Journal of Experimental Botany, Vol. 64, No. 17, pp. 5371–5381, 2013
doi:10.1093/jxb/ert331 Advance Access publication 31 October, 2013

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

The CEP family in land plants: evolutionary analyses, expression studies, and role in *Arabidopsis* shoot development

Ianto Roberts[1,2,†], Stephanie Smith[3,†], Bert De Rybel[1,2,*], Jana Van Den Broeke[1,2], Wouter Smet[1,2], Sarah De Cokere[1,2], Marieke Mispelaere[1,2], Ive De Smet[1,2,3,4,‡,$] and Tom Beeckman[1,2,‡]

[1] Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium
[2] Department of Plant Biotechnology and Genetics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium
[3] Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK
[4] Centre for Plant Integrative Biology, University of Nottingham, Sutton Bonington LE12 5RD, UK
* Present address: Laboratory of Biochemistry, Wageningen University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands.

† These authors contributed equally to this manuscript.
‡ These authors contributed equally to this manuscript.
$ To whom correspondence should be addressed. E-mail: ive.desmet@psb.vib-ugent.be

Received 18 June 2013; Revised 3 September 2013; Accepted 4 September 2013

Abstract

In *Arabidopsis*, more than 1000 putative small signalling peptides have been predicted, but very few have been functionally characterized. One class of small post-translationally modified signalling peptides is the C-TERMINALLY ENCODED PEPTIDE (CEP) family, of which one member has been shown to be involved in regulating root architecture. This work applied a bioinformatics approach to identify more members of the CEP family. It identified 10 additional members and revealed that this family only emerged in flowering plants and was absent from extant members of more primitive plants. The data suggest that the CEP proteins form two subgroups according to the CEP domain. This study further provides an overview of specific *CEP* expression patterns that offers a comprehensive framework to study the role of the CEP signalling peptides in plant development. For example, expression patterns point to a role in aboveground tissues which was corroborated by the analysis of transgenic lines with perturbed *CEP* levels. These results form the basis for further exploration of the mechanisms underlying this family of peptides and suggest their putative roles in distinct developmental events of higher plants.

Key words: *Arabidopsis*, small signalling peptides, phylogeny, evolutionary analyses, *CEP* expression, shoot development.

Introduction

In the last 20 years, the importance of small signalling peptides in plant cell-to-cell communication has become increasingly clear, with several families of small signalling peptides known at present to play vital roles in plant growth and development (Butenko *et al.*, 2009; Matsubayashi, 2011; Murphy *et al.*, 2012; Czyzewicz *et al.*, 2013). The majority of small signalling peptides falls into one of two broad groups: the cysteine-rich peptides, which are characterized by a typical mature peptide length of <160 amino acids with a cysteine-rich C-terminal domain; and the small post-translationally modified peptides, which are expressed as longer precursor proteins before undergoing post-translational modifications and subsequent cleavage to form an active, mature peptide <20 amino acids in length (Matsubayashi, 2011; Murphy *et al.*, 2012). Three types of post-translational modification—tyrosine sulfation, proline hydroxylation, and

hydroxyprolinearabinosylation—are known to occur in small post-translationally modified peptides, and these modifications appear to be crucial for optimal peptide bioactivity (Matsubayashi, 2011; Shinohara and Matsubayashi, 2013).

Identification of small signalling peptides in plants can be challenging due to the small size of their encoding genes (Murphy *et al.*, 2012). An *in silico* approach, specifically designed to identify genes encoding the hallmarks of small peptide signals, led to the discovery of the C-TERMINALLY ENCODED PEPTIDE (CEP) family of putative small post-translationally modified peptides (Ohyama *et al.*, 2008). In this initial study, five small genes encoding peptides of 82–126 amino acids in length were identified. The five different members of the CEP family display considerable sequence diversity in the majority of the expressed protein, with the exception of a conserved domain at the C-terminus, which represents the mature, active peptide following proteolytic cleavage from the expressed precursor. This was confirmed using mass spectrometry on *CEP1* (At1g47485) overexpression lines, further revealing that the mature product of 15 amino acids contains two hydroxyprolinated residues (Ohyama *et al.*, 2008). Initial analyses showed that overexpression of *CEP1* arrests root growth through repression of meristematic cell division and expansion, while no effects were observed in the quiescent centre and adjacent stem cells (Ohyama *et al.*, 2008).

Apart from these observations restricted to one member of the CEP family, hardly anything is known about the other members and their potential importance for plant growth and development. Therefore, this study further explored this family of small signalling peptides. First, the phylogenetic results showed that the CEP family contains more than five members and is conserved throughout higher land plants. Second, analysis of the expression of CEP family members revealed distinct patterns throughout plant development. Third, altering the expression levels of CEP family members resulted in dramatic growth and developmental phenotypes in aboveground plant parts.

Materials and methods

Plant growth

For analyses of aboveground parts, plants were grown on Levington M3 compost (Everris, Ipswich UK) in a glasshouse at 20–22 °C under long-day conditions (16/8 light/dark). Measurements were taken after 3 weeks of growth, immediately post-bolt. For rosette area quantification and leaf series analysis, plants were grown horizontally on square Petri plates (12 cm × 12 cm, Greiner Labortechnik) containing 50 ml solid half-strength MS growth medium (per litre: 2.15 g MS salts, 0.1 g *myo*-inositol, 0.5 g MES, 10 g plant tissue culture agar; pH adjusted to 5.7 with KOH) in a growth room at 22 °C under continuous light. Measurements were taken at 18 d after germination. For *GUS* expression analyses, seedlings were grown at 22 °C under continuous light (110 µE m^{-2} s^{-1} photosynthetically active radiation, supplied by cool-white fluorescent tungsten tubes, Osram) on square Petri plates containing 50 ml solid half-strength MS growth medium supplemented with 1% sucrose (per litre: 2.15 g MS salts, 0.1 g *myo*-inositol, 0.5 g MES, 10 g sucrose, 8 g plant tissue culture agar; pH adjusted to 5.7 with KOH), and flowering plants were grown in a greenhouse at 21 °C under long-day conditions.

Sequence identification and conserved motif analysis of CEP proteins

New members of the CEP family in *Arabidopsis thaliana* were identified using the 15-amino-acid mature region from known CEP peptides as input sequences for a TBLASTN search in all six open reading frames of the complete *A. thaliana* genome nucleotide sequence (http://blast.ncbi.nlm.nih.gov/Blast.cgi), and CEP assignment and naming was aligned with the results from Delay *et al.* (2013). The position on the genome of each hit was determined (using the SeqViewer browser from TAIR, http://tairvm09.tacc.utexas.edu/servlets/sv) and was screened in all six possible open reading frames (using the translate tool from ExPASy; http://web.expasy.org/translate/) for a peptide of approximately 75–250 amino acids (Supplementary Dataset S1, available at *JXB* online). If these proteins contained an N-terminal signal peptide (SignalP4.1; http://www.cbs.dtu.dk/services/SignalP; with standard settings, except for a D-cutoff of 0.45) and a 'CEP-like' sequence in the C-terminal region of the peptide, they were classified as CEP peptides. To identify potential members of the CEP gene family in the plant lineage, *A. thaliana* CEP family members were used as a query in BLAST searches against Phytozome version 9.0 (http://www.phytozome.net/search.php), which contains the most up-to-date list of genomes (13 December 2012). The BLAST analyses using full-length protein sequences were performed using standard settings (except for an E threshold of 10). The program MEME (Bailey *et al.*, 2009) (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) was used to identify motifs in the candidate CEP protein sequences. MEME was run with the following parameters: number of repetitions = any, maximum number of motifs = 5, and with optimum motif widths constrained to between 6 and 50 residues. Weblogo (http://weblogo.berkeley.edu/logo.cgi) was used with standard settings to represent the consensus CEP domains.

