Planta (2018) 247:1133–1148
https://doi.org/10.1007/s00425-018-2849-2

ORIGINAL ARTICLE

Genome-wide identification, classification, and expression of phytocyanins in *Populus trichocarpa*

Shuangshuang Luo[1,2] · Wenfang Hu[1,2] · Yue Wang[1,2] · Bin Liu[1,2] · Hanwei Yan[1,2,3] · Yan Xiang[1,2,3]

Received: 3 August 2017 / Accepted: 17 January 2018 / Published online: 30 January 2018

Abstract

Main conclusion **74 phytocyanin genes were identified in the *Populus trichocarpa* genome. Phylogenetic analysis grouped the PC proteins into four subfamilies (UCs, PLCs, SCs, and ENODLs). Closely related PC proteins share similar motifs, implying similar functions. Expression profiles of *PtPC* genes were analyzed in response to drought and salt-stress.**

Phytocyanins (PCs) are blue copper proteins associated with electron carrier activity that have a large influence on plant growth and resistance. The majority of PCs are chimeric arabinogalactan proteins (AGPs). In this work, we identified 74 *PC* genes in *Populus trichocarpa* and analyzed them comprehensively. Based on the ligands composition of copper-binding sites, glycosylation state, the domain structure and spectral characteristics of *PC* genes, PCs were divided into four subfamilies [uclacyanins (UCs), plantacyanins (PLCs), stellacyanins (SCs) and early nodulin-like proteins (ENODLs)], and phylogenetic relationship analysis classified them into seven groups. All *PtPCs* are randomly distributed on 17 of the 19 poplar chromosomes, and they appear to have undergone expansion via segmental duplication. Eight *PtPCs* do not contain introns, and each group has a similar conserved motif structure. Promoter analysis revealed *cis*-elements related to growth, development and stress responses, and established orthology relationships of PCs between *Arabidopsis* and poplar by synteny analysis. Expression profile analysis and qRT-PCR analysis showed that *PtPCs* were expressed widely in various tissues. Quantitative real-time RT-PCR analysis of *PC* genes expression in response to salt and drought stress revealed their stress-responses profiles. This work provides a theoretical basis for a further study of stress resistance mechanisms and the function of *PC* genes in poplar growth and development.

Keywords Evolution · Expression profile · Heat map · Phytocyanins · Poplar

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00425-018-2849-2) contains supplementary material, which is available to authorized users.

Shuangshuang Luo and Wenfang Hu contributed equally to this work.

✉ Hanwei Yan
hwyanahau@163.com

✉ Yan Xiang
xiangyan@ahau.edu.cn; xiangyanahau@sina.com

[1] Key Laboratory of Crop Biology of Anhui Province, School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei 230036, China

[2] Laboratory of Modern Biotechnology, School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, China

[3] National Engineering Laboratory of Crop Stress Resistance Breeding, Anhui Agricultural University, Hefei, China

Abbreviations

AG	Arabinogalactan
AGPs	Arabinogalactan proteins
ENODLs	Early nodulin-like proteins
PCs	Phytocyanins
PCLD	Plastocyanin-like domain
PLCs	Plantacyanins
SCs	Stellacyanins
SP	Signal peptide
UCs	Uclacyanins
K_s	Number of synonymous substitutions per synonymous site
K_a	Number of non-synonymous substitutions per non-synonymous site

Introduction

Blue copper proteins containing a single type I mononuclear copper site are known as phytocyanins (PCs) in plants which they are associated with electron carrier activity (Giri et al. 2004). The phytocyanin domain has a core β-sandwich comprising seven β-strands, and a disulfide bridge closing the metal centre is a characteristic feature (Hart et al. 1996). Based on the ligand composition of the copper-binding site, the glycosylation state, the domain structure, and the spectral characteristics, PCs can be grouped into four subfamilies: uclacyanins (UCs), plantacyanins (PLCs), stellacyanins (SCs) and early nodulin-like proteins (ENODLs) (Cao et al. 2015).

Residues ligating the copper ion in SCs consist of two His, one Cys, and one Gln, whereas PLCs and UCs also have two His and one Cys, but the Gln is replaced by a Met (Nersissian et al. 1998). Although PLCs have the same four conserved residues as UCs, they lack putative glycosylation sites on the backbone (Nersissian et al. 1998). Interestingly, early nodulins (ENODLs) might be involved in Cu-independent processes since they lack key copper-binding residues (Nersissian et al. 1998). PC genes containing AG (arabinogalactan) glycomodules and signal peptides (SPs) are believed to be members of the arabinogalactan proteins (AGP) superfamily (Mashiguchi et al. 2009).

PCs exert an important part in growth and development of plants, in addition to the influence of their spectroscopic and redox properties. *OsUCL29* and *ZmUC22* are significantly expressed under various stresses. *BrUCL16* and *ZmUC19* are specifically expressed in the stem and silique, respectively, which show that the first of the four PC subfamilies (UCs) appear to function in the evolution of polyploid plants (Ma et al. 2011; Li et al. 2013; Cao et al. 2015). Six SCs are induced by oxidative stress and Al toxicity in *Arabidopsis* (Ezaki et al. 2001, 2005). *PeSCL1* and *PeSCL3* are strongly expressed in the stem and roots of *Phalaenopsis equestris* (Xu et al. 2017). The PLC subfamily participates in the growth processes of several specific plants, including pollination through S-Rnase binding in tobacco (Cruz-Garcia et al. 2005). *AtPLCs* were also regarded as miR408, the member of the microRNAs targets, which were related to plant growth as transcription factors (Sunkar and Zhu 2004). Dong et al. (2005) found that *AtPLCs* were strongly expressed in the pistil to prevent pollination and destroy the endothecium structure, thereby influencing the growth of the anther in promoter-β-glucuronidase transgenic plants, as was also shown by an immunohistochemical analysis of wild-type pistil tissues. The last subfamily (ENODLs) is related to many aspects of plant development and important function on transport nutrients, solutes, amino acids or hormones, and improves the fitness to pathogens during host colonization at the plant–microbe interface activities (Denancé et al. 2014). Additionally, ENODLs have high expression levels in the inflorescence of some plants, such as *AtENODL3/4* (Mashiguchi et al. 2009), *BrENODL22/27* (Li et al. 2013) and *PeENODL5/7* (Xu et al. 2017). ENODLs are also relevant for the defence responses of plants. Mashiguchi et al. (2009) found that *AtENODL2/18* were induced by osmotic and salt stress. Using hybridisation analysis, Yoshizaki et al. (2000) demonstrated that ENODLs were specifically expressed in tissues including apical buds in *Pharbitis nil* and root nodules of legumes, and the expression of these genes was distinctly down-regulated during floral induction. Furthermore, these genes might be involved in the organ differentiation of plants (Yoshizaki et al. 2000). *AtFLA3,* another type of chimeric AGP, is involved in microspore development and the formation of the pollen intine via the deposition of cellulose. Overexpression of this gene restricted the progress of pollination to lower the rate of seed production (Li et al. 2010). *AtENODL14* localized at the synergid cell surface strongly and specifically interacts with the extracellular domain of the receptor-like kinase FERONIA, which could minutely control the reception of pollen tube (Escobar-Restrepo et al. 2007). Wild-type pollen tubes can not prevent growth and cause rupture after entering the ovules of quintuple ENODL mutants which loss the function, implying the core function ENODLs being in male–female communication and pollen tube reception (Hou et al. 2016). Furthermore, the overexpression of *AtENODL15* by the endogenous promoter results in disturbed pollen tube guidance and reduced fertility (Hou et al. 2016).

To date, the features and functions of the *PC* gene family have been identified and investigated in several plant species, including *Arabidopsis*, rice, Chinese cabbage, maize, and orchid. However, no comprehensive analyses of the *PC* gene family in poplar have been conducted. In the present study, we identified 74 probable *PtPC* genes in *Populus* and performed comprehensive phylogenetic, structural, promoter, gene expansion and microsynteny analyses. Additionally, we selected 18 *PtPCs* to investigate their behaviour under drought and salt treatments. The results provide valuable information about their biological functions and stress responses. Furthermore, analysis of tissue-specific expression of the *PtPC* genes during development showed differences in their spatiotemporal expression patterns, and many were expressed at high levels in roots and xylem. Genome-wide analysis of the *PC* genes in *Populus trichocarpa* will facilitate a better understanding of the role of this gene family during poplar growth and development.

Materials and methods

Identification of *PC* family genes in poplar

The sequences of previously identified *PC* genes in *Arabidopsis* were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/). We performed BLASTP searches (*E* value < $1e^{-6}$) with the *Arabidopsis* PC proteins as queries to identify *PtPCs* in the Phytozome database (http://www.phytozome.net, PF02298). Each protein sequence identified by BLASTP was checked for the existence of a plastocyanin-like domain (PCLD) to confirm membership of the *PC* gene family. The signal peptide (SP), glycosylphosphatidylinositol (GPI) anchor signal, and N-glycosylation sites of PtPCs were predicted by the SignalP 4.1 server (Petersen et al. 2011; Jeßberger et al. 2015), Big-PI Plant Predictor (Eisenhaber et al. 2003) and the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/), respectively. Potential arabinogalactan glycomodules (AGs) were predicted based on previously reported criteria (Mashiguchi et al. 2004). The subcellular localization of all PCs was predicted using the CELLOv2.5 server (http://cello.life.nctu.edu.tw/).

Phylogenetic analysis

Protein sequences and alignments were analyzed using the DNAMAN program, and a phylogenetic tree was constructed with default parameters using the neighbour-joining (NJ) method in MEGA6.0 with 1000 bootstrap replicates (Hu et al. 2010; Tamura et al. 2013).

Analysis of exons/introns, conserved motifs, and chromosomal location

We analyzed the exon/intron structure of *PC* genes by comparing the coding DNA sequence (CDS) and the corresponding genomic DNA sequence using the online GSDS server (http://gsds.cbi.pku.edu.cn/). Conserved motifs were predicted using the online MEME program (http://meme.nbcr.net/meme/cgi-bin/meme.cgi). An image of the chromosomal location was constructed using MapInspect software based on the initial positional information provided in the Phytozome database.

Analysis of microsynteny and gene duplication

A syntenic block is defined as an area in which exons/introns in orthologs are located within 15 genes upstream or downstream in both genomes (Wang et al. 2015). Syntenic blocks within the *PtPC* gene family among chromosomes were acquired from the PGDD database (http://chibba.agtec.uga.edu/duplication). Microsynteny analysis was performed using MicroSyn software, and the online OrthoMCL program (http://orthomcl.org/orthomcl/) was used to analyze duplicated genes.

