



## Comparative analyses of phytochelatin synthase (PCS) genes in higher plants

Ertugrul Filiz, Ibrahim Adnan Saracoglu, Ibrahim Ilker Ozyigit & Bahattin Yalcin

To cite this article: Ertugrul Filiz, Ibrahim Adnan Saracoglu, Ibrahim Ilker Ozyigit & Bahattin Yalcin (2019) Comparative analyses of phytochelatin synthase (PCS) genes in higher plants, *Biotechnology & Biotechnological Equipment*, 33:1, 178-194, DOI: [10.1080/13102818.2018.1559096](https://doi.org/10.1080/13102818.2018.1559096)

To link to this article: <https://doi.org/10.1080/13102818.2018.1559096>



© 2018 The Author(s). Published by Taylor & Francis Group on behalf of the Academy of Forensic Science.



Published online: 10 Jan 2019.



Submit your article to this journal [↗](#)



Article views: 2193



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 23 View citing articles [↗](#)

## Comparative analyses of phytochelatin synthase (PCS) genes in higher plants

Ertugrul Filiz<sup>a</sup>, Ibrahim Adnan Saracoglu<sup>b</sup>, Ibrahim Ilker Ozyigit<sup>c,d</sup> and Bahattin Yalcin<sup>b</sup>

<sup>a</sup>Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, Duzce, Turkey; <sup>b</sup>Department of Chemistry, Faculty of Science and Arts, Marmara University, Istanbul, Turkey; <sup>c</sup>Department of Biology, Faculty of Science and Arts, Marmara University, Istanbul, Turkey; <sup>d</sup>Department of Biology, Faculty of Science, Kyrgyz-Turkish Manas University, Bishkek, Kyrgyzstan

### ABSTRACT

Plants employ various defence strategies to ameliorate the effects of heavy metal exposures, leading to re-establishment of metal homeostasis. One of the strategies includes the biosynthesis of main heavy metal detoxifying peptides phytochelatins (PCs) by phytochelatin synthase (PCS). In the present study, 14 PCS homologues were identified in the genomes of 10 selected plants. The size of these PCSs was 452–545 amino acid residues, with characteristic phytochelatin and phytochelatin\_C domains. The N-terminal site of the proteins is highly conserved, whereas the C-terminal site is less conserved. Further, the present study also identified fully conserved Cys residues involved in heavy metal binding reported earlier. In addition, other preserved cysteines, with minor substitutions Cys(C)→Ser(S) or Tyr(Y) or Trp(W), were also identified in the PCS sequences that might be associated with metal binding. The reported catalytic triad residues from *Arabidopsis*, Cys56, His162 and Asp180, are all conserved at the respective sites of PCSs. A clear monocot/dicot separation was revealed by phylogenetic analysis and was further corroborated by the exon–intron organisations of the PCS genes. Moreover, gene ontology terms, co-expression network, cis-regulatory motif and miRNA analyses indicated that the complex as well as dynamic regulation of PCSs has significant involvement in different metabolic pathways associated with signalling, defence, stress and phytohormone, in addition to metal detoxification. Moreover, variations in protein structure are suggested to confer the functional divergence in PCS proteins.

### ARTICLE HISTORY

Received 13 August 2018  
Revised 26 November 2018  
Accepted 11 December 2018

### KEYWORDS

phytochelatin; metal homeostasis; heavy metal; bioinformatics

## Introduction

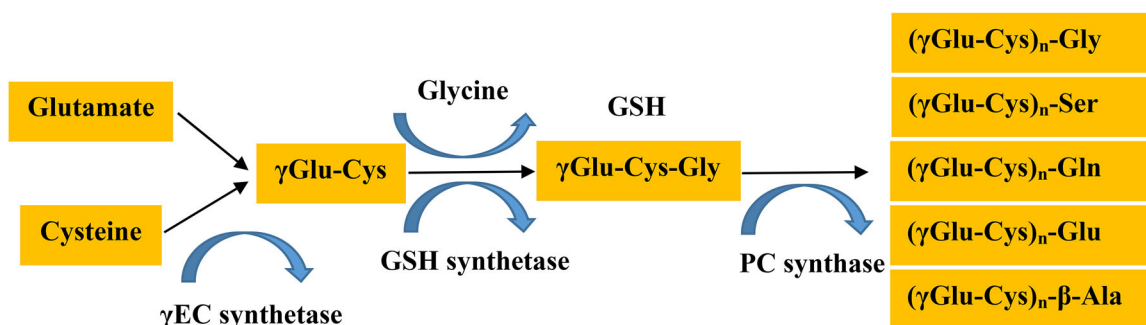
Beginning from the mid-20th century, the anthropogenic heavy metal pollution, from traffic, metal industries and mining, has been reported to pose serious threats to all living organisms [1]. Further, plants are exposed to heavy metals, particularly in contaminated environments [2]. However, plants utilize their defence strategies to ameliorate the effects of these adversities. This in turn helps plants to re-establish their homeostasis and progress with their stages of development [3]. Some strategies constitute the formation of complexes with organic molecules to reduce heavy metal availability [4]. These organic molecules include organic acids, malate, citrate, low molecular weight protein, metallothionein (MT), low molecular weight peptides, phytochelatins (PCs) and glutathione (GSH)

[4,5]. Phytochelatins are heavy metal-binding peptides in plants with a  $(\gamma\text{-Glu-Cys})_n\text{-Gly}$  structure ( $n=2\text{--}11$ ) [6]. However, in some plants, the C-terminal Gly can be replaced by serine as  $(\gamma\text{-Glu-Cys})_n\text{-Ser}$ , glutamine as  $(\gamma\text{-Glu-Cys})_n\text{-Gln}$ , glutamate as  $(\gamma\text{-Glu-Cys})_n\text{-Glu}$  and alanine as  $(-\gamma\text{-Glu-Cys})_n\text{-}\beta\text{-Ala}$  [7]. They are enzymatically synthesized from GSH by phytochelatin synthase (PCS) (Figure 1). PCS proteins demonstrate high similarity in their N-terminal domains, whereas their C-terminal domains are less conserved. The N-terminal core domain has been reported to confer the PCS activity, whereas C-terminus ensures improved protein stability, higher PCS activity and broader heavy metal spectrum [6,9,10].

PCS is constitutively expressed; however, heavy metals are major inducers for its activity [11]. PCs are easily induced when exposed to heavy metal ions

**CONTACT** Ibrahim Ilker Ozyigit  [ilkozyigit@gmail.com](mailto:ilkozyigit@gmail.com); Ertugrul Filiz  [ertugrulfiliz@gmail.com](mailto:ertugrulfiliz@gmail.com)  Department of Biology, Faculty of Science and Arts, Marmara University, Istanbul, 34722, Turkey

© 2018 The Author(s). Published by Taylor & Francis Group on behalf of the Academy of Forensic Science. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Phytochelatin (PC) biosynthesis molecular pathway in higher plants (Modified from Inouhe [8]). GSH is synthesized by two sequential enzymatic reactions by  $\gamma$ EC synthetase and GSH synthetase from precursors, glutamine, cysteine and glycine. Subsequently, PC synthase catalyses the biosynthesis of PCs from GSH. PCs are related to glutathione (GSH) with a  $(\gamma$ -Glu-Cys) $_n$ -Gly structure. However, in some plants, C-terminal Gly can be replaced by serine as  $(\gamma$ -Glu-Cys) $_n$ -Ser, glutamine as  $(\gamma$ -Glu-Cys) $_n$ -Gln, glutamate as  $(\gamma$ -Glu-Cys) $_n$ -Glu and alanine as  $(\gamma$ -Glu-Cys) $_n$ - $\beta$ -Ala.

such as cadmium (Cd), nickel (Ni), copper (Cu), zinc (Zn), silver (Ag), mercury (Hg), lead (Pb) and arsenic (As). Actually, Cd is one of the most effective activators for PC [9]. It was first proposed that the variable C-terminal site could bind the heavy metals via conserved Cys residues and translocate them to the catalytic N-terminal domain [12]. Also, Cd-binding assays revealed five conserved Cys residues in the N-terminus of AtPCS1, involved in the Cd<sup>2+</sup> binding, which is essential for PCS activity [13]. Later, another model suggested that the protein does not directly bind to Cd<sup>2+</sup> but instead forms a PC $_n$ -heavy metal thiolate complex as an acceptor molecule [11]. Site-directed mutagenesis demonstrated three conserved residues (Cys56, His162 and Asp180) in the N-terminal site and their substitutions resulted in the complete loss of AtPSC1 activity [6,11,14].

PCS genes have been identified from various plant species and some microorganisms. These include *Triticum aestivum* [15], *Oryza sativa* [16], *Arabidopsis thaliana* [17], *Brassica juncea* [18], *A. halleri* [19], *Thlaspi caerulescens* [11], *Schizosaccharomyces pombe* [20] and *Caenorhabditis elegans* [21]. Multiple earlier studies have demonstrated the role of PCS in heavy metal detoxification (reviewed in [5]). For instance, AtPCS1 shows slightly higher expression under Cd exposure during the early stages of development [22]. A similar expression pattern is reported for TaPCS1 in the roots of 4-day-old wheat seedlings under Cd treatment [20]. One-week Cd applications in adult plants cause a significant rise in AtPCS1 expression [23]. Transgenic Indian mustard (*B. juncea*) with moderate AtPCS1 expression has indicated significantly higher tolerance to Cd and Zn stresses. However, they showed lesser accumulation than wild types in both shoot as well as root tissues [24]. In the model legume *Lotus japonicus*, alternatively spliced variants of PCSs, LjPCS1-3 show variable responses towards Cd

treatments [25]. The overexpressed PCS (*NtPCS1*) cause elevation in the tolerance to Cd and arsenite in transgenic tobacco plants [26]. In addition to metal detoxification activity, new functions have been also reported for both the PCS and the PC [19]. PCs could be involved in Zn-sequestration [27] and long-distance transport of Cd<sup>2+</sup> [28]. The catabolism of GS conjugates in plants has been linked to the peptidase as a second enzymatic activity of PCS [29,30]. Further, another study suggests that AtPCS1 is involved in the plant innate immunity [31].