Data mining analyses

For environmental and hormonal effects on CEP genes, the transcriptome meta-analysis tool Genevestigator (Hruz *et al.*, 2008) was used, with a significance level of <0.05. For cell, tissue, and organ *CEP* expression data, the eFP browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) was used with standard settings.

Alignment and phylogenetic analysis

Phylogenetic analyses of the CEP proteins and CEP domains based on amino acid sequences were carried out using UPGMA methods in the CLC Main Workbench version 6.8.1 (www.clcbio.com). Support for each node was tested with 10 000 bootstrap replicates.

CEP *constructs*

Gateway cloning was used for every construct. Entry clones containing the *CEP* promoter sequences (*CEP1*, 1997 bp; *CEP2*, 1400 bp; *CEP3*, 1560 bp; *CEP4*, 2000 bp; *CEP5*, 900 bp) were created by cloning PCR-fragments into the pDONRP4P1R vector. The *pCEPx::GUS* constructs were created by cloning the promoter fragment into the pEX-K7SNFm14GW destination vector. An entry clone containing the genomic coding sequence of CEP5 (318 bp) was created by cloning the PCR-fragment into the pDONR221 vector. The *p35S::CEP5* construct was created by cloning this genomic coding sequence in the pK7GW2 destination vector. The *CEP5* RNAi construct was created using the pAGRIKOLA constructs (Hilson *et al.*, 2004). Constructs were transformed in Col-0 using floral dip (Clough and Bent, 1998).

GUS *expression*

For the GUS assays, plants were put overnight in 90% acetone, then transferred to a GUS-solution [1 mM X-Glc, 0.5% (v/v) dimethylformamide (DMF), 0.5% (v/v) Triton X-100, 1 mM EDTA (pH 8), 0.5 mM potassium ferricyanide ($K_3Fe(CN)_6$), 0.5% potassium

ferrocyanide ($K_4Fe(CN)_6$), 500 mM phosphate buffer (pH 7)] and incubated at 37 °C for GUS staining, and finally washed in 500 mM phosphate buffer (pH 7). For microscopic analysis, samples were cleared with 90% lactic acid or as described in Malamy and Benfey (1997). Samples were analysed by differential interference contrast microscopy (Olympus BX53) and a stereomicroscope (Leica MZ16).

RNA extraction, cDNA synthesis, and qRT-PCR analysis

Arabidopsis RNA was isolated from 20 pooled seedlings at 7 d after germination using a Plant RNeasy Kit (Qiagen, Germany) according to the manufacturer's instructions. cDNA was subsequently prepared from a minimum of 250 ng RNA (determined by UV spectrophotometry) using a SuperScript II reverse transcriptase kit and Oligo(dT)12–18 primers (Invitrogen, USA), according to the manufacturer's instructions. Quantitative real-time PCR (qRT-PCR) was performed in a 384-well white dish format using a LightCycler 480 (Roche Applied Science, USA) with 40 PCR amplification cycles using SYBR Green I fluorescent dye (Quanta Biosciences, USA) and primers for *CEP5* (5′-CCATGGACGAACCCTAAAAG-3′ and 5′-TGCCATCATCGTCTTGCTAT-3′) and *ACTIN* (5′-CTGGAGGTTTTGAGGCTGGTAT-3′ and 5′-CCAAGGGTGAAAGCAAGAAGA-3′). Expression was determined from a minimum of three biological replicates, each with three technical repeats, and normalized against ACTIN.

Results and discussion

Extended CEP family in A. thaliana

Originally, the CEP family was described as containing five members in *A. thaliana* (Ohyama *et al.*, 2008). In addition to these five, the current work and Delay *et al.* (2013) identified 10 additional members. We used BLAST searches and filtered for proteins with an N-terminal signal peptide sequence, a C-terminal 'CEP-like' sequence, and a total length of 75–250 amino acids (Table 1). The genes are located on different chromosomes, but not chromosome 4, and *CEP3* and *CEP11* and *CEP5*, *CEP6*, *CEP7*, and *CEP8* are organized in tandem repeats (Supplementary Fig. S1). The phylogenetic relationships between the CEP family proteins are shown in Fig. 1A.

In agreement with the five original CEP family members, this study predicted the presence and location of putative signal peptide cleavage sites in all CEP pre-propeptides using SignalP4.1 (Fig. 1B and Supplementary Dataset S1). Although SignalP4.1 does not always predict this, it is likely that the cleavage occurs at a conserved arginine at the N-terminus (Supplementary Dataset S1).

Previously, a C-terminal conserved domain was identified in the original five members of the CEP family (Ohyama *et al.*, 2008). To identify the presence and distribution of this domain in all CEP family members, pattern analyses were performed on the full-length *A. thaliana* CEP proteins using MEME (with optimized settings following iterative analyses) (Bailey and Elkan, 1994). Indeed, a motif that is similar to the active CEP1 peptide sequence (Ohyama *et al.*, 2008) is present across all CEP protein sequences (Fig. 1B and Supplementary Dataset S1), and therefore this was called the CEP domain. Interestingly, in several instances the CEP domain occurs multiple times within one pre-propeptide and not exclusively at the C-terminus (Fig. 1B and Supplementary Dataset S2).

This work then used all the CEP domains to build a phylogenetic tree, which revealed two main branches within the CEP family (Fig. 2A). Based on these phylogenetic relationships, the CEP family was divided into two groups: group I, CEP1–CEP12; and group II, CEP13–CEP15 (Fig. 2A). The amino acid sequences for the CEP domains for groups I and II were separately aligned, which resulted in a consensus sequence for these two groups (Fig. 2B). For both groups, the C-terminal part of the CEP domain (SPGV/IGH) showed high amino acid similarity (Fig. 2B). The CEP domain in group I contains three prolines, while in group II it contains two prolines. This is important as LC-MS/MS analysis of CEP1 revealed hydroxylation of some of the prolines within the CEP domain (Ohyama *et al.*, 2008). This hydroxyprolination likely affects bioactivity and hydrophilic nature of the CEP peptides. In future, it will be important to assess to what extent these *in silico* results are supported by biological validation.

Table 1. *The 15 members of the* CEP *family*

Gene	Locus	Publication
CEP1	At1g47485	Ohyama *et al.* (2008)
CEP2	At1g59835	Ohyama *et al.* (2008)
CEP3	At2g23440	Ohyama *et al.* (2008)
CEP4	At2g35612	Ohyama *et al.* (2008)
CEP5	At5g66815	Ohyama *et al.* (2008)
CEP6	At5g66816	This study and Delay *et al.* (2013)
CEP7	Between At5g66817–At5g66820[a]	This study and Delay *et al.* (2013)
CEP8	Between At5g66817–At5g66820[a]	This study and Delay *et al.* (2013)
CEP9	At3g50610	This study and Delay *et al.* (2013)
CEP10	Between At1g36040–At1g36050[a]	This study and Delay *et al.* (2013)
CEP11	Between At2g23440–At2g23450[a]	This study and Delay *et al.* (2013)
CEP12	At1g31670[b]	This study and Delay *et al.* (2013)
CEP13	At1g16950	This study and Delay *et al.* (2013)
CEP14	At1g29290	This study and Delay *et al.* (2013)
CEP15	At2g40530	This study and Delay *et al.* (2013)

[a], not listed on TAIR; [b], likely incorrectly annotated on TAIR.

