a result, compared to

Table 3 The distribution of CDPKs in *Arabidopsis*, *Populus*, rice and wheat

	Arabidopsis thaliana	*Populus trichocarpa*	*Oryza sativa*	*Triticum aestivum*
Group I	10	11	11	4
Group II-a	10	4	2	2
Group II-b	3	4	6	2
Group III-a	1	1	3	0
Group III-b	7	8	5	5
Group IV	3	2	2	1

Table 4 The Ka/Ks ratios and estimated divergence time for paralogous PtCDPK and its closely related proteins

	Paralogous pairs	Identities (%)	Ka	Ks	Ka/Ks	Duplication date (MY)
PtCDPK	PtCDPK1/4	85	0.0646	0.256	0.2524	14.07
	PtCDPK2/3	93	0.0344	0.2959	0.1164	16.26
	PtCDPK5/6	93	0.0239	0.1908	0.1254	10.48
	PtCDPK7/12	91	0.0392	0.2082	0.1884	11.44
	PtCDPK8/9	89	0.0312	0.192	0.1625	10.55
	PtCDPK10/11	93	0.0317	0.2055	0.1543	11.29
	PtCDPK14/15	74	0.0584	0.2842	0.2054	15.62
	PtCDPK16/17	90	0.0454	0.2178	0.2084	11.97
	PtCDPK18/19	88	0.0557	0.2883	0.1934	15.84
	PtCDPK21/22	93	0.0305	0.1859	0.1642	10.21
	PtCDPK23/24	81	0.0792	0.4042	0.1959	22.21
	PtCDPK25/26	93	0.0277	0.1816	0.1525	9.98
	PtCDPK27/28	91	0.0413	0.18	0.2293	9.89
	PtCDPK29/30	92	0.0186	0.147	0.1262	8.08
PtCRK	PtCRK1/2	90	0.0345	0.239	0.1441	13.13
	PtCRK3/9	86	0.0698	0.2274	0.3067	12.49
	PtCRK4/5	91	0.0402	0.2644	0.152	14.53
	PtCRK7/8	90	0.0514	0.2401	0.2142	13.19
PtPPCK	PtPPCK1/2	85	0.0636	0.3278	0.194	18.01
PtPEPRK	PtPEPRK2/3	89	0.0543	0.2299	0.2362	12.63
	PtPEPRK5/6	90	0.0485	0.2367	0.2049	13.01
PtCCaMK	PtCCaMK1/2	93	0.0996	0.2979	0.3344	16.37

the molecular masses of CCaMKs from other plant species (around 57 kDa) [27, 30, 57, 58], the molecular masses of PtCCaMKs (43.7 and 44.6 kDa) was much smaller (Table 2).

Phylogenetic analysis of CDPK and its closely related gene families

To examine the phylogenetic relationship among the CDPKs and its closely related protein kinases in *Populus*, *Arabidopsis*, rice and wheat, an unrooted tree was constructed from alignments of the full-length kinase sequences using method by MEGA5.0, and bootstrap value was tested with 1,000 replicates for statistical reliability. As shown in Fig. 1, the phylogenetic tree formed eight groups: CDPK I–IV, CRKs, PPCKs, PEPRKs and CCaMKs, both CDPKII and CDPKIII could be further divided into two subgroups. The distribution of CDPKs from the four species was shown in Table 3.

In the CDPK family, *CDPK I* genes consisted of the largest subfamilies in *Populus* and rice. However, in *Arabidopsis*, the number of *CDPK I* genes was smaller than that of *CDPK II* genes, which could mainly due to the expansion of one *Arabidopsis*-specific branch in subgroup CDPK II-a. In group CDPK III and CDPK IV individually, the number of genes were approximately identical across different species investigated, among which, the CDPK IV comprised the fewest members of the four groups (Table 3). The *CDPK* related kinase genes were comprehensively investigated across three species (*Arabidopsis,*

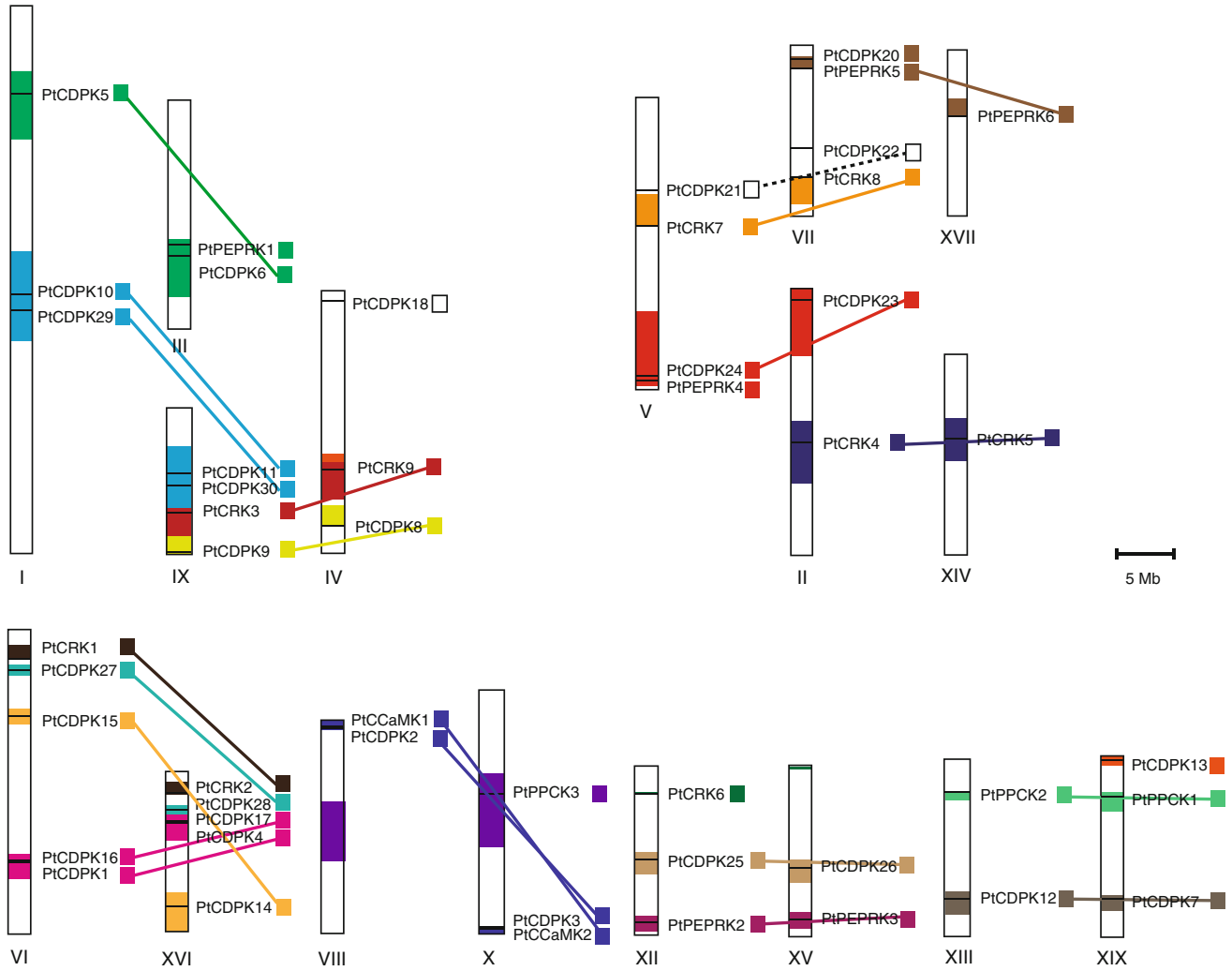


Fig. 2 Chromosomal locations and segmental duplication events of *Populus CDPK* and its closely related genes. The schematic diagram of genome-wide chromosome organization arisen from the salicoid genome duplication event in *Populus* was accomplished based on duplication coordinates from the *Populus* genome assembly v2.1. Segmental duplicated blocks are indicated with the *same colors*. The duplicated paralogous pairs of *PtCDPK* and its closely related genes are connected with *solid lines*. Genes located outside the duplicated blocks are connected by *dashed lines*. *Scale* represents a 5 Mb chromosomal distance

