agronomy

MDPI

Article

Soybean CEP6 Signaling Peptides Positively Regulate Nodulation

Shuai Wu †, Xiaoli Wang †, Jie Qin †, Wenqing Tian, Min Wang, Aiqin Yue, Lixiang Wang, Weijun Du * and Jinzhong Zhao

Houji Labortary in Shanxi Province, College of Agriculture, Shanxi Agricultural University, Jinzhong 030801, China; s20212185@stu.sxau.edu.cn (S.W.); z20213090@stu.sxau.edu.cn (X.W.); 20232043@mails.jlau.edu.cn (J.Q.); s20212119@stu.sxau.edu.cn (W.T.); wangmin@sxau.edu.cn (M.W.); yueaiqin@sxau.edu.cn (A.Y.); lxwang@sxau.edu.cn (L.W.); zhaojz@sxau.edu.cn (J.Z.)

* Correspondence: duweijun@sxau.edu.cn

† These authors contributed equally to this work.

Abstract: Nodulation is the most efficient nitrate assimilation system in the ecosystem, while excessive fertilization has an increased nitrate inhibition effect; deciphering the nitrate signal transduction mechanism in the process is of the utmost importance. In this study, genome-wide analyses of the *GmCEP* genes were applied to identify nodulation-related *CEP* genes; 22 *GmCEP* family members were identified, while *GmCEP6* was mainly expressed in nodules 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 that *GmCEP6* overexpression significantly increased the nodule number, while *GmCEP6* knock-out significantly decreased the nodule number, which suggests that *GmCEP6* functions as a positive regulator in soybean nodulation. qRT-PCR showed 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 numbers in the *GmCEP6*-overexpressed roots significantly increased under nitrogen treatments, which suggests that *GmCEP6* functions in the resistance to nitrate inhibition. The study helps 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 soybeans.

Keywords: CEP peptide; nitrogen inhibition; soybean; nodulation

Citation: Wu, S.; Wang, X.; Qin, J.; Tian, W.; Wang, M.; Yue, A.; Wang, L.; Du, W.; Zhao, J. Soybean CEP6 Signaling Peptides Positively Regulate Nodulation. *Agronomy* **2024**, *14*, 988. https://doi.org/10.3390/agronomy14050988

Academic Editor: Bertrand Hirel

Received: 5 March 2024
Revised: 11 April 2024
Accepted: 4 May 2024
Published: 8 May 2024

1. Introduction

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 soybean yields [2]. Legumes can not only absorb nitrogen from nitrogen-containing compounds such as ammonium and nitrate contained in the soil but also provide nitrogen by reducing free nitrogen to ammonia through symbiotic nitrogen fixation with rhizobia. The soybean, as an important symbiotic nitrogen fixation food crop, needs the rhizobia-soybean symbiotic system to fix the 50~90% nitrogen nutrition required for its growth [3]. Previous studies have 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 the 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 compensatory nitrate uptake in

a N-starvation side of the root system [6]. The 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 seven 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 the shoots [11]. The CEP family peptides are 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 are induced by different levels of nitrogen treatments [11]. However, the developmental role of CEP polypeptides in soybeans is not clear.

Here, 22 CEP family members were identified in soybeans through systematic bioinformatic study. We found that the expression of *GmCEP6* was higher in roots and nodules, and histochemical staining was applied to validate this result. The effect of *GmCEP6* on nodule development was evaluated, and we found that *GmCEP6* functioned as a positive regulator and was 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. Materials and Methods

2.1. Identification and Bioinformatic Analysis of GmCEP Genes

AtCEPs (*Arabidopsis thaliana* CEP genes family, AtCEPs) protein sequences were obtained from the *Arabidopsis* Information Resource database (https://www.arabidopsis.org/, (accessed on 2 January 2022)) [12]. The genome sequence, gff3 file, and protein sequence of soybean (*Glycine max*) were downloaded from the Ensembl database (http://plants.ensembl.org/index.html, (accessed on 4 January 2022)) [13]. The Blast wrapper tool in the bioinformatic analysis software TBtools v2.056 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 to the InterPro database (https://www.ebi.ac.uk/interpro/, (accessed on 14 January 2022)) [14] for protein domain prediction. Conserved CEP (C-terminal Encoded peptide) domains containing GmCEPs were screened for further analysis. The ProtParam (https://web.expasy.org/protparam/, (accessed on 11 February 2022)) [15] database was used for the physical and chemical properties analysis, including the number of amino acids, theoretical isoelectric point (pI), and relative molecular mass. The online website MEME (https://meme-suite.org/meme/, (accessed on 17 February 2022)) [16] was used for characteristic analysis of the GmCEPs motifs. The Muscle program of MEGA-X was applied to construct the CEPs' phylogenetic tree; the NJ (neighbor-joining) [17] adjacency method was used to analyze the evolution distance.