CEP family is evolutionarily conserved in monocot and eudicot plants

Notwithstanding the fact that CEP pre-proproteins are short and the CEP domain is only 15 amino acids (AAs) long (Figs. 1 and 2), BLAST with *A. thaliana* full-length CEP proteins and 15-amino-acid CEP domain sequences and phylogenetic analyses were used to identify CEP family genes within the supergroup Plantae (data not shown). This revealed that CEP peptides are present in eudicots and monocots, but absent in lower land plants (*Selaginella moellendorffii* and *Physcomitrella patens*) and in the green algae for which a genome sequence was available (Fig. 3).

CEP family members display limited transcriptional control by hormones and nutrients in Arabidopsis

Whilst small signalling peptides are often not well represented on available micro-arrays (Murphy *et al.*, 2012), *in*

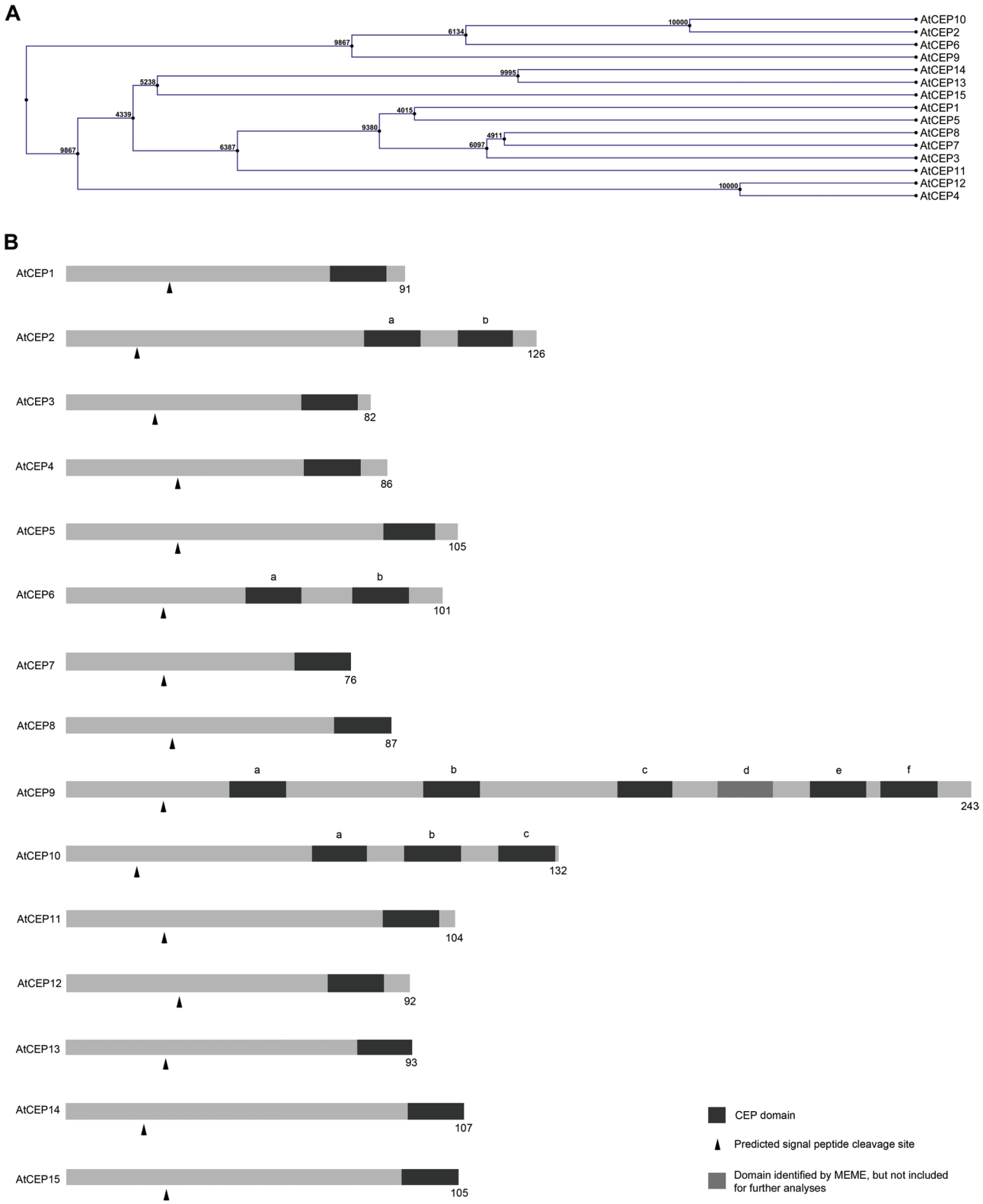

Fig. 1. The CEP family in *Arabidopsis*. (A) Phylogenetic tree of *Arabidopsis* CEP family members (based on full length protein sequences). Bootstrap values based on 10 000 replications are shown at branch nodes. (B) Schematic representation of *Arabidopsis* CEP family members. Numbers are number of amino acids). Dark grey, CEP domain; light grey, predicted CEP domain by MEME but with deviating sequence and not included for further analyses; arrowhead, predicted signal peptide cleavage site.

silico expression data for seven out of 15 *CEP*s were available. To examine, *CEP* family gene expression changes under several hormonal and nutritional stimuli, this study used the transcriptome meta-analysis tool Genevestigator (Hruz *et al.*, 2008) (Table 2). Both *CEP5* and *CEP15* were downregulated by salicylic acid treatment, gibberellic acid induced expression

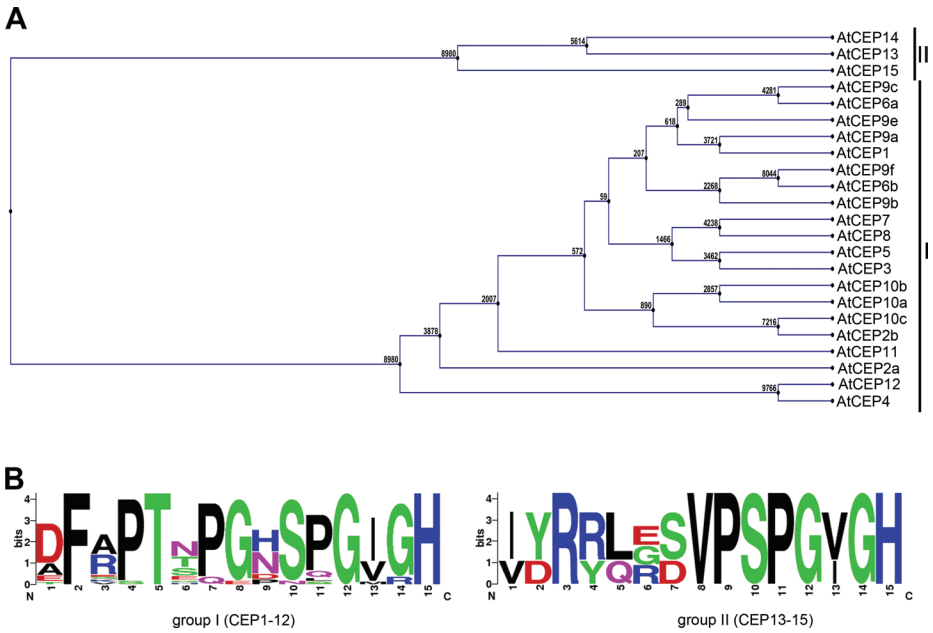

Fig. 2. The CEP family groups into two groups based on CEP domain. (A) Phylogenetic tree based on CEP domain sequences (see Fig. 1B, dark grey). Groups I and II are indicated. Bootstrap values are indicated on the tree. (B) Weblogo representation of group I and group II CEP domains.