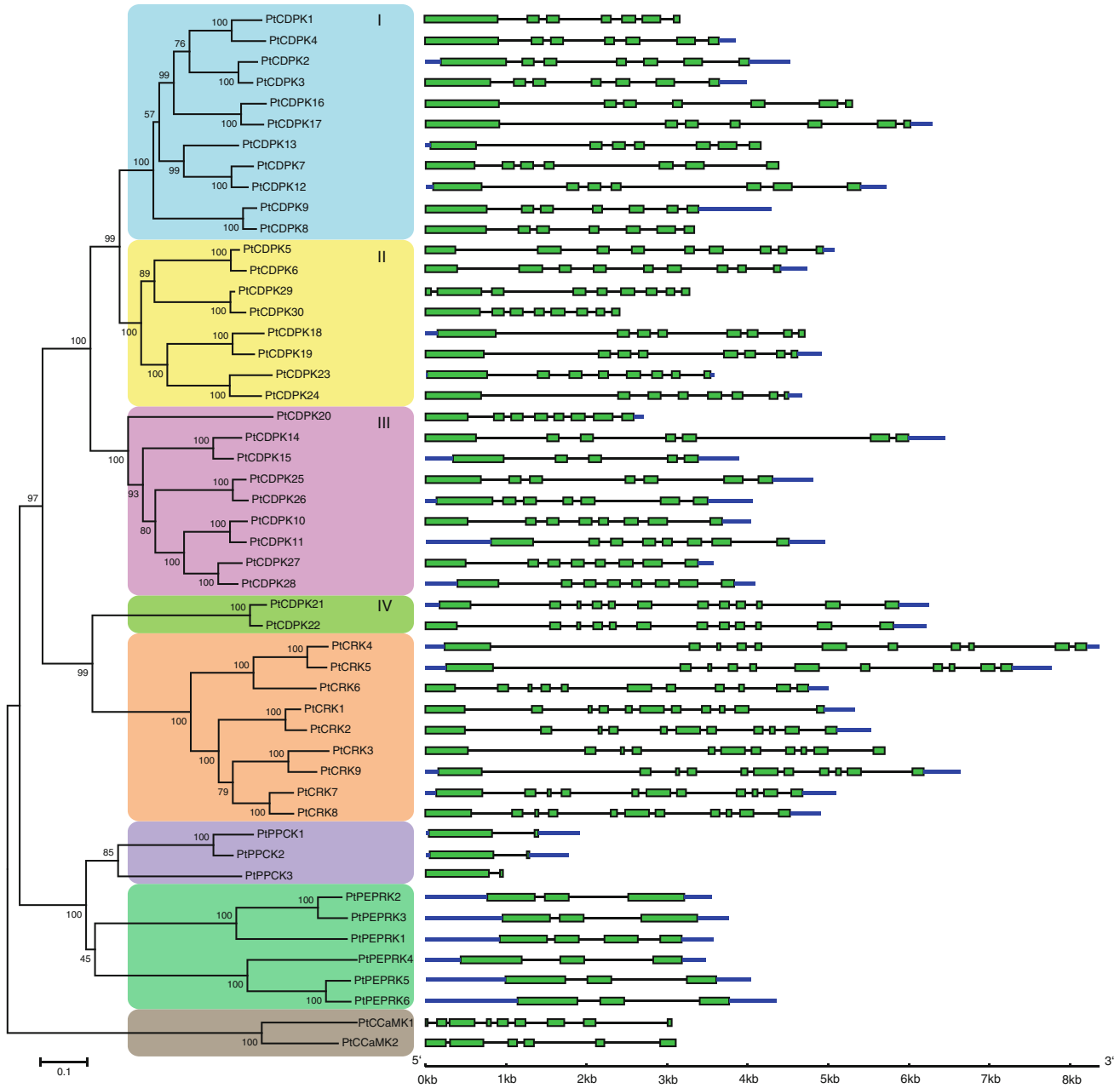


Fig. 3 Phylogenetic relationship and gene structure of *Populus CDPK* and its closely related genes. The phylogenetic tree was constructed using full-length protein sequences by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. The four PtCDPK phylogenetic subfamilies designated as I–IV as well as other four closely related families are marked with *different color backgrounds*. Exons and introns are represented by *green boxes* and *black lines*, respectively. The sizes of exons and introns are proportional to their sequence lengths. (Color figure online)

Populus and rice) except for the PPCKs from rice. Of which, the *CCaMK* members appeared to be lost during the evolution process of *Arabidopsis*, while the *PEPRKs* members from *Populus* were significantly expanded. In the other two *CDPK* related groups (*CRKs* and *PEPRKs*), the amount of genes from three species was roughly comparable. In general, as the CDPKs and its closely related kinases exhibited an interspersed distribution in each subgroup, it could be inferred that an ancestral set of the *CDPK* and its closely related kinase genes already existed before the monocot-eudicot divergence.

In addition, inspection of the phylogenetic tree revealed 14 *PtCDPK* paralogous gene pairs, four *PtCRK* paralogous gene pairs, one *PtPPCK* paralogous gene pair, two *PtPEPRK*

paralogous gene pairs and one *PtCCaMK* paralogous gene pair, with sequence identities ranging from 74.24 to 93.32 % (Table 4). Notably, the proportion (93.3 %) of paralogous gene pairs in *Populus* CDPK gene family was quite high, which was similar to the *Populus* HD-ZIP gene family [60], but much higher than NAC (60.1 %) [61] and GST (69.1 %) gene families [62]. This phenomenon might suggest that the CDPK gene family in *Populus* may have undergone multiple duplications during the evolution history.

Chromosomal location and gene duplication

Mapping of the CDPK gene location revealed that 29 of the 30 *CDPK* genes were distributed among 15 of the 19 *Populus* linkage groups (LG), while only one gene *PtCDPK19* was remained on yet unmapped scaffold. The distributed of the *PtCDPK* genes across the LGs appeared to be uneven: both LGVI and LGXVI encompassed the largest number of four *PtCDPK* genes, in contrast, no *PtCDPK* genes was found on LGXI, LGXIV, LGXVII and LGXVIII (Fig. 2).

Previous studies revealed that *Populus* genome had undergone at least three rounds of genome-wide duplications followed by multiple segmental duplication, tandem duplication, and transposition events such as retroposition and replicative transpositions [49]. Among them, the segmental duplication associated with the salicoid duplication event that occurred 65 million years ago remarkably contributed to the expansion of many multi-gene families [41, 60, 61, 63–66]. To determine the possible evolutionary relationship between the *PtCDPK* genes and potential segmental duplications, *Populus CDPK* along with the closely related kinase genes were mapped to the 163 recently identified duplicated blocks. The distribution of

