The soybean CEP6 signaling peptides positively regulates nodulation

Xiaoli Wang  
Shanxi Agricultural University

Jie Qin  
Shanxi Agricultural University

Wenqing Tian  
Shanxi Agricultural University

Bai miao  
Shanxi Agricultural University

Min Wang  
Shanxi Agricultural University

Weijun Du  
Shanxi Agricultural University

Lixiang Wang ( wanglixiang2@163.com )  
Shanxi Agricultural University

Research Article

Keywords: CEP peptide, Nitrogen inhibition, Soybean, Nodulation

Posted Date: April 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2794767/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Nodulation is the most efficient nitrate assimilation system in the ecosystem while excessive fertilization increased nitrate inhibition effect, deciphering the nitrate signal transduction mechanism in the process is of the utmost importance. In this study, genome-wide analysis of the GmCEP genes were applied to identify nodulation related CEP genes, 22 GmCEP family members were identified while GmCEP6 mainly expressed in nodule and significantly responded to nitrate treatment and rhizobium infection, especially in later stages. Overexpression and CRISPR-Cas9 were used to validate its role in nodulation. We found GmCEP6 overexpression significantly increased the nodule number while GmCEP6 knock out significantly decreased nodule number suggests GmCEP6 function as a positive regulator in soybean nodulation. qRT-PCR shown that alterations in the expression of GmCEP6 affected the expression of marker genes in the Nod factor signaling pathway. Lastly the function of GmCEP6 in nitrate inhibition of nodulation was analyzed, nodule number in the GmCEP6 overexpressed roots significantly increased under nitrogen treatments suggests GmCEP6 functions in resistance to nitrate inhibition. The study will help us understand that GmCEP6 promotes nodulation and participates in the regulation of nitrate inhibition of nodulation, which is of great significance for high efficiency utilization of nitrogen in soybean.

1. Introduction

The Nitrogen is one of the essential macroelements for plant growth, development, yield and quality formation [1]. Therefore, improving nitrogen utilization efficiency is an important guarantee for high and stable yield of soybean [2]. Legumes can not only absorb nitrogen nutrients from soil, but also provide nitrogen for themselves through symbiotic nitrogen fixation with rhizobia. Soybean, as an important symbiotic nitrogen fixation food crop, needs the rhizobia-soybean symbiotic system to fix 50% ~ 90% nitrogen nutrition required for its growth [3]. Previous studies shown that applying an appropriate amount of nitrogen fertilizer before sowing soybean can promote root nodule primordium formation and nodule organogenesis, improve the growth performance of rhizobia, promote plant growth, and provide effective carbon sinks and energy sinks for symbiotic nitrogen fixation [4]. Therefore, the symbiotic nitrogen fixation system between legume crops and rhizobia plays a very important role in nitrogen cycling.

Nitrogen uptake by plants from soil is mainly in the form of nitrate [5]. Nitrate, however, tends to be unevenly distributed in soils. Thus, plants have evolved a systematic long-distance signaling pathway (CEP-CEPR module) for the compensatory nitrate uptake in N starvation side of the root system [6]. CEP polypeptide is one of the largest peptide signal groups secreted by plants, biochemical and functional evidence suggests that 15 amino acid peptides derived from the C-terminal region of precursor peptides act as ligands to regulate various stages of plant growth and development. Although CEP peptides have long been known to play a role in local cell-to-cell communication within specific tissues, recent advances indicate their new role as long-distance mobile signals required for systemic nutritional responses [7–8].