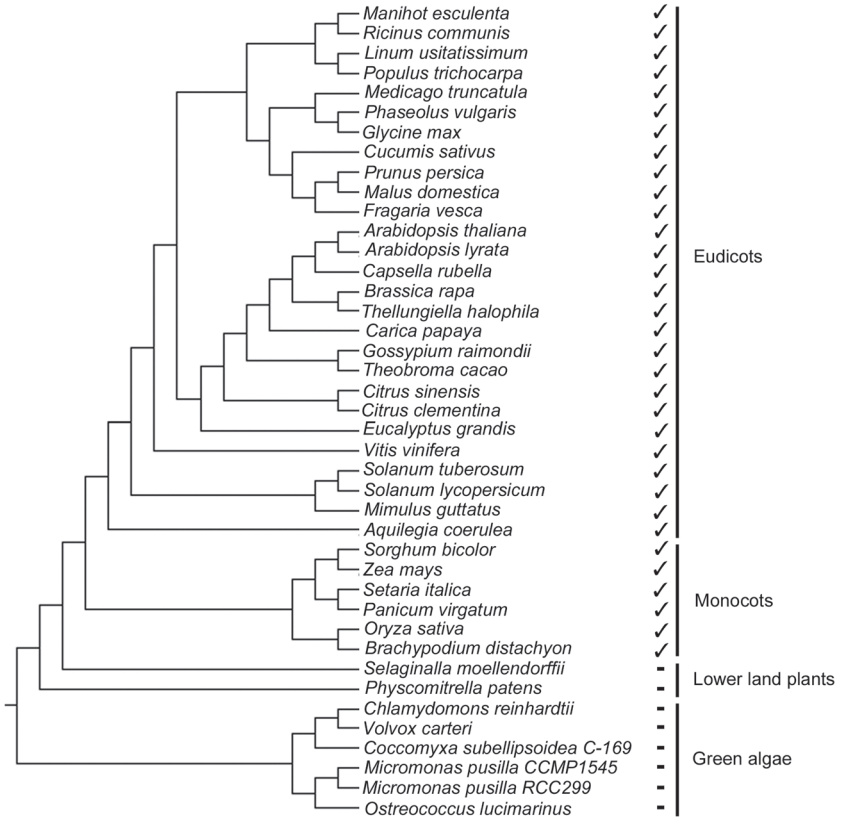

Fig. 3. Evolutionary analysis of the CEP family. Phylogenetic tree indicating the presence (√) or absence (–) of *CEP* family members in the indicated species.

of *CEP5* and *CEP13*, abscisic acid decreased expression of *CEP15*, whilst jasmonic acid increased expression of *CEP12*. Contrastingly, auxin (IAA) had opposite effects on the expression of some CEP family genes, namely an increase in expression of *CEP1* and *CEP3* but a decrease in *CEP5* and *CEP9* expression. Similarly, brassinolide reduced expression of *CEP5* and induced *CEP15* expression. Other hormones, such as ethylene, strigolactone, and cytokinin, did not have a significant effect on *CEP* expression. With respect to nutrients, a high nitrogen level downregulated *CEP3*, *CEP5*, and *CEP13*,

Table 2. *The effect of hormones and nutrients on* CEP *expression*

All data used were generated on AffymetrixGeneChip ATH1 22K platform. +, Increased expression following stimulus; –, reduced expression; =, no significant effect on expression (defined as $P \geq 0.05$ and/or a <1.5 fold change in expression levels compared with control); ND, *CEP* genes and/or stimuli for which no expression data are currently available in Genevestigator. ABA, abscisic acid; BR, brassinosteroids; GA, gibberellic acid; JA, jasmonic acid; SA, salicylic acid; Strigo, strigolactone; Cyto, cytokinin. Hormones and nutrients in *CEP2*, *CEP4*, *CEP6*, *CEP7*, *CEP8*, *CEP10*, *CEP11*, and *CEP14* are all ND.

Gene	Hormones									Nutrients		
	ABA	Ethylene[a]	BR[b]	GA	Auxin[c]	JA	SA	Strigo	Cyto[d]	N	P	K
CEP1	=	ND	ND	ND	+	ND	=	ND	ND	+	=	=
CEP3	=	=	=	=	+	=	=	ND	=	–	ND	–
CEP5	=	=	–	+	–	ND	–	ND	=	–	+	=
CEP9	=	=	=	=	–	=	=	=	=	+	=	+
CEP12	ND	ND	ND	ND	ND	+	ND	ND	ND	ND	ND	ND
CEP13	=	=	=	+	=	ND	=	ND	=	–	=	=
CEP15	–	=	+	=	=	=	–	=	=	=	=	=

[a–d]Measurements were taken using: [a]ACC (1-aminocyclopropane-1-carboxylic acid); [b]brassinolide; [c]indole-3-acetic acid, 1-naphthaleneacetic acid, and 2,4-dichlorophenoxyacetic acid; [d]zeatin.