Evaluation of K_a/K_s values

K_a (number of synonymous substitutions per synonymous site)/K_s (number of non-synonymous substitutions per non-synonymous site) ratios were evaluated using the DnaSP software. Sliding window analysis of K_a per nonsynonymous locus K_a/K_s ratio was also performed with a window size of 150 bp and a step size of 9 bp. Divergence time (*T*) was estimated by $T = K_s/(2 \times 9.1 \times 10^{-9}) \times 10^{-6}$ million years ago (Mya).

Promoter and microarray analysis

The 2000 bp upstream sequences of the *PtPCs* promoter regions were downloaded from the Phytozome database and used to identify the putative *cis*-elements in PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Goodstein et al. 2011). To better understand the expression levels of *PtPCs*, the GSE13990 array data from poplar expression profiling was downloaded from the Gene Expression Omnibus (GEO) database at the NCBI (https://www.ncbi.nlm.nih.gov/). The corresponding probes for *PC* genes were identified using the online ProbeMatch tool available at the NetAffx Analysis Center (http://www.affymetrix.com/analysis/index.affx). The final gene expression data was identified using the corresponding PC-Probe and the GSE13990 data.

Plant materials and stress treatments

The 8-week-old seedlings of *P. trichocarpa* (Torr. & Gray) used in all experiments were cultivated in the Tissue Culture Lab. Plants were treated with 20% PEG (polyethyleneglycol) and 200 mM NaCl for drought and salt stress, respectively. Untreated plants were used as controls. Leaves were collected for RNA extraction at four time points (4, 8, 12 and 24 h) after treatment.

RNA isolation and qRT-PCR analysis after different stress treatments

TRIzol reagent was used to extract RNA from young poplar leaves under different stress treatments and in different organs, and first-strand cDNAs were synthesized. Primer Premier 5.0 and the NCBI primer BLAST tool were used to design and check primers for amplifying *PtPC* genes (Table S1). The poplar housekeeping gene encoding ubiquitin (UBQ, gene ID no. Potri.001G418500) was used as

an internal control for normalizing experimental expression profile data (Hui et al. 2014). qRT-PCR was performed in a 20-µl volume, including 10 µl of 2 × SYBR® Premix Ex Taq™ (TaKaRa, Otsu, Japan), 0.4 µl of 50 × ROX Reference Dye, 2 µl diluted cDNA template, 0.8 µl of each specific primer, and 6 µl ddH_2O. The qPCR reaction conditions were as follows: 95 °C for 30 s, followed by 40 thermal cycles of denaturation at 95 °C for 5 s, and annealing at 55–60 °C for 34 s. The relative expression levels were calculated using the ΔΔCT method. It is noteworthy that relative gene expression [$2^{-\Delta\Delta CT,\ CK\ (0\ h)}$] for each gene in the control plants was normalized to 1 as described previously for stress treatments (Schmittgen and Livak 2008). GraphPad software was used for statistical analysis, and three biological and technical replicates were performed for each sample.

Results

Identification of *PC* family genes in poplar

We identified 77 potential PC protein sequences in poplar, of which three were paired with identical sequences, resulting in 74 putative *PtPCs* (Table 1). Based on multiple sequence alignments (Fig. 1), *PtPCs* were classified into three subfamilies: uclacyanin-like proteins (*PtUCs*, 7), stellacyanin-like proteins (*PtSCs*, 19) and plantacyanins (*PtPLCs*, 3) based on the predicted copper-binding ligands (His, Cys, His and Met/Gln). The remaining 45 *PtPCs* belonged to the ENODL family based on the modified copper-binding residues. Moreover, the prediction of subcellular localization indicated that 93.2% of *PtPCs* were associated with the plasma membrane or were extracellular (Table 1), and only five PCs (*PtENODL9*, *PtENODL27*, *PtUC6*, *PtSC8* and *PtSC18*) were localized in the nucleus.

Phylogenetic and structural analysis of *PtPCs*

To better understand their structure and function, we predicted N-terminal signal peptides (SPs), glycosylphosphatidylinositol (GPI) anchor signals (GASs), AG glycomodules and N-glycosylation (Fig. S1; Table 1). The results revealed that 60 *PtPCs* had a predicted SP involved in targeting to the endoplasmic reticulum. Additionally, 49 *PtPCs* had GASs related to the localization at the plasma membrane. About 58.1% of poplar genes contained hypothetical AG glycomodules in the PAST-rich region (Pro, Ala, Ser, Thr). There were 60 *PtPCs* with putative N-glycosylation sites in the PCLD and PAST-rich region.

According to the predicted domain structures, *PtPCs* were divided into six types (Fig. 2): Type I includes a N-terminal SP, a PCLD, an arabinogalactan-like region (ALR), and a C-terminal GAS; by contrast, GAS is absent from type II; type III resembles type I but lacks ALR; type IV only has SP and PCLD; type V lacks SP compared with type III; type VI only contains PCLD.

To investigate the evolutionary relationships of *PC* genes in *Arabidopsis*, *Zea mays* and poplar, we constructed a phylogenetic tree of *AtPC*, *ZmPC* and *PtPC* protein sequences. The 38 *AtPCs*, 60 *ZmPCs* and 74 *PtPCs* were divided into seven clades, consistent with previous *Arabidopsis*, rice and *Phalaenopsis equestris* studies (Fig. 3). The results indicated that clade VII had the most number of *PC* gene members (61), while clade VI contains the fewest PCs (12). All members of clade VI and clades VII entirely belong to the same subfamily, put it another way, the clade VI and the clades VII wholly contain SCs and ENODLs, respectively. Moreover, except for *ZmUC18/26* and *PtUC7*, all PCs in the clade IV belonged to the ENODL subfamily.

Gene structure and conserved motifs

It was well known that gene structural diversity resulted from the evolution of multi-gene families. The exon/intron structures of *PtPCs* are shown in Fig. S2. The gene structures of *PtPCs* are not complex and most include two introns. Like *PtSC8*, *PtENODL1*, *3*, *34, 35*, and *38*, have only one intron, and eight *PtPCs* contain no introns. In addition to exon/intron pattern, conserved motifs could also be important for the diversified functions of PCs. We identified 10 different conserved motifs (Fig. 4), suggesting that members of the same subfamily shared a similar motif structure. The details of each conserved motif are shown in Table S2. Motif 1, motif 2, motif 6 and motif 10 are related to intermolecular electron transfer reactions, while others have no functional annotation. Most *PtPCs* include motif 1, 3, and 6, in the order 1-6-3. Motif 1 is the most common motif, present in all *PtPC* genes. Motif 7, 9 and 10 are only present in *PtENODLs*. *PtUC1* differs from other *PtUCs* by lacking motif 6, and all *PtSCs* have motif 4 apart from *PtSC7*.

Chromosomal location, gene duplication and conserved microsynteny

Based on the chromosomal location map, 74 *PtPCs* were randomly distributed on the 17 poplar chromosomes (Fig. 5). Chromosome 1 has the largest number of *PC* genes (14), while chromosomes 8 and 19 only have one *PtPC* gene, and chromosomes 4, 11, 13, 15, 16 and 17 only contain ENODL subfamily genes.

Many gene families in plants appear to be generated by expansion through segmental or tandem duplication. To better comprehend the evolution of poplar *PC* genes, we investigated genome duplication events in this family. Duplicated genes were confirmed using the Vista Synteny browser (http://pipeline.lbl.gov/cgi-bin/gateway2), which indicated

Table 1 List of *PC* genes identified in poplar and their sequence characteristics

Name	Gene identifier	Chromosome location	Type b	SP	GPI	AG	N-glyco	Cu sites	Location	Reliable-index
PtENODL1	Potri.001G043600.1	Chr01:3153619..3155889 forward	V	−	+	−	+	Y,I,Q,M	PlasmaMembr	2.833*
PtENODL2	Potri.001G085100.1	Chr01:6745838..6747275 forward	I	+	+	+	+	D,S,H,–	PlasmaMembr	2.549*
PtENODL3	Potri.001G114200.1	Chr01:9138789..9140003 reverse	IV	+	−	−	−	H,–,H,–	Extracellular	1.997*
PtENODL4	Potri.001G187700.1	Chr01:16594525..16596114 reverse	I	+	+	+	+	D,S,H,Q	Extracellular	2.294*
PtENODL5	Potri.001G219800.1	Chr01:22559753..22560344 reverse	IV	+	−	−	+	Y,L,Y,M	PlasmaMembr	1.565*
PtENODL6	Potri.001G219900.1	Chr01:22560755..22561735 reverse	III	+	+	−	+	Y,I,S,M	Extracellular	1.418*
PtENODL7	Potri.001G273000.1	Chr01:27994628..27997398 forward	II	+	−	+	+	H,G,Q,Q	Extracellular	3.693*
PtENODL8	Potri.001G338800.1	Chr01:34229770..34230346 reverse	II	+	−	+	+	D,S,H,Q	Extracellular	2.572*
PtENODL9	Potri.001G398800.1	Chr01:41953025..41959574 forward	I	+	+	+	+	D,S,N,Q	Nuclear	2.921*
PtENODL10	Potri.001G419200.1	Chr01:44389327..44390703 forward	I	+	+	+	+	D,S,N,–	Extracellular	2.89*
PtENODL11	Potri.002G073800.1	Chr02:5106875..5107849 forward	IV	+	−	−	−	H,C,H,–	Extracellular	3.222*
PtENODL12	Potri.002G150600.1	Chr02:11346486..11347737 forward	I	+	+	+	+	Q,D,Q,–	Extracellular	3.66*
PtENODL13	Potri.002G241500.1	Chr02:23400033..23400930 reverse	VI	−	−	−	+	H,C,H,–	Extracellular	3.95*
PtENODL14	Potri.003G047300.1	Chr03:6843123..6843965 reverse	I	+	+	+	−	H,C,H,–	Extracellular	2.254*
PtENODL15	Potri.003G050500.1	Chr03:7416338..7417304 reverse	I	+	+	+	+	D,S,H,Q	PlasmaMembr	3.724*
PtENODL16	Potri.003G117900.1	Chr03:14120390..14121023 forward	I	+	+	+	+	H,C,H,–	Extracellular	2.331*
PtENODL17	Potri.003G183300.1	Chr03:18973467..18974902 reverse	V	−	+	NC	+	Y,I,Q,M	PlasmaMembr	3.379*
PtENODL18	Potri.004G121100.1	Chr04:11381069..11382693 reverse	III	+	+	−	+	M,I,F,M	Extracellular	2.601*
PtENODL19	Potri.004G169700.1	Chr04:18831956..18832471 forward	I	+	+	+	−	H,S,H,Q	Extracellular	3.178*
PtENODL20	Potri.004G171100.1	Chr04:19035084..19035571 forward	II	+	−	+	−	,–,S,H,Q	Extracellular	3.439*
PtENODL21	Potri.006G009000.1	Chr06:596909..598686 forward	I	+	+	+	−	G,S,K,–	Extracellular	1.57*
PtENODL22	Potri.006G184100.1	Chr06:19778884..19779912 forward	I	+	+	+	+	D,S,H,Q	Extracellular	1.949*
PtENODL23	Potri.006G264600.1	Chr06:26754833..26755600 forward	I	+	+	+	+	D,S,H,Q	Extracellular	2.083*
PtENODL24	Potri.007G104600.1	Chr07:12920703..12922699 reverse	IV	+	−	−	+	A,C,Q,M	Extracellular	3.194*
PtENODL25	Potri.009G067300.1	Chr09:6734134..6737755 forward	II	+	−	+	+	H,G,Q,Q	Extracellular	3.415*
PtENODL26	Potri.010G243600.1	Chr10:21935770..21936830 reverse	I	+	+	+	+	F,C,Y,–	PlasmaMembr	3.233*
PtENODL27	Potri.011G117800.1	Chr11:14318707..14320094 forward	I	+	+	+	+	D,S,N,Q	Nuclear	2.92*
PtENODL28	Potri.011G135400.1	Chr11:15924181..15925492 forward	I	+	+	+	+	D,S,H,–	Extracellular	3.187*
PtENODL29	Potri.013G030000.1	Chr13:1981875..1982471 forward	I	+	+	+	+	H,C,H,–	PlasmaMembr	2.937*
PtENODL30	Potri.013G054500.1	Chr13:4071268..4071889 forward	II	+	−	+	+	H,C,H,–	Extracellular	1.678*
PtENODL31	Potri.014G072700.1	Chr14:5870773..5871448 forward	II	+	−	+	+	Q,D,Q,–	Extracellular	2.389*
PtENODL32	Potri.015G052000.1	Chr15:6544233..6544871 forward	V	−	+	NC	+	D,S,H,Q	PlasmaMembr	2.735*
PtENODL33	Potri.015G113300.1	Chr15:12853236..12853705 reverse	IV	+	−	−	+	D,S,H,Q	PlasmaMembr	1.982*
PtENODL34	Potri.015G114300.1	Chr15:12983389..12984327 reverse	IV	+	−	−	−	D,S,H,Q	Extracellular	2.07*
PtENODL35	Potri.015G114700.1	Chr15:12995004..12995486 reverse	IV	+	−	−	−	D,S,H,Q	Extracellular	2.542*
PtENODL36	Potri.015G115600.1	Chr15:13099745..13100984 forward	VI	−	−	NC	−	D,S,H,Q	PlasmaMembr	1.965*