2.2. Plant Materials and Growth Conditions

Soybean (*G. max* L. cv. Williams 82) seeds (kindly provided by Professor Xia Li from Huazhong agriculture university for research only) were surface-sterilized in 95% alcohol for 1 min and in 5% NaClO for 5 min, then washed several times using ddH_2O water. The basic nutrient solution was referred to in a previous publication [18]. KNO_3 was selected as a nitrogen source to set different nitrate concentrations, with no nitrogen (0 N, 0 mM), low nitrogen (LN, 4 mM), and high nitrogen (HN, 16 mM). Soyabeans were planted in 12 cm

square boxes with a common vermiculite substrate. One soyabean plant was planted in each box and cultured in a growth house with photoperiod cycle (light/dark: 14 h/10 h) at 25 °C cultivation temperature, under the light intensity 10,000 lx and 70%. On the 10th day after seeding, the seedlings developed into the cotyledon stage and were inoculated with rhizobium bacteria, and the roots were harvested on the 7th day after seeding. After inoculation with rhizobia, HAI was the duration of different short-term treatments within 24 h, and DPI was the number of days of treatment. *Agrobacterium rhizogenes* strain K599 was used for the hairy-root transformation. The hairy-root transformation procedure was as previously described, with some modifications [14]. Soybeans sown with common vermiculite for 3 days were selected for hairy-root transformation and were inoculated with rhizobia on the 7th day after planting, and the roots were harvested on the 14th and 28th day after treatment, respectively. During the whole plant-growth period, nitrogen-free nutrient solution and water were required to be irrigated in rotation. The 14th and 28th days after inoculating the soybean seedlings with rhizobium were two important early nodulation development periods. For the nodulation assay, the plants were inoculated with a suspension of *B. japonicum* strain USDA110 (30 mL, OD600 = 0.08).

2.3. Vector Construction

For the *GmCEP6* promoter, a 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 a 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 two reliable sgRNAs (small guide RNA, sgRNA), CATGAACTACTCGGTAGTGAGGG and CCGTAGCATTAGAAGCCT AGGG 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.

2.4. RNA Extraction and Expression Analysis

RNAprep Pure Plant plus Trizol Kit was used to extract RNA from collected transgenic hairy roots, soybean leaves, roots, and nodules, and the first-strand cDNA was synthesized using a 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 Jump-Start Taq ReadyMix (Sigma-Aldrich, St. Louis, MO, USA). *GmCYP2* was used as an internal control [19]. (The primers used in this study are shown in Table S1).

2.5. Histochemical Analysis of GmCEP6 Expression

Composite transgenic roots expressing *GmCEP6*pro:GUS were inoculated with *B. japonicum* strain USDA110. The transformed hairy roots at different infection and nodulation stages 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).

2.6. Statistical Analysis

One-way analysis of variance (ANOVA) and Student's *t*-test were 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 data are means with ±SE (Standard Error) from three independent replicates. The statistical differences are marked as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

3. Results

3.1. Identification and Physicochemical Properties of the Soybean CEP Family Gene

22 soybean CEP gene family members were obtained from the 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), and the molecular weights of the 22 GmCEPs ranged from 8740.03 (GmCEP6)

to 17,533.65 (GmCEP20) Da. Theoretical isoelectric points of 22 GmCEPs family members ranged from 6.26 (GmCEP22) to 10.60 (GmCEP19) and belonged to alkaline proteins (Table 1). Six members of the *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 present in all CEP proteins. Motif 5 is the second motif in the GmCEP family, with 16 CEP proteins containing this motif. The least contained motif is Motif 9, with only two CEP proteins containing this motif (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 1. Characteristics of GmCEP family members.

Gene Name	Gene ID	Chromosomal Location	Number of Amino Acids	Theoretical pI	Molecular Weight (Average)	CEP Motif Number
GmCEP1	Glyma.01G184800	1	99	8.71	11,292.71	1
GmCEP2	Glyma.01G184900	1	87	9.82	9211.79	1
GmCEP3	Glyma.01G185000	1	150	8.74	16,377.5	2
GmCEP4	Glyma.01G185100	1	94	7.09	10,099.56	1
GmCEP5	Glyma.05G083900	5	86	7.8	9800.21	1
GmCEP6	Glyma.05G084000	5	80	10.24	8740.03	1
GmCEP7	Glyma.05G084100	5	156	9.3	16,969.18	2
GmCEP8	Glyma.05G161100	5	96	10.21	10,600.58	1
GmCEP9	Glyma.08G118500	8	93	9.92	10,181.12	1
GmCEP10	Glyma.09G218000	9	85	9.1	9335.88	1
GmCEP11	Glyma.11G057100	11	87	8.03	9270.58	1
GmCEP12	Glyma.11G057200	11	148	9.34	15,859.04	2
GmCEP13	Glyma.11G057300	11	87	9.83	9277.83	1
GmCEP14	Glyma.13G226600	13	94	10.14	10,213.95	1
GmCEP15	Glyma.15G085800	15	88	9.78	9634.2	1
GmCEP16	Glyma.16G108400	16	82	9.3	9010.27	1
GmCEP17	Glyma.17G176500	17	86	9.14	9865.22	1
GmCEP18	Glyma.17G176800	17	152	8.89	16,346.37	2
GmCEP19	Glyma.17G176900	17	87	10.6	9658.01	1
GmCEP20	Glyma.17G177000	17	163	9.27	17,533.65	2
GmCEP21	Glyma.17G177300	17	158	8.63	17,110.26	2
GmCEP22	Glyma.17G244700	17	108	6.26	11,742.32	1

3.2. Phylogenetic Analysis of CEPs in Soybeans

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 proteins, 11 *Medicago truncatula* (Mt) MtCEP proteins, and 15 *Oryza sativa* (Os) OsCEP sequences, collected for analysis. As shown in Figure 1, the CEP family members of the four species were divided into three subfamilies (I, II, and III). There were 24 gene family members in group I that contained 14 rice CEP family members. The remaining family members were from *Arabidopsis* (five), soybean (four), and one from *M. truncatula*. Except AtCEP16, the CEP family members in Group II were from soybean and *M. truncatula*, which suggests it may be a dicots- or even legume-specific group CEP. Ten *Arabidopsis* and eleven soybean CEP family members were distributed in Group III.