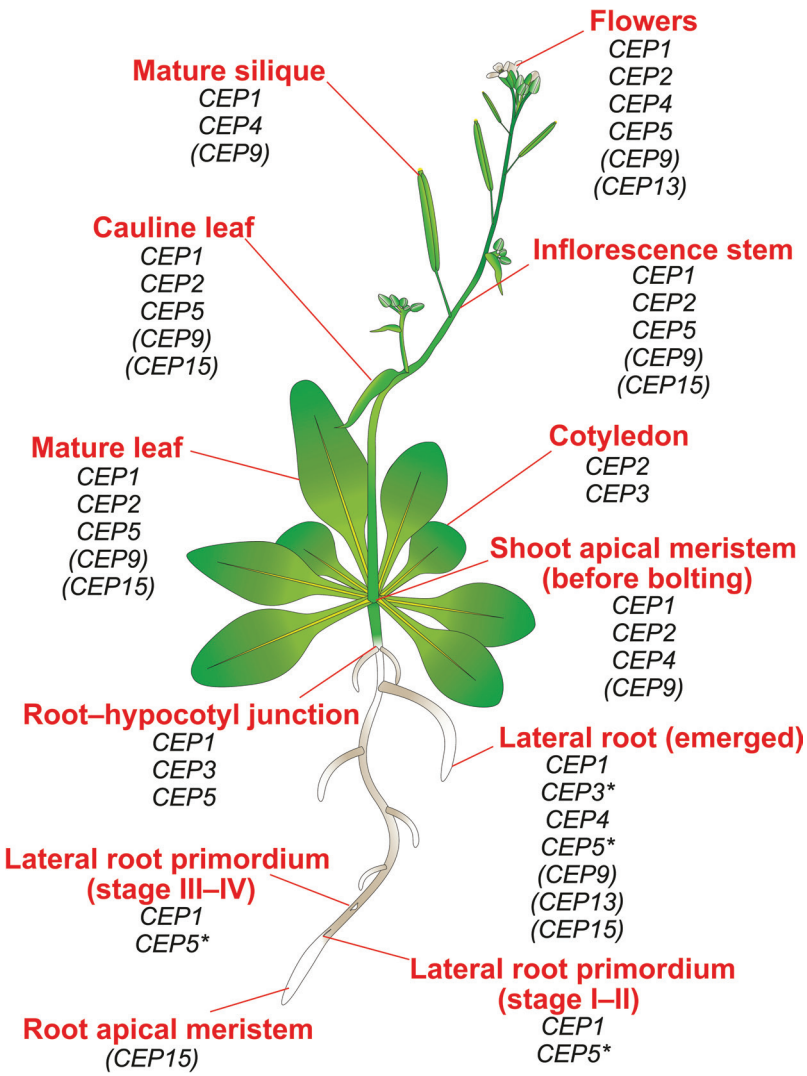

Fig. 4. Expression of *CEP* genes throughout the *Arabidopsis* plant, based on data from *in planta* (*GUS* expression; see Figs. 5–8) and *in silico* (eFP Browser; in parentheses) studies. *In silico* patterns were not included in the figure if there was a discrepancy with the *GUS* expression data. *, Associated with vasculature of primary root at these stages.

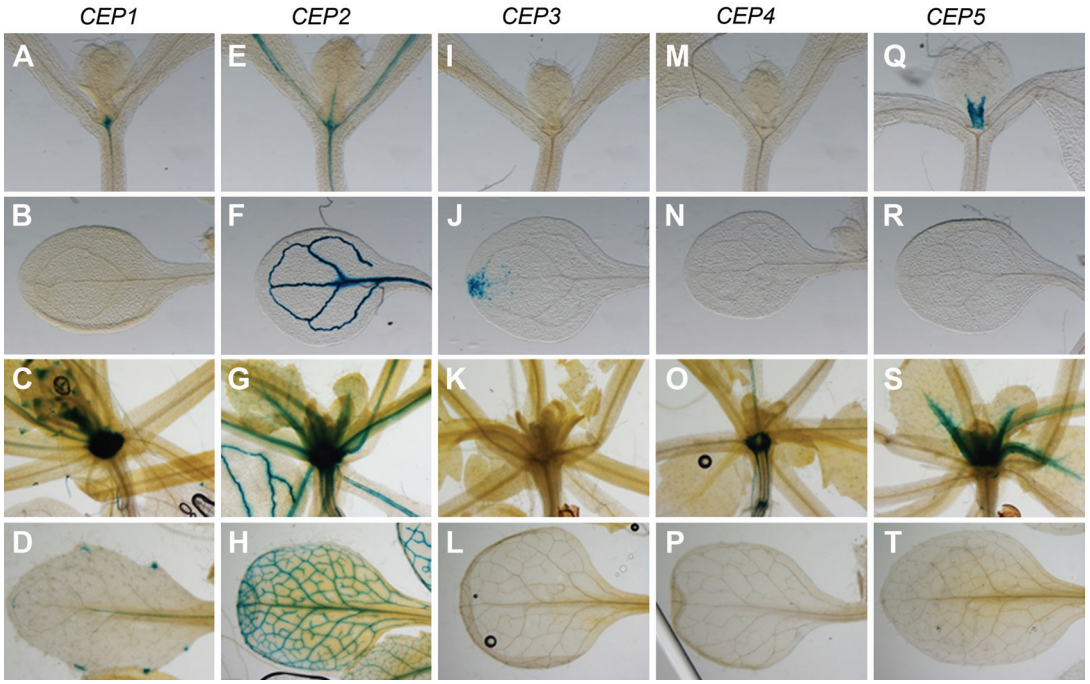

Fig. 5. *CEP* expression in the vegetative shoot. (A–D) *pCEP1::GUS* reporter line, (E–H) *pCEP2::GUS* reporter line, (I–L) *pCEP3::GUS* reporter line, (M–P) *pCEP4::GUS* reporter line, (Q–T) *pCEP5::GUS* reporter line. (A, E, I, M, Q) Shoot apical meristem region of 5-d-old seedling. (B, F, J, N, R) Cotyledon of 5-d-old seedling. (C, G, K, O, S) Shoot apical meristem region of 2-week-old seedling. (D, H, L, P, T) Leaf of 2-week-old seedling.

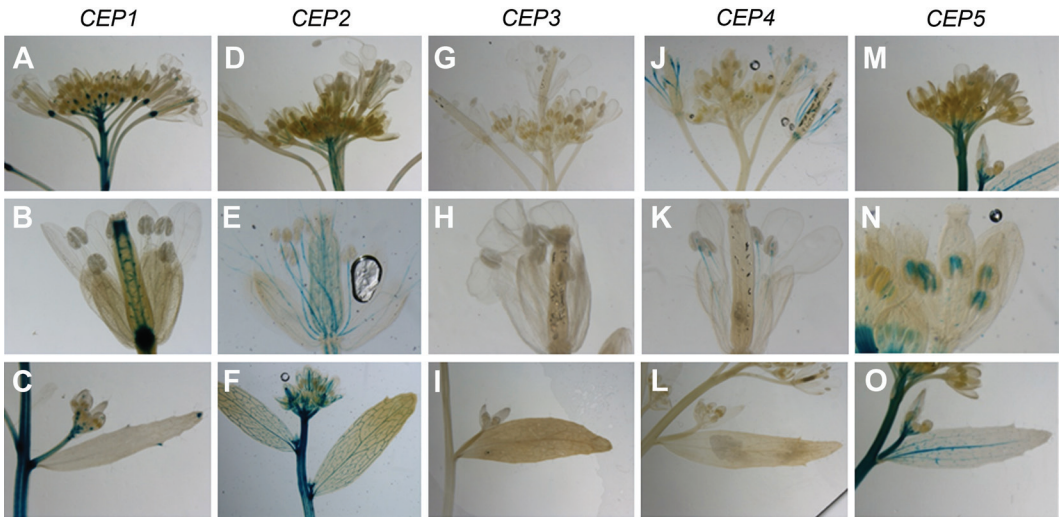

Fig. 6. *CEP* expression in the inflorescence. (A–C) *pCEP1::GUS* reporter line, (D–F) *pCEP2::GUS* reporter line, (G–I) *pCEP3::GUS* reporter line, (J–L) *pCEP4::GUS* reporter line, (M–O) *pCEP5::GUS* reporter line. (A, D, G, J, M) *CEP* expression in the apical part of inflorescence. (B, E, H, K, N) *CEP* expression during flower development. (C, F, I, L, O) *CEP* expression in a cauline leaf.

and upregulated *CEP1* and *CEP9*. Phosphorus upregulated expression of *CEP5*. In addition, potassium downregulated *CEP3* expression and upregulated *CEP9* expression.

