Crop genomics and breeding CLAVATA3/EMBRYO SURROUNDING REGION-RELATED (CLE) proteins belong to a small peptide family in plants. During plant development, CLE gene family members play a pivotal role in regulating cell-to-cell communication and stem cell maintenance. However, the evolutionary process and functional importance of CLEs are unclear in Brassicaceae. In this study, a total of 70 BnCLEs were identified in *Brassica napus* (2n = 4x = 38, A_Cn*), 32 from the A* subgenome, 36 from the Cn subgenome, and 2 from the unanchored subgenome. Meanwhile, 29 BrCLE and 32 BoCLE genes were explored in *Brassica rapa* (2n = 2x = 20, A_r) and *Brassica oleracea* (2n = 2x = 18, C_o). Phylogenetic analysis revealed that 163 CLEs derived from three *Brassica* species and *Arabidopsis thaliana* can be divided into seven subfamilies. Homology and synteny analyses indicated whole-genome triplication (WGT) and segmental duplication may be the major contributors to the expansion of CLE family. In addition, RNA-seq and qPCR analysis indicated that 19 and 16 BnCLEs were more highly expressed in immature seeds and roots than in other tissues. Some CLE gene pairs exhibited different expression patterns in the same tissue, which indicated possible functional divergence. Furthermore, genetic variations and regional association mapping analysis indicated that 12 BnCLEs were potential genes for regulating important agronomic traits. This study provided valuable information to understand the molecular evolution and biological function of CLEs in *B. napus* and its diploid progenitors, which will be helpful for genetic improvement of high-yield breeding in *B. napus*.

**KEYWORDS**

CLE peptide, *Brassica napus*, diploid progenitor, orthologous pairs, expression pattern, association mapping analysis, yield traits
Introduction

Peptide hormones are active molecules composed of many to several tens of amino acids and serve as signal molecules to exchange information between cells (Grienenberger and Fletcher, 2015). In plants, there are various polypeptides; CLAVATA3/EMBRYO SURROUNDING REGION-RELATED (CLE), one of the largest families of known polypeptides, is generally 12–13 amino acids in length, including a secretion signal peptide in N-terminus and a highly conserved CLE domain in C-terminus. Domain deletion and domain exchange experiments have indicated that the CLE domain of CLV3 in Arabidopsis thaliana plays an independent role in adjacent flanking sequences (Fletcher et al., 1999; Rojo et al., 2002). With the help of matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) mass spectrometry, an active 12-amino acid CLV3 polypeptide molecule (corresponding to the CLE domain, with one amino acid removed from each side), is isolated from A. thaliana over-expressing CLV3, and the peptide synthesized in vitro is functional (Tatsuhiko, 2006). Exogenous application of synthetic CLV3, CLE19, and CLE40 polypeptides exhibits a similar phenotype with overexpression of these CLE genes. CLV3 polypeptides also restore the clv3-2 mutant phenotype (Fiers et al., 2006). These results indicate that CLE polypeptides are the active form of CLE family proteins.

It is difficult to clearly identify the function of each CLE peptide owing to its small size and high sequence conservation. However, the application of new technologies such as gene editing makes it easier to study their functions. However, it is easier to study their functions based on the application of new technology, like gene editing. Now, many results show that CLE polypeptides play important roles in plant development and hormone and stress response. In shoot apical meristems (SAMs), the stem cell homeostasis is maintained by a dynamic negative feedback loop involved in the CLV3-WUSCHEL (WUS) pathway. The transcription factor gene WUS, which interacts with SHOOT MERISTEMLESS (STM), can promote the expression of CLV3; meanwhile, overmuch CLV3 will suppress the expression of STM and WUS to maintain the stem cell population in the SAM (Schoof et al., 2000).