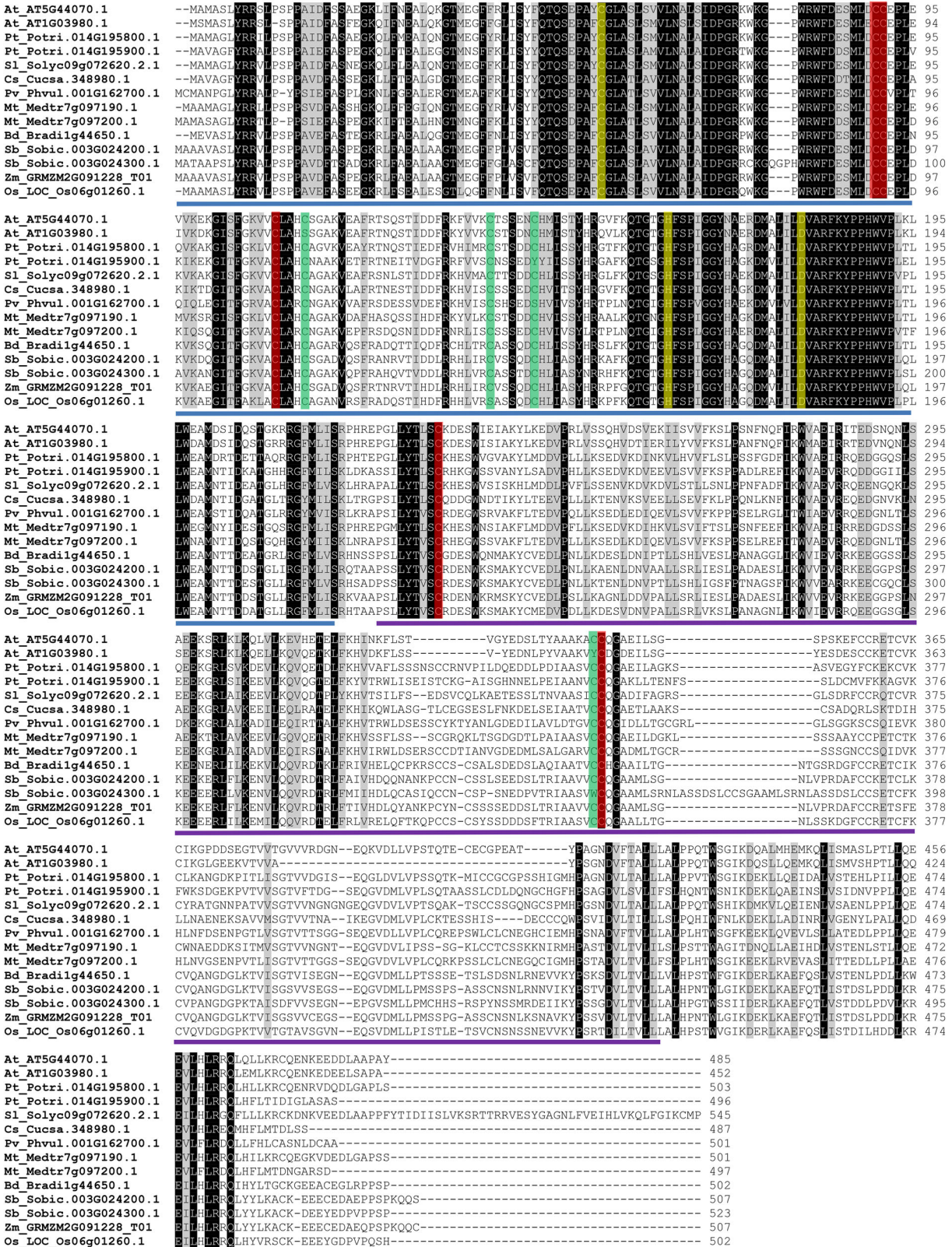
The prospect of preventing plant-based heavy metal toxicity and utilising plants for the bioremediation of heavy metal polluted sites is an emerging issue [32]. Thus, exploration and understanding of the molecular basis of major genes involved in heavy metal detoxification were the primary tasks of the present study. In this sense, this work attempted to understand the PCS gene, which is responsible for the biosynthesis of one of the main heavy metal detoxifying peptides—PCs—in plants. Moreover, potential PCS genes were identified at genome-wide scale in 10 different plant species and were comparatively investigated at the primary, secondary and tertiary level.

## Materials and methods

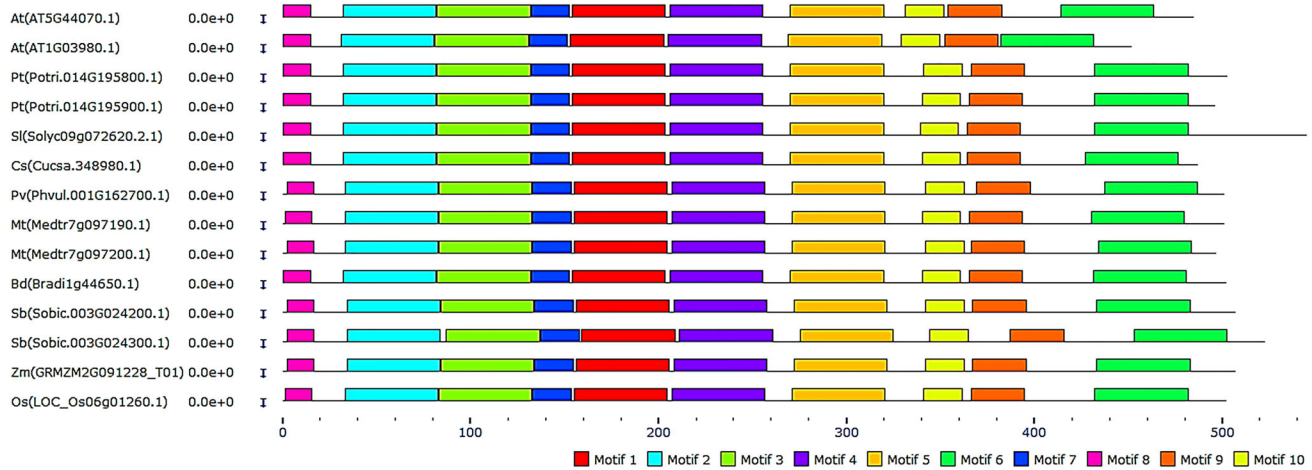
### Sequence analyses of PCSs

Two known *Arabidopsis* PCSs, PCS1 (Q9S7Z3) and PCS2 (Q9ZWB7), were obtained from the UniProt [33] database to be exploited as reference. These references were queried against the proteome datasets of 10 selected higher plants in the plant genomic resource Phytozome v11 [34,35] database. The studied species included *A. thaliana*, *Populus trichocarpa*, *Solanum lycopersicum*, *Medicago truncatula*, *Cucumis sativus*, *Phaseolus vulgaris*, *Brachypodium distachyon*, *Sorghum*

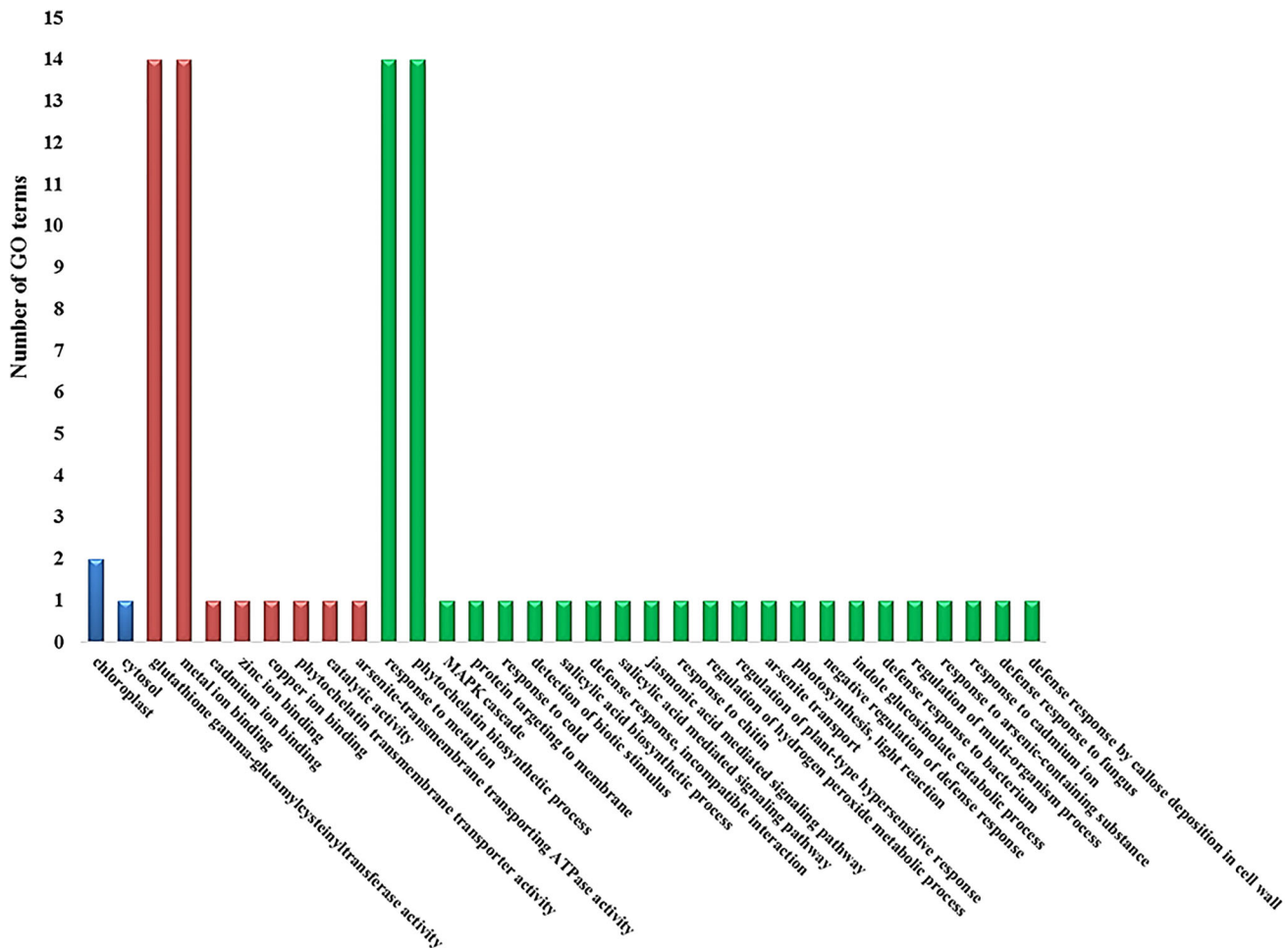




**Figure 2.** Multiple sequence alignment of PCS proteins in 10 plant species. Sequences are aligned by ClustalW, and identical and similar residues are shaded as black and grey, respectively, with a very strict (100%) threshold. Highlighted red and yellow columns show, respectively, the strictly conserved Cys residues and catalytic triad (Cys56, His162 and Asp180 in *Arabidopsis*) [6]. Highlighted green columns also show the conserved Cys residues but some amino substitutions are observed such as Cys(C)→Ser(S)/Tyr(Y)/Trp(W). Blue and purple lines below sequences indicate the approximate locations of Phytochelatin (PF05023) and Phytochelatin\_C (PF09328) domains, respectively.



**Figure 3.** Block diagram representation of the most conserved 10 motifs in PCS protein sequences. Motifs were more uniformly distributed in the N-terminal site of the enzyme. Motif 1 is indicated with red, motif 2 with cyan, motif 3 with light-green, motif 4 with purple, motif 5 with gold, motif 6 with green, motif 7 with blue, motif 8 with pink, motif 9 with orange and motif 10 with yellow.



**Figure 4.** Gene ontology (GO) enrichment of PCS sequences. The x-axis shows the GO terms such as 'Cellular components', 'Molecular functions' and 'Biological processes' with blue, red and green columns, respectively. The y-axis indicates the number of GO terms.