The Arabidopsis genome contains 15 CEP genes [9], of which 7 are upregulated about 10 times in response to N starvation [10]. They are expressed specifically in the stele of lateral roots and are loaded
into xylem vessels for transportation to shoot [11]. The CEP family peptides were then recognized by receptor kinase CEP receptor 1 (CEPR1), which is expressed in leaf vascular tissue and induces the production of shoot secondary signals that up-regulate nitrate transport genes, such as NRT2.1, at the distal end of the root to compensate for local N-starvation [6]. Because CEP family peptides and CEPR1 are widely presented in seed plants, the CEP-CEPR signaling module appears to be evolutionarily conserved [6, 8]. In Medicago truncatula, MtCEP1 is the homologous of AtCEP9, but there are two CEP domains in MtCEP1, which are mainly expressed in the root tip, root vascular tissue and lateral root meristem, and was induced by different levels of nitrogen treatments[11]. The phenotype of overexpressed MtCEP1 root was different from that of overexpressed AtCEP1, which resulted in a reduction in the number of lateral roots [12]. However, the developmental role of CEP polypeptides in soybean is not clear.

Here, 22 CEP family members were identified in soybean through systematic bioinformatic study. We found the expression of GmCEP6 was higher in roots and nodules, and the histochemical staining was applied to validate this result. The effect of GmCEP6 on nodule development was evaluated and we found that GmCEP6 function as a positive regulator and partially tolerant to nitrate inhibition of nodulation. The results of this study may be utilized for high nitrogen fixation efficiency soybean breeding in future.

2. Results

2.1. Identification and bioinformatics analysis of GmCEP genes

22 soybean CEP gene family members were obtained from soybean genome by using BLAST and HMMER search, and were named GmCEP1-22 according to their chromosomal positions, the amino acids residues encoded by them ranged from 80 (GmCEP6) to 163 (GmCEP20), with molecular weights of the 22 GmCEPs ranged from 8,740.03 (GmCEP6) to 17,533.65 (GmCEP20) Da.. Theoretical isoelectric points of 22 GmCEPs family member ranged from 6.26 (GmCEP22) to 10.60 (GmCEP19) and belonged to alkaline proteins (Table 1). There were 6 members of GmCEPs family contained two CEP motifs while 16 contained one. Conserved domains analysis of GmCEP family members showed that Motif 1 and Motif 2 were presented in all CEP proteins (Table 1; Figure S1). We found there were no introns in all 22 GmCEP gene family members according to the gene structure analysis (Figure S2).
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene ID</th>
<th>Chromosomal location</th>
<th>Number of amino acids</th>
<th>Theoretical pI</th>
<th>Molecular weight (average)</th>
<th>CEP Motif number</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmCEP1</td>
<td>Glyma.01G184800</td>
<td>1</td>
<td>99</td>
<td>8.71</td>
<td>11292.71</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP2</td>
<td>Glyma.01G184900</td>
<td>1</td>
<td>87</td>
<td>9.82</td>
<td>9211.79</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP3</td>
<td>Glyma.01G185000</td>
<td>1</td>
<td>150</td>
<td>8.74</td>
<td>16377.5</td>
<td>2</td>
</tr>
<tr>
<td>GmCEP4</td>
<td>Glyma.01G185100</td>
<td>1</td>
<td>94</td>
<td>7.09</td>
<td>10099.56</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP5</td>
<td>Glyma.05G083900</td>
<td>5</td>
<td>86</td>
<td>7.8</td>
<td>9800.21</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP6</td>
<td>Glyma.05G084000</td>
<td>5</td>
<td>80</td>
<td>10.24</td>
<td>8740.03</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP7</td>
<td>Glyma.05G084100</td>
<td>5</td>
<td>156</td>
<td>9.3</td>
<td>16969.18</td>
<td>2</td>
</tr>
<tr>
<td>GmCEP8</td>
<td>Glyma.05G161100</td>
<td>5</td>
<td>96</td>
<td>10.21</td>
<td>10600.58</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP9</td>
<td>Glyma.08G118500</td>
<td>8</td>
<td>93</td>
<td>9.92</td>
<td>10181.12</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP10</td>
<td>Glyma.09G218000</td>
<td>9</td>
<td>85</td>
<td>9.1</td>
<td>9335.88</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP11</td>
<td>Glyma.11G057100</td>
<td>11</td>
<td>87</td>
<td>8.03</td>
<td>9270.58</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP12</td>
<td>Glyma.11G057200</td>
<td>11</td>
<td>148</td>
<td>9.34</td>
<td>15859.04</td>
<td>2</td>
</tr>
<tr>
<td>GmCEP13</td>
<td>Glyma.11G057300</td>
<td>11</td>
<td>87</td>
<td>9.83</td>
<td>9277.83</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP14</td>
<td>Glyma.13G226600</td>
<td>13</td>
<td>94</td>
<td>10.14</td>
<td>10213.95</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP15</td>
<td>Glyma.15G085800</td>
<td>15</td>
<td>88</td>
<td>9.78</td>
<td>9634.2</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP16</td>
<td>Glyma.16G108400</td>
<td>16</td>
<td>82</td>
<td>9.3</td>
<td>9010.27</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP17</td>
<td>Glyma.17G176500</td>
<td>17</td>
<td>86</td>
<td>9.14</td>
<td>9865.22</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP18</td>
<td>Glyma.17G176800</td>
<td>17</td>
<td>152</td>
<td>8.89</td>
<td>16346.37</td>
<td>2</td>
</tr>
<tr>
<td>GmCEP19</td>
<td>Glyma.17G176900</td>
<td>17</td>
<td>87</td>
<td>10.6</td>
<td>9658.01</td>
<td>1</td>
</tr>
<tr>
<td>GmCEP20</td>
<td>Glyma.17G177000</td>
<td>17</td>
<td>163</td>
<td>9.27</td>
<td>17533.65</td>
<td>2</td>
</tr>
<tr>
<td>GmCEP21</td>
<td>Glyma.17G177300</td>
<td>17</td>
<td>158</td>
<td>8.63</td>
<td>17110.26</td>
<td>2</td>
</tr>
<tr>
<td>GmCEP22</td>
<td>Glyma.17G244700</td>
<td>17</td>
<td>108</td>
<td>6.26</td>
<td>11742.32</td>
<td>1</td>
</tr>
</tbody>
</table>