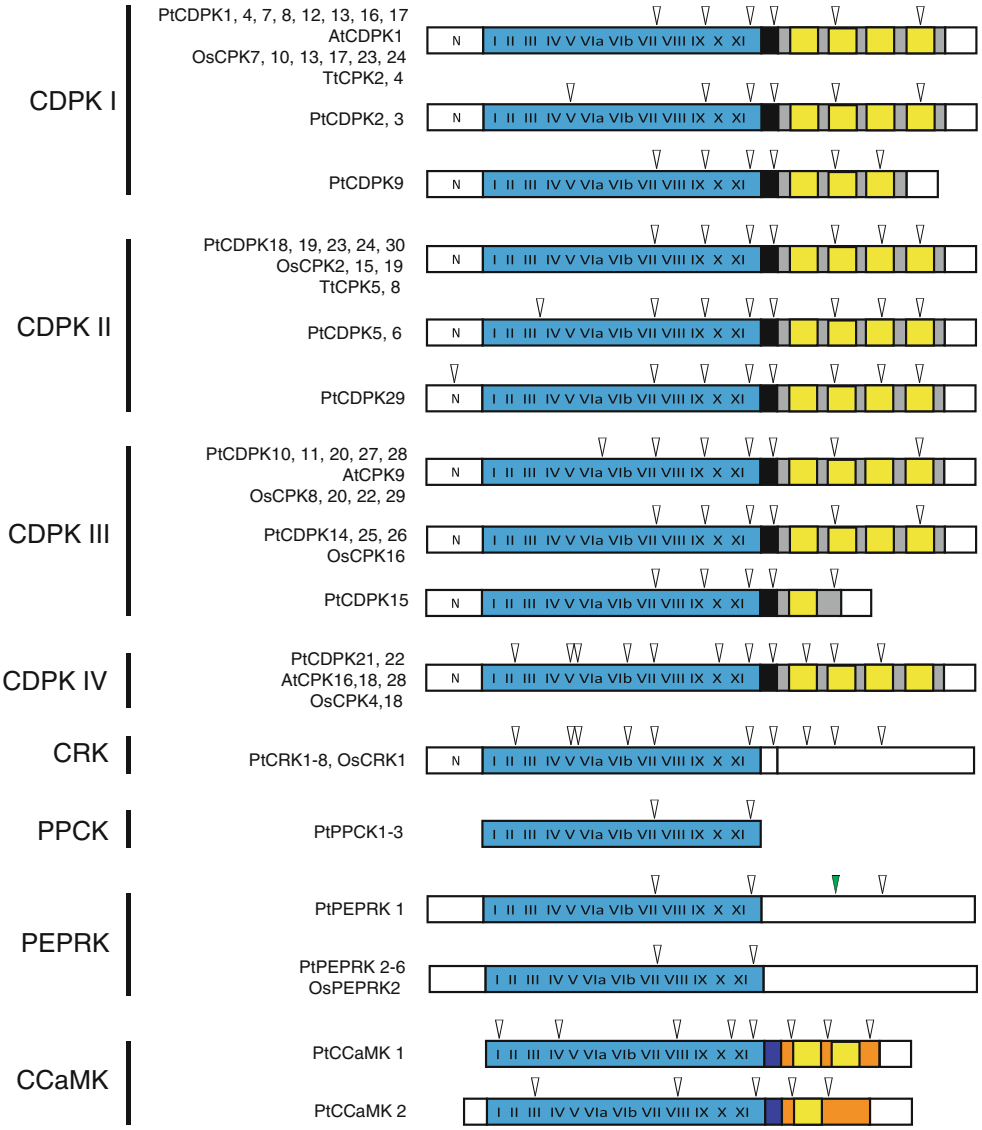


Fig. 4 Splicing sites of the *Populus CDPK* and its closely related genes. Introns locations relative the protein sequences are shown by *triangles*. A *green triangle* indicates the alternative splicing site in PtPEPRK1. *Blue boxes* protein kinase domains, *black boxes* autoinhibitory domains, *gray boxes* calmodulin-like domains, *yellow boxes* EF hand motifs, *orange box* visinin-like domain, *dark-blue boxes* autoinhibitory domains overlap with calmodulin-binding domains, *N* variable N-terminal domain, *I–XI* subdomains of the protein kinase domain. (Color figure online)

the genes relative to the corresponding duplicated blocks was illustrated in Fig. 2.

Of the 29 mapped *PtCDPK* genes, 90 % (26/29) were located within the duplicated regions while only *PtCDPK18*, *PtCDPK21* and *PtCDPK22* were located outside of any duplicated blocks. Within the identified duplicated blocks associated with the recent salicoid duplication event, 80 % (24/30) of the *Populus CDPK* genes were preferentially retained duplicates that located in both duplicated regions. In contrast, two duplicated blocks only contained *PtCDPK* genes (*PtCDPK13* and *PtCDPK20*) on one of the blocks and lacked duplicates on the corresponding block, suggesting that dynamic rearrangement might have occurred following the segmental duplication which led to the loss of a subset of genes.

PtCDPK1/4 and *PtCDPK16/17* were two of the paralogous gene pairs identified in the present study. Interestingly, *PtCDPK1* and *PtCDPK16*, together with *PtCDPK4* and *PtCDPK17*, were found within a distance of less than 9 kb on the duplication blocks with both identities higher than 60 % which could also be considered as tandem duplications. This arrangement, along with the closely related location in the phylogenetic tree, suggested that the two paralogous gene pairs (*PtCDPK1/4* and *PtCDPK16/17*) might originate from a common ancestor which firstly underwent tandem duplication prior to the segmental duplication.

Based on the genomic organization of *PtCDPK* genes, we may conclude that segmental duplications significantly contributed to the expansion of *Populus CDPK* gene family, which were also observed in other multi-gene families in *Populus* [41, 60, 61, 63–67]. In addition, our results indicated that *Populus CDPK* genes had been preferentially retained at a relatively high rate of 80 %, which was much higher than the average rate (33 %) following the salicoid genome-wide duplication in the *Populus* lineage [49]. The high retention rates of duplicated genes were also observed in other gene families of *Populus* [63, 65, 67]. These findings also corroborated previous results that genes involved in transcription regulations and signal transductions were preferentially retained following duplications [68–70].