*bicolor*, *Zea mays* and *Oryza sativa*. Subsequently, the sequences were retrieved by application of a Hidden Markov Model search for protein domains verification by Pfam [36,37] (Figure 2). The physicochemical

properties of the PCS proteins were determined by using the ProtParam tool [38]. The CELLO server was utilized further to predict the subcellular localisations [39]. The MEME tool [40] was exploited to perform a



**Table 1.** Putative phytochelatin synthase (PCS) genes in 10 plant species and their gene/protein features.

Species	Transcript ID (Phytozome)	Exon no.	Protein domains (Pfam)	Length (aa)	MW (kDa)	<i>pI</i>	Cellular localisation
<i>A. thaliana</i>	AT5G44070.1	9	Phytochelatin, Phytochelatin_C	485	54.47	6.18	Cytosol
<i>A. thaliana</i>	AT1G03980.1	9	Phytochelatin, Phytochelatin_C	452	51.55	6.58	Nucleus/cytosol
<i>P. trichocarpa</i>	Potri.014G195800.1	8	Phytochelatin, Phytochelatin_C	503	55.81	5.91	Cytosol
<i>P. trichocarpa</i>	Potri.014G195900.1	8	Phytochelatin, Phytochelatin_C	496	54.98	5.95	Cytosol
<i>S. lycopersicum</i>	Solyc09g072620.2.1	10	Phytochelatin, Phytochelatin_C	545	60.38	6.60	Cytosol/nucleus
<i>C. sativus</i>	Cucsa.348980.1	8	Phytochelatin, Phytochelatin_C	487	54.29	5.35	Cytosol
<i>M. truncatula</i>	Medtr7g097190.1	7	Phytochelatin, Phytochelatin_C	501	55.35	6.19	E. cellular/nucleus
<i>M. truncatula</i>	Medtr7g097200.1	7	Phytochelatin, Phytochelatin_C	497	55.04	5.54	E. cellular/P. membrane
<i>P. vulgaris</i>	Phvul.001G162700.1	8	Phytochelatin, Phytochelatin_C	501	55.80	5.29	Cytosol
<i>B. distachyon</i>	Bradi1g44650.1	6	Phytochelatin, Phytochelatin_C	502	55.55	6.09	E. cellular/cytosol
<i>S. bicolor</i>	Sobic.003G024200.1	6	Phytochelatin, Phytochelatin_C	507	56.01	6.00	E. cellular/nucleus
<i>S. bicolor</i>	Sobic.003G024300.1	6	Phytochelatin, Phytochelatin_C	523	57.55	5.99	E. cellular
<i>Z. mays</i>	GRMZM2G091228_T01	6	Phytochelatin, Phytochelatin_C	507	56.15	6.17	Nucleus
<i>O. sativa</i>	LOC_Os06g01260.1	6	Phytochelatin, Phytochelatin_C	502	55.88	5.87	E. cellular/nucleus

PF05023: Phytochelatin domain, PF09328: Phytochelatin\_C domain.

search for conserved protein motifs [41] with the following parameters: max number of motifs to find, 10; and min/max width of motifs, 6–50 (Figure 3). ClustalW was used for alignment of protein sequences [42]. Further, the phylogenetic tree was generated by using MEGA 7 [43] with the maximum-likelihood (ML) method that was based on the JTT matrix-based model [44] for 1000 bootstraps [45]. Initial tree(s) for the heuristic search were generated automatically by application of Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances. The distances were estimated using a JTT model. This was followed by selection of topology with the help of superior log likelihood value. The analysis involved 14 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 438 positions in the final dataset. Gene ontology (GO) terms such as cellular component, molecular function and biological process were retrieved based on the predefined 'Templates' using InterMine interface in the Phytozome database (Figure 4).

### Gene structure and promoter analyses

Exon/intron structures of PCS genes were extracted from Phytozome v11 database. From the transcription start site (TSS), 1000 bp upstream regions of PCSs were retrieved from Phytozome, where the database has an internal option to allow retrieving up- or down-stream regions of a given gene. Further, they were delivered to the PlantCARE database for promoter analysis [46,47]. PCS coding sequences were scanned for putative miRNA targets in the psRNATarget database [48,49]. We employed a maximum expectation score of 5 for higher prediction coverage.

### Co-expression network of *Arabidopsis* PCS

A co-expression network analysis was performed in order to understand the PCS genes in a complex cell

environment at the molecular level. The network was constructed for gene *AtPCS1* using ATTED-II server v8 [50,51]. KEGG (Kyoto Encyclopedia of Genes and Genomes) data of connected genes were extracted from the gene network. ATTED-II is a co-expression database for plant species using multiple co-expression data sets and network analysis tools [51,52]. For *Arabidopsis*, the database included 20,836 genes from microarray (15,275 samples) and 25,296 genes from RNA-seq (1401 samples) data.

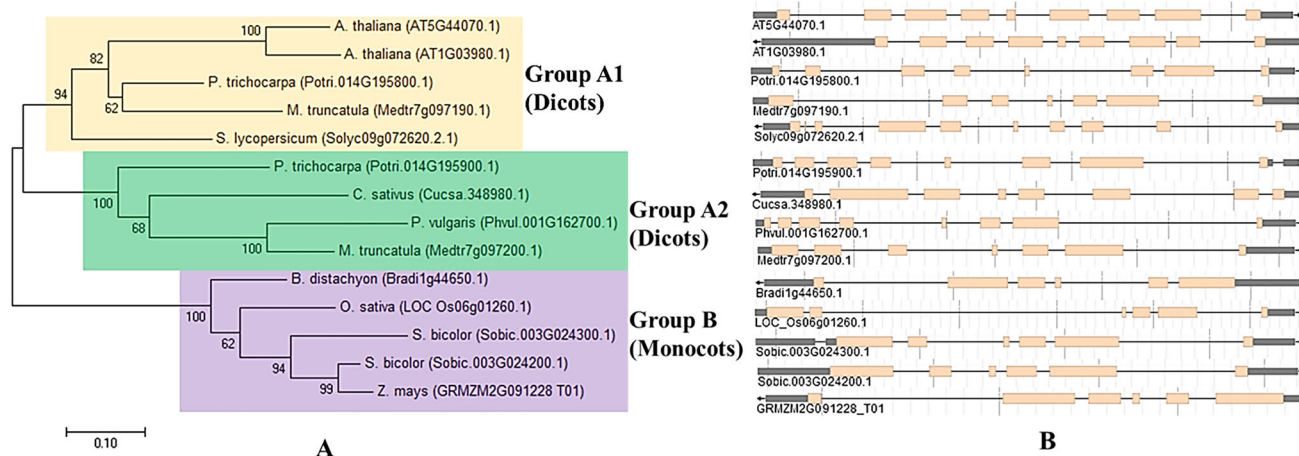
### Homology modelling of PCSs

Three-dimensional (3D) models were predicted using the Phyre<sup>2</sup> server [53,54] at intensive mode. The crystal structure of Alr0975 (2BTW, structures of transferase from *Nostoc* sp. PCC 7120) was used in homology modelling with resolution of 2.0 Å. The model quality was checked by Ramachandran plot analysis [55]. We used the VADAR 1.8 server [56] for secondary structure analysis [57]. Structural overlap percentage was calculated using the CLICK server based on the alpha-carbon superposition of models [58,59].

## Results and discussion

### Sequence analysis of PCSs

Using two known *Arabidopsis* *AtPCS1* and *AtPCS2* sequences as references, 15 putative PCS homologues were identified in the genomes of 10 selected plant species with high  $e^{-161}$  value (Table 1). Genome-wide search revealed that *A. thaliana*, *P. trichocarpa*, *M. truncatula* and *S. bicolor* harbour two PCS homologues in their genomes, whereas *S. lycopersicum*, *C. sativus*, *P. vulgaris*, *B. distachyon*, *Z. mays* and *O. sativa* only contain a single PCS gene. All PCS proteins were characterized with both Phytochelatin (PF05023) and Phytochelatin\_C (PF09328) domains (Figure 2). They are



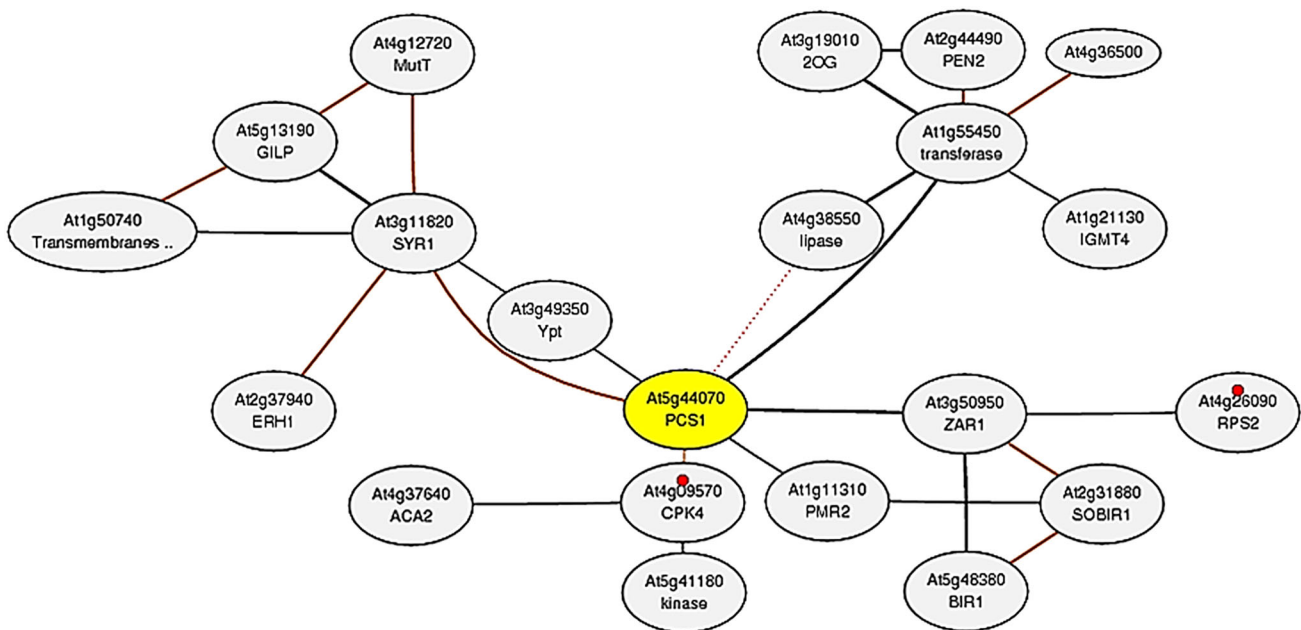
**Figure 5.** Phylogenetic tree showing 14 protein sequences of PCS from 10 higher plants. (A) Tree was constructed by MEGA7 using ML method based on the JTT matrix-based model and bootstrap consensus tree was generated with 1000 replicates. (B) Exon–intron organisations of PCS genes obtained from Phytosome database. Pink boxes and thin grey lines indicate the exons and introns, respectively.