2.2. Phylogenetic analysis and classification of soybean CEPs
In order to study the evolutionary relationships of soybeans with other plants, 22 soybean CEP amino acids sequences were aligned with 15 *Arabidopsis thaliana* (At) AtCEP protein, 11 *Medicago truncatula* (Mt) MtCEP protein and 15 *Oryza sativa* (Os) OsCEP sequences were collected for analysis. As shown in Fig. 1, the CEP family members of the four species were divided into 3 subfamily (I, II, and III). There were 24 gene family members in group I that contained 14 rice CEP family members. The left family members were from Arabidopsis (5), soybean (4), and 1 from *M. truncatula*. Except AtCEP16, the CEP family members in Group II were from soybean and *M. truncatula* suggests its may existed a dicots or even legume specific group CEP. Ten Arabidopsis and eleven soybean CEP family members were distributed in Group III.

### 2.3 Digital expression pattern of CEPs in soybean

A dynamic expression Heatmap was construct to dissect the soybean symbiotic related GmCEP genes, the online transcriptom data of soybean nine tissues covered the most organs and developmental stages were analyzed. The results shown that CEP genes had diverse expression patterns in different developmental stages of organs and tissues, for instance, most GmCEPs had similar expression level in various tissues while GmCEP22, GmCEP15 and GmCEP5 had extremely low expression in other tissues except the higher expression in one or two special tissues, which suggest that they might be a special role for their corresponding biological processes. In particularly, we found that GmCEP6 was highest expressed gene in root nodules among 22 soybean CEPs, indicating that might be involved in soybean and rhizobia interaction. Nitrogen including nitrate, ammonia and urea had significant effect for rhizobium infection and nodule initiation, the expression change of above three nitrogen treatment of root and leaf were compared to standard value. We found GmCEP6 was the only gene which were dramatically affected by three kinds of nitrogen treatment both in leaf and root.