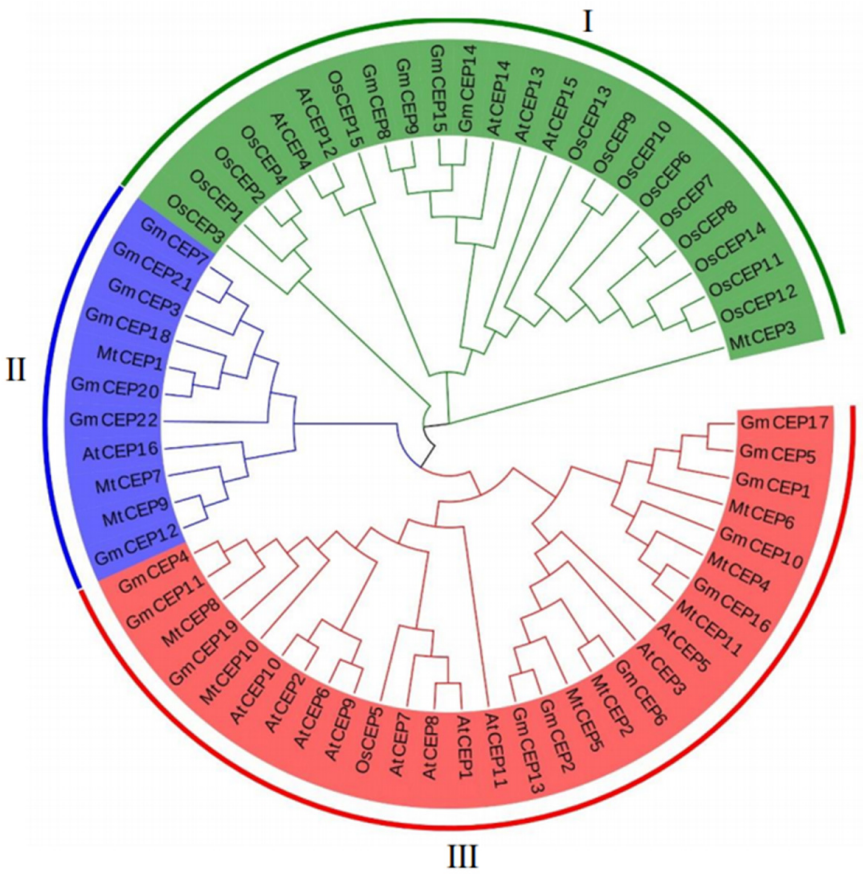

Figure 1. Phylogenetic analysis of CEPs in *G. max* (Gm), *O. sativa* (Os), *A. thaliana* (At), and *M. truncatula* (Mt). The amino acid sequences of GmCEPs, OsCEPs, AtCEPs, and MtCEPs were downloaded and submitted into MEGA-X v10.0.0 software for alignment and phylogenetic tree construction; the phylogenetic tree was constructed using the NJ (neighbor-joining) adjacency method with 1000 bootstrap replicates.

3.3. Digital Expression Pattern of CEPs in Soybeans

A dynamic expression Heatmap was constructed to dissect the soybean symbiotic related *GmCEP* genes; the online transcriptome data of nine soybean tissues covered most organs, and the developmental stages were analyzed. As shown in Figure 2, the results show that *CEP* genes had diverse expression patterns in different developmental stages of organs and tissues; for instance, most *GmCEPs* had similar expression levels in various tissues, while *GmCEP22*, *GmCEP15*, and *GmCEP5* had extremely low expression in other tissues, except for a higher expression in one or two special tissues, which suggests that they might play a special role in their corresponding biological processes. In particular, we found that *GmCEP6* was the highest expressed gene in root nodules among the 22 soybean *CEPs*, indicating that it might be involved in soybean and rhizobia interaction. Nitrogen, including nitrate, ammonia, and urea, had a significant effect on rhizobium infection and nodule initiation; the expression changes of the above three nitrogen treatments of roots and leaves were compared to standard values. We found that *GmCEP6* was the only gene which was dramatically affected by the three kinds of nitrogen treatment of both the leaf and root.