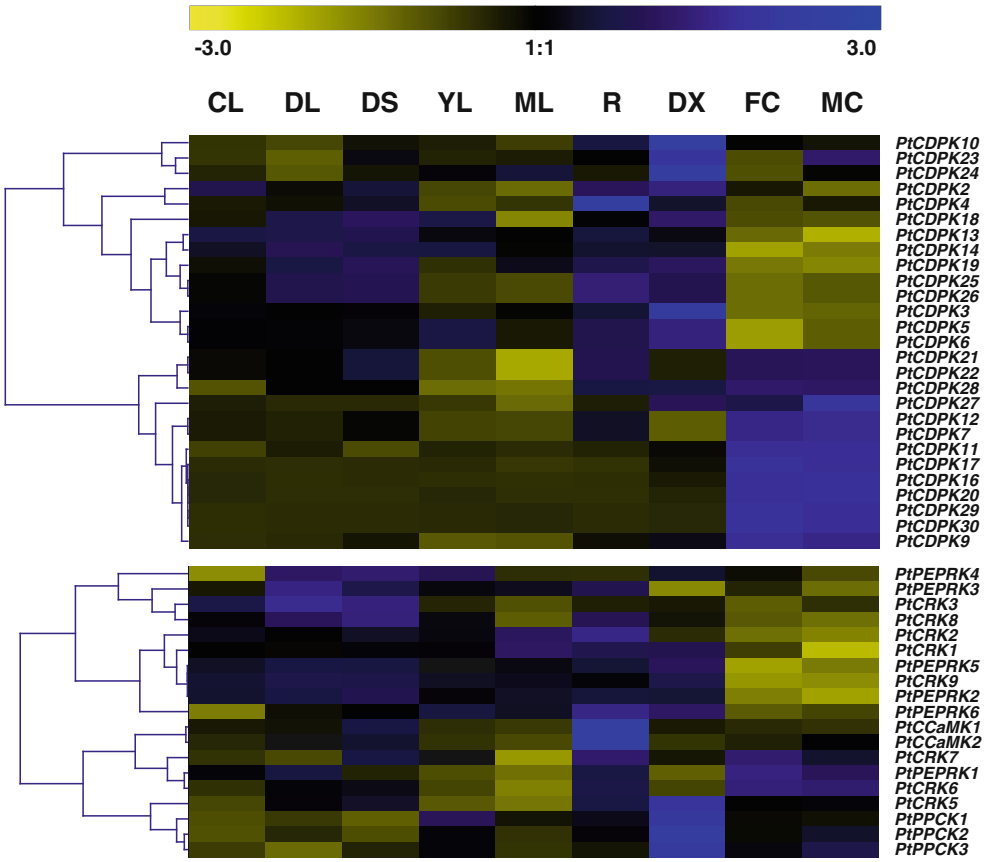
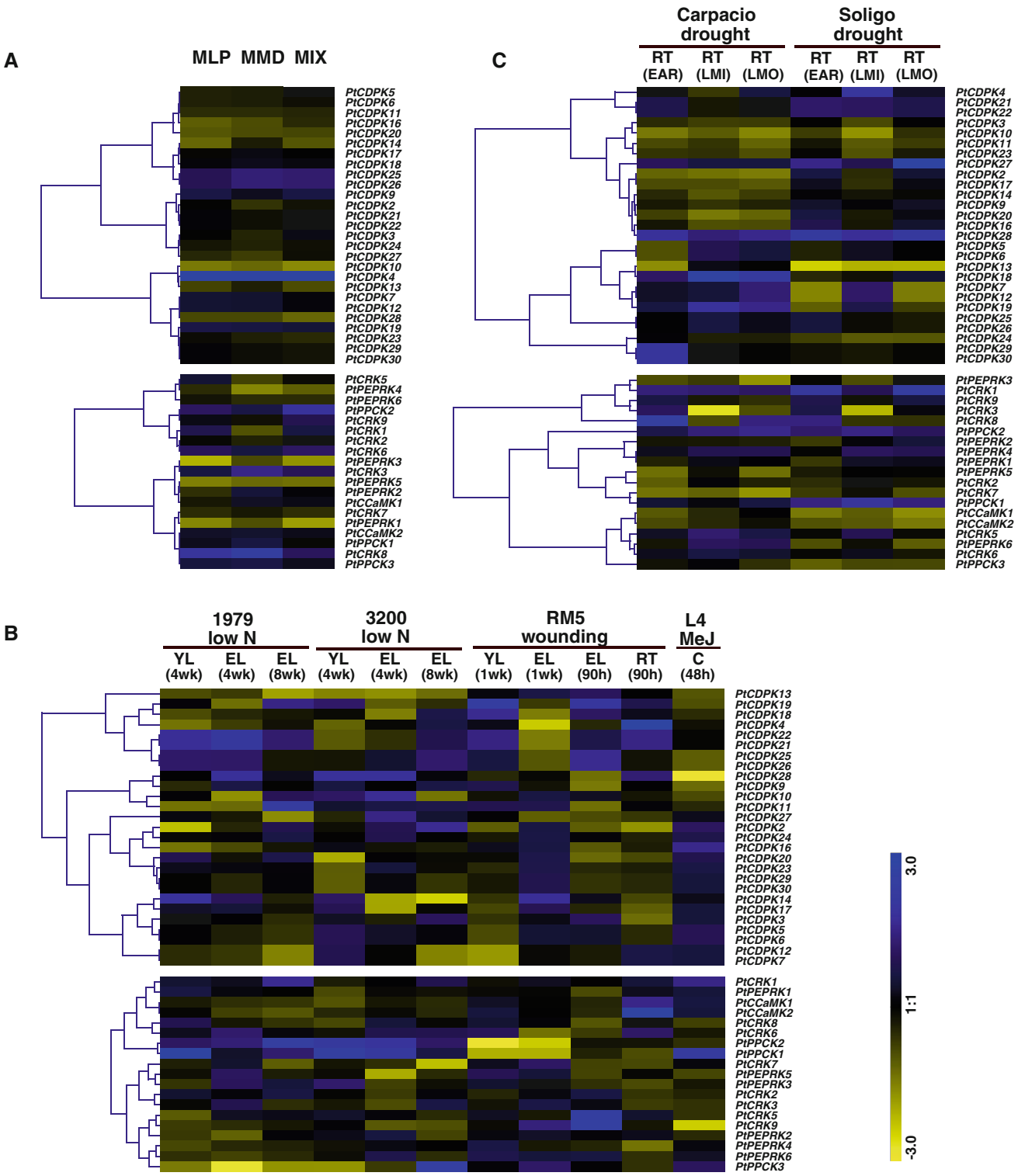


Fig. 5 Expression profiles of *Populus CDPK* and its closely related genes across different tissues. Background corrected expression intensities were log-transformed and visualized as heatmaps (see Materials and Methods). The Affymetrix microarray data were obtained from NCBI Gene Expression Omnibus (GEO) database under the series accession number GSE13990. *CL* continuous light-grown seedling, *DL* etiolated dark-grown seedling transferred to light for 3 h, *DS* dark-grown seedlings, *YL* young leaf, *ML* mature leaf, *R* root, *DX* differentiating xylem, *FC* female catkins, *MC* male catkins



Analysis of paralogous *PtCDPK* gene pairs demonstrated that 12 out of 14 gene pairs retained in conserved positions on the segmental duplicated blocks (Fig. 2), suggesting that these 12 paralogous pairs might derive from segmental duplications. While no traceable duplication event could be inferred for two paralogous pairs (*PtCDPK18/19* and *PtCDPK21/22*), *PtCDPK18* was located outside of any duplicated regions with its counterpart (*PtCDPK19*) not mapped on LGs yet, and neither *PtCDPK21* nor *PtCDPK22* was located on the duplicated blocks (Fig. 2).

All *PtCDPK* closely related kinase genes were located within the duplicated blocks (Fig. 2). Furthermore, all of

◀ **Fig. 6** Differential expression of *Populus CDPK* and its closely related genes under different stresses. Expression is indicated as fold-change of experimental treatments relative to control samples and visualized in heatmaps (see Materials and Methods). *Color scale* represents log2 expression values, *yellow* represents low level and *blue* indicates high level of transcript abundances. **a** Heatmap showing hierarchical clustering of *Populus CDPK* and its closely related genes under short-term and long-term water deficit in root tips. Microarray data under the series accession number GSE17223 was obtained from NCBI GEO database. *EAR* early response (EAR) to water deficit by 36 h; LMI, long-term (10-day) response to mild stress with soil relative extractable water (REW) at 20–35 %; *LMO* long-term (10-day) response to moderate stress with soil relative extractable water (REW) at 10–20 %. **b** Heatmap showing hierarchical clustering of *Populus CDPK* and its closely related genes under fungal infection in leaves. Microarray data under the series accession number GSE9673 was obtained from NCBI GEO database. *MLP Populus* leaves infected with *Melampsora larici-populina*; *MMD Populus* leaves infected with *Melampsora medusae f.* sp. *Deltoidae*, *MIX Populus* leaves infected with mixture of the two fungal species. **c** Heatmap showing hierarchical clustering of *Populus CDPK* and its closely related genes across various tissues and genotypes analyzed. Microarray data under the series accession number GSE16786 was obtained from NCBI GEO database. Genotypes analyzed included: *P. fremontii* × *angustifolia* clones 1979, 3200, and RM5, *P. tremuloides* clones 271 and L4, and *Populus deltoids* clones Soligo and Carpaccio. Tissues analyzed included: *YL* young leaves, *EL* expanding leaves, *ML* mature leaves, *RT* root tips, *C* suspension cell cultures. Stress treatments included: *low N* nitrogen limitation, *MeJ* methyl jasmonate elicitation; wounding, sampled either 1 week or 90 h after wounding. (Color figure online)

the eight paralogous pairs (80 % of the identified genes) were retained duplicates that located in both duplicated blocks. Although none of the rest four genes had paralogous counterpart, they also located within the duplicated blocks, which suggested dynamic rearrangements may have occurred after the segmental duplication. The above results indicated that the expansion of the PtCDPK closely related gene families also originated from segmental duplications.