### 2.4 GmCEP6 is preferentially expressed in soybean nodules

Bioinformatics analysis revealed that the full length transcript of GmCEP6 was 730 bp with a entire exon, the cDNA contained a 5' untranslated region (UTR) of 101 nucleotides and 3' UTR of 386 bp, the gene contained a 243 bp open reading frame (ORF) (Fig. 3A), encoding a predicted 80 amino acid residues protein with a conserved C-terminally encoded peptides (CEP) (66-80aa) (Fig. 3B), similar to its homologs in other plant species. A comparative analysis of GmCEP6 transcript levels was performed. Firstly, the relative expression level of GmCEP6 in nodules, root, stem, and leaves was determined using qRT-PCR, as shown in Fig. 3C, GmCEP6 was mainly expressed in roots and nodules, which indicating a possible role of GmCEP6 in nodulation (Fig. 3C). To further check the GmCEP6 expression in response to rhizobium infection, soybean seedlings was inoculated with *Bradyrhizobium japonicum* USDA110, and the transcript abundance of GmCEP6 at different stage was confirmed, as shown in Fig. 3d, GmCEP6 was weakly induced by USDA110 treatment and peaked at 16 DPI in infected root, while dramatically decreased at 28 DPI. In addition, GmCEP6 was markedly higher expressed in nodules than in roots at the checked time points (Fig. 3D). Finally, to visually determine the expression profile of GmCEP6 in soybean nodulation, transgenic hairy roots harboring the 2-kb promoter region upstream of GmCEP6 ATG was
fused to the β-glucuronidase (GUS) reporter (pCEP6:GUS). Histochemical GUS staining was performed in transgenic hairy roots inoculated with rhizobia at 10 DPI. We found that GmCEP6 was mainly expressed in pericycle, nodule primordium, lateral root primordium and root nodule (Fig. 3E-3I). In addition, GUS signaling was mainly detected in the infection zone of mature nodules (28 DPI). These results suggest that GmCEP6 may play a vital role in soybean nodulation and nitrogen fixation.

2.5 GmCEP6 plays a key role in soybean nodulation

To determine the roles of GmCEP6 in soybean nodulation, we generated transgenic hairy roots carrying GmCEP6 overexpressing (OE) or GmCEP6-CRISPR cas9 (KO). As shown in Fig. 4, both GmCEP6 overexpressing (OE) and GmCEP6-CRISPR cas9 (KO) significantly affected soybean nodulation. In GmCEP6 overexpressing roots, qRT-PCR was applied to check the overexpression efficiency, the result showed that the transcript of GmCEP6 was about 20-fold in the GmCEP6-OE roots than that in the control roots (Fig. 4D). Then the nodule numbers per root was quantified at 14 and 28 days after inoculation, we found in GmCEP6 overexpressing hairy roots, nodule number increased by 2.875 times (14 DPI) and 4 times(28 DPI), respectively. This data suggested that GmCEP6 plays a positive role in regulating soybean nodulation.

To further check this, the effect of GmCEP6 on soybean nodulation was evaluated in GmCEP6-CRISPR cas9 transgenic roots, gene editing and knock out efficiency were validated by sequencing, deletions and mutations can be detected in GmCEP6-KO roots (Figure S3). As shown in Fig. 4e-f, in GmCEP6-KO root lines, nodule number at 14 and 28 days were decreased by 17 times and 44 times, respectively, compared with the control. Combined with the overexpression results, it is suggested that GmCEP6 is critical for the regulation of soybean nodulation.