452–545-residue proteins with molecular weight in the range of 51.55–60.38 kDa and pI values in the range of 5.29–6.60. The transcripts of PCS genes included 6–10 exons (Figure 5A–B). Interestingly all grass species possessed six exons, suggesting that PCS gene structures might be well conserved in the monocots during the evolutionary history of the PCS genes. In variations of exon–intron organisations, three main mechanisms have been proposed, including exon–intron gain/loss, exonisation/pseudoexonisation and insertion/deletion [60]. In addition, alternative splicing is also considered to be another important factor responsible for tissue localisation differences or stress responses [61]. A recent study reported alternatively spliced transcripts of rice *OsPCS2* gene to be involved in alleviation of Cd and As stresses in a tissue-specific way [62]. In another recent work, four rice *OsPCS1* transcript variants and one *OsPCS2* were isolated and their expression in PCS-deficient *Arabidopsis* and *Schizosaccharomyces pombe* mutants revealed PCS activity in response to Cd and As in the longest *PCS1* variant, *OsPCS1full* [63]. In this regard, it was quite possible that the evolutionary forces could attribute exon–intron variations between monocot and dicot species. Then, a detailed search in NCBI's RefSeq database was conducted to gain insights about the PCS orthologues in various other plants. Some entries included *S. tuberosum* (451–503 aa), *Glycine max* (437–498 aa), *Beta vulgaris* (416–541 aa), *Camelina sativa* (452–485 aa), *Tarenaya hassleriana* (488 aa), *Nelumbo nucifera* (456–510 aa), *Eucalyptus grandis* (398–505 aa), *Musa acuminata* (498–500 aa), *Nicotiana tomentosiformis* (501 aa), *Pyrus x bretschneideri* (489–509 aa), *Brasica rapa* (492 aa), *Phoenix dactylifera* (418–505 aa), *Vigna angularis* (497–499 aa), *Malus domestica* (476–497 aa), *C. melo*

(337–549 aa), *N. tabacum* (492–501 aa), *Prunus mume* (497–575 aa), *Arachis duranensis* (226–497 aa), *Ziziphus jujube* (371–506 aa), *O. brachyantha* (494–505 aa), *Ricinus communis* (494–502 aa), *Citrus sinensis* (452–523 aa), *Gossypium raimondii* (435–505) and *Populus euphratica* (496–503 aa). The above-database entries and findings of the present study confirmed that PCS isoforms were fairly similar in length among different plant species. Moreover, a similarity search was also performed between identified *Arabidopsis* and other PCSs using the blastp algorithm. AtPCS1 (AT5G44070.1) and 2 (AT1G03980.1), respectively, showed highest similarity to poplar (Potri.014G195800.1) with 67% and 64%, *Medicago* (Medtr7g097190.1) with 64% and 61% and tomato (Solyco09g072620.2.1) with 64% and 60%. On the other hand, they had lowest similarity with *Sorghum* (Sobic.003G024300.1) with 53% and 49%, respectively. This might point to the presence of a monocot/dicot divergence in terms of PCS sequences. Besides, even two PCS variants in the same genome demonstrated relatively less identity in *Arabidopsis* (78%), *P. trichocarpa* (62%), *M. truncatula* (60%) and *S. bicolor* (76%). This implicated that PCS homologues could have different functional roles in addition to heavy metal detoxification.

### Conserved motif/domain analysis

PCS proteins were multiple-aligned to figure out the potential conserved residues or domains that were involved in the enzyme activity and stability. The strict shading of the identical and similar residues on the alignment revealed highly conserved residues (Figure 2). Further, PCS proteins have been reported



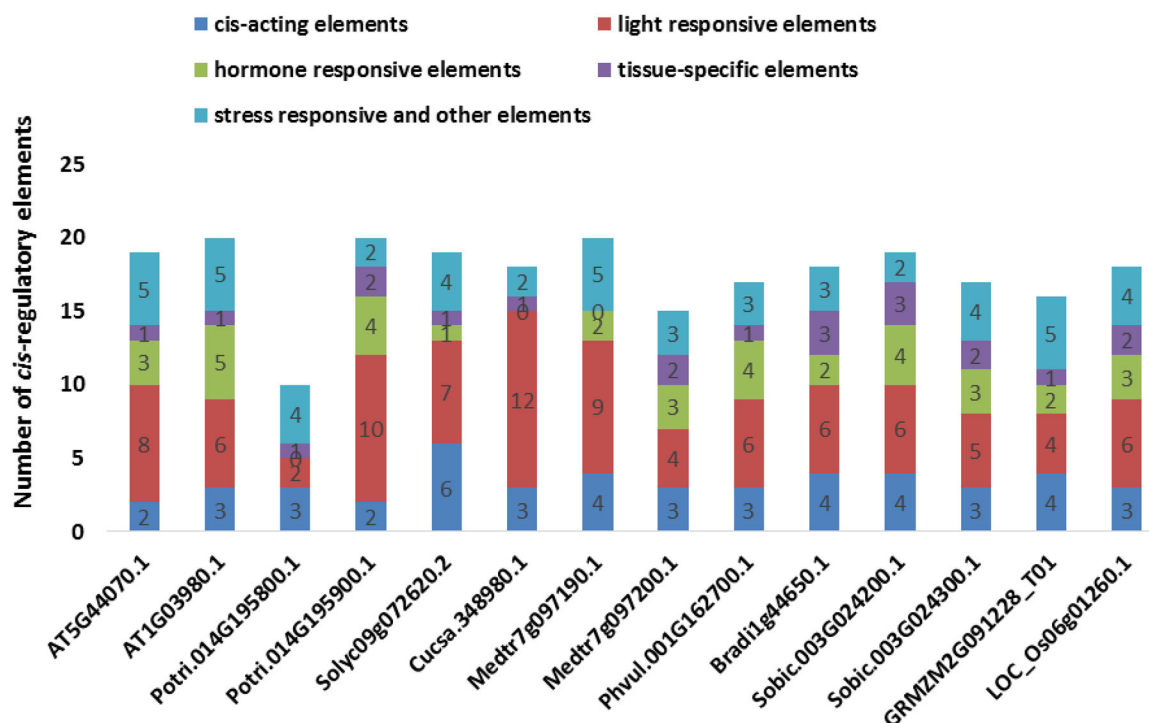
**Figure 6.** Co-expression network of *AtPCS1* (At5g44070; yellow node). The network was constructed by using microarray and RNA-seq datasets in the ATTED-II server. In the network, grey circles represent the nodes or genes, whereas interconnecting black lines show the edges or associations. Based on KEGG pathway (ID: ath04626), small red circles show the plant–pathogen interaction with two genes (At4g09570 and At4g26090).

to have high similarity in their N-terminal domains, whereas C-terminal domains remained less conserved as confirmed in earlier studies. The N-terminal core domain is also reported to confer the PCS activity, whereas the C-terminus provides improved protein stability, higher PCS activity and broader heavy metal spectrum [9,10]. In addition, conserved Cys residues at N- (five in *AtPCS1*) and C-terminal regions have been observed to be involved at the sites of the heavy metal (e.g.  $\text{Cd}^{2+}$ ) binding [12,13]. Moreover, studies from site-directed mutagenesis revealed three conserved residues, Cys56, His162 and Asp180, as a catalytic triad at the N-terminal site of the enzyme. Substitution of these residues caused the complete loss of *AtPSC1* activity [6,11,14]. In light of the above facts, alignment analysis showed that the N-terminal site of the studied PCS proteins was highly conserved, whereas the C-terminal site turned out to be less conserved (Figure 2). Six columns of fully conserved Cys residues including five in the N-terminus and one in the C-terminus (red highlights on alignment) were identified in the aligned sequences. In addition, other near full columns of Cys residues but with some substitutions Cys(C)→Ser(S) or Tyr(Y) or Trp(W) were also present in the PCS sequences (green highlights on alignment). These minor substitutions at Cys residues might be associated with the metal-binding properties of sequences, considering the roles of conserved Cys residues during heavy metal binding [12,13].

Moreover, reported catalytic triad residues (Cys56, His162 and Asp180 in *Arabidopsis*) were also fully conserved in the corresponding sites of all sequences (yellow highlights on alignment). The approximate locations of the PCS domains such as phytochelatin (PF05023) and phytochelatin\_C (PF09328) also indicated the sequences given below with the blue and purple lines, respectively, where the catalytic triad localized in the phytochelatin domain. Thus, inferring from the above-given studies, the presence of the catalytic triad and the conserved Cys residues implied that they may be somehow involved in PCS activity and heavy metal binding.

Furthermore, most conserved 10 motif sequences in PCS proteins were analysed using the MEME tool (Figure 3). The identified motifs 1–6 were 50 amino acids in length, whereas motifs 7–10 were 15–29 amino acids in length. Motif 1 comprised of sequences 'FKQTGTGHFSPIGGYHAGQDMA LILDVARFKYPPHWVPLTLL WEAMNTID', motif 2 of 'QNGTMEGFFRLISYFQTQSE PAYCGLASLSMVLNLAIDPGRKWKGPWRW', motif 3 of 'FDESMLDCCEPLDKV KA KGITFGKVACLAHCNGA KVQAFRTNQSTIDDFR', motif 4 of 'TGQHRGFMLISRHH RAPSILYTVSCRHESWKSMAKYCMEDVPLLKSENV', motif 5 of 'PANFNFIK WVA EVRRQEDGNQSLSKKEEKRLAIKENV LKQVRDTRLFKH', motif 6 of 'DVLTVLLL ALHPHTWSG IKDEKLKAEFQSLISTENLPLLQEEILHLRRQ', motif 7 of 'KHVIRCSSS QDCHMISSYHRG', motif 8 of 'MAM AGLYRRVLPSP', motif 9 of 'DGYCCRETCV KCWNANG





**Figure 7.** *Cis*-regulatory elements distribution at + 1000 bp upstream regions of the 14 PCS genes. Many regulatory elements were revealed in promoter sites of PCS genes but they were broadly categorized as *cis*-acting elements (dark blue segment), light responsive elements (red segment), hormone-responsive elements (green segment), tissue-specific elements (purple segment) and stress-responsive and other elements (light blue segment).

DNPKTVISGTV' and motif 10 of 'DSLTYIAASVCCQ GAEMLTGN'. Motifs 1–6 were related with the phytochelatin domain (PF05023) structure, motifs 9–10 with the phytochelatin\_C domain (PF09328), whereas motifs 7–8 did not relate to the domain structures. Particularly, motifs 1 and 2 were of significance due to the localisation of the catalytic triad in these motifs. The patterns of distributed motifs along sequences were also relatively more uniform at the N-terminal site. This also corroborated the earlier reports [9,10]. In addition, all 10 motifs were also available in all the PCSs and were distributed along the sequences. This confirmed that the PCSs were well conserved in higher plants.

### GO annotations of PCSs

GO terms, 'cellular component', 'molecular function' and 'biological process' provide the controlled vocabulary for genes or gene products. These are quite essential for the understanding of the diverse molecular functions of proteins [64]. Herein, enrichment of the PCSs with GO terms was performed using the InterMine interface of the Phytozome database (Figure 4). The inferred 'Molecular functions' included 'glutathione gamma-glutamylcysteinyltransferase activity' (GO:0016756), 'metal ion binding' (GO:0046872), 'cadmium ion binding' (GO:0046870), 'zinc ion binding'

(GO:0008270), 'copper ion binding' (GO:0005507), 'phytochelatin transmembrane transporter activity' (GO:0071992), 'catalytic activity' (GO:0003824) and 'arsenite-transmembrane transporting ATPase activity' (GO:0015446). In particular, 'glutathione gamma-glutamylcysteinyltransferase activity' and 'metal ion binding' terms as molecular function have been identified for all PCS sequences. For 'biological process', the PCS sequences were inferred with the terms 'response to metal ion' (GO:0010038), 'phytochelatin biosynthetic process' (GO:0046938), 'MAPK cascade' (GO:0000165), 'protein targeting to membrane' (GO:0006612), 'response to cold' (GO:0009409), 'detection of biotic stimulus' (GO:0009595), 'salicylic acid biosynthetic process' (GO:0009697), 'defence response, incompatible interaction' (GO:0009814), 'salicylic acid mediated signalling pathway' (GO:0009862), 'jasmonic acid mediated signalling pathway' (GO:0009867), 'response to chitin' (GO:0010200), 'regulation of hydrogen peroxide metabolic process' (GO:0010310), 'regulation of plant-type hypersensitive response' (GO:0010363), 'arsenite transport' (GO:0015700), 'photosynthesis, light reaction' (GO:0019684), 'negative regulation of defence response' (GO:0031348), 'indole glucosinolate catabolic process' (GO:0042344), 'defence response to bacterium' (GO:0042742), 'regulation of multi-organism process' (GO:0043900), 'response to arsenic-containing

substance' (GO:0046685), 'response to cadmium ion' (GO:0046686), 'defence response to fungus' (GO:0050832) and 'defence response by callose deposition in cell wall' (GO:0052544). However, the categories of 'response to metal ion' and 'phytochelatin biosynthetic process' as biological processes were common in the all the PCS sequences. As for the 'cellular components', only some sequences were inferred with terms 'chloroplast' (GO:0009507) and/or 'cytosol' (GO:0005829).