The substitution rate ratio of nonsynonymous (dN or Ka) versus synonymous (dS or Ks) is an indicator of selection history on genes or gene regions. Generally, Ka/Ks <1 indicates the functional constraint with negative or purifying selection of the genes, Ka/Ks >1 means accelerated evolution with positive selection, and Ka/Ks = 1 suggests neutral selection [71]. In this study, the Ka/Ks ratio of 22 putative paralogous gene pairs identified were calculated to reveal the divergence fate after duplication of *Populus CDPK* and its closely related genes. As the results showed that the Ka/Ks ratios of all paralogous pairs were no larger than 0.4 (Table 4), we could conclude the *Populus* CDPK and its closely related gene families had undergone great purifying selection pressure with limited functional divergence after segmental duplications. In addition, based on the divergence rate of 9.1×10^{-9} synonymous mutations per synonymous site year proposed for *Populus* [38], duplications of the paralogous gene pairs were estimated to occur between 8.08 and 22.21 million years (MY) ago (Table 4).

Structures of the *Populus* CDPK and its closely related genes

In order to gain further insight into the structural diversity of *Populus CDPK* genes, we constructed a separate phylogenetic tree exclusively using the full-length *Populus* CDPK protein sequences and compared the exon/intron organization in the coding sequences of each *Populus CDPK* genes (Fig. 3). In addition, the splicing sites were also mapped to the open reading frame (ORF) of the *Populus CDPKs* (Fig. 4).

As indicated in Figs. 3 and 4, most *Populus CDPK* members within the same subfamilies shared vary similar gene structures in terms of intron numbers and exon lengths, and although the length varied, introns inserted into nearly the same locations of the gene ORF. All members from subgroup I possessed six introns, and most shared the same intron/exon organization except for three genes (*PtCDPK2*, *PtCDPK3* and *PtCDPK9*) (Fig. 3). The first introns of the two paralogous genes, *PtCDPK2* and *PtCDPK3*, were both shifted to the location between subdomain IV and V instead of subdomain VII and VIII, and the CaM-like domain of *PtCDPK9* appeared to be truncated and lost one EF-hand motif (Fig. 4). Five members of subgroup II (*PtCDPK18*, *PtCDPK19*, *PtCDPK23*, *PtCDPK24* and *PtCDPK30*) shared the similar splicing patterns with seven introns, whereas the other three (*PtCDPK5*, *PtCDPK6* and *PtCDPK29*), which all belonged to subgroup II-b, had one additional splicing site between subdomain III and IV or at the N-variable domain (Fig. 4). In subgroup III, the splicing patterns of five genes (*PtCDPK10*, *PtCDPK11*, *PtCDPK20*, *PtCDPK27* and *PtCDPK28*) were largely identical with seven introns. All of the remaining four genes (*PtCDPK14*, *PtCDPK15*, *PtCDPK25* and *PtCDPK26*) lacked the first intron inserted between subdomain VIa and VIb, furthermore, *PtCDPK15* also had a truncated C-terminus leading to the losing of another intron (Fig. 4).

Unlike subgroup I, II and III, which shared gene structures largely in agreement, the splicing sites of two genes (*PtCDPK21* and *PtCDPK22*) in subgroup IV were largely in line with the *PtCRKs* rather than other *PtCDPK* members. What's more, subgroup IV was located on a branch most closely related to the PtCRKs in the phylogenetic tree. Thus, according to the above results, and along with the degenerated EF-hand motifs in the PtCRKs, we could infer that *PtCDPKs* of the subgroup IV and the *PtCRKs* were originated from a common ancestor comparatively recently, which was also proposed in *Arabidopsis* and rice [9, 27]. Conversely, the gene structures of the other *Populus* CDPK closely related kinases showed relatively lower

degree of similarities with the *Populus* CDPKs (Figs. 3, 4), suggesting they could diverged a long time ago.

Expression profiles of *Populus CDPK* and its closely related kinase genes

Whole genome microarray is considered as a useful tool of studying gene expression profiles in *Populus* [60, 61, 63]. To gain insight into the expression profiles of *Populus CDPK* genes in different tissues, we reanalyzed the *Populus* microarray data generated by Wilkins and coworkers [41] (Fig. 5). Three (*PtCDPK1*, *PtCDPK8* and *PtCDPK15*) of the 30 *Populus CDPK* genes did not have the corresponding probe sets in the dataset (Table S1). Of the 27 *Populus CDPK* genes whose expression profiles were further analyzed, 13 showed preferentially high expression level in root and/or differentiating xylem, 13 in female and male catkins, and only one in young leaf (Fig. 5). The genes closely related to the *Populus CDPK* genes were also investigated for their expression patterns across various tissues. The results demonstrated that the *PtCCaMKs* were preferentially expressed in root and the *PtPPCKs* had the highest transcription abundance in differentiating xylem, whereas, the *PtCRKs* and the *PtPEPRKs* showed relatively diversified expression profiles (Fig. 5).

To further investigate the response of *Populus CDPK* and its closely related genes to biotic and abiotic stresses, we examined their expression patterns under fungal infection, low nitrogen limitation, mechanical wounding, drought and methyl jasmonate (MeJ) treatment (Fig. 6). Fungal infection caused up-regulation of 7 *PtCDPK* genes and down-regulation of 6 *PtCDPK* genes. Among them, *PtCDPK4* showed a remarkably high expression level under all three infection treatments, suggesting a high probability for it to participate in the anti-fungal signal transduction pathway. Further analysis of the *PtCDPK* closely related genes also revealed six up-regulated genes and six down-regulated genes. Interestingly, a few genes such as *PtCRK1* seem to have different expression profiles under different pathogen infections (Fig. 6a). The responses of *Populus CDPK* genes to nitrogen deficit stress differ between two *Populus* genotypes examined. For instance, *PtCDPK21* and *PtCDPK22* were significantly induced at all three tissues examined in genotype 1979, which was not the case in genotype 3200. However, among the *PtCDPK* closely related genes, *PtPPCK1* and *PtPPCK2* were generally up-regulated across all tissues examined in both genotypes (Fig. 6b). Mechanical wounding induced up-regulation of six *PtCDPK* genes (*PtCDPK18*, *PtCDPK19*, *PtCDPK21*, *PtCDPK22*, *PtCDPK25* and *PtCDPK26*) and down-regulation of eight *PtCDPK* genes (*PtCDPK2*, *PtCDPK7*, *PtCDPK12*, *PtCDPK17*, *PtCDPK23*, *PtCDPK28*, *PtCDPK29* and *PtCDPK30*) in young leaves 1 week after wounding as well as expanded leaves 90 h after wounding. Interestingly, gene expression profiles in expanded leaves one week after wounding were distinctively different, which revealed one significantly up-regulated gene (*PtCDPK14*) and four significantly down-regulated genes (*PtCDPK4*, *PtCDPK18*, *PtCDPK21* and *PtCDPK22*). In root tips, four up-regulated *PtCDPK* genes (*PtCDPK4*, *PtCDPK21*, *PtCDPK22* and *PtCDPK28*) and five down-regulated genes (*PtCDPK2*, *PtCDPK3*, *PtCDPK14*, *PtCDPK17* and *PtCDPK20*) were also identified 90 h after mechanical wounding. Analysis of the *PtCDPK* closely related genes proposed two sharply down-regulated genes (*PtPPCK1* and *PtPPCK2*) in young and expanded leaves 1 week after wounding and two up-regulated genes (*PtCRK5* and *PtCRK9*) in expanded leaves 90 h after wounding. In root tips, both *PtCCaMK* genes (*PtCCaMK1* and *PtCCaMK2*) appeared to be up-regulated 90 h after wounding (Fig. 6b). In response to MeJ feeding in cell culture, two *PtCDPK* genes, *PtCDPK2* and *PtCDPK16*, were shown to be significantly up-regulated, whereas *PtCDPK28* was obviously down-regulated. In addition, analysis of the *PtCDPK* closely related genes proposed two sharply up-regulated genes (*PtCRK1* and *PtPPCK1*) and one down-regulated gene (*PtCRK9*) (Fig. 6b).