CEP *family members display distinct expression patterns during* Arabidopsis *development*

To gain further insight in the expression patterns of the *CEP* family during development, the *CEP* expression data were first compiled and visualized from online repositories, namely eFP browser (Winter *et al.*, 2007) and Genevestigator v3 (Hruz *et al.*, 2008). These *in silico* expression patterns suggested that CEP peptides are expressed throughout the plant (Fig. 4; Supplementary Tables S1 and S2). *CEP1*, *CEP3*, and *CEP9* were expressed in the shoot apical meristem of the vegetative shoot, and *CEP3* was also expressed in the shoot apical meristem of the inflorescence stem. Only *CEP1* and *CEP15* were expressed in the primary root apical meristem, making these

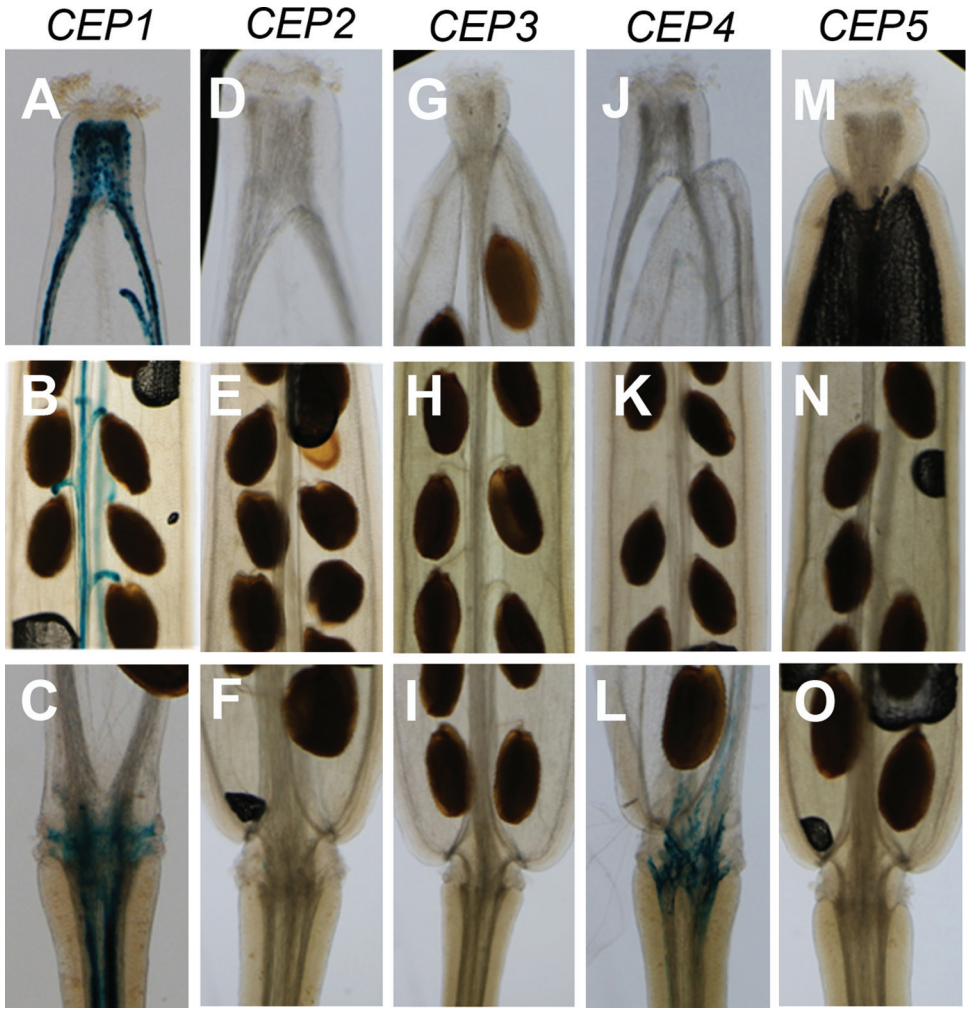

Fig. 7. *CEP* expression in a mature silique. (A–C) *pCEP1::GUS* reporter line, (D–F) *pCEP2::GUS* reporter line, (G–I) *pCEP3::GUS* reporter line, (J–L) *pCEP4::GUS* reporter line, (M–O) *pCEP5::GUS* reporter line. (A, D, G, J, M) *CEP* expression in the tip of a mature silique. (B, E, H, K, N) *CEP* expression in middle part of mature silique. (C, F, I, L, O) *CEP* expression at the base of a mature silique.

likely candidates for controlling root apical meristem maintenance. *CEP3*, *CEP9*, *CEP13*, and *CEP15* were expressed in cotyledons and/or leaves. During flower development, *CEP1*, *CEP3*, *CEP9*, and *CEP13* were expressed. During lateral root development, *CEP1*, *CEP3*, *CEP5*, *CEP9*, *CEP13*, and *CEP15* were expressed.

To further explore *CEP* expression patterns *in planta*, this study selected the five CEPs from group I that were also identified by Ohyama *et al.* (2008), generated *promoter::GUS* reporter lines, and characterized the reported lines throughout plant development (Figs. 4–8).

In the shoot apical meristem of the vegetative shoot of a 5-d-old seedling, only *CEP1* and *CEP2* were expressed (Fig. 5A and E), while in a 2-week-old plant, *CEP4* was also expressed (Fig. 5C, G, and O). In cotyledons and leaves, the expression domains of the CEP peptides were remarkably distinct and restricted to specific regions. In the cotyledon of 5-d-old seedlings, *CEP2* was expressed in the leaf veins and *CEP3* was expressed only in the tip of the cotyledons (Fig. 5F and J). In the leaves of 2-week-old plants, *CEP1* was expressed in the small dentations at the leaf margin and *CEP2* was expressed in the leaf veins (Fig. 5D and H). Both *CEP2* and *CEP5* are expressed in the leaf petioles (Fig. 5Q and S).

In the cauline leaves of the inflorescence, a similar expression pattern for *CEP1*, *CEP2*, and *CEP5* as in the mature vegetative leaves was observed (Fig. 6C, F, and O). In the shoot apical meristem of the inflorescence shoot, *CEP1*, *CEP2*, and *CEP5* were expressed (Fig. 6A, D, and M). During flower development, *CEP1* and *CEP2* were expressed in the gynoecium (Fig. 6B). In the androecium, both *CEP2* and *CEP4* were expressed in the filaments (Fig. 6E and K) and *CEP5* was expressed in the anthers (Fig. 6N).

In maturing siliques, *CEP1* and *CEP4* were expressed (Fig. 7A, B, C, and L), and both were expressed in the abscission zone (Fig. 7C and L). None of the five investigated CEP peptides showed a signal in the root apical meristem of the primary root (Supplementary Fig. S2). During adventitious root development at the root–hypocotyl junction, expression of *CEP1*, *CEP3*, *CEP4*, and *CEP5* was observed (Supplementary Fig. S3).

During lateral root development *CEP1*, *CEP3*, and *CEP5* were expressed at various stages (Fig. 8). For example, *CEP1* was expressed in the inner layer of the stage II primordium (Fig. 8B), and later in the central core of the developing