Overall, the presence of terms related to phytochelatin biosynthesis and metal-binding activities in the studied PCSs was reasonably understandable. However, ontology terms associated with different functions, such as signalling, defence, stress and phytohormone, indicated that PCSs could also have some other functions in addition to heavy metal detoxification. This was in agreement with some recent studies, suggesting novel functions for PCS and PC [19,65,66]. For example, PCs could be involved in Zn-sequestration [27] and long distance Cd<sup>2+</sup> transport [28]. The catabolism of GS conjugates has recently been linked to the peptidase as a second enzymatic activity of PCS [29,30]. In another study, AtPCS1 was suggested to be involved in the plant innate immunity [31]. All these observations implicated that PCSs could also be involved in various other metabolic processes. Therefore, further elucidation in terms of better understanding of their roles in metal homeostasis/detoxification is awaited.

### Phylogenetic analysis of PCS proteins

Another major objective of the present study was to comparatively analyse the phylogenetic distribution of PCS proteins. A phylogenetic tree was constructed by MEGA 7 with the ML method (Figure 5A). The tree demonstrated two major clusters as groups A and B, and was based on the clustering topology. Group A was further subdivided into two subgroups, A1 and A2. Group A (A1 and A2) only contained dicot species, whereas group B had monocots. Subgroup A1 included the sequences of both *Arabidopsis* PCSs, poplar (Potri.014G195800.1), tomato (Solyc09g072620.2.1) and *Medicago* (Medtr7g097190.1), whereas subgroup A2 had poplar (Potri.014G195900.1), *Medicago* (Medtr7g097200.1), common bean and cucumber sequences. However, all monocots including *Brachypodium*, *Sorghum*, maize and rice sequences were clustered in group B. Thus, a clear monocot/dicot separation was apparent in the analysed sequences. This divergence was also obvious in the exon–intron

organisations of the PCS genes where all monocots included a fixed number of six exons (Figure 5B). These variations between monocots and dicots might also be the results of exon–intron gains or losses during the time of monocot/dicot divergence. Ramos et al. [25] reported that phylogenetic tree with 20 PCS sequences showed separate clades for cyanobacteria, yeasts, nematodes, ferns and higher plants, as well as with divergence of monocots and dicots. Further, monocot–dicot divergence of 27 PCS protein sequences from bacteria and eukaryotes was identified with a higher bootstrap value. Moreover, AtPCS1 and AtPCS2 were grouped together [29]. In addition, PCS homologues of the same species, *Medicago* and poplar, were distributed in separate clusters. One variant was observed in subgroup A1 and another in subgroup A2. This suggested that in addition to metal homeostasis/detoxification, PCSs could also participate in different metabolic processes. Inferred PCS GO terms associated with signalling, defence, stress and phytohormones as well as some reports [19,27,28,30,31] support this suggestion.

### Co-expression analysis of PCS genes

To understand the interactions of PCS genes in the complex cell environment, a co-expression network analysis was performed. The network was constructed for the *AtPCS1* (At5g44070) gene using the microarray and RNA-seq datasets in ATTED-II (Figure 6). Six genes directly connected with *AtPCS1* on the network were identified, including syntaxin of plants 121 (At3g11820), S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (At1g55450), HOPZ-ACTIVATED RESISTANCE 1 (At3g50950), seven-transmembrane MLO family protein (At1g11310), calcium-dependent protein kinase 4 (At4g09570) and Ypt/Rab-GAP domain of gyp1p superfamily protein (At3g49350). The syntaxin family proteins played important roles in vesicle sorting, docking and fusion in the secretory pathway [67]. The vesicle-trafficking protein SYP121 (SYR1/PEN1) was associated with ion channel control at the plasma membrane of stomatal guard cells [68]. Plant S-adenosyl-L-methionine-dependent methyltransferases (SAM-Mtases) were the major enzymes in the phenylpropanoid, flavonoid and many other metabolic pathways [69]. Plant resistance (R) proteins involved in the plant defence against potential pathogens, e.g. *Arabidopsis* R protein HOPZ-ACTIVATED RESISTANCE 1 (ZAR1; At3g50950) [70]. The seven-transmembrane domain proteins were homologues of barley mildew resistance locus o (*MLO*) and

Table 2. Putative miRNAs for PCS transcripts, targeting positions and inhibition types.

miRNA Type	Target PCS	miRNA length	Target start-end	miRNA aligned fragment	Target aligned fragment	Inhibition type
ath-miR2934-5p	AT5G44070.1	1-20	719-737	GGUCCGCAACAGCUUUCU	CCAAG-UUUUUAGGAAGA	Cleavage
ath-miR5639-5p	AT5G44070.1	1-20	971-989	GGAUUCUGGUGACCCUGAU	UCUUU-CCACAGUGGUUA	Cleavage
ath-miR773b-3p	AT5G44070.1	1-20	696-714	UCUUUUUUGGACUUUAGUU	GGAUAAAAGCUGGA-UCGAA	Cleavage
ath-miR773b-3p	AT5G44070.1	1-20	1280-1299	UCUUUUUUGGACUUUAGUU	AGACUAGGUCAGGUAUCAA	Translation
ath-miR830-5p	AT5G44070.1	1-20	577-596	GGAUUUGAUAACCCUUCU	CUUAAACUUUUUUGGAGC	Translation
ath-miR837-5p	AT5G44070.1	1-20	1157-1176	CUUUGUUUUUUUUGACUA	GGAUAAAACAAAGGUUAGU	Translation
ath-miR838	AT5G44070.1	1-21	279-299	ACACGUUUUUUAUUUUUU	UCUGAAAGUAGUAGAAAGAA	Cleavage
ath-miR156a-i-5p	ATTG03980.1	1-20	1150-1170	CACGAGUGA-GAGAAGACAGU	GUGUUACGUCUUCUGUUG	Cleavage
ptc-miR482c-3p	Potri.014G195800.1	1-20	214-233	AUACCCUUCGAGCCUUUCU	CCUGGAGAAAGGAAAGG	Translation
ptc-miR6462c-5p	Potri.014G195800.1	1-20	962-981	AAUACGGUAAAACAGGAA	UUUGGCAUUUUUGUCUUA	Cleavage
ptc-miR6467	Potri.014G195800.1	1-21	1126-1146	UGUUUUUGGAGUCGAAUA	GUGAAUUCUAAAUCUCAA	Cleavage
ptc-miR395a	Potri.014G195900.1	1-20	405-424	UCAAGGAGUUUGGAAUC	GUUUUUUUUUUUUCUAC	Translation
sly-miR156a-c	Solyc09g072620.2.1	1-21	784-805	CACGAGAGAU-AGAAGACAGU	GUGUUUUUUUUUUUCUAC	Cleavage
mtr-miR156g-5p	Medtr7.g097190.1	1-21	964-982	CACGGGAGUAGAAGACAGU	GUGUCCUC-UUUUCUGCAA	Cleavage
mtr-miR160a.b.d.e	Medtr7.g097190.1	1-20	657-676	CCGUUUGUCCUCCGUCUCC	GCCACUAGGAAACCCUGCA	Cleavage
mtr-miR164a-c	Medtr7.g097190.1	1-20	960-978	CGUGCAGGGACGAAGAGU	ACACGUGUCU-CUUUUCUG	Translation
mtr-miR2087-3p	Medtr7.g097190.1	1-21	992-1012	CUUCAUUUUUGGUCGACGUC	GAAGCAGAAACAAUCUACG	Cleavage
mtr-miR2654	Medtr7.g097190.1	1-19	674-693	CGUGGAAACA-GGACUUA	GCAUGUUUUUACUCUGAGU	Translation
mtr-miR2673a.b	Medtr7.g097190.1	1-20	918-937	UCUUCUUUUUUUUUUUCU	AGAGGAGUUUUUGGACAG	Translation
mtr-miR5225c	Medtr7.g097190.1	1-20	962-981	UCCACAGAGAGAGAGCGU	ACGUGUCCUUUUUCUGCA	Cleavage
mtr-miR5251	Medtr7.g097190.1	1-21	1431-1451	UUCUGUUUUUUUUUUUUU	GAGACGUAACUACAUUUUC	Cleavage
mtr-miR5256	Medtr7.g097190.1	1-20	819-838	AUUAGAUAUUUAGGUAAU	UAAUUUUGAAGAAUUAUCA	Translation
mtr-miR5298b.c	Medtr7.g097190.1	1-23	797-819	AGUAGAUAUUUAGAGAGU	UUAAUUCACAUUUUGCCAU	Cleavage
mtr-miR5298d	Medtr7.g097190.1	1-20	797-816	AGUAGAUAUUUAGAGAGU	UUAAUUCACAUUUUGCCAU	Translation
mtr-miR679a-c	Medtr7.g097200.1	1-20	1063-1082	GGUUUUUUUAAAACUUAUU	GCAGAAUUUUUUAAGUGAAA	Translation
mtr-miR2615a-c	Medtr7.g097200.1	1-20	183-202	AAAAACUACCAAGAAUUGA	UCUUUCUGUUUUUUUAUG	Translation
mtr-miR2641	Medtr7.g097200.1	1-20	953-972	GAAAUUUUUUACGCUAGUCC	UUUUUUUUUUUUUUUUUAUG	Translation
mtr-miR2679a-c	Medtr7.g097200.1	1-20	1120-1139	UAUUUGCAUUUUUUUUUUU	UUUUUUUUUUUUUUUUUAUG	Translation
mtr-miR5229a.b	Medtr7.g097200.1	1-20	216-235	UGGGCAAAUUUUUUUUUUU	CCUUGUAGAAAUUGGAAAG	Cleavage
mtr-miR5554a-c-3p	Medtr7.g097200.1	1-21	174-194	GUUUCAGUAGAGAGAGCGAU	CUUAGCCACUUUUUCUUGU	Cleavage
bdi-miR395d-3p	Bradi1.g44650.1	1-21	1253-1272	ACUCUAGAGUUUUUUUUU	UUGUUCUCUGUAAUUGAAGU	Cleavage
bdi-miR722-3p	Bradi1.g44650.1	1-21	565-585	GGUUCAGAGUUUUUUUUU	UUUUGUUUUUUUUUUUAAGU	Cleavage
bdi-miR776-5p.1	Bradi1.g44650.1	1-20	569-587	CUCAAGGAGUUUUUUUUU	GGUUUUUUUUUUUUUUUAAGU	Cleavage
bdi-miR7780-3p	Bradi1.g44650.1	1-21	400-419	GGAGUAGGGUUAUUGGAAAG	GGUUUUUUUUUUUUUUUAAGU	Cleavage
sbi-miR397-5p	Bradi1.g44650.1	1-20	1175-1194	GGUUUUUUUUUUUUUUUUU	GGUUUUUUUUUUUUUUUAAGU	Translation
sbi-miR821a.c	Sobic.003G024200.1	1-19	308-327	AGAAUUCUUUCAG-GGAUG	CCUUGGCAAGUCCUUCG	Cleavage
sbi-miR821e	Sobic.003G024200.1	1-21	674-694	GUAGUUUGGACGUGAGUUAU	UUUUCGUCUUCGAGGCAAU	Cleavage
sbi-miR159b	Sobic.003G024300.1	1-20	727-746	UUUAAAUAACAAUCUGAA	UUUUCGUCUUCGUGGCUU	Cleavage
sbi-miR564c-5p	Sobic.003G024300.1	1-21	1307-1327	GUUAAAUAACAAUCUGAA	UUUUCGUCUUCGUGGCUU	Translation
sbi-miR6220-3p	Sobic.003G024300.1	1-21	1303-1322	UCCUAGGAAUUUAGGUUC	AGAG-UCCAAAUUUAAG	Translation
sbi-miR6223-3p	Sobic.003G024300.1	1-21	425-445	CGACGUCGACAAAGUUAUA	GCUGCGUUUUUUUUUUUAAG	Translation
sbi-miR6228-3p	Sobic.003G024300.1	1-20	458-477	GUAGUUUUUUUUUUUUUUU	CUUCCUAAAUAAGGAGCAU	Translation
sbi-miR821a.c	Sobic.003G024300.1	1-21	641-661	GAGAAUCCUUCUUCGUAUC	UUUCAGGGUUUUUUUUUUU	Translation
zma-miR169n-3p	Sobic.003G024300.1	1-23	575-598	GGAAUAUUU-AGAUGCGGUG	CACCUAUUUUUUUUUUUUU	Translation
osa-miR167h-3p	GRMZM2G091228.T01	1-20	1367-1387	GAUUCGUUUUUUUUUUUU	UUUUCGUCUUCGUGGCUU	Translation
osa-miR1846d-3p	LOC_Os06g01260.1	1-20	315-334	UAUUUUUUUUUUUUUUUUU	UUUUGGCAAGUCCUUCG	Translation
osa-miR1858a.b	LOC_Os06g01260.1	1-20	922-941	UACUUUUUUUUUUUUUUUU	AUGAAUUAACAAAGCAAGC	Cleavage
osa-miR2093-3p	LOC_Os06g01260.1	1-20	33-52	GGAGGAGCCGGGCGCUAU	CCUUUUUUUUUUUUUUUUU	Translation
osa-miR5544	LOC_Os06g01260.1	1-20	181-200	GGGGUAGGCGGAGGAGAG	UUUUUUUUUUUUUUUUUUU	Translation
osa-miR5819	LOC_Os06g01260.1	1-21	724-743	UAGUAAUUUAACUUUUAUA	UCCUUCUCCGUCUCCUUA	Translation
osa-miR821a-c	LOC_Os06g01260.1	1-21	125-145	GGUUUAGAGUAGGCAACA	AAGUUUUGCAUUGAAAGU	Translation
	LOC_Os06g01260.1	1-21	177-197	GCGGCGCAAGGGAGCAGGA	UCAACCUUUCUCCGUGUCC	Cleavage
	LOC_Os06g01260.1	1-23	1302-1324	AAGUUUAAAACAAACUACUGAA	UUUAAACUUUUUUUUUUUU	Translation