Water-deficit is a primary environment stress that most land plants would encounter during life spans. As quite a subset of *CDPK* genes identified in various plant species had been proven to play crucial roles in drought stress response, we further investigated the expression profiles of *Populus CDPK* genes under water-deficit conditions by examining two Affymetrix microarray datasets (GSE17223 and GSE17230) [43] (Fig. 6c). Under drought stress, most *PtCDPK* and its closely related genes showed more significant and organized expression profiles in root tips than in mature leaves, which was consistent with the global expression changes of the drought-driven genes in *Populus* [43]. Thus the microarray dataset GSE17223 in which gene expression profiles in root tips of two *Populus* genotypes (Carpaccio and Soligo) under water-deficit conditions was further analyzed. According to the dataset, two genes (*PtCDPK27* and *PtCDPK28*) were up-regulated and one gene (*PtCDPK10*) was down-regulated uniformly in both genotypes across all drought condition tested which included an early response (EAR) to water deficiency by 36 h and long-term (10-day) response to mild stress (LMI) and moderate stress (LMO). What's more, the transcription abundance of a subset of *PtCDPKs* was only altered in one separate genotype or under certain water-deficit conditions. For instance, *PtCDPK18* showed up-regulation of transcription abundance in the genotype "Carpaccio" rather than "Soligo", whereas *PtCDPK13* was significantly down regulated across all conditions in "Soligo" but only under EAR condition in "Carpaccio" (Fig. 6c). Analysis of the

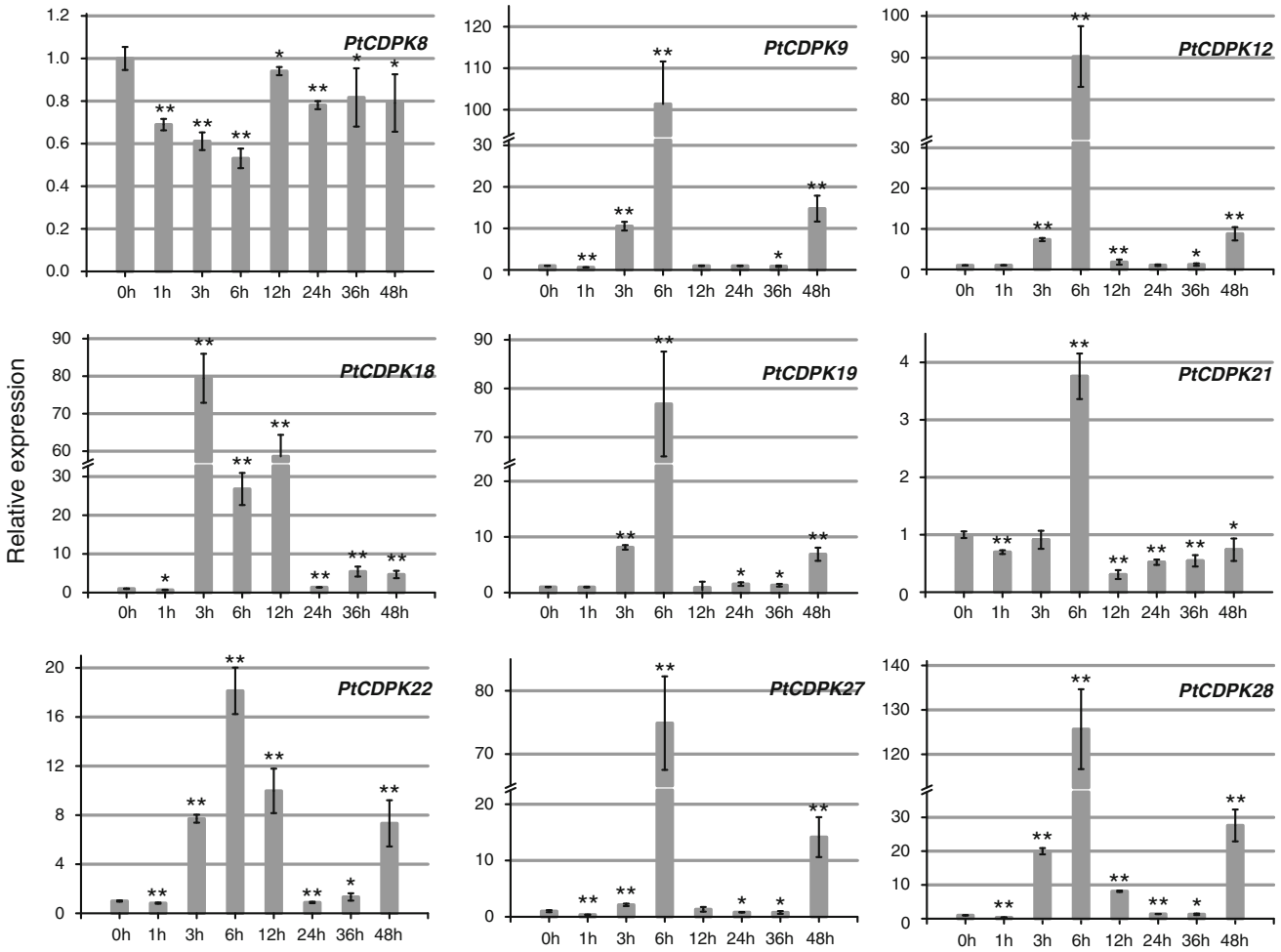


Fig. 7 Expression analysis of nine selected *PtCDPK* genes under drought stresses using qRT-PCR. The relative mRNA abundance of nine selected *PtCDPK* genes was normalized with respect to reference genes *UBQ10* in drought stress treatments. Two biological replicates each with three technique replicates were performed and *bars* represent standard deviations (SD) of the replicates. *X-axis* is time courses of stress treatments for each gene. Stress treatment groups showed significant difference in transcript abundance compared to the control group were indicated with *asterisks*, with *single* or *double asterisks* indicating a significant difference of $P < 0.05$ or $P < 0.01$ between controls and treatments respectively

PtCDPK closely related genes also revealed that two genes (*PtCRK1* and *PtPPCK2*) were up-regulated in both genotypes and across all conditions. Notably, in both genotypes, *PtCRK3* was up-regulated under EAR but down-regulated dramatically under LMI, which suggested a term-based response mode to water-deficiency stress (Fig. 6c).

Duplicated genes might undergo different evolution fates including nonfunctionalization, neofunctionalization and subfunctionalization, which could be demonstrated by the divergence of the gene expression profiles [72]. Of the 22 *PtCDPK* and its closely related paralogous gene pairs, ten pairs were either lack of corresponding probe sets, shared same probe sets, or located outside of any of the duplicated blocks. As a result, 12 paralogous pairs remained to be further analyzed. Among them, four paralogous pairs (*PtCDPK2/3*, *PtCDPK16/17*, *PtCDPK23/24* and *PtCCaMK1/2*) shared almost the same expression profiles with respect to various tissues and stresses. On the contrary, the expression pattern of three paralogous pairs (*PtCRK1/2*, *PtCRK3/9* and *PtCRK7/8*) diverged dramatically, suggesting substantial neofunctionalization during subsequent evolution process. The rest of the duplicated genes demonstrated partially redundant expression pattern with distinct shifts, which indicated that they might have undergone subfunctionalization. These findings above suggested that the expression patterns of the *PtCDPK* and its closely related genes had diversified substantially, which might due to the different divergent fates after segmental duplications.