In addition to participating in the division and differentiation of meristem cells, CLE polypeptides also play a key role in the development of seeds (Fiume, 2010). With the use of the method of promoter fusion GUS, CLE8 and WUSCHEL-related homeobox 8 (WOX8) are expressed in the embryo and endosperm during the early stages of seed development in Arabidopsis (Fiume and Fletcher, 2012). The number of embryo and endosperm cells decreases in the cle8-1 mutant, suggesting that CLE8 promotes the proliferation of embryo and endosperm cells. The length and width of seeds produced by CLE8-overexpressing plants and wox8-1 mutant seeds significantly increase and decrease, respectively, compared with those of wild-type seeds. WOX8 expression significantly increases in CLE8-overexpression lines, indicating that CLE8 promotes its expression (Fiume and Fletcher, 2012; Song et al., 2013). Some CLE peptides likely interact with hormones. For example, CLE26 expressed at the phloem pole, regulated root architecture, and its expression is significantly enhanced by auxin treatment. CLE26 can affect the activity of the polar auxin transporter in the auxin signaling (Czyzewicz et al., 2015). The shoot growth is related to the root-expressed CLE6 under the gibberellin effect (Bidadi et al., 2014). In addition to development and hormone response, CLE also mediates responses to various abiotic stress. The expression of CLE25 is increased when the root is hydropenic, and then the root-derived CLE25 peptide moves into the leaves to modulate the closure of the stoma (Takahashi et al., 2018). In response to low-sulfate conditions, the expression of CLE2 and CLE3 in roots is reduced, which could inhibit lateral root development (Dong et al., 2019).

CLE genes are widely present in plants and even in several plant-parasitic nematodes. Three CLEs expressed in the maize endosperm (Esr1, Esr2, and Esr3) are supposed to be involved in signal transduction between embryo and endosperm during early development (Bonello et al., 2000). ZmCLE7 and ZmFCP1 are CLV3 homologs, and promoter editing performed by CRISPR-Cas9 increases many grain-yield-related traits in maize (Liu et al., 2021). Two CLEs from Lotus japonicus (LjCLE-RS1 and LjCLE-RS2) are repressors of excess root nodulation (Okamoto et al., 2013). An AtCLE19-like gene with high expression in the flower bud, pistil, and embryo of Brassica napus was ectopically expressed in A. thaliana, resulting in large heads (Fiers et al., 2004). The mutations of BrCLE3s induced by CRISPR/Cas9 could result in multicellular siliques and an increase in seed production (Yang et al., 2018). Similarly, BrCLV3 mutations conferred multicellular pods in Brassica rapa (Fan et al., 2014). In Raphanus sativus, overexpression of RsCLE2 and RsCLE19 increased the number of xylem elements (Gancheva et al., 2016). Moreover, CLEs, which may assist the infection process, are identified in cyst nematodes (Wang J. et al., 2011). Current research suggests that the CLEs are highly conserved and involved in biological evolution, especially in plants.

Recent genome-wide analyses have identified CLE genes in many plants (Zhang et al., 2014; Hastwell et al., 2015; Gancheva et al., 2016; Han et al., 2016; Han et al., 2020). The allotetraploid species B. napus (2n = 4x = 38, A1B1C0), an important oil crop, was formed from the hybridization between B. rapa and Brassica oleracea (2n = 2x = 18, C0) at about 7,500 years ago; therefore, the phylogenetic relationship of Brassica provides a good basis for studying the evolution of gene family (Allender and King, 2010). Based on bioinformatics and comparative genomic approaches, this study performed a multidimensional investigation for CLEs in rapeseed and its diploid progenitors, including genome-wide identification, molecular characterization, phylogenetic analysis,
synteny analysis, expression profiling in different tissues, and regional association mapping analysis. The results will provide useful information for biological functions and molecular evolution in Brassica, which also supply candidate genes for genetic improvement in rapeseed breeding.

Materials and methods

Identification of CLEs in Brassicaceae

Multiple TBLASTN and BLASTN searches were performed for CLEs identification in B. napus reference genome (http://brassicadb.org/brad/downloadOverview.php) (Wang et al., 2015) based on A. thaliana CLE [expected threshold (e–10)]. The results were then validated by the Conserved Domain Database (CDD) in the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/cdd) (Lu et al., 2020) and Modular Architecture Research Tool SMART (http://smart.embl-heidelberg.de/) (Letunic et al., 2020) to confirm the authenticity of the CLE domain in the open reading frame. Open reading frames of homologous chromosome regions were confirmed for potential unannotated or truncated duplicates of CLEs. BrCLEs in B. rapa and BoCLEs in B. oleracea (http://brassicadb.org/brad/downloadOverview.php) were obtained as described above to explore the evolution of CLEs in Brassica.