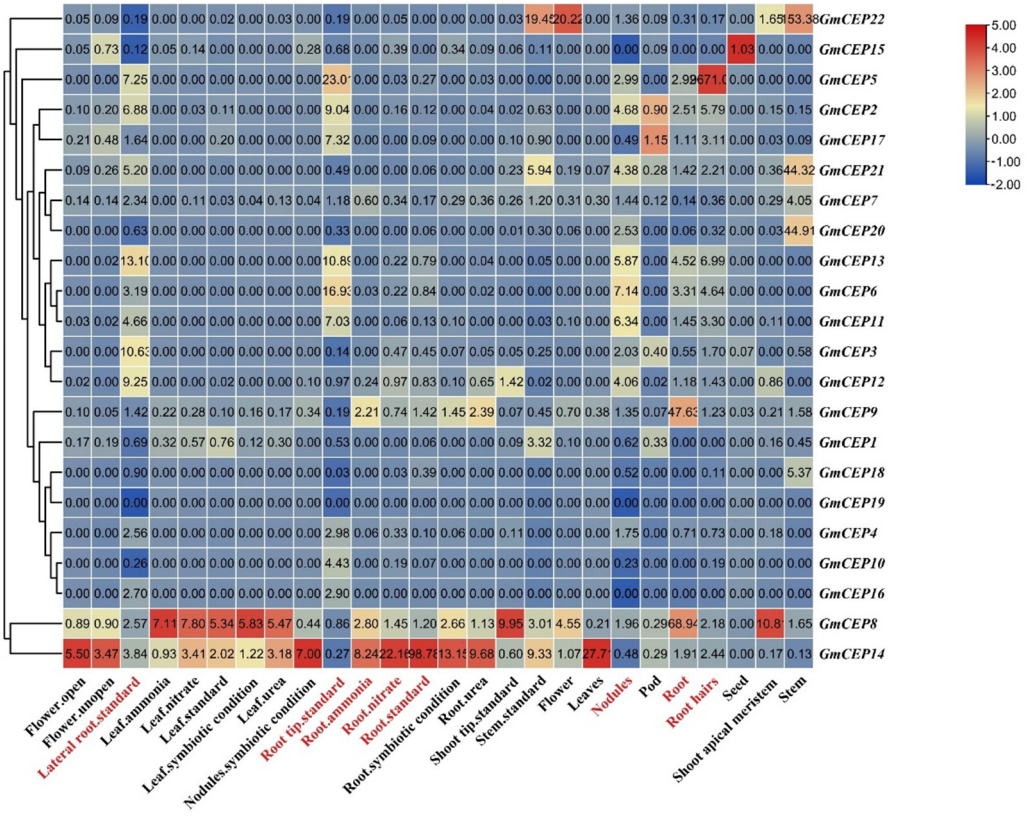

Figure 2. The *GmCEP* expression profiles in different organs with or without treatments. The transcriptome data of *GmCEP* genes in nine tissues (root, stem, bud, leaf, flower, nodule, root tip, pod, and seed) and three different treatment development stages (ammonia treatment, nitrogen treatment, and urea treatment of root and leaf) were obtained from the phytozome database. The Heatmap package in Tbtools [20] software was applied to show the expressed FPKM value in different tissues.

3.4. GmCEP6 Is Preferentially Expressed in Soybean Nodules

Bioinformatics analysis revealed that the full length transcript of *GmCEP6* was 730 bp with an entire exon, the cDNA contained a 5′ untranslated region (UTR) of 101 nucleotides and 3′ UTR of 386 bp, and the gene contained a 243 bp open reading frame (ORF) (Figure 3A), encoding a predicted 80 amino acid residues proteins with conserved C-terminally encoded peptides (CEP) (66-80aa) (Figure 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 Figure 3C; *GmCEP6* was mainly expressed in roots and nodules, which indicates a possible role of *GmCEP6* in nodulation (Figure 3C). To further check the *GmCEP6* expression in response to rhizobium infection, soybean seedlings were inoculated with *Bradyrhizobium japonicum* USDA110, and the transcript abundance of *GmCEP6* at different stages was confirmed, as shown in Figure 3D; *GmCEP6* was weakly induced by USDA110 treatment and peaked at 16 DPI in infected roots, while dramatically decreased at 28 DPI. In addition, *GmCEP6* was markedly higher expressed in nodules than in roots at the checked time points (Figure 3D). Finally, to visually determine the expression profile of *GmCEP6* in soybean nodulation, transgenic hairy roots harboring the -2kb promoter region upstream of *GmCEP6* ATG were 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 the pericycle, nodule primordium, lateral root primordium, and root nodule (Figure 3E–I). 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.

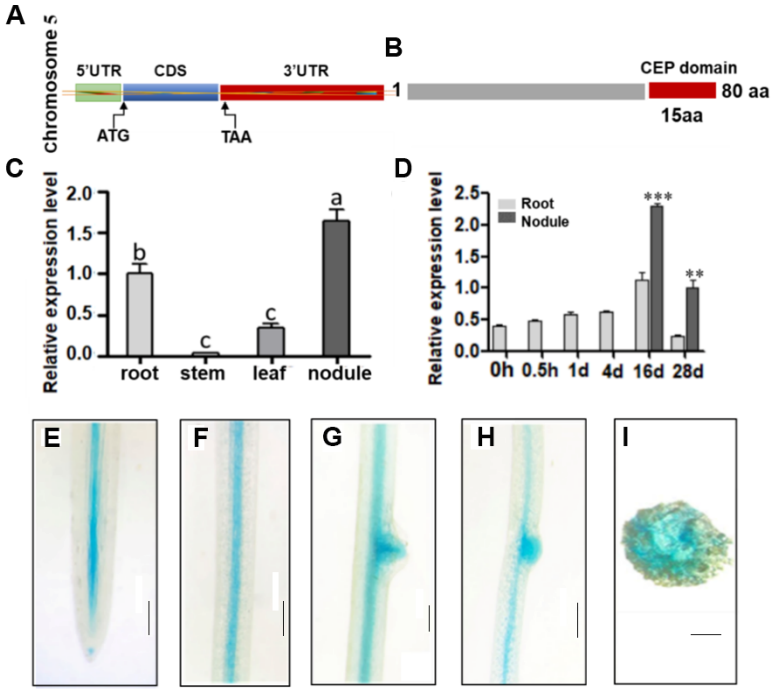

Figure 3. Expression pattern of *GmCEP6* in soybean nodulation. (**A**) Schematic gene structure of *GmCEP6*. (**B**) Domain analysis of *GmCEP6*. The conserved C-terminally encoded peptide (CEP) (66–80 aa) is shown in red. (**C**) Relative expression of *GmCEP6* in soybean root, nodule, stem and leaf 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 (Standard Error); 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**), emerged pericycle (**F**), lateral root primordium (**G**), nodule primordium (**H**), and nodule (**I**). Scale bar in (**E**–**I**) = 1 cm.