Examination of *Populus CDPK* gene expression profiles under drought stress by qRT-PCR

In order to verify the expression profiles of putative drought-responsive *Populus CDPK* genes which were speculated by the microarray analysis or orthologous to the previously identified *Arabidopsis CDPK* genes, qRT-PCR analysis was performed for nine selected *PtCDPK* genes including four paralogous pairs in *Populus* root tips under drought stresses (Fig. 7).

PtCDPK8 and *PtCDPK9* were orthologous to *Arabidopsis AtCPK6* which had been proven to be involved in both drought and salinity stresses [26]. Notably, although *PtCDPK9* was significantly induced after dehydration stress, the transcription abundance of *PtCDPK8* decreased gradually and reached its lowest point 6 h after treatment, suggesting their function may have been diversified during subsequent evolution process. *PtCDPK18* and *PtCDPK19*, the orthologous to *Arabidopsis AtCPK21* involved in abiotic stress responses [51], showed up-regulation under drought stress according to the microarray analysis. This paralogous gene pair were both up regulated following dehydration treatments although they reached the highest transcription abundance at different time points (3 or 6 h after treatment). In consistent with the microarray analysis, two paralogous pairs, *PtCDPK21/22* and *PtCDPK27/28*, together with *PtCDPK12* which was orthologous to *Arabidopsis AtCDPK2/CPK11* involved in ABA-related drought and salinity stress responses [23, 73], were all induced after dehydration treatments (Fig. 7).

Taken together, although different *Populus* genotype (*Populus deltoides*) were used, the gene expression patterns detected by qRT-PCR were largely consistent with the results of the microarray analysis. Furthermore, all *Populus CDPK* genes except *PtCDPK18* characterized by qRT-PCR had either highest or lowest transcription abundance in root tips 6 h after dehydration treatments. Therefore, we attempted to speculate that 6 h after environmental changes was the time point at which the *Populus* drought-responsive calcium-dependent signal pathway reached its maximum activity.

Acknowledgments This study was supported by grants from the National Basic Research Program of China (2012CB114501), the National High-Tech Research and Development Program of China (2011AA100209), the National Natural Science Foundation of China (31070272, 31000311 and 30901157), the Program of 100 Distinguished Young Scientists of the Chinese Academy of Sciences (to Gongke Zhou), the Promotive Research Fund for Young and Middle-aged Scientists of Shandong Province (BS2011SW055), and the Qingdao Municipal Science and Technology Plan Project (11-2-4-8-(1)-jch, 11-2-4-8-(2)-jch and 12-1-4-9-(2)-jch). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Trewavas AJ, Malhó R (1998) Ca^{2+} signalling in plant cells: the big network! Curr Opin Plant Biol 1(5):428–433
2. Knight H, Knight MR (2001) Abiotic stress signalling pathways: specificity and cross-talk. Trends Plant Sci 6(6):262–267
3. Luan S, Kudla J, Rodriguez-Concepcion M, Yalovsky S, Gruissem W (2002) Calmodulins and calcineurin B-like proteins: calcium sensors for specific signal response coupling in plants. Plant Cell 14:S389–S400
4. Reddy ASN (2001) Calcium: silver bullet in signaling. Plant Sci 160(3):381–404
5. Sanders D, Pelloux J, Brownlee C, Harper JF (2002) Calcium at the crossroads of signaling. Plant Cell 14:S401–S417
6. Snedden WA, Fromm H (2001) Calmodulin as a versatile calcium signal transducer in plants. New Phytol 151(1):35–66
7. Sanders D, Brownlee C, Harper JF (1999) Communicating with calcium. Plant Cell 11(4):691–706
8. Hernández Sebastià C, Hardin SC, Clouse SD, Kieber JJ, Huber SC (2004) Identification of a new motif for CDPK phosphorylation in vitro that suggests ACC synthase may be a CDPK substrate. Arch Biochem Biophys 428(1):81–91
9. Hrabak EM, Chan CWM, Gribskov M, Harper JF, Choi JH, Halford N, Kudla J, Luan S, Nimmo HG, Sussman MR, Thomas M, Walker-Simmons K, Zhu JK, Harmon AC (2003) The *Arabidopsis* CDPK-SnRK superfamily of protein kinases. Plant Physiol 132(2):666–680
10. Kolukisaoglu U, Weinl S, Blazevic D, Batistic O, Kudla J (2004) Calcium sensors and their interacting protein kinases: genomics of the *Arabidopsis* and rice CBL-CIPK signaling networks. Plant Physiol 134(1):43–58
11. McCormack E, Braam J (2003) Calmodulins and related potential calcium sensors of *Arabidopsis*. New Phytol 159(3):585–598
12. Zielinski RE (1998) Calmodulin and calmodulin-binding proteins in plants. Annu Rev Plant Phys 49:697–725
13. Cheng SH, Willmann MR, Chen HC, Sheen J (2002) Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. Plant Physiol 129(2):469–485
14. Yang TB, Poovaiah BW (2003) Calcium/calmodulin-mediated signal network in plants. Trends Plant Sci 8(10):505–512
15. Wernimont AK, Amani M, Qiu W, Pizarro JC, Artz JD, Lin YH, Lew J, Hutchinson A, Hui R (2011) Structures of parasitic CDPK domains point to a common mechanism of activation. Proteins 79(3):803–820
16. Wernimont AK, Artz JD, Finerty P Jr, Lin YH, Amani M, Allali-Hassani A, Senisterra G, Vedadi M, Tempel W, Mackenzie F, Chau I, Lourido S, Sibley LD, Hui R (2010) Structures of apicomplexan calcium-dependent protein kinases reveal mechanism of activation by calcium. Nat Struct Mol Biol 17(5):596–601
17. Harper JF, Harmon A (2005) Plants, symbiosis and parasites: a calcium signalling connection. Nat Rev Mol Cell Biol 6(7):555–566
18. Saijo Y, Hata S, Kyozuka J, Shimamoto K, Izui K (2000) Overexpression of a single Ca^{2+}-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. Plant J 23(3):319–327
19. Abbasi F, Onodera H, Toki S, Tanaka H, Komatsu S (2004) OsCDPK13, a calcium-dependent protein kinase gene from rice, is induced by cold and gibberellin in rice leaf sheath. Plant Mol Biol 55(4):541–552
20. Asano T, Hakata M, Nakamura H, Aoki N, Komatsu S, Ichikawa H, Hirochika H, Ohsugi R (2011) Functional characterisation of OsCPK21, a calcium-dependent protein kinase that confers salt tolerance in rice. Plant Mol Biol 75(1):179–191