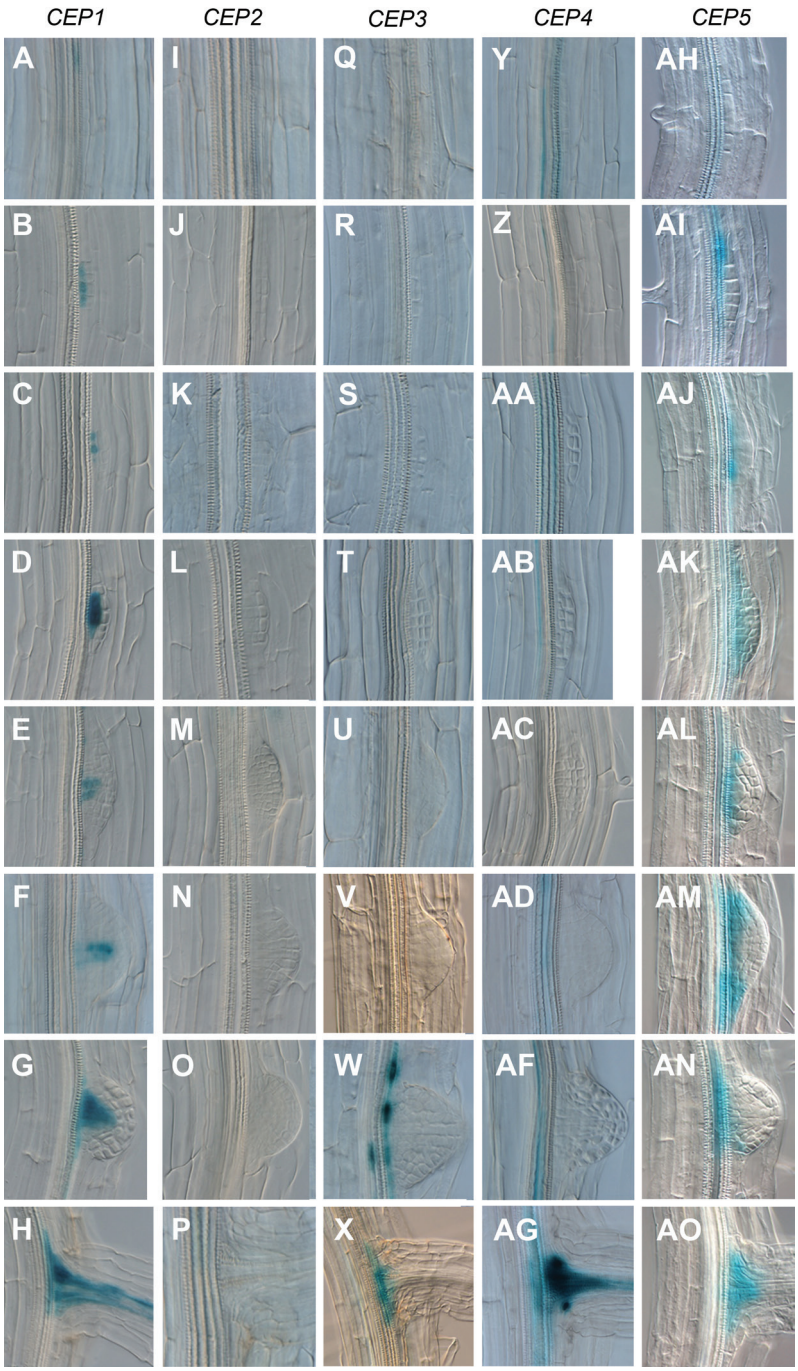

Fig. 8. *CEP* expression during lateral root development. (A–H) *pCEP1::GUS* reporter line, (I–P) *pCEP2::GUS* reporter line, (Q–X) *pCEP3::GUS* reporter line, (Y–AG) *pCEP4::GUS* reporter line, (AH–AO) *pCEP5::GUS* reporter line. Various stages of lateral root development, from initiation (asymmetric cell division) (A, I, Q, Y, AH) to an emerged lateral root (H, P, X, AG, AO), are shown.

lateral root primordium, coinciding with the (future) vasculature (Fig. 8E–H). Both *CEP3* and *CEP5* were expressed at the base of the lateral roots, where *CEP5* seemed to be expressed from an earlier time point in development compared to *CEP3* (Fig. 8W, X, and AH–AO). *CEP4* was also expressed at the base and in the vasculature of the emerged lateral root (Fig. 8AG).

The combination of *in planta* and *in silico* expression patterns showed that *CEP* genes are expressed throughout plant development (Fig. 4). This study observed some variation between *in planta* and *in silico* data (Supplementary Table S1), which could be due to experimental conditions (e.g. responsiveness of *CEPs* to external and/or environmental stimuli). It is also interesting to note that the *CEP* expression patterns were often associated with the vasculature during a specific developmental process. At present, there is no functional evidence for any evolutionary reason behind these vasculature-associated expression patterns. However, CEPs are absent in the more primitive groups of the green lineage, such as green algae, the non-vascular land plant *P. patens*, and *S. moellendorffii*, which is a representative of the earliest vascular plants that has a simple protostele (phloem surrounding the xylem).