2.6 GmCEP6 regulate soybean nodulation through modulating NF pathway

As soybean nodule number was significantly affected by GmCEP6 overexpression and GmCEP6 knock out, we questioned whether GmCEP6 regulates soybean nodulation through the NF signaling pathway. To this end, we examined the expression pattern of several NF pathway marker genes in GmCEP6 overexpressed or knocked out soybean roots. We selected GmENOD40 (Early nodulin), GmNINa (Nodule Inception), GmNSP1 (Nodulation Signaling Pathway 1), NF-YA1 (GmHAP2-1), and NF-YA2 (GmHAP2-2) to verify this[13–16]. As shown in Fig. 5, the expression of GmNINA, GmENOD40, GmNSP1, GmHAP2-1, and GmHAP2-2 in GmCEP6-OE roots were significantly increased compared with that in empty vector control roots at 7 DPI. Meanwhile, we found the expression levels of these genes in GmCEP6-KO hairy roots were markedly reduced. These results suggested that GmCEP6 regulate soybean nodulation and nodule number controlling via modulating these symbiosis-related nod factor signaling pathway genes.

Nodule number is regulated by an autoregulatory mechanism and by the N status of the root, previous study in Medicago truncatula shown that MtCEP1 increased nodulation and promotes nodule
development at different nitrate concentrations. To investigate whether \textit{GmCEP6} overexpression increase nitrate tolerance of soybean in nodulation. Firstly, we found that the expression of \textit{GmCEP6} was significantly induced under high nitrogen (16mM) concentration than that under 4mM concentration (normal concentration). Then, nodule performance of \textit{GmCEP6-OE} and EV under different concentration nitrate was evaluated, as shown in Fig. 5d, we found nodule number in GmCEP6-overexpressed roots increased significantly compared with empty vector at different nitrate concentrations. These results indicated that soybean nodulation is enhanced by overexpression of \textit{GmCEP6} under different nitrogen concentrations, and this tolerance to nitrogen inhibition for nodulation engaged by GmCEP6 could have beneficial outcomes in soybean breeding.

3. Discussion

CEP peptides plays multiple roles in various plant biological process. The firstly identified C-Terminally encoded secreted Peptide \textit{AtCEP1} significantly arrests root growth [17]. The following reports proved the \textit{CEP} genes were respond to nitrogen deficiency and other stress [18]. The CEP peptides were percepted by shoot expressed LRR-RLK CEPR, to mediate a systematic regulating of nitrogen deficiency [6]. Moreover, the CEP peptides and cytokinin converge on CEPD glutaredoxins to inhibit root growth through a local system [19]. CEP peptides family number varied from different plants, there were 15 CEP peptides \textit{A. thaliana}, 11 in \textit{M. truncatula}, 15 in \textit{O. sativa}, 6 CEP in \textit{C. sativus} and 21 in \textit{P. sativum}, the function of several CEPs in above plants have been identified. However, little is known about the CEP peptide family in soybean. In this study, a systematic bioinformatics analysis were applied to identify soybean CEP peptides. A total of 22 GmCEPs were characterized from soybean genome, GmCEP proteins shown similar features to the previous discovered CEP family (Fig. 1; Table 1). By analyzing the expression patterns of \textit{GmCEPs} in the transcriptome, the diverse expression patterns in different developmental stages of organs and tissues of GmCEPs, implying multiple roles of GmCEPs in regulating different biological processes of soybean (Fig. 2).