21. Asano T, Wakayama M, Aoki N, Komatsu S, Ichikawa H, Hirochika H, Ohsugi R (2010) Overexpression of a calcium-dependent protein kinase gene enhances growth of rice under low-nitrogen conditions. Plant Biotechnol 27(4):369–373
22. Zou JJ, Wei FJ, Wang C, Wu JJ, Ratnasekera D, Liu WX, Wu WH (2010) *Arabidopsis* calcium-dependent protein kinase CPK10 functions in abscisic acid- and Ca^{2+}-mediated stomatal regulation in response to drought stress. Plant Physiol 154(3): 1232–1243
23. Zhu SY, Yu XC, Wang XJ, Zhao R, Li Y, Fan RC, Shang Y, Du SY, Wang XF, Wu FQ, Xu YH, Zhang XY, Zhang DP (2007) Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in *Arabidopsis*. Plant Cell 19(10):3019–3036
24. Zhao R, Sun HL, Mei C, Wang XJ, Yan L, Liu R, Zhang XF, Wang XF, Zhang DP (2011) The *Arabidopsis* Ca^{2+}-dependent protein kinase CPK12 negatively regulates abscisic acid signaling in seed germination and post-germination growth. New Phytol 192(1):61–73
25. Ma SY, Wu WH (2007) AtCPK23 functions in *Arabidopsis* responses to drought and salt stresses. Plant Mol Biol 65(4):511–518
26. Xu J, Tian YS, Peng RH, Xiong AS, Zhu B, Jin XF, Gao F, Fu XY, Hou XL, Yao QH (2010) AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in *Arabidopsis*. Planta 231(6):1251–1260
27. Asano T, Tanaka N, Yang G, Hayashi N, 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(2):356–366
28. Ray S, Agarwal P, Arora R, Kapoor S, Tyagi AK (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(5):493–505
29. Li AL, Zhu YF, Tan XM, Wang X, Wei B, Guo HZ, Zhang ZL, Chen XB, Zhao GY, Kong XY, Jia JZ, Mao L (2008) Evolutionary and functional study of the CDPK gene family in wheat (*Triticum aestivum* L.). Plant Mol Biol 66(4):429–443
30. Levy J, Bres C, Geurts R, Chalhoub B, Kulikova O, Duc G, Journet EP, Ane JM, Lauber E, Bisseling T, Denarie J, Rosenberg C, Debelle F (2004) A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science 303(5662):1361–1364
31. Leclercq J, Ranty B, Sanchez-Ballesta MT, Li Z, Jones B, Jauneau A, Pech JC, Latche A, Ranjeva R, Bouzayen M (2005) Molecular and biochemical characterization of LeCRK1, a ripening-associated tomato CDPK-related kinase. J Exp Bot 56(409):25–35
32. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K, Bairoch A (2002) The PROSITE database, its status in 2002. Nucleic Acids Res 30(1):235–238
33. Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X (2008) CSS-Palm 2.0: an updated software for palmitoylation sites prediction. Protein Eng Des Sel 21(11):639–644
34. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25(24):4876–4882
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739
36. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612
37. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586–1591
38. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151–1155
39. Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDS: a gene structure display server. Yi Chuan 29(8):1023–1026
40. Barrett T, Edgar R (2006) Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol 411:352–369
41. Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. Plant Physiol 149(2):981–993
42. Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: cluster analysis of microarray data. Bioinformatics 18(1):207–208
43. Cohen D, Bogeat-Triboulot MB, Tisserant E, Balzergue S, Martin-Magniette ML, Lelandais G, Ningre N, Renou JP, Tamby JP, Le Thiec D, Hummel I (2010) Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. BMC Genomics 11:630
44. Yuan Y, Chung JD, Fu X, Johnson VE, Ranjan P, Booth SL, Harding SA, Tsai CJ (2009) Alternative splicing and gene duplication differentially shaped the regulation of isochorismate synthase in *Populus* and *Arabidopsis*. Proc Natl Acad Sci USA 106(51):22020–22025
45. Ma HS, Liang D, Shuai P, Xia XL, Yin WL (2010) The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in *Arabidopsis thaliana*. J Exp Bot 61(14):4011–4019
46. Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11(2):113–116
47. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods 25(4):402–408
48. Hrabak EM, Dickmann LJ, Satterlee JS, Sussman MR (1996) Characterization of eight new members of the calmodulin-like domain protein kinase gene family from *Arabidopsis thaliana*. Plant Mol Biol 31(2):405–412
49. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313(5793):1596–1604
50. Christodoulou J, Malmendal A, Harper JF, Chazin WJ (2004) Evidence for differing roles for each lobe of the calmodulin-like domain in a calcium-dependent protein kinase. J Biol Chem 279(28):29092–29100
51. Franz S, Ehlert B, Liese A, Kurth J, Cazale AC, Romeis T (2011) Calcium-dependent protein kinase CPK21 functions in abiotic stress response in *Arabidopsis thaliana*. Mol Plant 4(1):83–96

52. Johnson DR, Bhatnagar RS, Knoll LJ, Gordon JI (1994) Genetic and biochemical studies of protein *N*-myristoylation. Annu Rev Biochem 63:869–914
53. Mehlmer N, Wurzinger B, Stael S, Hofmann-Rodrigues D, Csaszar E, Pfister B, Bayer R, Teige M (2010) The $Ca^{(2+)}$-dependent protein kinase CPK3 is required for MAPK-independent salt-stress acclimation in *Arabidopsis*. Plant J 63(3):484–498
54. Luan S (2011) Coding and decoding of calcium signals in plants, vol. 10. Springer, Berlin
55. Lindzen E, Choi JH (1995) A carrot cDNA encoding an atypical protein kinase homologous to plant calcium-dependent protein kinases. Plant Mol Biol 28(5):785–797
56. Zhang L, Liu BF, Liang S, Jones RL, Lu YT (2002) Molecular and biochemical characterization of a calcium/calmodulin-binding protein kinase from rice. Biochem J 368(Pt 1):145–157
57. Liu Z, Xia M, Poovaiah BW (1998) Chimeric calcium/calmodulin-dependent protein kinase in tobacco: differential regulation by calmodulin isoforms. Plant Mol Biol 38(5):889–897
58. Patil S, Takezawa D, Poovaiah BW (1995) Chimeric plant calcium/calmodulin-dependent protein kinase gene with a neural visinin-like calcium-binding domain. Proc Natl Acad Sci USA 92(11):4897–4901
59. Takezawa D, Ramachandiran S, Paranjape V, Poovaiah BW (1996) Dual regulation of a chimeric plant serine/threonine kinase by calcium and calcium/calmodulin. J Biol Chem 271(14):8126–8132
60. Hu R, Chi X, Chai G, Kong Y, He G, Wang X, Shi D, Zhang D, Zhou G (2012) Genome-wide identification, evolutionary expansion, and expression profile of homeodomain-leucine zipper gene family in poplar *(Populus trichocarpa)*. PLoS ONE 7(2):e31149
61. Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G (2010) Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biol 10:145
62. Lan T, Yang ZL, Yang X, Liu YJ, Wang XR, Zeng QY (2009) Extensive functional diversification of the *Populus* glutathione *S*-transferase supergene family. Plant Cell 21(12):3749–3766
63. Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, DiLoreto DS, Yellanki P, Carlson JE (2009) The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. BMC Plant Biol 9:26
64. Barakat A, Choi A, Yassin NB, Park JS, Sun Z, Carlson JE (2011) Comparative genomics and evolutionary analyses of the *O*-methyltransferase gene family in *Populus*. Gene 479(1–2):37–46
65. Kalluri UC, Difazio SP, Brunner AM, Tuskan GA (2007) Genome-wide analysis of Aux/IAA and ARF gene families in *Populus trichocarpa*. BMC Plant Biol 7:59
66. Zhuang J, Cai B, Peng RH, Zhu B, Jin XF, Xue Y, Gao F, Fu XY, Tian YS, Zhao W, Qiao YS, Zhang Z, Xiong AS, Yao QH (2008) Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa*. Biochem Biophys Res Commun 371(3):468–474
67. Tuominen LK, Johnson VE, Tsai CJ (2011) Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. BMC Genomics 12:236
68. Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16(7):1679–1691
69. Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16(7):1667–1678
70. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA 102(15):5454–5459
71. Yang X, Tuskan GA, Cheng MZ (2006) Divergence of the Dof gene families in poplar, *Arabidopsis*, and rice suggests multiple modes of gene evolution after duplication. Plant Physiol 142(3):820–830
72. Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3(11):827–837
73. Urao T, Katagiri T, Mizoguchi T, Yamaguchi-Shinozaki K, Hayashida N, Shinozaki K (1994) Two genes that encode $Ca^{(2+)}$-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*. Mol Gen Genet 244(4):331–340