Table 1 (continued)

Name	Gene identifier	Chromosome location	Type b	SP	GPI	AG	N-glyco	Cu sites	Location	Reliable-index
PtENODL37	Potri.015G117100.1	Chr15:13225870..13227122 forward	VI	–	–	NC	–	D,S,H,Q	PlasmaMembr	2.107*
PtENODL38	Potri.016G015200.1	Chr16:829100..830322 reverse	V	–	+	NC	–	G,S,K,–	Extracellular	2.007*
PtENODL39	Potri.016G050700.1	Chr16:3224432..3224818 reverse	IV	+	–	–	+	H,I,E,–	Extracellular	3.724*
PtENODL40	Potri.017G011200.1	Chr17:951902..952811 reverse	I	+	+	+	+	D,S,D,–	Extracellular	1.645*
PtENODL41	Potri.017G012300.1	Chr17:1038196..1039065 reverse	I	+	+	+	+	D,S,D,–	Extracellular	1.597*
PtENODL42	Potri.017G088500.1	Chr17:10587935..10589774 reverse	I	+	+	+	+	Y,L,Y,M	PlasmaMembr	2.909*
PtENODL43	Potri.017G088600.1	Chr17:10596482..10598961 reverse	III	+	+	–	+	M,I,F,M	Extracellular	2.459*
PtENODL44	Potri.018G018200.1	Chr18:1393715..1394619 reverse	I	+	+	+	+	D,S,H,Q	Extracellular	1.647*
PtENODL45	Potri.019G037800.1	Chr19:4280478..4281067 reverse	I	+	+	+	+	H,C,H,–	Extracellular	2.747*
PtUC1	Potri.001G080700.1	Chr01:6385809..6386749 forward	I	+	+	+	+	H,C,H,M	Extracellular	4.24*
PtUC2	Potri.001G209300.1	Chr01:21013646..21014419 reverse	IV	+	–	–	+	H,C,H,M	Extracellular	2.817*
PtUC3	Potri.002G101200.1	Chr02:7311304..7312368 reverse	I	+	+	+	+	H,C,H,M	Extracellular	2.097*
PtUC4	Potri.002G101300.1	Chr02:7320271..7321432 reverse	I	+	+	+	+	H,C,H,M	Extracellular	3.688*
PtUC5	Potri.003G150300.1	Chr03:16446774..16447642 reverse	I	+	+	+	+	H,C,H,M	Extracellular	4.113*
PtUC6	Potri.007G120200.1	Chr07:13942892..13943817 reverse	I	+	+	+	+	H,C,H,M	Nuclear	2.383*
PtUC7	Potri.009G136200.1	Chr09:10959367..10960342 forward	IV	+	–	–	+	H,C,H,M	PlasmaMembr	2.363*
PtPLC1	Potri.001G332200.1	Chr01:33568341..33569287 reverse	I	+	+	+	–	H,C,H,M	Extracellular	3.038*
PtPLC2	Potri.002G074000.1	Chr02:5115230..5116446 forward	IV	+	–	–	–	H,C,H,M	Extracellular	3.482*
PtPLC3	Potri.014G049600.1	Chr14:3976460..3977486 forward	I	+	+	+	–	H,C,H,M	Extracellular	2.279*
PtSC1	Potri.001G192100.1	Chr01:17483564..17484349 reverse	I	+	+	+	+	H,C,H,Q	Extracellular	1.839*
PtSC2	Potri.001G268700.1	Chr01:27654231..27654728 forward	I	+	+	+	+	H,C,H,Q	Extracellular	3.885*
PtSC3	Potri.002G052500.1	Chr02:3480730..3481467 forward	I	+	+	+	+	H,C,H,Q	Extracellular	3.329*
PtSC4	Potri.002G156100.1	Chr02:11751550..11751981 forward	V	–	+	NC	+	H,C,H,Q	Extracellular	3.496*
PtSC5	Potri.002G156400.1	Chr02:11764955..11765386 forward	V	–	+	NC	+	H,C,H,Q	Extracellular	3.442*
PtSC6	Potri.002G161300.1	Chr02:12175435..12175944 forward	I	+	+	+	+	H,C,H,Q	Extracellular	4.138*
PtSC7	Potri.006G067200.1	Chr06:5035410..5037170 reverse	VI	–	–	NC	+	H,C,H,Q	Extracellular	2.268*
PtSC8	Potri.006G067300.1	Chr06:5041892..5043953 reverse	V	–	+	NC	+	H,C,H,Q	Nuclear	2.42*
PtSC9	Potri.006G067400.1	Chr06:5050605..5051796 reverse	III	+	+	–	+	H,C,H,Q	Extracellular	2.43*
PtSC10	Potri.006G259000.1	Chr06:26257485..26257937 forward	V	–	+	NC	+	H,C,H,Q	Extracellular	3.919*
PtSC11	Potri.006G259100.1	Chr06:26261015..26261502 forward	VI	–	–	NC	+	H,C,H,Q	Extracellular	2.792*
PtSC12	Potri.008G151000.1	Chr08:10210912..10213390 reverse	II	+	–	+	+	H,C,H,Q	Extracellular	1.625*
PtSC13	Potri.010G089900.1	Chr10:11367693..11370059 forward	II	+	–	+	+	H,C,H,Q	PlasmaMembr	2.394*
PtSC14	Potri.013G061300.1	Chr13:4639750..4640351 reverse	II	+	–	+	+	H,C,H,Q	PlasmaMembr	1.458*
PtSC15	Potri.018G128800.1	Chr18:15080350..15081733 reverse	III	+	+	–	+	H,C,H,Q	PlasmaMembr	3.336*
PtSC16	Potri.018G128900.1	Chr18:15086508..15088125 reverse	III	+	+	–	+	H,C,H,Q	PlasmaMembr	3.263*
PtSC17	Potri.018G129000.1	Chr18:15095358..15096492 reverse	V	–	+	NC	+	H,C,H,Q	PlasmaMembr	4.234*

Table 1 (continued)

Name	Gene identifier	Chromosome location	Type b	SP	GPI	AG	N-glyco	Cu sites	Location	Reliable-index
PtSC18	Potri.018G129200.1	Chr18:15101044..15102292 reverse	I	+	+	+	+	H,C,H,Q	Nuclear	2.098*
PtSC19	Potri.018G129400.1	Chr18:15108397..15109586 reverse	I	+	+	+	+	H,C,H,Q	PlasmaMembr	2.071*

SP signal peptide, *GPI* glycosylphosphatidylinositol, *AG* arabinogalactan, *N-glyco* N-glycosylation, *NC* not checked as for lacking a signal peptide in its precursor protein backbone

* indicates that each gene has the highest reliable-index of subcellular localization in different cellular

that 11 pairs of genes originated from segmental duplication (Table S3). Neighbouring genes were also analyzed to determine whether tandem duplication had occurred (Ye et al. 2009), but there was no evidence of this phenomenon in *PtPCs*.

To examine the evolutionary selection process, we calculated K_a/K_s ratios of 10 pairs of *PtPC* paralogs (Table 2). Almost all K_a/K_s ratios were < 0.5, but the *PtSC8/PtSC17* pair was > 1. Moreover, the duplication events of the 10 gene pairs were estimated to have occurred between 10.20 and 55.80 million years ago (Mashiguchi et al. 2004), based on sliding-window analysis of K_a/K_s ratios for all *PC* paralog pairs (Fig. S3).