were localized in the plasma membrane [71]. The MLO gene family has been reported to be crucial in the resistance to the powdery mildew disease [72]. Calcium-dependent protein kinases have been reported to be present in all plants. They are confirmed in an earlier study to play important roles in metabolism, osmosis, hormone response and stress signalling pathways [73]. Ypt/Rab proteins modulate the vesicular protein transport in all eukaryotic cells [74]. In addition, based on KEGG pathway data, *AtPCS1* is connected to the plant-pathogen interaction pathway with two genes (*At4g09570* and *At4g26090*). The primary interaction partners of *AtPCS1* were associated with various pathways including defence, stress, signalling, secondary metabolite production and vesicle trafficking, indicating that PCS proteins might also participate in different metabolic processes, in addition to their metal detoxification activities. Some reports from recent studies also corroborated this, but further studies under various perturbations are essential to gain further insights into the PCS proteins.

### Promoter site analyses

The promoter sequence analysis was performed to obtain insights about the regulation of identified PCS genes. The 1000 bp upstream flanks of PCSs from the transcription start site (TSS) were supplied to the PlantCARE database for *cis*-regulatory element analysis (Figure 7). A number of *cis*-regulatory elements were revealed in the promoter sites of PCS genes and were broadly categorized as *cis*-acting elements, light responsive elements, hormone-responsive elements, tissue-specific elements, stress-responsive elements and other elements. Ten different types of *cis*-acting elements were present in PCS promoters including CAAT-box, TATA-box, 5UTR Py-rich stretch, TA-rich region, AT-rich element, AT-rich sequence, Box III, CCAAT-box, OBP-1 site and A-box. However, all PCS genes shared CAAT- and TATA-box motifs. The light responsive elements formed the largest number of *cis*-regulatory motifs in the PCS promoters with 27 members. The seed germination, seedling photomorphogenesis, shade-avoidance and photoperiod responses are some light-regulated processes. In addition, recent genomic studies also revealed that light induces massive reprogramming in the plant transcriptome with involvement of various transcription factors [75]. Therefore, this provided some clues about the presence of a high number of light responsive motifs in the promoter regions of the genes. Eleven types of identified hormone-responsive elements included

CGTCA-motif, TGACG-motif, TCA-element, ERE, TGA-element, AuxRR-core, ABRE, motif IIb, P-box, TATC-box and GARE-motif. They were associated with plant hormones, methyl jasmonate, salicylic acid, ethylene, auxin, abscisic acid and gibberellin. Phytohormones have been reported to be involved in the regulation of numerous metabolic processes, e.g. seed germination, flowering, senescence, stress response and fruit ripening [76]. Therefore, it appeared that PCS genes might be modulated by various internal and external factors. Five types of tissue-specific elements were identified in the PCS promoters including CCGTCC-box, CAT-box, GCN4\_motif, Skn-1\_motif and as-2-box. These regulatory elements were associated with meristem, endosperm and shoot tissues. In addition, PCS promoters also included stress-responsive elements related with drought, fungal elicitor, low temperature, wound and cold. Heavy metals in plants could cause oxidative stress, leading to cellular damage via lipid peroxidation, protein oxidation and DNA damage. PCs are one of the major molecules chelating heavy metals for detoxification [77]. Herein, the existence of various types and number of stress-responsive elements in PCS promoters indicated that PCS genes might also play important roles in stress conditions. Taken together, the presence of diverse *cis*-regulatory elements connected with various metabolic processes confirmed the complex as well as dynamic regulation of PCS genes in the plant genomes.

### Predicted miRNA-PCS targets

MicroRNAs (miRNAs) are non-coding endogenous RNAs, often consisting of 21–25 nucleotides. They function either by suppressing the translation of a target gene or by degrading the target mRNAs post-transcriptionally. miRNAs play important roles in many metabolic activities such as morphogenesis, signal transduction, various biotic/abiotic stresses, cell development and differentiation [78,79]. In the present study, a total of 53 miRNA types were identified for the adopted search parameters (described in *Materials and methods*). Further, their distribution per species was as follows: one for *S. lycopersicum* and *Z. mays*, four for *P. trichocarpa*, five for *B. distachyon*, seven for *O. sativa*, eight for *A. thaliana*, nine for *S. bicolor* and 18 for *M. truncatula* (Table 2). *MtPCS* were found to be targeted by 18 miRNAs. On the other hand, *SIPCS* and *ZmPCS* were targeted by a single miRNA. Ding et al. [80] reported that miR156, 166, 168, 171 and 390 were involved in rice plants in response to Cd stress. miR395 was involved in Cd detoxification in *Brassica*

**Table 3.** Distribution of secondary structure elements,  $\alpha$ -helices,  $\beta$ -sheets, coils and turns in PCS proteins.\*

Species	Transcript ID (Phytozome)	$\alpha$ -helices (aa/%) <sup>1</sup>	$\beta$ -sheets (aa/%)	Coils (aa/%)	Turns (aa/%)
<i>A. thaliana</i>	AT5G44070.1	187/38	50/10	248/51	108/22
<i>A. thaliana</i>	AT1G03980.1	188/41	54/11	210/46	96/21
<i>P. trichocarpa</i>	Potri.014G195800.1	191/37	56/11	256/50	108/21
<i>P. trichocarpa</i>	Potri.014G195900.1	168/33	59/11	269/54	116/23
<i>S. lycopersicum</i>	Solyc09g072620.2.1	206/37	34/6	305/55	128/23
<i>C. sativus</i>	Cucsa.348980.1	144/29	60/12	283/58	132/27
<i>M. truncatula</i>	Medtr7g097190.1	156/31	48/9	297/59	112/22
<i>M. truncatula</i>	Medtr7g097200.1	173/34	68/13	256/51	116/23
<i>P. vulgaris</i>	Phvul.001G162700.1	149/29	71/14	281/56	104/20
<i>B. distachyon</i>	Bradi1g44650.1	181/36	44/8	277/55	116/23
<i>S. bicolor</i>	Sobic.003G024200.1	199/39	50/9	258/50	132/26
<i>S. bicolor</i>	Sobic.003G024300.1	182/34	63/12	278/53	104/19
<i>Z. mays</i>	GRMZM2G091228_T01	174/34	65/12	268/52	108/21
<i>O. sativa</i>	LOC_Os06g01260.1	193/38	57/11	252/50	116/23

<sup>1</sup>'aa' stands for amino acids and indicates the number of amino acid residues in the given secondary structure. '%' stands for percentage and indicates the percentage of residues in the given secondary structure.

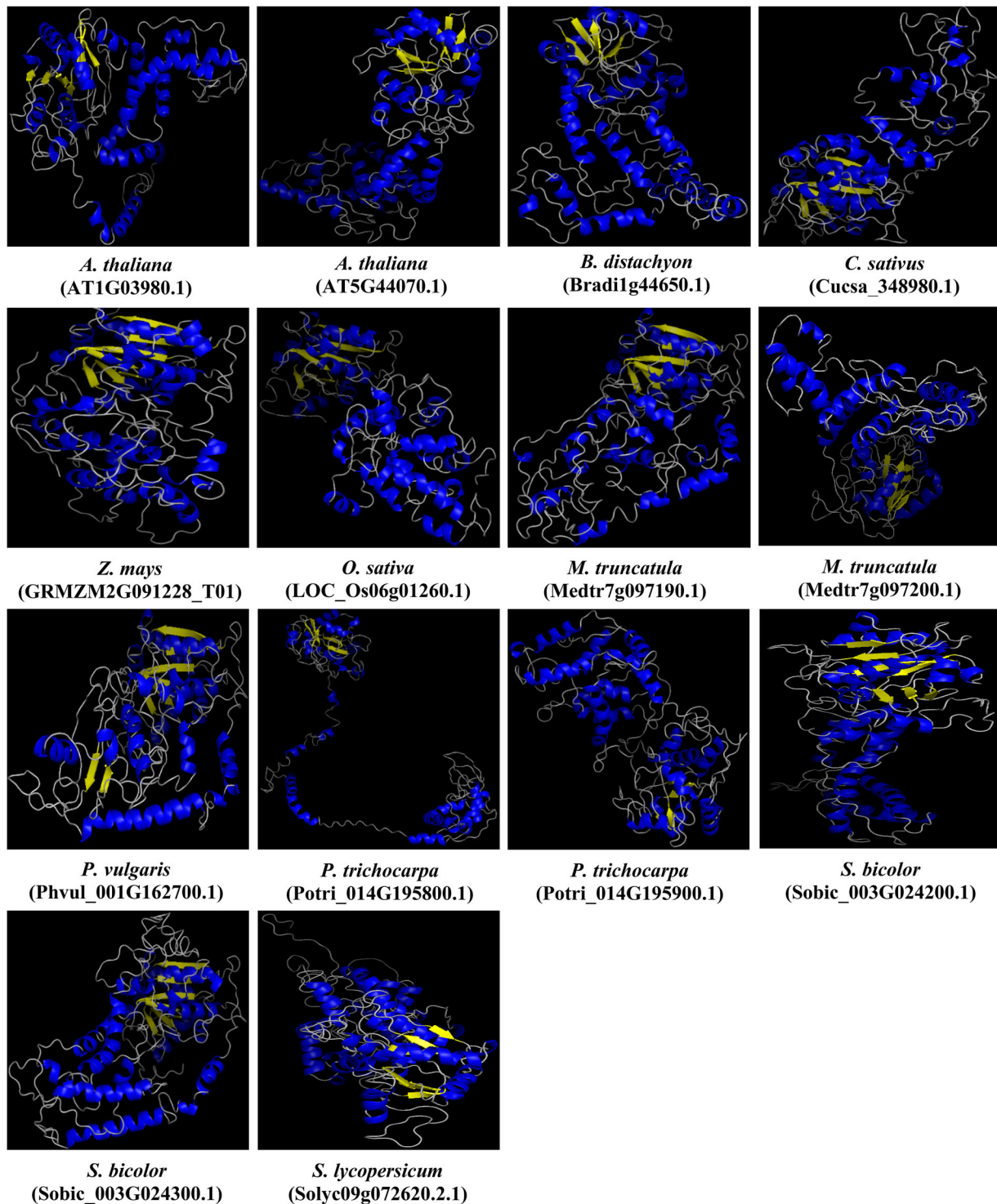
\*The percentage of secondary structure elements was calculated by using the VADAR 1.8 server.