3.5. 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 Figure 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 that in the control roots (Figure 4D). Then, the nodule numbers per root was quantified at 14 and 28 days after inoculation; we found, in *GmCEP6* overexpressing hairy roots, the nodule number increased by 2.875 times (14 DPI) and four times (28 DPI), respectively. These 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, and deletions and mutations can be detected in *GmCEP6*-KO roots (Figure S3). As shown in Figure 4E,F, in *GmCEP6*-KO root lines, nodule numbers 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.

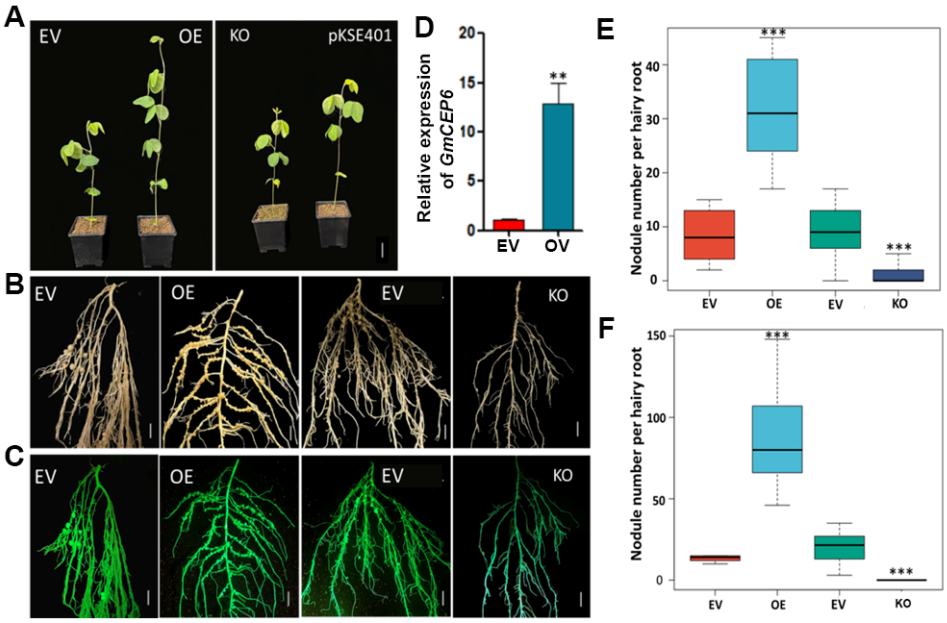

Figure 4. *GmCEP6* regulates soybean symbiotic nodulation. (**A**) Soybean growth status of composite plants harboring empty vector, *GmCEP6*-OE, and *GmCEP6*-KO. (**B**) Nodulation performance of transgenic hairy root harboring empty vector, *GmCEP6*-OE, and *GmCEP6*-KO at 28 DPI. (**C**) Transgenic 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 GmCEP6 in 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**) Quantitative data of nodule number per hairy root at 14 DPI. EV (empty vector for overexpressed transgenic roots) is the control. pKSE401 was the empty vector for CRISPR cas9 (KO). (**F**) Quantitative data of nodule number per hairy root at 28 DPI. EV: empty vector for overexpressed transgenic roots. pKSE401 was the empty vector for CRISPR cas9 (KO). These experiments were conducted on more than three dependent biological replicates. Data are means with ±SE (Standard Error) from three independent replicates ($n = 12$). Asterisks indicate significant difference within a p level in *t*-tests, ** $p < 0.01$; *** $p < 0.001$.

3.6. GmCEP6 Affects the Expression of Related Genes in Nodulation Signal Pathway

As soybean nodule numbers were significantly affected by *GmCEP6* overexpression and *GmCEP6* knock-out, we questioned whether *GmCEP6* regulates soybean nodulation through the NF (Nodular Factor, 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 [21–24]. As shown in Figure 5, the expression of *GmNINa*, *GmENOD40*, *GmNSP1*, *GmHAP2-1*, and *GmHAP2-2* in *GmCEP6-OE* roots was 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 suggest that *GmCEP6* regulates 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 nitrogen state of roots; a previous study on *Medicago truncatula* has shown that *MtCEP1* increased nodulation and promoted nodule development at different nitrate concentrations. In order to investigate the response degree of CEP6 to nitrogen, we selected a medium concentration (4 mM) of nitrogen and high concentration (16 mM) of nitrate for treatment. It was found that CEP6 decreased with the increase of nitrogen concentration. However, we observed *GmCEP6* overexpression in the hairy roots of 35s where CEP6 was treated with a high concentration of

nitrate. Compared with WT, the number of overexpressed nodules was significantly higher. These results indicated that soybean nodulation is enhanced by overexpression of *GmCEP6*, and this tolerance to nitrogen inhibition for nodulation engaged by *GmCEP6* could have beneficial outcomes in soybean breeding.