To probe the relationships between homologous genes, we investigated the microsynteny between poplar and *Arabidopsis* sequences (Fig. S4). Pairwise comparison of flanking genes in chromosomal regions containing *PC* genes revealed three or more pairs displaying conserved microsynteny. Analysis of intraspecies microsynteny identified 23 collinear gene pairs in poplar (Fig. S5). Microsynteny analysis was also performed to evaluate the relationship between paralogous and orthologous *PC* genes, and 15 orthologous gene pairs with two-for-one microsynteny were identified between *Populus* and *Arabidopsis* sequences, but only six pairs were identified with one-for-one microsynteny, including *AtENODL20-PtENODL5*, and *At1g22480-PtUC3/AtENODL8-PtENODL2*.

Identification of *cis*-regulatory elements in the promoters of *PtPCs*

To investigate gene function and regulation, we analyzed *cis*-regulatory elements in the promoters of *PtPCs* (Table S4). In previous studies, we found that *cis*-regulatory elements play a pivotal role in controlling physical and reproductive growth, phytohormone responses, and abiotic and biotic stress responses. As shown in Fig. 6, endosperm expression elements (Skn-1_motif and GCN4_motifs) were found in 17 *PtPCs* and 6 *PtPCs*, respectively (Washida et al. 1999). The Skn-1_motif was the most abundant *cis*-element in the promoter sequences of 17 *PtPCs*, and the Skn-1_motif was the most populous in *PtSC6*. The CAT-box, RY-element and CCGTCC -box are present in two *PtPCs*, but MBSII, HD-Zip1 and HD-Zip2 are only present in one *PtPC* gene (Sessa et al. 1993; Bobb et al. 1997). We also identified the circadian control element and the O_2 site, which were involved in zein metabolism regulation, in the promoters of *PtPCs*, and ten *cis*-elements related to phytohormone responses were detected (Anderson et al. 1994). The CGTCA and TGACG motifs that are related to MeJA responsiveness were found in seven *PtPCs* (Nejad et al. 2012). The TCA element, ABRE, ERE and TGA element were found in the

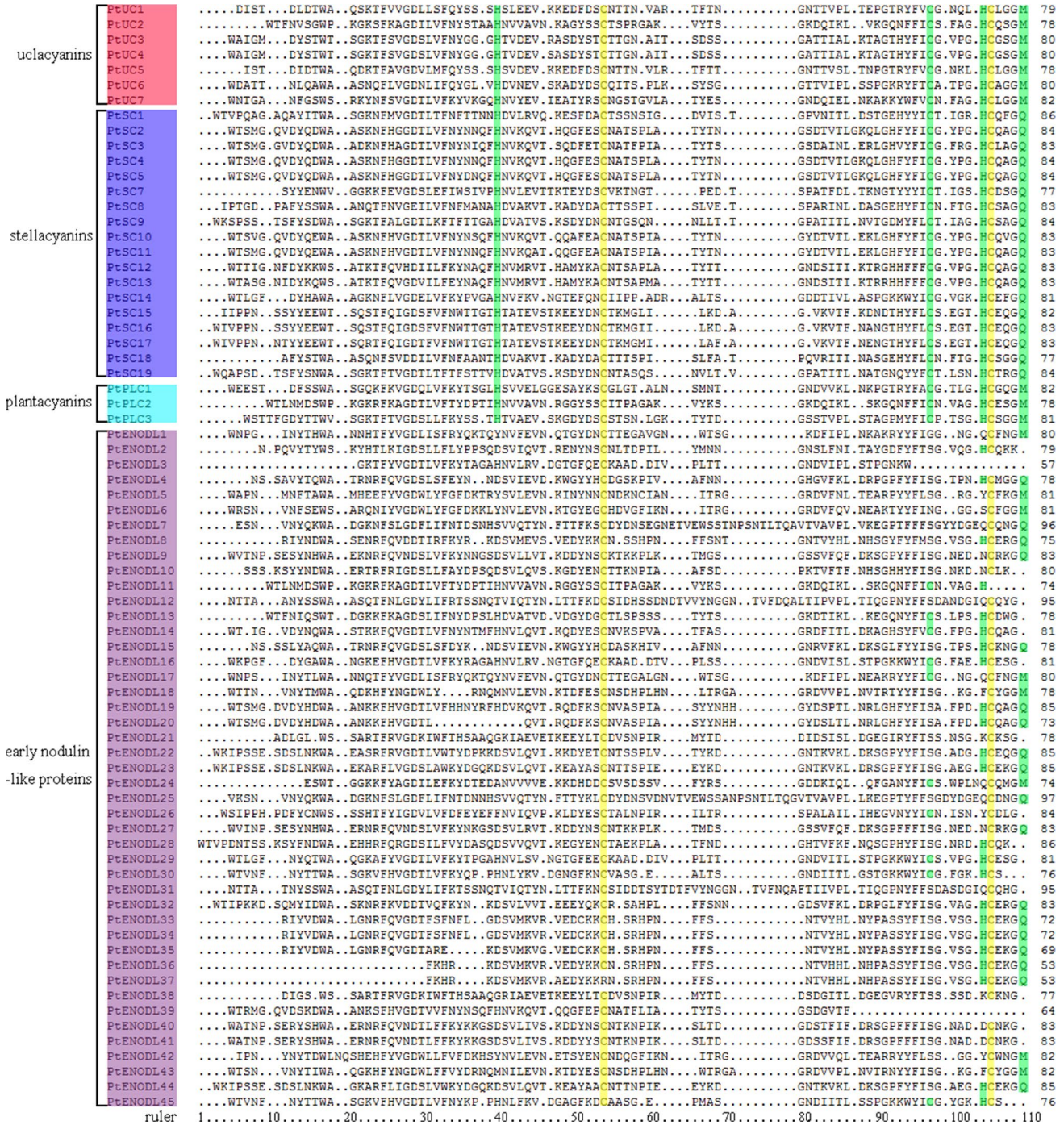

Fig. 1 Multiple sequence alignment of the amino acid sequences of the plastocyanin-like domains (PCLD) of *PtPC* proteins in *Populus trichocarpa*. Red, uclacyanins (UCs); blue, stellacyanins (SCs); cyan, plantacyanin (PLCs); purple, early nodulin-like proteins (ENODLs). The conversed amino acids involved in copper binding are highlighted by a green background (His, Cys, His, and Gln/Met), while the Cys residues involved in disulfide linkage are indicated by a yellow background

promoters of 13, 8, 6 and 3 *PtPCs*, respectively (Goldsbrough et al. 1993; Shen and Ho 1995). Other elements, including the AuxRR core, TATC-box, GARE motif and P-box were also observed in 1, 2, 3 and 4 *PtPCs*, respectively (Washida et al. 1999). In addition, we also identified some *cis*-elements related to abiotic and biotic stresses, including HSE and TC-rich repeats in 14 *PtPCs*, but the WUN motif was only present in *PtSC6*.

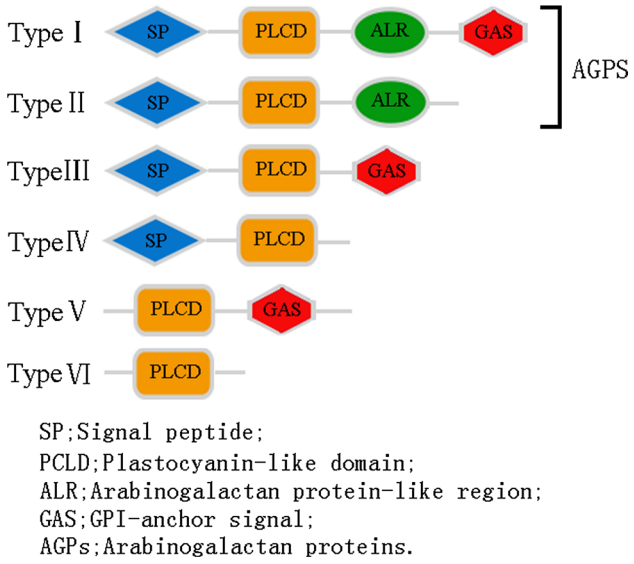

Fig. 2 Schematic representation of five groups of *PtPCs*. The diagram showing the features of *PtPC* domains was generated with MyDomains (http://prosite.expasy.org/cgi-bin/prosite/mydomains/). The figure is not drawn to scale

Expression of *PtPC* genes

To reveal the different evolutionary fates of duplicated genes, we analyzed the expression patterns of *PtPCs* in six different tissues of poplar: young leaves, roots, xylem, female catkin, male catkin, and mature leaves (Fig. S6; Table S5). The heatmap showed that 11 *PtPCs* (*PtPLC1, PtENODL43, PtENODL7, PtSC6, PtUC2, PtENODL45, P9tSC, PtENODL41, PtSC10, PtENODL25, PtUC1*) were highly expressed in six organs. Notably, *PtPLC1* and *PtUC4* have relatively high expression levels only in roots and xylem, respectively.

Examination of *PC* gene expression by qRT-PCR

To confirm the expression of *PtPCs* in response to stresses, we used qRT-PCR to analyze relative expressions of 18 *PtPCs* under drought and salt treatments (Figs. 7, 8). The results of drought treatments showed that three *PtPCs* (*PtSC9*, *PtENODL24* and *PtENODL41*) were highly expressed at all four-time points measured, and some *PtPCs* were up-regulated at particular time points. For instance, *PtPLC1* was only up-regulated at 4 h after treatment. *PtUC7*, *PtSC13* and *PtSC17* were distinctly up-regulated at 24 h. By contrast, expression of *PtENODL9* remained low at all time points, while *PtENODL45* was observably down-regulated at 24 h after treatment.

Subsequently, we analyzed the expression patterns of *PtPCs* under salt treatments. The expression levels of *PtEDNOL9* remained low at all time points, but *PtPLC1* was down-regulated at 4 h yet up-regulated at other time points. Expression of *PtENODL24* and *PtENODL41* obviously showed a high expression under both, drought and salt treatment, at all time points, and expression of three *PtPCs* (*PtUC7, PtSC13* and *PtSC17*) was significantly increased at 24 h after salt treatment. Additionally, *PtUC2* was distinctly up-regulated at all time points except at 12 h. *PtUC7*, *PtSC13* and *PtSC17* had the similar expression after both drought and salt treatments at all time points.

To predict possible functions of *PtPC* genes in organ development, we also performed qRT-PCR analyses to examine the relative expression of the 18 *PtPCs* in five organs, including young leaves (YL), roots (RT), xylem (XY), mature leaves (ML), and phloem (PH) (Fig. 9; Table S6). As shown in Fig. 9, *PtPCs* showed the highest mRNA accumulation in all the tissues, including three in the roots (*PtUC1, PtENODL45, PtENODL41*) and young leaves (*PtENODL24, PtENODL9, PtSC6*), four in the mature leaves (*PtENODL24, PtSC17, PtENODL7, PtUC7)* and the phloem (*PtENODL9, PtENODL25, PtENODL33, PtENODL43*), five in the xylem (*PtUC2, PtUC4, PtSC10, PtSC13, PtPLC1*). However, most of the remaining genes had intensively different expression patterns. For example, *PtENODL45* had low expression levels in mature leaves. It is worth noting that *PtSC17* and *PtENODL24* were both up-regulated in the mature leaves while they were expressed at a relatively low level in young leaves and xylem, respectively.