Gene structure, conserved motifs, and cis-acting regulatory elements analysis

Gene structures and conserved motifs for the CLEs were constructed via TBtools. Logo diagrams used to define consensus sequences were obtained using multiple sequence alignments for each BnCLE peptide group (I–V), including A. thaliana, B. rapa, and B. oleracea by TEShade (Beitz, 2000). The signal peptides of CLEs were identified via SMART (Letunic et al., 2020). PlantCARE was carried out to predict the promoters in the 2-kb region before the start codon for the cis-element identification (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Magali, 2002). Gene Structure Display Server (GSDS 2.0) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to exhibit the gene structure (Hu et al., 2015), and the heatmap was visualized by R package (https://cran.r-project.org/).

CLE gene duplication pattern and synteny analysis

In order to investigate the synteny relationship of CLEs in Brassicaceae, all the CLEs were searched as “syntenic genes” in Brassicaceae Database (BRAD) (http://brassicadb.cn/#/). At the same time, TBtools (Chen et al., 2020) was used to detect gene duplication patterns and verify the synteny relationship, as well as calculate the ratio of non-synonymous to synonymous substitutions (Ka/Ks) of syntenic gene pairs. Orthologous CLEs located on syntenic chromosome blocks were displayed by Circos software (Krzywinski et al., 2009).

Phylogenetic analysis of CLEs

Multiple alignments of the CLE peptide sequences from the four Brassicaceae species were performed by the ClustalW (Larkin, 2007). Phylogenetic analysis was generated using the MEGA7 soft with the maximum likelihood (ML) method, 1,000 bootstrap replications, and the JTT+G model (Kumar et al., 2016). The tree was visualized using Evolview (https://www.evolgenius.info/evolview/) (He et al., 2016).

Prediction of protein–protein interactions

Protein–protein interactions in A. thaliana were obtained in the STRING database (https://www.string-db.org/). The analysis and demonstration of CLEs interactions in B. napus followed the description based on the previous study (Xie et al., 2022).

Transcriptome expression pattern of CLEs

RNA-seq raw data of siliques, leaves, flowers, and stems for B. napus, B. rapa, and B. oleracea were downloaded from NCBI (ProjectID: PRJNA489323); after being filtered with Trimomatic (Bolger et al., 2014), the clean data without adapters and low-quality bases were aligned with the reference genome (http://brassicadb.org/brad/downloadOverview.php) using hisat2 (Kim et al., 2015). Based on the mapping results, the gene expressions (fragments per kilobase million (FPKM)) were counted by Stringtie (Pertea et al., 2015), and the heatmaps were drawn by R.

Quantitative real-time PCR analysis

The B. napus ZS11 was used in this study. Tissues like flowers, roots, stems, leaves, immature pods, immature seeds, and apical meristems were collected and extracted using an RNasy Extraction Kit (Invitrogen, Carlsbad, CA, USA). Quantitative real-time PCR (qRT-PCR) was carried out by referring to a formerly described protocol (Zhao et al., 2021). The expression level was displayed by heatmap in R. Expression
patterns of five genes were shown by TBtools-eFP (Chen et al., 2020) (http://yanglab.hzau.edu.cn/BnTIR/eFP).

Genetic variation of CLEs in Brassica napus core accessions

To investigate the genetic variation of CLE genes in B. napus, a panel of 204 rapeseed accessions was selected for this work (Zhao et al., 2022). The single-nucleotide polymorphism (SNP) information of CLE genes was retained from the previous study. After annotation with SnpEff (Cingolani et al., 2012), the distribution was analyzed to inspect the position of variations. The agronomic traits for three consecutive years (2014–2016) including plant height, branch number, initial branch height, length, silique number, and silique density of main inflorescence, main inflorescence, silique length, seed number per silique, main inflorescence seed density, thousand seed weight, seed weight per silique, and seed weight of the main inflorescence were surveyed and handled with the best linear unbiased prediction (BLUP). To study the potential impact of CLE genes on agronomic traits in B. napus, SNPs within 30 kb upstream and downstream of CLEs were used for the investigation of its potential impact. Regional association analysis considering population structure and relative kinship was conducted by EMMAX (Kang et al., 2010).