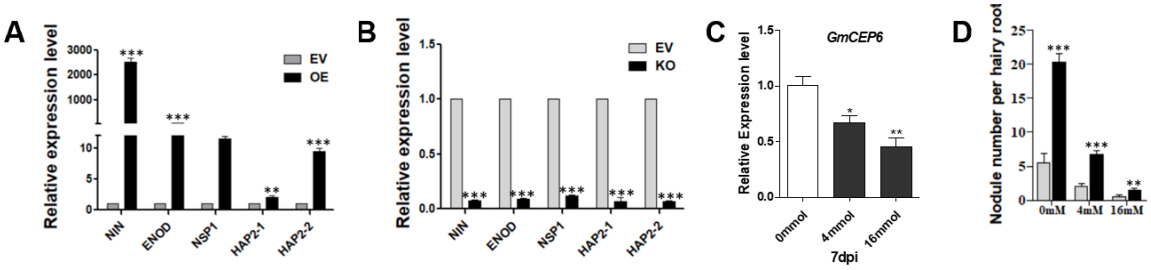

Figure 5. *GmCEP6* regulates Nod factor signaling pathway genes and response to nitrate. (**A**,**B**) *GmCEP6* regulates soybean symbiotic nodulation through Nod factor signaling pathway genes. (**A**) qRT-PCR analysis of the expression of *GmNINa*, *GmENOD40*, *GmNSP1*, *GmHAP2-1*, and *GmHAP2-2* in 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 ±SE. Asterisks indicate significant difference within a P level in *t*-tests. ** $p < 0.01$; *** $p < 0.001$. (**C**) qRT-PCR analysis of the expression of GmCEP6 in different nitrogen concentrations. (**D**) Quantification of nodule number at 28 DAI under different nitrate concentrations. Asterisks indicate significant difference within a P level in *t*-tests. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

4. Discussion

CEP peptides play multiple roles in various plant biological processes. The first identified C-Terminally encoded secreted peptide AtCEP1 significantly arrests root growth [25]. The following reports proved that *CEP* genes responded to nitrogen deficiency, drought stress, and salt stress [26]. 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 [27]. The CEP peptides family number varied in different plants; there were 15 CEP peptides in *A. thaliana*, 11 in *M. truncatula*, 15 in *O. sativa*, 6 in *C. sativus*, and 21 in *P. sativum*. The function of several CEPs in the above plants have been identified. However, little is known about the CEP peptide family in soybeans. In this study, a systematic bioinformatics analysis was applied to identify soybean CEP peptides. A total of 22 *GmCEPs* were characterized from the soybean genome; the GmCEP proteins showed similar features to the previously discovered CEP family (Figure 1; Table 1). On analyzing the expression patterns of *GmCEPs* in the transcriptome, there were diverse expression patterns of *GmCEPs* in different developmental stages of organs and tissues, implying multiple roles of *GmCEPs* in regulating different biological processes in soybeans (Figure 2).

Legumes can specifically interact with their compatible rhizobia in the surrounding soil to form nodules. However, nodulation and nitrogen fixation in mature nodules is an high 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 [28,29]. NODULE INCEPTION (NIN) induced the expression of CLE ROOT SIGNAL1 (*CLE-RS1*) and *CLE-RS2* to activate AON [30]. Another phenomenon in legume nodulation is sensitivity to soil nitrogen content; interestingly, recent studies have shown plants to also have a long-distance system (CEP-CEPR) in the nitrogen assimilation signaling pathway [6]. In legumes, the key transcriptional factor NIN coordinates CEP and CLE signaling peptides, combining these two long-distance signaling pathways to balance nitrogen absorption and symbiotic nitrogen fixation in order to meet high nitrogen demands [31]. In this study, another symbiosis-related *CEP* gene was characterized; we first identified a nodulation-related *Gm-*

CEP6 as regulating soybean nodulation (Figures 3 and 4) and as sharing a phenotype with reported legume *CEP* overexpression; the nodule number of its overexpression being close to the super-noduling phenotype of the NARK mutant in soybeans [18,30]. Moreover, *MtCEP1* promoted *MtNRT2.1* expression and nodulation dependent on compact root architecture 2 (*MtCRA2*) in low nitrate conditions [32]; *MtCEP1*/*MtCRA2* balances root and nodule development by reducing auxin and ethylene responses [33]. Another study reports that *MtCEP1*, *2*, and *12* redundantly regulate lateral root number and nodulation. Further study is needed to clarify the function diversification of GmCEPs family members, to construct the relationship between cytokinin and GmCEP, and to determine the relationship between the AON shoot-center component and CEPR [34]. We also need to clarify the roles of carbon signals in balancing the AON pathway and CEP-CEPR pathways that regulate nodule numbers.