Discussion

Genome-wide analysis of *PC* genes in plants has been performed previously in various species including *Arabidopsis thaliana* (Mashiguchi et al. 2009), *Oryza sativa* (Ma et al. 2011), *Brassica rapa* (Li et al. 2013), *Zea mays* (Cao et al. 2015) and *Phalaenopsis equestris* (Xu et al. 2017). All of these studies indicated a role in plant growth and development (Fedorova et al. 2002; Ozturk et al. 2002; Diab et al. 2004; Ma and Jie 2010; Wu et al. 2011), but *PC* genes have not been investigated extensively in poplar. In the present study, we identified 74 *PtPCs* and grouped them into seven clades based on phylogenetic tree analysis. The phylogenetic tree showed that most genes of each subfamily in poplar grouped tightly with these genes of subfamily in *Arabidopsis* and *Zea mays*, which is consistent with the fact that maize, *Arabidopsis* and poplar diverged from a common ancestor before the divergence of monocot and dicot lineages. AGPs belong to a subfamily of hydroxyproline-rich glycoproteins (HRGPs), and are involved in the growth and development of the plant, such as stem strength, somatic embryogenesis in cotton, cell culture and extracellular signals transduction (Tan et al. 2004; Seifert and Roberts 2007; MacMillan et al. 2010; Poon et al. 2013; Ma et al.

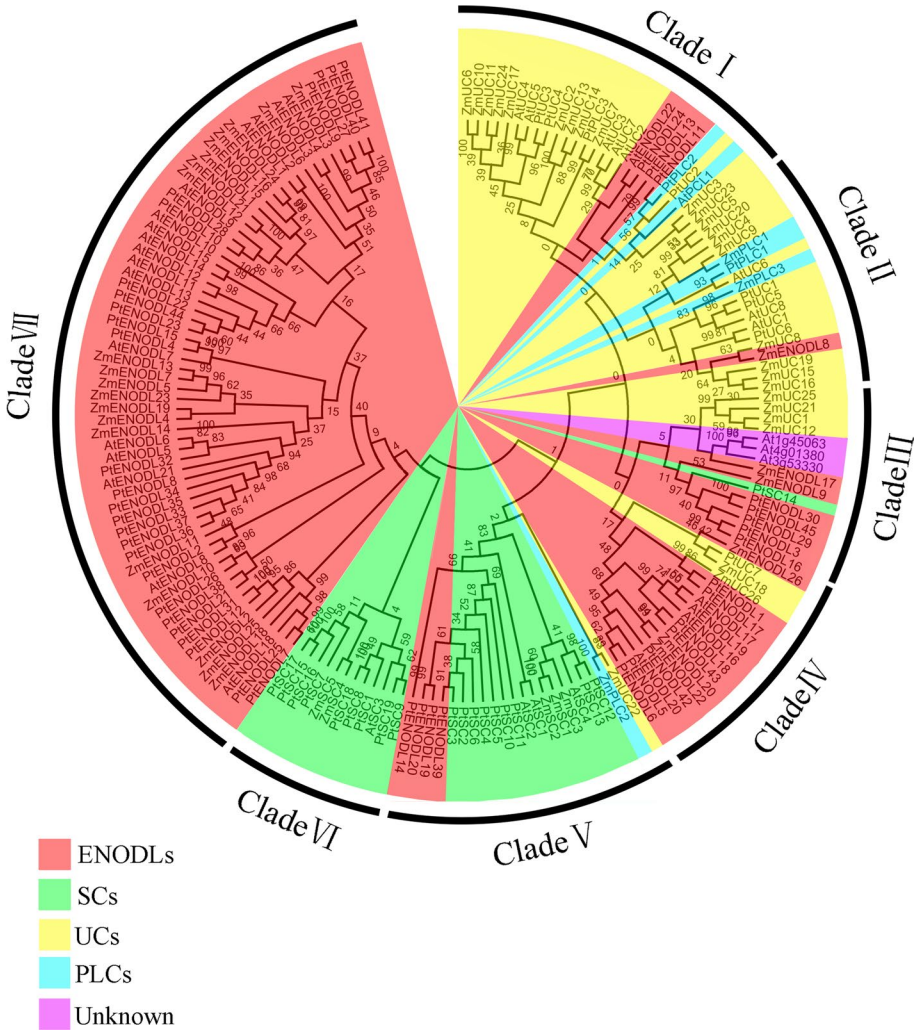

Fig. 3 Phylogenetic tree for Phytocyanins (PCs) from *Populus trichocarpa*, *Arabidopsis*, and *Zea mays*. The phylogenetic tree was constructed using PC protein sequences from the three plant species by the neighbour-joining method in MEGA6.0, with 1000 bootstrap replicates displayed at each node. PC proteins are divided into seven groups, and different subfamilies are indicated by different colours. Red, early nodulin-like proteins (ENODLs); green, stellacyanins (SCs); yellow, uclacyanins (UCs); cyan, plantacyanin (PLCs); purple, unknown

2017). Several bioinformatics studies were used to identify the AGP gene family in plants. For instance, the PAST amino acid bias was calculated for AGPs in *Arabidopsis* as well rice (Schultz et al. 2002; Ma and Zhao 2010), the well-designed BIO OHIO program was employed for HRGPs in *Arabidopsis* and poplar (Showalter et al. 2010, 2016), a Python script named Finding-AGP was utilized in 47 plant specie (Ma et al. 2017), MAAB bioinformatics pipeline was conducted to classify HRGPs (Johnson et al. 2017), and BLAST searches were performed for *AtENODLs*, *OsPCs* and *PePCs* (Mashiguchi et al. 2009; Ma et al. 2011; Xu et al. 2017). Through these methods, the classical AGPs, lysine-rich AGPs, AG peptides, fasciclin-like AGPs, plastocyanin AGPs and other chimeric AGPs have been identified in *Arabidopsis*, rice and poplar. Chimeric AGPs were determined if the protein sequences contained at least one arabinogalactosylated domain and a domain with an unrelated motif (Schultz et al. 2002). In our study, 43 *PtPCs* contain AG glycomodules and SPs, including 27 *PtENODLs*, five *PtUCs*, nine *PtSCs* and two *PtPLCs* that might belong to the chimeric AGP group. Showalter et al. (2016) have identified 39 *PtPAGs* in poplar by BIO OHIO 2.0 bioinformatics program, and all *PtPAGs* belonged to the AGPs we identified in the present study except two *PtPAGs* (*PtPAG25* and *PtPAG27*). Moreover, eight additional *PtPCs* (*PtENODL2/7/8/12/19/21/25/31*) might be the member of a AGP superfamily due to the presence of putative arabinogalactan glycomodules found in our study. The previous report has identified 18 ENOD-like, seven UC-like and four SC-like AGPs in *Arabidopsis* (Li et al. 2013). There were about 1.5 times more ENOD-like, 0.71 times more UC-like and 2.25 more SC-like AGPs in poplar than in *Arabidopsis*, which may result from the evolutionary diversification of the two plants.

Multiple sequence alignment confirmed that Cys residues involved in the formation of disulfide linkages are highly conserved in all 74 *PtPCs*, suggesting they are essential for maintaining PCLD structure, function and stability. PCs are involved in electron transport in the cytomembrane. This process is related to photophosphorylation and can affect

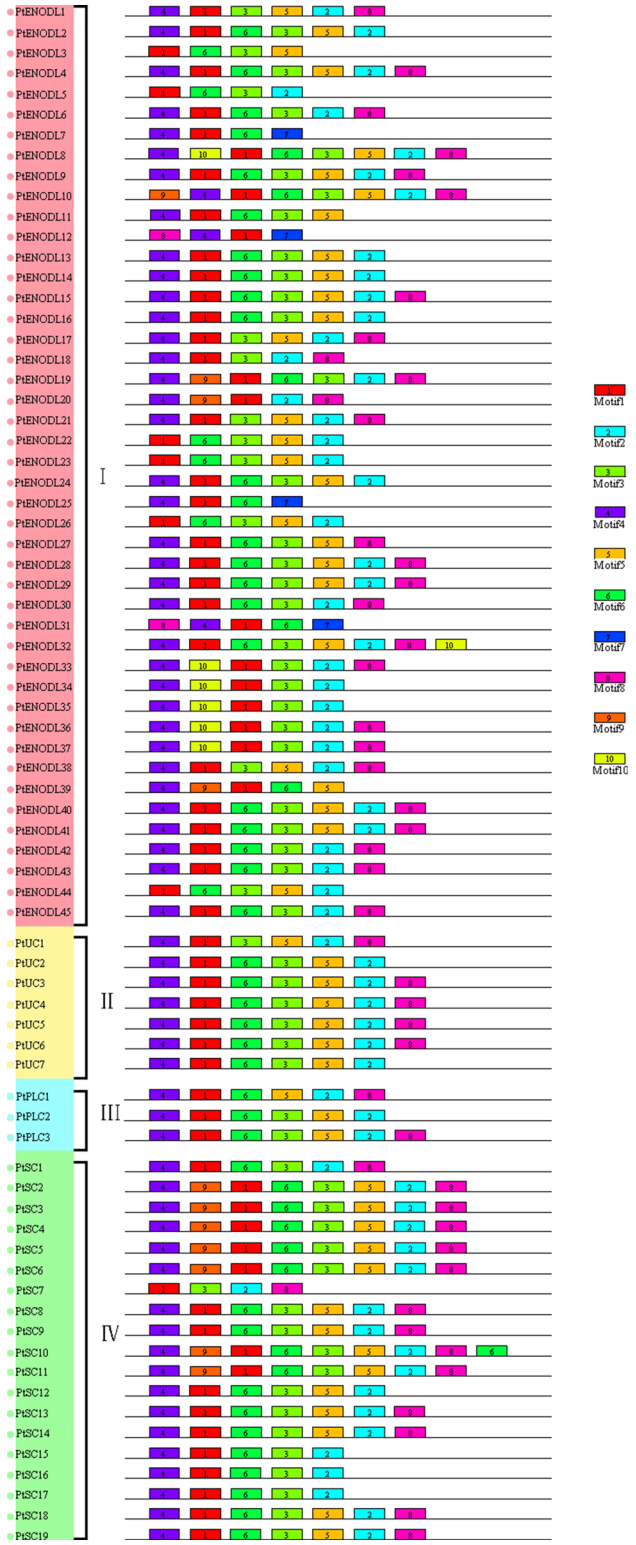

Fig. 4 Schematic representation of the 10 conserved motifs in *PtPC* proteins. Motifs of the *PtPC* proteins were identified by MEME online tool. Each motif is represented by a differently coloured block, with their numbers in the centre of the motifs. The number in boxes (1–10) represents motif 1–motif 10, respectively. The position and length of each coloured box represents the actual motif size

ATP generation, and therefore influences physiological processes in plants.