Legumes can specifically interact with its compatible rhizobia in the surrounding soil to form nodules. However, nodulation and nitrogen fixation in mature nodule is an highly energy consumption process, thus, host legumes have evolved a root-shoot-root long distance auto-regulation of nodulation (AON) system to refine the number of nodules [20–21] NODULE INCEPTION (NIN) induced the expression of CLE ROOT SIGNAL1 (\textit{CLE-RS1}) and \textit{CLE-RS2} to activate AON [22]. Another phenomenon in legume nodulation is sensitive to soil nitrogen content, interestingly, recent studies have shown plant also existed a long-distance system (CEP-CEPR) in nitrogen assimilation signaling pathway [6]. In legume, The key transcriptional factor NIN coordinates CEP and CLE signaling peptides, combining this two long distance signaling pathways to balance nitrogen absorption and symbiotic nitrogen fixation, to meet highly nitrogen demands [23]. In this study, another symbiosis-related \textit{CEP} genes was characterized, we firstly identified a nodulation related GmCEP6 in regulating soybean nodulation (Figs. 3 and 4), shared the phenotype as reported legume \textit{CEP} overexpression, nodule number of its overexpression close to supernoduling phenotype of NARK mutant in soybean [12, 22]. Moreover, MtCEP1 promoted \textit{MtNRT2.1}
expression and nodulation dependent on compact root architecture 2 (MtCRA2) at low nitrate condition [24]. MtCEP1/MtCRA2 balances root and nodule development by reducing auxin and ethylene response [25]. Another study report MtCEP1, 2, and 12 redundantly regulate lateral root number and nodulation. The further study is needed to clarify the function diversification of GmCEPs family members, construct the relationship between cytokin and GmCEP, determine the relationship of AON shoot centre component and CEPR [26]. We also need to clarify the roles of carbon signals in balancing AON pathway and CEP-CEPR pathways that regulate nodule number.

4. Materials And Methods

4.1 Identification and bioinformatic analysis of GmCEP Genes

AtCEPs protein sequences were obtained from the Arabidopsis Information Resource database (https://www.arabidopsis.org/), Genome sequence, gff3 file and protein sequence of soybean (Glycine max) were downloaded from the Ensembl database (http://plants.ensembl.org/index.html). Blast wrapper tool in bioinformatic analysis software Tbtools was applied to retrieve the GmCEPs based on the AtCEP protein sequence. After removing the duplicates of the GmCEP sequences, the amino acid sequences of remaining GmCEPs were submitted into InterPro database (https://www.ebi.ac.uk/interpro/) for protein domain prediction. Conserved CEP (C-terminal Encoded peptide) domain containing GmCEPs were screened for further analysis. ProtParam (https://web.expasy.org/protparam/) database was used for the physical and chemical properties analysis. Online website MEME (http://meme-suite.org/tools/meme/) was used for characteristic analysis of GmCEPs motifs. Muscle program of MEGA-X was applied to construct CEPs phylogenetic tree, NJ (neighbor-joining) adjacency method was used to analyze the evolution distance.

4.2 Plant materials and growth conditions

Soybean (G. max L. cv. Williams 82) seeds (kindly provided by Professor Xia Li from Huazhong agriculture university and the seeds are for research only) were surface sterilized in 95% alcohol for 1 minutes and in 5% NaClO for 5 minutes, then washed several times using ddH2O water. The basic nutrient solution was refers to the previously publication [12], KNO3 was selected as nitrogen source to set different nitrate concentrations, with no nitrogen (0N, 0 mM), low nitrogen (LN, 4 mM) and high nitrogen (HN, 16 mM). Soybean were planted in vermiculite, and cultured in growth house with photoperiod cycle (light/dark: 14 h/10 h) at 25°C cultivation temperature. Under the light intensity 10000 lx and 70%. Agrobacterium rhizogenes strain K599 was used for the hairy root transformation. The hairy root transformation procedure was previously described with some modifications [14]. For nodulation assay, the plants inoculated with a suspension of B. japonicum strain USDA110 (30 mL, OD600 = 0.08).

4.3 Vector Construction
For the GmCEP6 promoter:GUS reporter fusion construct, 2369 bp upstream ATG of GmCEP6 region was selected and amplified from cv. Williams 82 genomic DNA, and cloned into pMDC162 though gateway system. The GmCEP6 full-length coding sequence was cloned into pB7RWG2.0 using the same strategy for the overexpression construction. For the CRISPR-Cas9 construction, the top 2 reliable sgRNAs, CATGAACTACTCGGTAGTGAGGG and CCGTAGCATTAGAAGCCTAGGG were selected, Then, vector pCBC-DT1T2 was used as a template to clone the two CRISPR fragments and the two obtained products were inserted into vector pKSE401-GFP.