5. Conclusions

In this study, a comprehensive analysis of the *GmCEP* genes was conducted at the whole-genome level; a total of 22 members in the CEP gene family were identified, and the structural features and evolutionary relationships of the GmCEPs were systematically analyzed. Further expression pattern analysis found that *GmCEP6* was mainly expressed in nodules and showed significant responses to nitrate treatment and rhizobial infection, indicating its involvement in nodulation-related processes. Both overexpression and CRISPR-Cas9 were used to validate its role in nodulation. The results demonstrated that overexpression of *GmCEP6* significantly increased nodule numbers, while knock-out of *GmCEP6* led to a significant decrease, suggesting that *GmCEP6* acts as a positive regulator in soybean nodulation. qRT-PCR results showed that changes in *GmCEP6* expression positively influenced the expression of marker genes in the Nod factor signaling pathway. Finally, an analysis of the role of *GmCEP6* in nitrate inhibition of nodulation revealed that, under nitrogen treatment, overexpression of *GmCEP6* significantly increased nodule numbers, indicating its involvement in resistance to nitrate inhibition.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy14050988/s1, Figure S1: Motifs analysis of GmCEPs. Each block represented the position and strength of a motif. The blocks of GmCEPs motif were predicted using MEME. The motif sequence were listed in lower panel; Figure S2: Gene structure of 22 *GmCEP* genes. The gene structure information of *GmCEPs* were obtained through gff file of soybean, the diagram was constructed by Tbtools software; Figure S3: Validation of the mutation of *GmCEP6*-edited hairy roots. (A) Sequence of a region of soybean *GmCEP6* with two target sites indicated. (B) Alignment of sequences of target-1 mutated alleles identified from cloned PCR fragments from crispr cas9 *GmCEP6* (KO) transgenic root lines. Highlighting blue denotes the degree of homology of the aligned fragments, and only aligned regions of interest are displayed. Each trait represents a different mutation type. The most mutation was a base shift, represented by a green triangle, with a total of 5 (n = 13); Table S1: Primer sequences used in this study.

Author Contributions: L.W., W.D. and J.Z., conceived the project; X.W., J.Q., W.T. and M.W. performed the most experiments; X.W., A.Y. and L.W. prepared the original draft; S.W. supplemented the experimental data; 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.

Funding: This research was funded by The Central Guidance for Local Science and Technology Development Projects (grant number YDZJSX2022A035), Technological Innovation 2030 Major Project (grant number 2023ZD040350404), 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), Shanxi Breeding Innovation Joint research and development projects (Grant No.2023xczx03), and National Laboratory of Minor Crops Germplasm Innovation and Molecular Breeding (in preparation) (Grant No. 202105D121010-23, 202204010910001-33).

Data Availability Statement: The RNA-seq data used in this study were download from the Phytozome database (https://phytozome-next.jgi.doe.gov/, (accessed on 1 March 2022)). Sequence data from this article can be found in the GenBank/EMBL or *Glycine max* Wm82.a4.v1 database, with the following entry number shown in Table 1. AtCEPs protein sequences were obtained from the Arabidopsis Information Resource database (https://www.arabidopsis.org/, (accessed on 6 March 2022)). The genome sequence, gff3 file, and protein sequence of soybean (*Glycine max*) were downloaded from the Ensembl database (http://plants.ensembl.org/index.html, (accessed on 9 March 2022)). The Muscle program of MEGA-X was applied to construct the CEPs' phylogenetic tree; the 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.

Acknowledgments: We thank Xia Li and Zhaosheng Kong for technical support and valuable suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kim, D.G.; Isaac, M.E. Nitrogen dynamics in agroforestry systems. A review. *Agron. Sustain. Dev.* **2022**, *42*, 60. [CrossRef]
2. Terpolilli, J.J.; Hood, G.A.; Poole, P.S. Chapter 5. What Determines the Efficiency of N2-Fixing Rhizobium-Legume Symbioses. *Adv. Microb. Physiol.* **2012**, *60*, 325–389. [CrossRef]
3. Herridge, D.F.; Peoples, M.B.; Boddey, R.M. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **2008**, *311*, 1–18. [CrossRef]
4. Gopalakrishnan, S.; Sathya, A.; Vijayabharathi, R.; Varshney, R.K.; Gowda, C.L.; Krishnamurthy, L. Plant growth promoting rhizobia: Challenges and opportunities. *3 Biotech* **2015**, *5*, 355–377. [CrossRef]
5. Murray, J.D.; Liu, C.-W.; Chen, Y.; Miller, A.J. Nitrogen sensing in legumes. *J. Exp. Bot.* **2016**, *68*, 1919–1926. [CrossRef]
6. Tabata, R.; Sumida, K.; Yoshii, T.; Ohyama, K.; Shinohara, H.; Matsubayashi, Y. Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. *Science* **2014**, *346*, 343–346. [CrossRef]
7. Taleski, M.; Imin, N.; Djordjevic, M.A. CEP peptide hormones: Key players in orchestrating nitrogen-demand signalling, root nodulation, and lateral root development. *J. Exp. Bot.* **2018**, *69*, 1829–1836. [CrossRef]
8. Roberts, I.; Smith, S.; Stes, E.; De Rybel, B.; Staes, A.; van de Cotte, B.; Njo, M.F.; Dedeyne, L.; Demol, H.; Lavenus, J.; et al. CEP5 and XIP1/CEPR1 regulate lateral root initiation in Arabidopsis. *J. Exp. Bot.* **2016**, *67*, 4889–4899. [CrossRef]
9. Ogilvie, H.A.; Imin, N.; Djordjevic, M.A. Diversification of the C-TERMINALLY ENCODED PEPTIDE (CEP) gene family in angiosperms, and evolution of plant-family specific CEP genes. *BMC Genom.* **2014**, *15*, 870. [CrossRef]
10. Roberts, I.; Smith, S.; De Rybel, B.; Broeke, J.V.D.; Smet, W.; De Cokere, S.; Mispelaere, M.; De Smet, I.; Beeckman, T. The CEP family in land plants: Evolutionary analyses, expression studies, and role in Arabidopsis shoot development. *J. Exp. Bot.* **2013**, *64*, 5371–5381. [CrossRef]
11. Zhu, F.; Ye, Q.; Chen, H.; Dong, J.; Wang, T. Multigene editing reveals that MtCEP1/2/12 redundantly regulate lateral root and nodule number in Medicago truncatula. *J. Exp. Bot.* **2021**, *72*, 3661–3676. [CrossRef]
12. Huala, E.; Dickerman, A.W.; Garcia-Hernandez, M.; Weems, D.; Reiser, L.; LaFond, F. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **2001**, *29*, 102–105. [CrossRef]
13. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [CrossRef]
14. Mitchell, A.L.; Attwood, T.K.; Babbitt, P.C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S.D.; Chang, H.Y.; El-Gebali, S.; Fraser, M.I.; et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2018**, *47*, D351–D360. [CrossRef]
15. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef]
16. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, w202–w208. [CrossRef]
17. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
18. Imin, N.; Mohd-Radzman, N.A.; Ogilvie, H.A.; Djordjevic, M.A. The peptide-encoding CEP1 gene modulates lateral root and nodule numbers in *Medicago truncatula*. *J. Exp. Bot.* **2013**, *64*, 5395–5409. [CrossRef]
19. Jian, B.; Liu, B.; Bi, Y.; Hou, W.; Wu, C.; Han, T. Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol. Biol.* **2008**, *9*, 59. [CrossRef]
20. Chen, C.J.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.H.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef]
21. Franssen, H.; Yang, W.C.; Katinakis, P.; Bisseling, T. *Characterization of GmENOD40, a Gene Expressed in Soybean Nodule Primordia*; Springer: Amsterdam, The Netherlands, 1993. [CrossRef]