*napus* [81]. In a different study, miR396, 397, 398 and 408 were identified as related to Cd exposure in *B. napus* [82]. In radish (*Raphanus sativus*), Cd-responsive conserved microRNAs such as miR156, 159, 160, 164, 167, 169, 395, 397 and their target genes were identified in Cd-free (CK) and Cd-treated (Cd200) libraries from radish roots. These authors also proposed that target genes of Cd-responsive miRNAs were phytochelatin synthase 1 and iron and ABC transporter proteins [83]. In two soybean genotypes (Huaxia3 and Zhonghuang24), gma-miR397a was observed to be a cadmium tolerance-associated miRNA [84]. Huang et al. [85] reported that the expression levels of bnamir393 in leaves, bna-miR156a, bna-miR167a/c in roots and leaves, and bna-miR164b and bna-miR394a/b/c in all tissues of rapeseed (*Brassica napus*) under Cd exposure showed increase. The above-reported miRNAs from Cd treatments were one of the most effective activators for PC [9]. Thereby, PCS were also identified for targeting some PCS sequences: *Arabidopsis* (AT1G03980.1) was targeted by ath-miR156a-i-5p with cleavage, *Solanum* was targeted by sly-miR156a-c with cleavage, *Medicago* (Medtr7g097190.1) targeted by mtr-miR156g-5p and mtr-miR160a,b,d,e with cleavage, and by mtr-miR164a-c with translation, *Populus* (Potri.014G195900.1) targeted by ptc-miR395a with translation, *Brachypodium* targeted by bdi-miR395d-3p and bdi-miR395p-3p with cleavage, *Sorghum* (Sobic.003G024200.1) targeted by sbi-miR397-5p with cleavage and Sobic.003G024300.1 targeted by sbi-miR159b with translation, *Zea* targeted by zma-miR169n-3p with translation and *Oryza* targeted by osa-miR167h-3p with cleavage. The above observations suggested that the regulation of PCS genes in heavy metal detoxification is a very dynamic process. Moreover, it is transcriptionally controlled by variable miRNAs either by suppressing the translation of the

target gene or by degrading the target mRNAs. Also, various miRNA types implied a highly complex regulation of metal homeostasis at a species-dependent level [86,87].

### Secondary and tertiary structures of PCSs

Finally, PCSs were investigated at secondary and tertiary structural levels to gain further insights about the functional forms of proteins. It is well documented that divergence of protein structures was observed less in comparison to sequence divergence. This indicated that selective constraints favoured to preserve the protein structure [88]. In addition, many studies show relation of the protein structural information to the amino acid replacement process, thereby resulting in protein divergence or evolution. For example, Goldman et al. [89] reported that there is a significant relationship between secondary structure environment and amino acid replacement process. Another study introduced an evolutionary model combining the protein secondary structure and amino acid replacement [88]. In another study on protein structural information, transmembrane proteins were employed for evolutionary inference [90]. In the present study, the secondary structures of PCS proteins ranged from 29 to 41% for  $\alpha$ -helices, from 6 to 14% for  $\beta$ -sheets, from 46 to 59% for coils and from 19 to 27% for turns (Table 3). Although the above secondary structure entities did not show much divergence, some variations were also present. Considering the reports from earlier studies, as well as the findings of this work, PCSs were suggested to play different metabolic roles in the plants. Therefore, it is quite reasonable to claim that these variations in secondary structures might account for the functional diversities of PCS proteins.



**Figure 8.** Predicted 3D models of PCS proteins in 10 different plants. Models were generated by using the Phyre2 server at intensive mode. Blue and yellow colours, respectively, show the  $\alpha$ -helices and  $\beta$ -sheets, whereas other structures such as turns and coils are represented in grey.

Moreover, using the Phyre2 server at intensive mode for all 14 PCS sequences, we generated 3D models of the PCS proteins in the present study. Template '2BTW' from structures of transferase from *Nostoc* sp. PCC 7120 with resolution of 2.0 Å was used in homology modelling (Figure 8). The qualities of

generated models were satisfactorily verified by Ramachandran plot analysis where  $\geq 90\%$  of residues were in the allowed region. The models of *Arabidopsis* AtPCS1 (AT5G44070.1) and 2 (AT1G03980.1) were then superimposed with models of other PCSs on the basis of the alpha carbon superposition in order to figure



out the degree of conservancy, thereby protein evolution. The superposed structural overlaps ranged from 41.44 to 56.91% for AtPCS1 and 45.58 to 60.84% for AtPCS2. The highest structural overlap (60.84%) was between dicot AtPCS2 and monocot ZmPCS, whereas the lowest value (41.44%) was noticed between dicots AtPCS1 and MtPCS (Medtr7g097200.1). Interestingly, contrary to the general acceptance that suggested divergence of protein structures occurs much less rapidly in comparison to divergence of protein sequences [88], in the present work, the 3D structures of PCS proteins were relatively less conserved when compared with primary protein sequences (refer to 'sequence analysis' section). So, it could be speculated that these variations in protein folding state/conformation might be somehow related to the diverged protein functions, however further experimental analysis is required to confirm this.

## Conclusions

The present study identified 14 putative PCS homologues in the genomes of 10 selected plant species (*A. thaliana*, *P. trichocarpa*, *S. lycopersicum*, *M. truncatula*, *C. sativus*, *P. vulgaris*, *B. distachyon*, *S. bicolor*, *Z. mays* and *O. sativa*) responsible for the biosynthesis of one of the main groups of heavy-metal detoxifying peptides phytochelatins (PCs) in plants. PCS proteins showed variations in protein length and some differences in the C-terminal site, indicating functional diversities of PCSs in plants. In addition, vital residues such as Cys56, His162 and Asp180 are well conserved in PCSs, and PCSs were found to be related to many metabolic pathways, proving the importance of PCS genes in plant metabolism as well as metal detoxification. Thus, this study was an initial, theoretical step that can provide a basis for future studies into the PCS genes in higher plants.

## Disclosure statement

The authors declare that they have no conflict of interest.

## References

- [1] Ozturk A, Yarci C, Ozyigit II. Assessment of heavy metal pollution in Istanbul using plant (*Celtis australis* L.) and soil assays. *Biotechnol Biotechnol Equip.* 2017;31: 948–954.
- [2] Osma E, Ozyigit II, Leblebici Z, et al. Determination of heavy metal concentrations in tomato (*Lycopersicon esculentum* Miller) grown in different station types. *Rom Biotech Lett.* 2012;17:6962–6974.
- [3] Dogan I, Ozyigit II, Demir G. Influence of aluminium on mineral nutrient uptake and accumulation in *Urtica pilulifera* L. *J Plant Nutr.* 2014;37:469–481.
- [4] Peterson AG, Oliver DJ. Leaf-targeted phytochelatin synthase in *Arabidopsis thaliana*. *Plant Physiol Biochem.* 2006;44:885–892.
- [5] Cobbett C, Goldsbrough P. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. *Annu Rev Plant Biol.* 2002;53:159–182.
- [6] Vivares D, Arnoux P, Pignol D. A papain-like enzyme at work: native and acyl-enzyme intermediate structures in phytochelatin synthesis. *Proc Natl Acad Sci USA.* 2005;102:18848–18853.
- [7] Kanaujia SP. 13 understanding toxic metal-binding. In: Surajit Das, Hirak Ranjan Dash, editors. *Handbook of metal-microbe interactions and bioremediation.* Boca Raton, FL, USA: CRC Press, Taylor and Francis Group; 2017. p. 201.
- [8] Inouhe M. Phytochelatins. *Braz J Plant Physiol.* 2005; 17:65–78.
- [9] Ha SB, Smith AP, Howden R, et al. Phytochelatin synthase genes from *Arabidopsis* and the yeast *Schizosaccharomyces pombe*. *Plant Cell.* 1999;11: 1153–1164.
- [10] Tsuji N, Nishikori S, Iwabe O, et al. Comparative analysis of the two-step reaction catalyzed by prokaryotic and eukaryotic phytochelatin synthase by an ion-pair liquid chromatography assay. *Planta.* 2005;222: 181–191.
- [11] Vatamaniuk OK, Mari S, Lang A, et al. Phytochelatin synthase, a dipeptidyltransferase that undergoes multisite acylation with  $\gamma$ -glutamylcysteine during catalysis. *J Biol Chem.* 2004;279:22449–22460.
- [12] Cobbett CS. Phytochelatins and their roles in heavy metal detoxification. *Plant Physiol.* 2000;123:825–832.
- [13] Maier T, Yu C, Kullertz G, et al. Localization and functional characterization of metal-binding sites in phytochelatin synthases. *Planta.* 2003;218:300–308.
- [14] Romanyuk ND, Rigden DJ, Vatamaniuk OK, et al. Mutagenic definition of a papain-like catalytic triad, sufficiency of the N-terminal domain for single-site core catalytic enzyme acylation, and C-terminal domain for augmentative metal activation of a eukaryotic phytochelatin synthase. *Plant Physiol.* 2006;141:858–869.
- [15] Wang F, Wang Z, Zhu C. Heteroexpression of the wheat phytochelatin synthase gene (*TaPCS1*) in rice enhances cadmium sensitivity. *Acta Bioch Bioph Sin.* 2012;44:886–893.
- [16] Li JC, Guo JB, Xu WZ, et al. RNA interference-mediated silencing of phytochelatin synthase gene reduce cadmium accumulation in rice seeds. *J Integr Plant Biol.* 2007;49:1032–1037.
- [17] Vatamaniuk OK, Mari S, Lu YP, et al. AtPCS1, a phytochelatin synthase from *Arabidopsis*: isolation and in vitro reconstitution. *Proc Natl Acad Sci USA.* 1999;96: 7110–7115.
- [18] Heiss S, Wachter A, Bogs J, et al. Phytochelatin synthase (PCS) protein is induced in *Brassica juncea* leaves after prolonged Cd exposure. *J Exp Bot.* 2003; 54:1833–1839.