Several reports on subcellular localization of PCs have provided information on their specific functions in plants. For instance, *AtSC3* is located in the plasma membrane and related to aluminum as well oxidation stresses (Ezaki et al. 2001, 2005), which is possibly the response to other abiotic stresses. Khan et al. (2007) found that *AtENODL9-1*, one of the double mutants of *AtENODL9*, which is located in the sieve element plasma membrane, plays a more significant role in reproductive processes than irregular physiological processes, as revealed by analysis of the phenotype of homozygous T-DNA insertion mutants. All known PCs have secretion signals, which are essential to localize in the extracellular space. According to reports, more than half of ENODLs were predicted to be GPI-anchored proteins, implying theirs functions at the plasma membrane or extracellular matrix. Many proteins exist in the cell in a dynamic form, and they can perform diverse cell functions in different subcellulars. In addition, the plants often grow under a variety of environmental conditions, which may make some proteins, such as transcription factors, appear in different organelles during signal transduction. Our current study revealed that most *PtPCs* were localized in the plasma membrane or extracellular, and the phenomenon demonstrated that the different subcellular localization of phytocyanins may be due to diverse function in poplar. *PtPCs* may exert a significant part in the growth and the development of poplar, and they were divided into four subfamilies.

The exon/intron structure and motif arrangement are highly conserved in all subfamilies. The 74 identified *PtPCs* include different numbers of introns, indicating diverse functions during their evolution. The similar exon/intron and motif composition of members within subfamilies imply conservation of function, and the phylogenetic analysis can, therefore, be beneficial. The specific motifs are ordered 1-6-3 in all sequences, indicating strong conservation of domain structure. Motifs 7, 9 and 10 are only present in members of the ENODL subfamily, indicating a role in the functional divergence of *PtENODLs*. Notably, all PLCs lack motif 7 in poplar, demonstrating divergence from other subfamilies during evolution and its probable specific-role in plant development and growth, which could be important to their functions during electron transportation. The identified subfamily has different motifs which may play a crucial role in the different functions that their genes perform. These specific motifs present in the *PtPCs* are most probably the structural basis for their diverse functions.

The endosperm offers diverse nutrients to seedling growth after germination before it turns into photosynthetic and self-sufficiency (Zhang et al. 2016). Most *PtPCs* had Skn-1_motif *cis*-regulatory elements were relevant to endosperm expression. Analysis of promoter regions showed

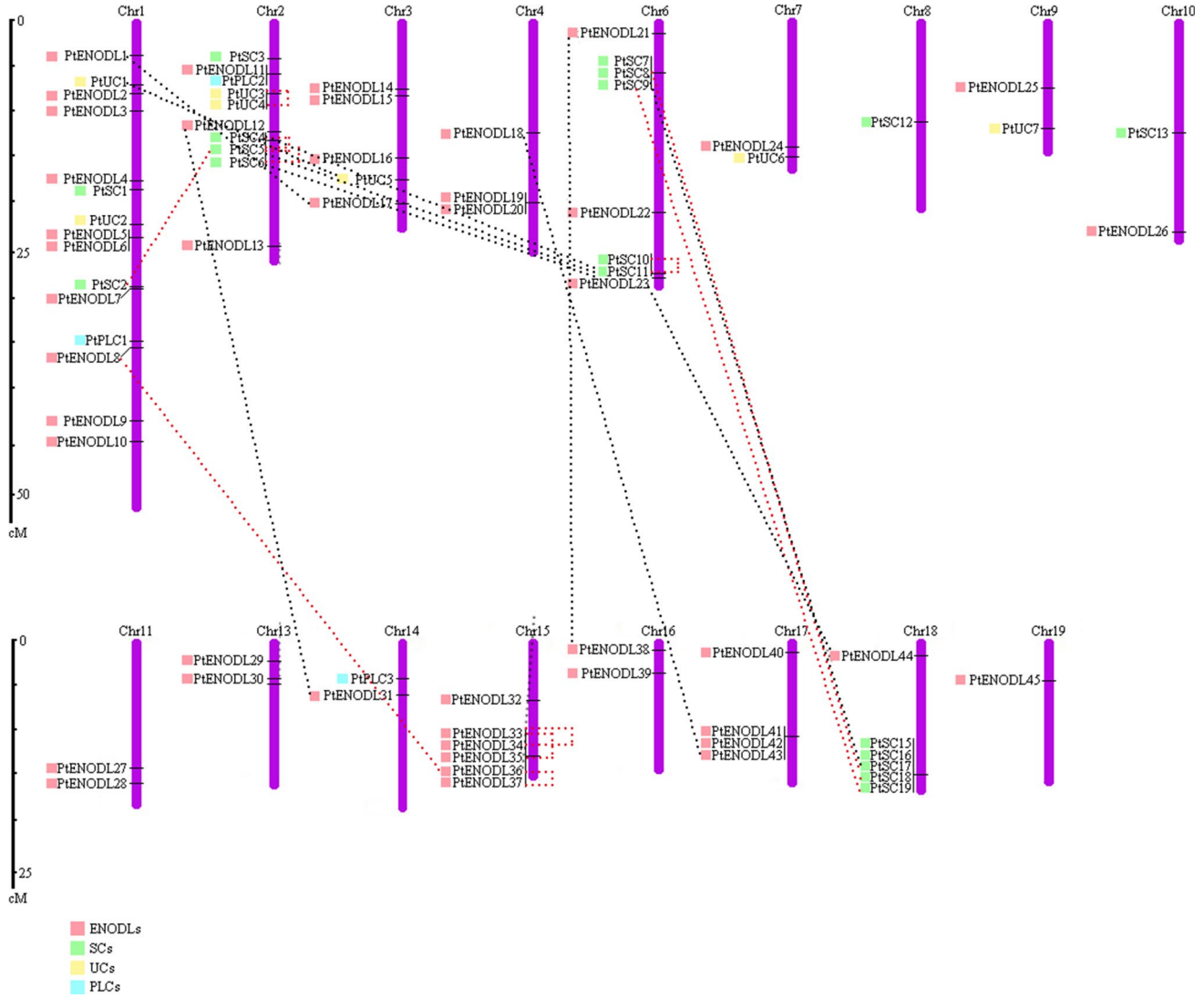

Fig. 5 Chromosomal locations of the 74 predicted *PtPC* genes. The chromosome number is indicated above of each chromosome. Duplicated paralogous pairs of *PtPC* genes and segmental duplication genes are connected by black and red dashed lines in different colours

that the majority of *PtPCs* contained TC-rich repeats *cis*-regulatory elements, which suggested that *PtPC* genes also played a significant role in stress responses. In *Boea crassifolia*, *BcBCP1*, an ENOD-like gene, increases tolerance to osmotic stress in transgenic tobacco under the control of the *CaMV* 35S promoter. Ezaki et al. (2005) found that the *Arabidopsis* blue-copper binding gene restrained aluminum absorption so as to keep plants and *Saccharomyces cerevisiae* (yeast) out of aluminum toxicity. In the present study, *PtENODL24* containing the MYB *cis*-element was strongly induced under drought treatment, which was supported by the promoter *cis*-element analysis.

Gene duplication is a vital source of new genes in the course of evolution (Lynch and Conery 2000; Gu et al. 2003; Khan et al. 2007). Gene duplication can also help organisms adapt to different environments during development and growth (Bowers et al. 2003). Gene duplication mechanisms such as unequal exchange, reverse transcription, or entire gene duplication produce a gene or base sequence that resembles the original gene (Zhang 2003). However, previous studies indicated that half of *OsPCs* in rice were the result of segmental duplication, and the others derived from tandem duplication, which indicated that these two types of duplication event played an equivalent role in the expansion of *OsPC* genes (Ma et al. 2011). The results of the present study indicated that the conserved region of 10 *PtPC* paralogs resulted from segmental duplication events, suggesting that segmental duplication exerted a crucial part in the expansion of the *PtPC* gene family.

K_a/K_s ratio can be used to measure the historical choice of coding sequences (Vandepoele et al. 2003; Wu et al. 2016). In this study, K_a/K_s sliding window analysis indicated that

Table 2 K_a, K_s and K_a/K_s values calculated for paralogous *PC* gene-pairs (Pt–Pt) in the *Populus trichocarpa* genome

Paralogous pairs	K_a	K_s	K_a/K_s ratio	Date (MY)
PtENODL1-PtENODL17	0.13114	0.31199	0.42	17.14
PtENODL12-PtENODL31	0.11411	0.33971	0.336	18.67
PtENODL18-PtENODL43	0.05937	0.21822	0.272	11.99
PtENODL21-PtENODL38	0.11769	0.31169	0.378	17.13
PtENODL23-PtENODL44	0.08612	0.33559	0.257	18.44
PtSC4-PtSC11	0.10877	0.38728	0.281	21.28
PtSC5-PtSC11	0.11514	0.39735	0.29	21.83
PtSC6-PtSC11	0.0918	0.43306	0.212	23.79
PtUC1-PtUC5	0.10047	0.25478	0.394	14
PtSC8-PtSC17	1.68599	1.01526	1.661	55.78

most *PtPCs* were under disadvantageous selection, imply a strong selection constraint and purifying selection in the *PtPC* genes. This suggests that functional divergence of duplicated genes might have been promoted by positive selection during evolution, which may facilitate the adaptation of plants in different environments. By evaluating the duplication time of paralogs, we inferred that all large-scale duplication events involving *PtPCs* occurred within the last 10.20–55.80 million years. We observed microsynteny in *PtPCs* and identified 23 collinear *PC* gene pairs, indicating a low degree of divergence during evolution. Simultaneously, we observed microsynteny between *Populus* and *Arabidopsis PC* genes and identified 15 pairs orthologs, hence orthologs derived from common ancestral genes are present in different species.