22. Wang, L.; Sun, Z.; Su, C.; Wang, Y.; Yan, Q.; Chen, J.; Ott, T.; Li, X. A GmNINa-miR172c-NNC1 Regulatory Network Coordinates the Nodulation and Autoregulation of Nodulation Pathways in Soybean. *Mol. Plant* **2019**, *12*, 1211–1226. [CrossRef]
23. Wang, Z.; Wang, L.; Wang, Y.; Li, X. The NMN Module Conducts Nodule Number Orchestra. *iScience* **2020**, *23*, 100825. [CrossRef]
24. He, C.; Gao, H.; Wang, H.; Guo, Y.; He, M.; Peng, Y.; Wang, X. GSK3-mediated stress signaling inhibits legume-rhizobium symbiosis by phosphorylating GmNSP1 in soybean. *Mol. Plant* **2021**, *14*, 488–502. [CrossRef]
25. Ohyama, K.; Ogawa, M.; Matsubayashi, Y. Identification of a biologically active, small, secreted peptide in Arabidopsis by in silico gene screening, followed by LC-MS-based structure analysis. *Plant J.* **2010**, *55*, 152–160. [CrossRef]
26. Aggarwal, S.; Kumar, A.; Jain, M.; Sudan, J.; Singh, K.; Kumari, S.; Mustafiz, A. C-terminally encoded peptides (CEPs) are potential mediators of abiotic stress response in plants. *Physiol. Mol. Biol. Plants* **2020**, *26*, 2019–2033. [CrossRef]
27. Taleski, M.; Chapman, K.; Novák, O.; Schmülling, T.; Frank, M.; Djordjevic, M.A. CEP peptide and cytokinin pathways converge on CEPD glutaredoxins to inhibit root growth. *Nat. Commun.* **2023**, *14*, 1683. [CrossRef]
28. Krusell, L.; Madsen, L.H.; Sato, S.; Aubert, G.; Genua, A.; Szczyglowski, K.; Duc, G.; Kaneko, T.; Tabata, S.; de Bruijn, F.; et al. Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* **2002**, *420*, 422–426. [CrossRef]
29. Okamoto, S.; Kawaguchi, M. Shoot HAR1 mediates nitrate inhibition of nodulation in Lotus japonicus. *Plant Signal. Behav.* **2015**, *10*, e1000138. [CrossRef]
30. Soyano, T.; Hirakawa, H.; Sato, S.; Hayashi, M.; Kawaguchi, M. NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14607–14612. [CrossRef]
31. Laffont, C.; Ivanovici, A.; Gautrat, P.; Brault, M.; Djordjevic, M.A.; Frugier, F. The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically. *Nat. Commun.* **2020**, *11*, 3167. [CrossRef]
32. Luo, Z.; Wang, J.; Li, F.; Lu, Y.; Fang, Z.; Fu, M.; Mysore, K.S.; Wen, J.; Gong, J.; Murray, J.D.; et al. The small peptide CEP1 and the NIN-like protein NLP1 regulate NRT2.1 to mediate root nodule formation across nitrate concentrations. *Plant Cell* **2023**, *35*, 776–794. [CrossRef] [PubMed]
33. Zhu, F.; Deng, J.; Chen, H.; Liu, P.; Zheng, L.; Ye, Q.; Li, R.; Brault, M.; Wen, J.; Frugier, F.; et al. A CEP Peptide Receptor-Like Kinase Regulates Auxin Biosynthesis and Ethylene Signaling to Coordinate Root Growth and Symbiotic Nodulation in Medicago truncatula. *Plant Cell* **2020**, *32*, 2855–2877. [CrossRef] [PubMed]
34. Magori, S.; Kawaguchi, M. Long-distance control of nodulation: Molecules and models. *Mol. Cells* **2009**, *27*, 129–134. [CrossRef] [PubMed]