4.4 RNA Extraction and Expression Analysis

RNAsup Pure Plant plus Trizol Kit was used to extracting RNA from collected transgenic hairy roots, soybean leaves, roots and nodules, and the first-strand cDNA was synthesized by using super Mix Kit (Hifair II 1 strand cDNA Synthesis SuperMix, gDNA digester plus) (Yeasen Biotech co. Ltd, Shanghai, China). qPCR was performed using SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich). GmCYP2 was used as an internal control (Jian et al., 2008). (The primers used in this study are shown in Table S1)

4.5 Histochemical analysis of GmCEP6 expression

Composite transgenic roots expressing GmCEP6pro:GUS were inoculated with B. japonicum strain USDA110. The transformed hairy roots at different infection and nodulation stage were stained with X-Gluc at 37°C for 8 h to test for β-glucuronidase activity. GUS activity was observed with a light microscope (OLYMPUS U-TV0.5XC-3).

4.6 Statistical Analysis

One-way analysis of variance (ANOVA) and Student's t-test was used to performed P values. The gene expression and nodule numbers were analyzed using IBM SPSS 22.0 and GraphPad Prism 5 (GraphPad Software). The statistical differences are marked as follows: *, P< 0.05; **, P< 0.01; ***, P< 0.001.

Declarations

Ethics approval and consent to participate

All experimental research methods in this study were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable

Data Availability Statement

The RNA-seq data used in this study were download from Phytozome database (https://phytozome-next.jgi.doe.gov/), Sequence data from this article can be found in the GenBank/EMBL or Glycine max Wm82.a4.v1 database with the following entry number was shown in Table 1. AtCEPs protein sequences
were obtained from the Arabidopsis Information Resource database (https://www.arabidopsis.org/), Genome sequence, gff3 file and protein sequence of soybean (Glycine max) were downloaded from the Ensembl database (http://plants.ensembl.org/index.html). Muscle program of MEGA-X was applied to construct CEPs phylogenetic tree, NJ (neighbor-joining) adjacency method was used to analyze the evolution distance. The data that support the findings of this study are available from the corresponding author, L. W., upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Funding:** This research was funded by The Central Guidance for Local Science and Technology Development Projects (grant number YDZJSX2022A035), Natural Science Foundation of Henan Province (Grant No. 202300410056), Basic research program of Shanxi Province (20210302123365, 202103021224146), National Natural Science Foundation of China (Grant No. 32241046), National Key Research and Development Program (Grant No. 2021YFD1600605-10). National Laboratory of Minor Crops Germplasm Innovation and Molecular Breeding (in preparation) (Grant No. 202105D121010-23, 202204010910001-33).

**Author Contributions:** L.W. and W.D. conceived the project; X.W., J.Q. W.T. and M.B. performed the most experiments; X.W. and L.W. prepared the original draft; L.W. and W.D. reviewed and finalized the manuscript, L.W. and W.D. supervised the project. All authors have read and agreed to the published version of the manuscript.

**Acknowledgments:** We thank professor Xia Li and Zhaosheng Kong for technical supports and valuable suggestions.

**References**


Figures
Figure 1

Phylogenetic analysis of CEPs in G. max (Gm), O. sativa(Os), A. thaliana (At) and M. truncatula(Mt). The amino acid sequences of GmCEPs, OsCEPs, AtCEPs and MtCEPs were downloaded and submitted into MEGA-X software for alignment and phylogenetic tree construction, phylogenetic tree was constructed using NJ (neighbor-joining) adjacency method with 1000 bootstrap replicates.
Figure 2