Since expression analysis can provide valuable information for further exploration of the relative expression levels and better understand the *PC* gene function in *P. trichocarpa*, we examined the expression of *PtPCs* in poplar in different tissues using microarray data. The results demonstrated that most genes were highly expressed in roots and xylem, indicating roles in vegetative growth. Through comparisons of

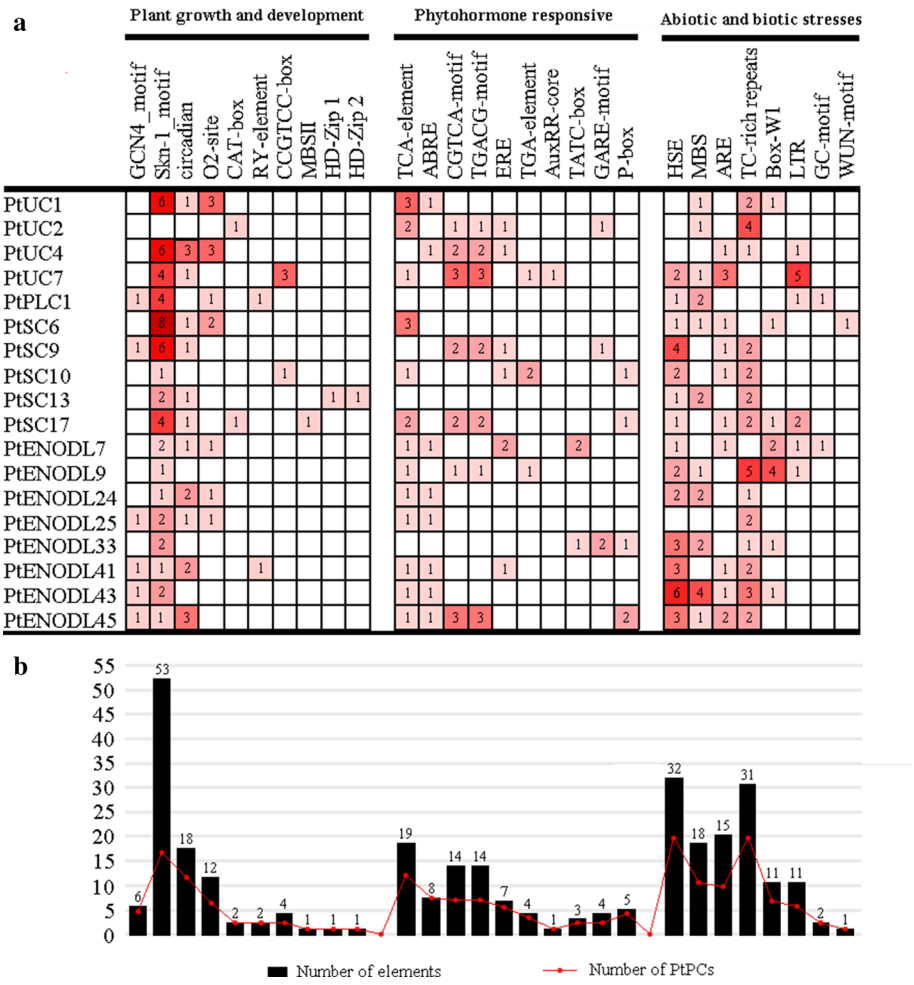

Fig. 6 *Cis*-acting elements in the promoter regions of poplar *PC* genes. **a** Number of each *cis*-acting element in the promoter region (2 kb upstream of the translation start site) of *PtPC* genes. **b** Statistics for all *PtPC* genes, including the corresponding *cis*-acting elements (red dots) and the total number of *cis*-acting elements in the *PtPC* gene family (black boxes) are given. Based on functional annotation, the *cis*-acting elements were classified into three major classes: plant growth and development, phytohormone responsive, or abiotic and biotic stresses-related (detailed results shown in Table S4)

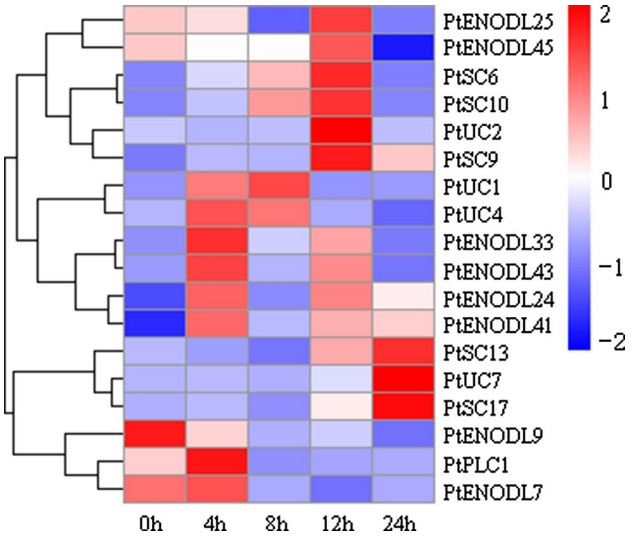

Fig. 7 Expression profiles of *PtPC* genes in response to drought treatment as determined by qRT-PCR. A heat-map shows the hierarchical clustering of the relative expression of 74 *PtPC* genes under drought treatment. Blue indicates lower and red represents higher transcript abundance compared to the relevant control. The leaves were sprayed with 20% PEG-6000 and sampled after 4, 8, 12 and 24 h of treatment. Relative expression levels of 74 *PC* genes were examined by qRT-PCR and normalized with respect to the reference gene UBQ (Potri.001G418500) under drought stress treatment

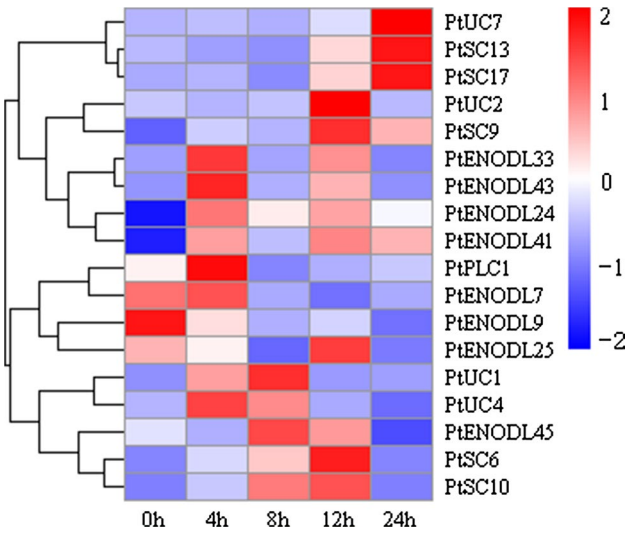

Fig. 8 Expression profiles of *PtPC* genes under NaCl treatment as determined by qRT-PCR. A heat-map shows the hierarchical clustering of the relative expression of 74 *PtPC* genes under NaCl treatment. Blue indicates lower and red represents higher transcript abundance compared to the relevant control. Salt-stress was carried out by watering the plants with a 200 mM solution of NaCl. Plants were sampled after 4, 8, 12 and 24 h of treatment. Relative expression levels of 74 *PC* genes were examined by qRT-PCR and normalized with respect to the reference gene UBQ (Potri.001G418500) under NaCl treatment

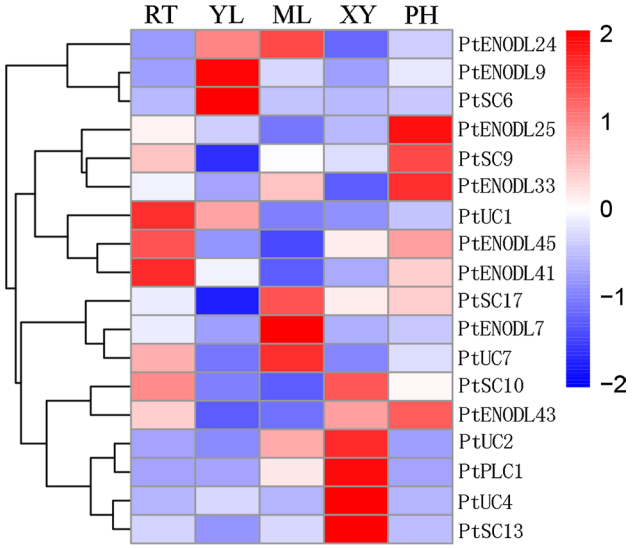

Fig. 9 The qRT-PCR analysis of expression profiles. A heat-map shows the hierarchical clustering of the relative expression of 18 *PtPC* genes across the five different tissues analyzed. The vertical colour scale at the right of the image represents $\log_2$ expression values: red indicates a high level and blue represents a low level of transcript abundance. Relative expression levels of all *PtPCs* were examined by qRT-PCR and normalized with respect to the reference gene UBQ (Potri.001G418500) in different tissues. *RT* roots, *YL* young leaves, *ML* mature leaves, *XY* xylem, *PH* phloem

microarray analysis and qRT-PCR analysis, we found that the majority of *PtPCs* was similarly expressed in the same organs. However, there were some different results. For example, *PtENODL41* was expressed at low levels in roots according to microarray analysis but relatively high in the same tissue according to qRT-PCR, which might result from differences in the experiments, such as conditions, poplar ages, sample collection times and so on. In addition, abiotic stresses such as high salinity and high temperature influence plant growth and development. Many stress-response genes are activated to help plants to deal with stress in adverse circumstances. Hence, it is important to identify the major regulatory pathways of stress responses in poplar. The results of qRT-PCR experiments indicated an important function for *PtPCs* under drought or salt stresses. Previous reports showed that *OsUC23/26/27* and *BrUC6/16* were highly expressed under drought or salt stresses. Differences of gene expression among species indicated that *PC* genes in different plants may generate diverse responses to abiotic stress.

Conclusions

In the current study, we identified 74 *PtPCs* and comprehensively and systematically analyzed their PC domains, gene structure, gene replication, chromosomal distribution,

and conserved microsynteny. The evolutionary relationships between PCs in *Arabidopsis* and poplar were also investigated, and the results revealed different expression patterns, indicating differentiated functions. To further understand the function and role of each *PC* family gene, multiple methods including molecular genetic analysis should be employed. The qRT-PCR analysis indicated a central role of *PtPCs* in many aspects of plant growth and development and plant stress response. Our findings provide a theoretical basis for further research on the function of *PtPCs*.

Author contribution statement SSL and WFH designed and conceived the experiment, carried out the principal bioinformatics analysis, drafted the manuscript. Performed the experiments: SSL, WFH. Edited the data, figures and tables: YW, BL. Contributed reagents/materials/analysis tools: YX. All authors read and approved the final manuscript.