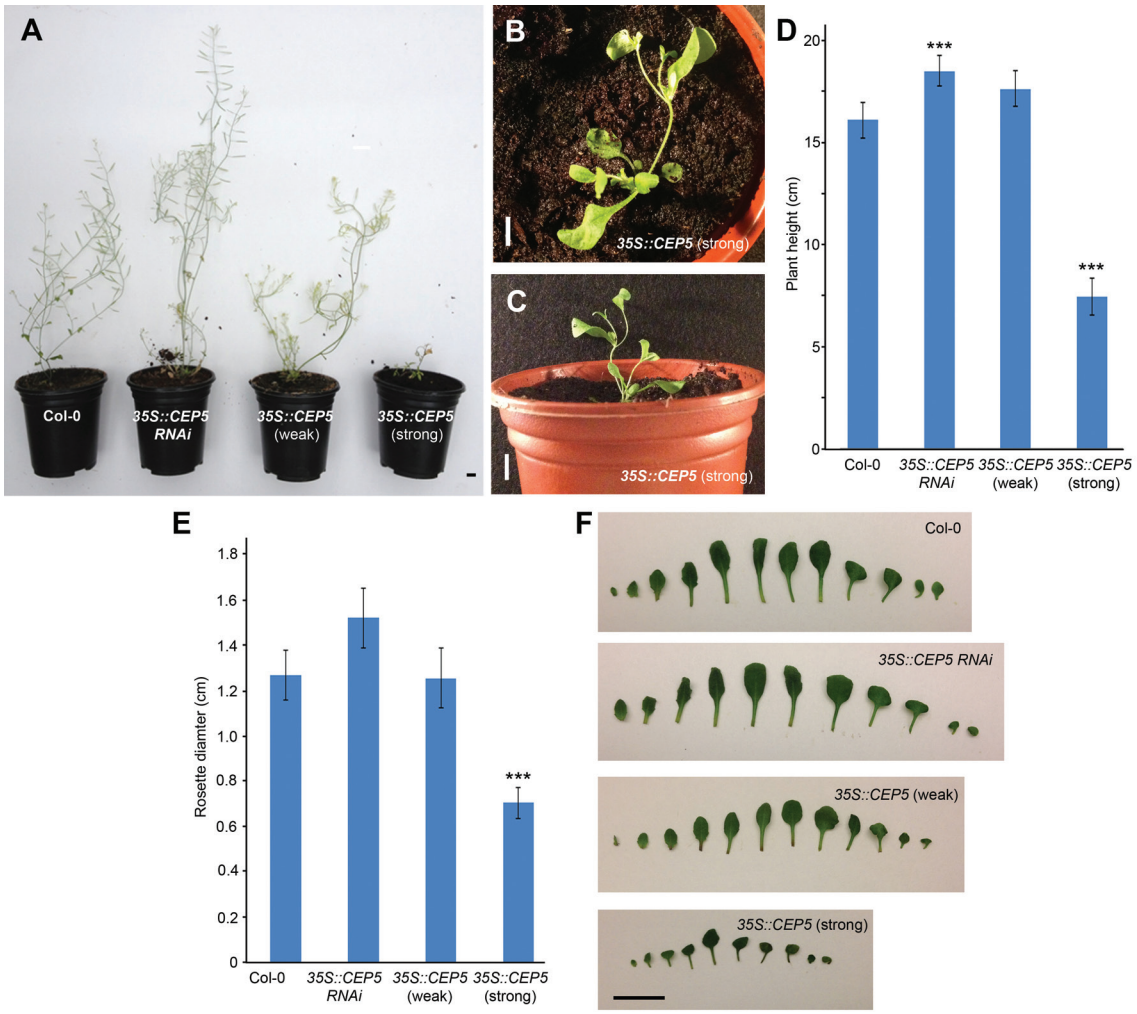

Fig. 9. Characterization of aboveground growth in lines with perturbed *CEP5* expression levels. (A) Representative shoot of Col-0 (background line), 35S::CEP5 RNAi knockdown line, and two overexpression lines [35S::CEP5 (weak) and 35S::CEP5 (strong)] after 3 weeks of growth post-bolting. (B and C) Severely stunted growth of 35S::CEP5 (strong) plants after 3 weeks of growth post-bolting. (D) Quantification of the plant height. (E) Rosette area 18 d after germination. (F) Representative leaf series of Col-0 (background line), 35S::CEP5 RNAi, 35S::CEP5 (weak), and 35S::CEP5 (strong). Data in D and E are means±standard errors of at least 20 plants. ***, Student's *t*-test with a *P*-value <0.05. Bars, 1 cm.

Therefore, one explanation could be that the appearance of CEPs coincides with the formation of more complex vascular tissues, such as actinosteles and eusteles in which the vascular tissue becomes more fragmented in separate bundles.

CEP5 is involved in aboveground growth

Since a role for CEP1 in root growth and development had been described previously (Ohyama *et al.*, 2008), this work focused on the effects of perturbing *CEP* levels on aboveground parts to determine if CEPs play a role in shoot development. Overexpression and knockdown lines were generated and analysed for *CEP5*, which is expressed in the shoot (Fig. 4). Multiple lines with a range of *CEP5* expression levels and displaying similar phenotypes were generated (data not shown), but this work selected representative knockdown and overexpression lines. The obtained phenotypes for knockdown or overexpression lines (see Fig. 9) were correlated with reduced or increased *CEP5* expression levels, respectively (Supplementary Fig. S4).

RNAi knockdown plants for *CEP5* after 3 weeks of growth post-bolting demonstrated a slight increase in plant height compared to Col-0 (Fig. 9A and D). In contrast, a strong overexpression line for *CEP5* was extremely stunted and displayed a loss of shoot gravitropic response (Fig. 9A–D). A weaker *CEP5* overexpression line did not display the same dramatic phenotype but appeared mildly defected in stem gravitropism (Fig. 9A). The rosette diameter was significantly reduced in the strong *CEP5* overexpression line, compared to the control (Fig. 9E). In addition, strong *CEP5* overexpression resulted in smaller, often curled, leaves (Fig. 9F). However, given the *CEP5* expression pattern, the leaf size phenotype might be due to non-specific effects and might reflect a role for another, highly similar, CEP peptide.

Conclusion

In conclusion, in *A. thaliana*, more than 1000 small signalling peptides have been predicted, but very few have been

functionally characterized (Lease and Walker, 2006; Butenko *et al.*, 2009; Matsubayashi, 2011; Murphy *et al.*, 2012; Czyzewicz *et al.*, 2013). One class of small post translationally modified signalling peptides is the CEP family (Ohyama *et al.*, 2008). One member of this family has already been shown to be involved in regulating root architecture (Ohyama *et al.*, 2008). Here, a bioinformatics approach was applied to identify more members of this family and to reveal that this family only emerged from higher land plants onward. The data further suggest that the CEP proteins form two subgroups according to their CEP domain. The specific *CEP* expression patterns offer a comprehensive framework to study the role of the CEP signalling peptides in plant development and hint to a possible role in cell communication mechanisms in the more complex vasculature of flowering plants. Expression patterns and perturbing levels of CEP family peptides pointed to a role in aboveground tissues, such as leaf and flower development. These results form the basis for further exploration of the mechanisms underlying this family of peptides and suggest that this family of small signalling peptides has a distinct role associated with developmental events associated with higher plants.

Supplementary material

Supplementary data are available at *JXB* online.

Supplementary Dataset S1. *Arabidopsis* CEP proteins.

Supplementary Dataset S2. CEP domains predicted by MEME.

Supplementary Fig. S1. Position of CEP family members on the five *Arabidopsis* chromosomes.

Supplementary Fig. S2. Expression patterns of *CEP1–CEP5* in the root apical meristem.

Supplementary Fig. S3. *CEP1–CEP5* expression in the root–hypocotyl junction.

Supplementary Fig. S4. *CEP5* expression levels in 35S::CEP5RNAi and 35S::CEP5 lines

Supplementary Table S1. Comparison of *in silico* (eFP browser) and *in planta* (*pCEPx::GUS*) *CEP* expression.

Supplementary Table S2. Genevestigator data on *CEP* expression.

Acknowledgements

The authors thank Michael Djordjevic and co-workers for useful discussions and Maria Njo for help with the figures. This work was supported by a BBSRC David Phillips Fellowship (BB_BB/H022457/1) and a Marie Curie European Reintegration Grant (PERG06-GA-2009–256354) to I.D.S. S.S. received a Biotechnology and Biological Science Research Council doctoral training grant studentship. I.R. was supported by the Agency for Innovation by Science and Technology. B.D.R. was funded by the Special Research Fund of Ghent University, a long-term Federation of European Biochemical Societies fellowship, and a Marie Curie long-term FP7 Intra-European Fellowship (IEF-2009–252503). This work was in part financed by grants of the Interuniversity Attraction Poles Programme (IAP VI/33 and IUAP P7/29 'MARS') from the Belgian Federal Science Policy Office.

References

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37,** W202–W208.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology* **2,** 28–36.

Butenko MA, Vie AK, Brembu T, Aalen RB, Bones AM. 2009. Plant peptides in signalling: looking for new partners. *Trends in Plant Science* **14,** 255–263.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16,** 735–743.

Czyzewicz N, Yue K, Beeckman T, De Smet I. 2013. Message in a bottle: small signalling peptide outputs during growth and development *Journal of Experimental Botany* **64,** 5281–5296.

Delay C, **Imin N**, **Djordjevic MA**. 2013. *CEP* genes regulate root and shoot development in response to environmental cues and are specific to seed plants. *Journal of Experimental Botany* **64,** 5383–5394.

Hilson P, Allemeersch J, Altmann T, *et al.* 2004. Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. *Genome Research* **14,** 2176–2189.

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P. 2008. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics* **2008,** 420747.

Lease KA, Walker JC. 2006. The *Arabidopsis* unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiology* **142,** 831–838.

Malamy JE, Benfey PN. 1997. Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. *Development* **124,** 33–44.

Matsubayashi Y. 2011. Small post-translationally modified Peptide signals in *Arabidopsis*. *The Arabidopsis Book* **9,** e0150.

Murphy E, Smith S, De Smet I. 2012. Small signaling peptides in *Arabidopsis* development: how cells communicate over a short distance. *The Plant Cell* **24,** 3198–3217.

Ohyama K, Ogawa M, Matsubayashi Y. 2008. Identification of a biologically active, small, secreted peptide in *Arabidopsis* by in silico gene screening, followed by LC-MS-based structure analysis. *The Plant Journal* **55,** 152–160.

Shinohara H, Matsubayashi Y. 2013. Chemical synthesis of *Arabidopsis* CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. *Plant Cell Physiology* **54,** 369–374.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2,** e718.