The GmCEP Expression Profiles in Different Organs with or without Treatments. The transcriptom data of GmCEP genes in nine tissues (root, stem, bud, leaf, flower, nodule, root tip, pod and seed) and three different treatment development stages (ammonia treatment, nitrogen treatment and urea treatment in root and leaf) were obtained from phytozome database. The Heatmap package in Tbtools software was applied to show the expressed FPKM value in different tissues.
Expression Pattern of GmCEP6 in Soybean Nodulation. (A) Schematic gene structure of GmCEP6. (B) Domain analysis of GmCEP6. The conserved C-terminally encoded peptide (CEP) (66-80 aa) was shown in red color. (C) Relative expression of GmCEP6 in soybean root, nodule, stem and leaf at 28 DPI. (D) Relative expression of GmCEP6 in inoculated soybean roots (0, 0.5 HAI and 1, 4, 16, 28 DPI) and 16, 28 DPI nodules. Gene expression level was normalized based on the expression of housekeeping gene GmCYP2. Error bar represents the mean of four biological replicates with ± SE, the different letters indicate significant differences, $P < 0.05$. Asterisks in (D) indicate significant difference within a $P$ level in t-tests. **, $P < 0.01$; ***, $P < 0.001$. (E-I) Histochemical analysis of GmCEP6 expression in transgenic composite soybean roots and nodules, root tip region (E), and emerged pericycle (F), lateral root primordium (G), nodule primordium (H) and (I) nodule. Scale bar in (E-I) = 1 cm.
Figure 4

GmCEP6 Regulate Soybean Symbiotic Nodulation. (A) Soybean growth status of composite plants harboring Empty vector, GmCEP6-OE and GmCEP6-KO. (B) Nodulation performance of transgenic hairy root harboring Empty vector, GmCEP6-OE and GmCEP6-KO at 28 DPI. (C) Transgenic validation of hairy root harboring Empty vector, GmCEP6-OE and GmCEP6-KO using LUYOR-3415RG Hand-Held Lamp, GFP positive roots were selected for further phenotype analysis. Scale bar in (B, C) =1 cm. (D) Relative expression level of GmCEP6in empty vector and GmCEP6-OE transgenic roots at 28 DPI were conducted to check the overexpression ratio, the expression value was normalized based on the expression of reference gene GmCYP2. (E-F) Quantitative data of nodule number per hairy root at 14 DPI and 28 DPI. EV: empty vector for overexpressed transgenic roots. pKSE401 was the empty vector for CRISPR cas9 (KO). These experiments were conducted more than three dependent biological replicates. Data are means with SE from three independent replicates (n=12), Asterisks indicate significant difference within a P level in t-tests, ***, p< 0.001.
Figure 5

GmCEP6 Regulates Nod Factor Signaling Pathway Genes and Tolerant to Nitrate Inhibition. (A-B) GmCEP6 Regulates Soybean Symbiotic Nodulation through Nod Factor Signaling Pathway Genes. (A) qRT-PCR analysis of the expression of GmNINa, GmENOD40, GmNSP1, GmHAP2-1 and GmHAP2-2 in roots carrying EV and GmCEP6-OE at 2 DAI (n =10). (B) qRT-PCR analysis of GmNINa, GmENOD40, GmNSP1 and GmHAP2-2 in roots harboring pKSE401 and GmCEP6-KO at 28 DAI (n =10). We set all of the transcript profiles of these genes in EV hairy roots at 28 DAI as “1”. The transcript amounts in each sample were normalized to the expression of reference gene GmCYP2. The expression levels are means ±SD. Asterisks indicate significant difference within a P level in t-tests. **, p < 0.01; ***, p < 0.001. (C) Gus staining shown GmCEP6 was induced by high nitrogen treatment, Left: 4 mM nitrate treatment, Right: 16 mM nitrate treatment, Bar=1 cm. (D) Quantication of nodule number at 28 DAI under different nitrate concentrations. Asterisks indicate significant difference within a P level in t-tests. **, p < 0.01; ***, p < 0.001.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarydata.pdf