- [19] Meyer CL, Peisker D, Courbot M, et al. Isolation and characterization of *Arabidopsis halleri* and *Thlaspi caerulescens* phytochelatin synthases. *Planta*. 2011;234:83–95.
- [20] Clemens S, Kim EJ, Neumann D, et al. Tolerance to toxic metals by a gene family of phytochelatin synthases from plants and yeast. *EMBO J*. 1999;18:3325–3333.
- [21] Clemens S, Schroeder JI, Degenkolb T. *Caenorhabditis elegans* expresses a functional phytochelatin synthase. *Eur J Biochem*. 2001;268:3640–3643.
- [22] Lee S, Korban S. Transcriptional regulation of *Arabidopsis thaliana* phytochelatin synthase (AtPCS1) by cadmium during early stages of plant development. *Planta*. 2002;215:689–693.
- [23] Semane B, Cuypers A, Smeets K, et al. Cadmium responses in *Arabidopsis thaliana*: glutathione metabolism and antioxidative defence system. *Physiol Plant*. 2007;129:519–528.
- [24] Gasic K, Korban SS. Expression of *Arabidopsis* phytochelatin synthase in Indian mustard (*Brassica juncea*) plants enhances tolerance for Cd and Zn. *Planta*. 2007;225:1277–1285.
- [25] Ramos J, Clemente MR, Naya L, et al. Phytochelatin synthases of the model legume *Lotus japonicus*. A small multigene family with differential response to cadmium and alternatively spliced variants. *Plant Physiol*. 2007;143:1110–1118.
- [26] Lee BD, Hwang S. Tobacco phytochelatin synthase (NtPCS1) plays important roles in cadmium and arsenic tolerance and in early plant development in tobacco. *Plant Biotechnol Rep*. 2015;9:107–114.
- [27] Tennstedt P, Peisker D, Bottcher C, et al. Phytochelatin synthesis is essential for the detoxification of excess zinc and contributes significantly to the accumulation of zinc. *Plant Physiol*. 2008;149:938–948.
- [28] Mendoza-Cózatl DG, Springer F, et al. Identification of high levels of phytochelatin, glutathione and cadmium in the phloem sap of *Brassica napus*. A role for thiol-peptides in the long-distance transport of cadmium and the effect of cadmium on iron translocation. *Plant J*. 2008;54:249–259.
- [29] Clemens S. Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. *Biochimie*. 2006;88:707–719.
- [30] Blum R, Meyer KC, Wunschmann J, et al. Cytosolic action of phytochelatin synthase. *Plant Physiol*. 2010;153:159–169.
- [31] Clay NK, Adio AM, Denoux C, et al. Glucosinolate metabolites required for an *Arabidopsis* innate immune response. *Science*. 2009;323:95–101.
- [32] DalCorso G. Heavy metal toxicity in plants. In: Antonella Furini, editor. *Plants and heavy metals*. Netherlands: Springer; 2012. p.1–25.
- [33] UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–D169. <https://www.uniprot.org/help/publications>
- [34] Phytozome, The Plant Comparative Genomics Portal. The Regents of the University of California; c.1997–2017 [Accessed 2018]. Available from: <https://phytozome.jgi.doe.gov/pz/portal.html>
- [35] Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40:1178–1186.
- [36] El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. Forthcoming 2019:gky995. DOI:10.1093/nar/gky995, <https://pfam.xfam.org/>
- [37] Punta M, Coghill PC, Eberhardt RY, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012;44:D279–D285.
- [38] Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. *The proteomics protocols handbook*. Louisville, KY (USA): Humana; 2005. p. 571–607.
- [39] Yu CS, Chen YC, Lu CH, et al. Prediction of protein subcellular localization. *Proteins*. 2006; 64:643–651.
- [40] Motif-based analysis of DNA, RNA and protein sequences. 2018. Available from: <http://meme-suite.org/tools/meme>
- [41] Timothy L, Mikael B, Buske FA, et al. Meme Suite: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:202–208.
- [42] Thompson JD, Higgins DG, Gibson TJ. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22:4673–4680.
- [43] Kumar S, Stecher G, Tamura K. Mega7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;3:1870–1874.
- [44] Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. 1992;8:275–282.
- [45] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39:783–791.
- [46] PlantCARE, a database of plant cis-acting regulatory elements and a portal. 2018. Available from: <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>
- [47] Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res*. 2002;30:325–327.
- [48] A Plant Small RNA Target Analysis Server. 2018. Available from: <https://plantgrn.noble.org/psRNATarget/>
- [49] Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res*. 2018;46:W49–W54.
- [50] Gene coexpression database. 2018. Available from: <http://atted.jp/>
- [51] Aoki Y, Okamura Y, Tadaka S, et al. ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. *Plant Cell Physiol*. 2016;57:e5.
- [52] Obayashi T, Okamura Y, Ito S, et al. ATTED-II in 2014: evaluation of gene co-expression in agriculturally important plants. *Plant Cell Physiol*. 2014;55:e6. [7. [cited 2018 Sep 19] p.] DOI: 10.1093/pcp/pct178
- [53] Structural Bioinformatics Group - Imperial College London. 2018. Protein Homology/analogy Recognition Engine V 2.0. Available from: <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>

- [54] Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10:845.
- [55] Ramachandran Plot Analysis. 2018. Available from: <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>
- [56] Single (or Multiple) Model Protein Structure Analysis - Volume, Area, Dihedral Angle Reporter. 2018. Available from: <http://vadar.wishartlab.com/>
- [57] Willard L, Ranjan A, Zhang H, et al. Vadar: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* 2003;31:3316–3319.
- [58] Topology Independent Comparison of Biomolecular 3D Structures. 2018. Available from: <http://mspc.bii.a-star.edu.sg/minhn/pairwise.html>
- [59] Nguyen MN, Tan KP, Madhusudhan MS. CLICK - Topology-independent comparison of biomolecular 3D structures. *Nucleic Acids Res.* 2011;39:W24–W28.
- [60] Xu L, Zhu L, Tu L, et al. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *J Exp Bot.* 2011;62:5607–5621.
- [61] Lorkovic ZJ, Wiczeorek KDA, Lambermon MHL, et al. Pre-mRNA splicing in higher plants. *Trends Plant Sci.* 2000;5:160–167.
- [62] Das N, Bhattacharya S, Bhattacharyya S, et al. Identification of alternatively spliced transcripts of rice phytochelatin synthase 2 gene *OsPCS2* involved in mitigation of cadmium and arsenic stresses. *Plant Mol Biol.* 2017;94:167–183.
- [63] Uraguchi S, Tanaka N, Hofmann C, et al. Phytochelatin synthase has contrasting effects on cadmium and arsenic accumulation in rice grains. *Plant Cell Physiol.* 2017;58:1730–1742.
- [64] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000; 25:25–29.
- [65] Kuhnlenz T, Westphal L, Schmidt H, et al. Expression of *Caenorhabditis elegans* PCS in the *AtPCS1*-deficient *Arabidopsis thaliana* cad1-3 mutant separates the metal tolerance and non-host resistance functions of phytochelatin synthases. *Plant Cell Environ.* 2015;38: 2239–2247.
- [66] De Benedictis M, Brunetti C, Brauer EK, et al. The *Arabidopsis thaliana* knockout mutant for phytochelatin synthase1 (*cad1-3*) is defective in callose deposition, bacterial pathogen defense and auxin content. *Front Plant Sci.* 2018;9:19.
- [67] Zheng H, Bassham DC, da Silva Conceição A, et al. The syntaxin family of proteins in *Arabidopsis*: a new syntaxin homologue shows polymorphism between two ecotypes. *J Exp Bot.* 1999;50:915–924.
- [68] Eisenach C, Chen ZH, Grefen C, et al. The trafficking protein SYP121 of *Arabidopsis* connects programmed stomatal closure and  $K^+$  channel activity with vegetative growth. *Plant J.* 2012;69:241–251.
- [69] Joshi CP, Chiang VL. Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant Mol Biol.* 1998;37:663–674.
- [70] Lewis JD, Wu R, Guttman DS, et al. Allele-specific virulence attenuation of the *Pseudomonas syringae* HopZ1a type III effector via the Arabidopsis ZAR1 resistance protein. *PLoS Genet.* 2010;6:e1000894.
- [71] Devoto A, Hartmann HA, Piffanelli P, et al. Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *J Mol Evol.* 2003;56: 77–88.
- [72] Miao LX, Jiang M, Zhang YC, et al. Genomic identification, phylogeny, and expression analysis of *MLO* genes involved in susceptibility to powdery mildew in *Fragaria vesca*. *Genet Mol Res.* 2016;15:1–12.
- [73] Valmonte GR, Arthur K, Higgins CM, et al. Calcium-dependent protein kinases in plants: evolution, expression and function. *Plant Cell Physiol.* 2014;55: 551–569.
- [74] Albertš, WE, Gallwitz D. Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J.* 1999;18: 5216–5225.
- [75] Lee JH, Jung JH, Park CM. Light inhibits COP1-mediated degradation of ICE transcription factors to induce stomatal development in *Arabidopsis*. *Plant Cell.* 2017;29:2817–2830.
- [76] McAtee P, Karim S, Schaffer RJ, et al. A dynamic interplay between phytohormones is required for fruit development, maturation, and ripening. *Front Plant Sci.* 2013;4:79.
- [77] Yadav SK. Heavy metals toxicity in plants: an overview on the role of glutathione and phytochelatin in heavy metal stress tolerance of plants. *S Afr J Bot.* 2010;76:167–179.
- [78] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 2004;116:281–297.
- [79] Lv S, Nie X, Wang L, et al. Identification and characterization of MicroRNAs from barley (*Hordeum vulgare* L.) by high-throughput sequencing. *Int J Mol Sci.* 2012;13:2973–2984.
- [80] Ding Y, Chen Z, Zhu C. Microarray-based analysis of cadmium responsive microRNAs in rice (*Oryza sativa*). *J Exp Bot.* 2011;62:3563–3573.
- [81] Zhang LW, Song JB, Shu XX, et al. miR395 is involved in detoxification of cadmium in *Brassica napus*. *J Hazard Mater.* 2013;250:204–211.
- [82] Zhou ZS, Song JB, Yang ZM. Genome-wide identification of *Brassica napus* microRNAs and their targets in response to cadmium. *J Exp Bot.* 2012;63: 4597–4613.
- [83] Xu L, Wang Y, Zhai L, et al. Genome-wide identification and characterization of cadmium-responsive microRNAs and their target genes in radish (*Raphanus sativus* L.) roots. *J Exp Bot.* 2013; 64:4271–4287.
- [84] Fang X, Zhao Y, Ma Q, et al. Identification and comparative analysis of cadmium tolerance-associated miRNAs and their targets in two soybean genotypes. *PLoS One.* 2013;8:e81471.
- [85] Huang SQ, Xiang AL, Che LL, et al. A set of miRNAs from *Brassica napus* in response to sulfate deficiency and cadmium stress. *Plant Biotechnol J.* 2010;8: 887–899.
- [86] Zhou ZS, Zeng HQ, Liu ZP, et al. Genome-wide identification of *Medicago truncatula* microRNAs and their



- targets reveals their differential regulation by heavy metal. *Plant Cell Environ.* 2012;35:86–99.
- [87] Yang ZM, Chen J. A potential role of microRNAs in plant response to metal toxicity. *Metallomics.* 2013;5:1184–1190.
- [88] Thorne JL, Goldman N, Jones DT. Combining protein evolution and secondary structure. *Mol Biol Evol.* 1996;13:666–673.
- [89] Goldman N, Thorne JL, Jones DT. Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Mol Biol Evol.* 1999;149:445–458.
- [90] Lio P, Goldman N. Using protein structural information in evolutionary inference: transmembrane proteins. *Mol Biol Evol.* 1999;16:1696–1710.