Acknowledgements We thank the members of the Laboratory of Modern Biotechnology for their assistance in this study.

Funding National Natural Science Foundation of China (31370561) and National Science and Technology Support Plan Corpus (2015BAD07B070104).

Compliance with ethical standards

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

References

Anderson SL, Teakle GR, Martino-Catt SJ, Kay SA (1994) Circadian clock- and phytochrome regulated transcription is conferred by a 78 bp *cis*-acting domain of the *Arabidopsis CAB2* promoter. Plant J 6(4):457–470

Bobb AJ, Chern MS, Bustos MM (1997) Conserved RY-repeats mediate transactivation of seed-specific promoters by the developmental regulator PvALF. Nucleic Acids Res 25(3):641–647

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422(6930):433–438

Cao J, Li X, Lv Y, Ding L (2015) Comparative analysis of the phytocyanin gene family in 10 plant species: a focus on *Zea mays*. Front Plant Sci 6:515. https://doi.org/10.3389/fpls.2015.00515

Cruz-Garcia F, Nathan HC, Kim D, Mcclure B (2005) Stylar glycoproteins bind to S-RNase in vitro. Plant J 42(3):295–304

Denancé N, Szurek B, Noël LD (2014) Emerging functions of nodulin-like proteins in non-nodulating plant species. Plant Cell Physiol 55(3):469–474

Diab AA, Teulat-Merah B, This D, Ozturk NZ, Benscher D, Sorrells ME (2004) Identification of drought-inducible genes and differentially expressed sequence tags in barley. Theor Appl Genet 109(7):1417–1425

Dong J, Kim ST, Lord EM (2005) Plantacyanin plays a role in reproduction in Arabidopsis. Plant Physiol 138(2):778–789

Eisenhaber B, Wildpaner M, Schultz CJ, Borner GH, Dupree P, Eisenhaber F (2003) Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. Plant Physiol 133(4):1691–1701. https://doi.org/10.1104/pp.103.023580

Escobar-Restrepo JM, Huck N, Kessler S, Gagliardini V, Gheyselinck J, Yang WC, Grossniklaus U (2007) The FERONIA receptor-like kinase mediates male–female interactions during pollen tube reception. Science 317(5938):656–660

Ezaki B, Katsuhara M, Kawamura M, Matsumoto H (2001) Different mechanisms of four aluminum (Al)-resistant transgenes for Al toxicity in Arabidopsis. Plant Physiol 127(3):918–927

Ezaki B, Sasaki K, Matsumoto H, Nakashima S (2005) Functions of two genes in aluminium (Al) stress resistance: repression of oxidative damage by the *AtBCB* gene and promotion of efflux of Al ions by the *NtGDI1*gene. J Exp Bot 56(420):2661–2671

Fedorova M, Mortel JVD, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP (2002) Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. Plant Physiol 130(2):519–537

Giri AV, Anishetty S, Gautam P (2004) Functionally specified protein signatures distinctive for each of the different blue copper proteins. BMC Bioinform 5(1):1–8

Goldsbrough AP, Albrecht H, Stratford R (1993) Salicylic acid-inducible binding of a tobacco nuclear protein to a 10 bp sequence which is highly conserved amongst stress-inducible genes. Plant J 3(4):563–571

Goodstein DM, Shu S, Russell H, Rochak N, Hayes RD, Joni F, Therese M, William D, Uffe H, Nicholas P (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:1178–1186

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH (2003) Role of duplicate genes in genetic robustness against null mutations. Nature 421(6918):63–66

Hart PJ, Nersissian AM, Herrmann RG, Nalbandyan RM, Valentine JS, Eisenberg D (1996) A missing link in cupredoxins: crystal structure of cucumber stellacyanin at 1.6 A resolution. Protein Sci 5(11):2175–2183

Hou Y, Guo X, Cyprys P, Zhang Y, Bleckmann A, Cai L, Huang Q, Luo Y, Gu H, Dresselhaus T, Dong J, Qu L (2016) Maternal ENODLs are required for pollen tube reception in *Arabidopsis*. Curr Biol 26(17):2343–2350

Hu R, Qi G, Kong Y, Kong D, Qian G, Zhou G (2010) Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biol 10(1):145. https://doi.org/10.1186/1471-2229-10-145

Hui M, Lin F, Zhu C, Xue C, Hualin Z, Yan X (2014) Genome-wide identification and expression analysis of the IQD gene family in *Populus trichocarpa*. Plant Sci 229:96–110

Jeßberger N, Krey VM, Rademacher C, Böhm ME, Mohr AK, Ehlingschulz M, Scherer S, Märtlbauer E (2015) From genome to toxicity: a combinatory approach highlights the complexity of enterotoxin production in *Bacillus cereus*. Front Microbiol 6(6):560

Johnson KL, Cassin AM, Lonsdale A, Bacic A, Doblin MS, Schultz CJ (2017) A motif and amino acid bias bioinformatics pipeline to identify hydroxyproline-rich glycoproteins. Plant Physiol 174(2):886–903

Khan JA, Wang Q, Sjölund RD, Schulz A, Thompson GA (2007) An early nodulin-like protein accumulates in the sieve element plasma membrane of Arabidopsis. Plant Physiol 143(4):1576–1589

Li J, Yu M, Geng LL, Zhao J (2010) The fasciclin-like arabinogalactan protein gene, *FLA3*, is involved in microspore development of Arabidopsis. Plant J 64(3):482–497

Li J, Gao G, Zhang T, Wu X (2013) The putative phytocyanin genes in Chinese cabbage (*Brassica rapa L.*): genome-wide identification,

classification and expression analysis. Mol Genet Genom 288(1–2):1–20
Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151–1155
Ma H, Jie Z (2010) Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). J Exp Bot 61(10):2647–2668
Ma H, Zhao J (2010) Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). J Exp Bot 61(10):2647–2668
Ma H, Zhao H, Liu Z, Zhao J (2011) The phytocyanin gene family in rice (*Oryza sativa* L.): genome-wide identification, classification and transcriptional analysis. PLoS One 6(10):e25184
Ma Y, Yan C, Li H, Wu W, Liu Y (2017) Bioinformatics prediction and evolution analysis of arabinogalactan proteins in the plant kingdom. Front Plant Sci 8:1–17
MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. Plant J 62(4):689–703
Mashiguchi K, Yamaguchi I, Suzuki Y (2004) Isolation and identification of glycosylphosphatidylinositol-anchored arabinogalactan proteins and novel β-glucosyl Yariv-reactive proteins from seeds of rice (*Oryza sativa*). Plant Cell Physiol 45(12):1817–1829
Mashiguchi K, Asami T, Suzuki Y (2009) Genome-wide identification, structure and expression studie, and mutant collection of 22 early nodulin-like protein genes in *Arabidopsis*. Biosci Biotechnol Biochem 73(11):2452–2459
Nejad ES, Askari H, Soltani S (2012) Regulatory TGACG-motif may elicit the secondary metabolite production through inhibition of active cyclin-dependent kinase/cyclin complex. Plant Omics 5(6):553–558
Nersissian AM, Immoos C, Hill MG, Hart PJ, Williams G, Herrmann RG, Valentine JS (1998) Uclacyanins, stellacyanins, and plantacyanins are distinct subfamilies of phytocyanins: plant-specific mononuclear blue copper proteins. Protein Sci 7(9):1915–1929
Ozturk ZN, Talamé V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, Tuberosa R, Bohnert HJ (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 48(5–6):551–573
Petersen TN, Brunak S, Von HG, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786
Poon S, Heath R, Clarke A (2013) A chimeric arabinogalactan protein promotes somatic embryogenesis in cotton cell culture. Plant Physiol 160(2):684–695
Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3(6):1101–1108
Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. Plant Physiol 129(4):1448–1463
Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. Annu Rev Plant Biol 58(1):137–161
Sessa G, Morelli G, Ruberti I (1993) The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. EMBO J 12(9):3507–3517
Shen Q, Ho TH (1995) Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel *cis*-acting element. Plant Cell 7(3):295–307
Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiol 153(2):485–513
Showalter AM, Keppler B, Liu X, Lichtenberg J, Welch LR (2016) Bioinformatic identification and analysis of hydroxyproline-rich glycoproteins in *Populus trichocarpa*. BMC Plant Biol 16(1):229
Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell 16(8):2001–2019
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725–2729
Tan L, Qiu F, Lamport DTA, Kieliszewski MJ (2004) Structure of a hydroxyproline (Hyp)-arabinogalactan polysaccharide from repetitive Ala-Hyp expressed in transgenic *Nicotiana tabacum*. J Biol Chem 279(13):1315656–1315665
Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15(9):2192–2202
Wang Y, Feng L, Zhu Y, Li Y, Yan H, Xiang Y (2015) Comparative genomic analysis of the WRKY III gene family in *populus*, grape, *arabidopsis* and rice. Biol Direct 10:48. https://doi.org/10.1186/s13062-015-0076-3
Washida H, Wu CY, Suzuki A, Yamanouchi U, Akihama T, Harada K, Takaiwa F (1999) Identification of *cis*-regulatory elements required for endosperm expression of the rice storage protein glutelin gene *GluB-1*. Plant Mol Biol 40(1):1–12
Wu H, Shen Y, Hu Y, Tan S, Lin Z (2011) A phytocyanin-related early nodulin-like gene, *BcBCP1*, cloned from *Boea crassifolia* enhances osmotic tolerance in transgenic tobacco. J Plant Physiol 168(9):935–943
Wu M, Li Y, Chen D, Liu H, Zhu D, Xiang Y (2016) Genome-wide identification and expression analysis of the IQD gene family in moso bamboo (*Phyllostachys edulis*). Sci Rep UK 6:24520
Xu L, Wang XJ, Wang T, Li LB (2017) Genome-wide identification, classification, and expression analysis of the phytocyanin gene family in *Phalaenopsis equestris*. Biol Plant 61(3):1–8
Ye H, Du H, Tang N, Li X, Xiong L (2009) Identification and expression profiling analysis of TIFY family genes involved in stress and phytohormone responses in rice. Plant Mol Biol 71(3):291–305
Yoshizaki M, Furumoto T, Hata S, Shinozaki M, Izui K (2000) Characterization of a novel gene encoding a phytocyanin-related protein in morning glory (*Pharbitis nil*). Biochem Biophys Res Commun 268(2):466–470
Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18(6):292–298
Zhang R, Tucker MR, Burton RA, Shirley NJ, Little A, Morris J, Milne L, Houston K, Hedley PE, Waugh R, Fincher GB (2016) The dynamics of transcript abundance during cellularisation of developing barley endosperm. Plant Physiol 170(3):1549–1565