Result

Identification of CLEs in Brassica napus (Bn), Brassica rapa (Br), and Brassica oleracea (Bo)

A genome-wide analysis of CLE genes in B. napus and its diploid progenitors was performed involving multiple BLAST queries and iterative queries, followed by domain validation and removal of false positives (i.e., no CLE domain). Finally, a total of 29 (BrCLEs), 32 (BoCLEs), and 70 (BnCLEs) genes were identified (Table S2). The total number of BrCLEs and BoCLEs in the two diploid progenitors was lower than that of BnCLEs in the allotetraploid rapeseed, indicating that CLE gene expansion event has occurred in B. napus during polyploidization.

The CLEs in the three species of Brassica species were renamed according to the AtCLEs in Arabidopsis based on the naming conventions for the B. genus (Ostergaard and King, 2008), with the last letter “a” indicating the highest homology with Arabidopsis, and then “b”, and so on. In B. napus, the capital letters A and C following “Bn” represented the A_n and C_n subgenomes, respectively.

Phylogenetic analysis of CLEs in Brassicaceae

To explore the evolution of the CLEs in Brassica, we constructed a phylogenetic tree using full-length amino acid sequences from Brassicaceae (Figure 1). The results showed CLEs clearly divided into seven subfamilies (I to VII). Subfamily V was the largest (42 members), followed by subfamily I (38 members). There were 17, 11, 23, 17, and 14 members in subfamilies II, III, IV, VI, and VII, respectively. The CLEs of the four species were distributed in all subfamilies. Most AtCLEs in each subfamily matched multiple sets of orthologs from B. napus and its two progenitors. A sister pair demonstrated the closest genetic relationship in a phylogenetic tree. A total of 56 sister pairs were observed. The majority of the sister pairs were orthologous gene pairs between the A_n (or C_n) subgenomes of B. napus and B. rapa (or B. oleracea), with 18 A_n–A_n, pairs and 19 C_n–C_n pairs. These results supported that the gene duplication events happened in the B. napus genome and indicated that CLE orthologous genes of distinct subfamilies were highly conserved in the respective genome.

Duplication pattern and chromosome localization analysis of CLEs

The chromosomal locations of BnCLEs, BrCLEs, and BoCLEs were investigated according to their physical positions (Figure 2). The BrCLEs were asymmetrically distributed on the 19 chromosomes in B. napus. There were a total of 32 BrCLEs in the A_n subgenome and 36 in the C_n subgenome, which were similar to those in B. rapa (A_r, 29) and B. oleracea (C_r, 32). The remaining two BnCLE genes were located on unanchored scaffolds (Table S2). Each chromosome harbored at least one CLE gene. Chromosome A_n07 in B. napus carried the most CLEs (seven CLEs). On chromosomes A_n01, C_n01, and C_n08, only one CLE gene was found. Furthermore, many CLEs retained their relative position in A_n and A_r, whereas only a portion of CLEs retained their relative position in C_n and C_r. For example, the same number of CLEs in Ar01–Ar01, Ar05–Ar05, Ar06–Ar06, and Ar08–Ar08 was observed, which also showed similar locations in the chromosomes (Figure 2).

We searched four duplicated types in each Brassicaceae species, including dispersed, proximal, tandem, and whole-genome duplication (WGD) (Table S3). We found that 56 of 70 BnCLEs were derived from segmental duplication/whole-genome triplication (WGT). Therefore, it appeared that segmental duplication/WGT played an important role in the BnCLE expansion. In addition, we examined CLE gene expansion patterns in B. oleracea and B. rapa, finding that
most BoCLEs (53.1%; 17/32) and BrCLEs (89.7%; 26/29) were also derived from segmental duplication/WGT (Table S3). One cluster of CLE tandem repeat genes was only present in B. rapa and B. oleracea, but not in B. napus.

Structure analysis of the CLEs

A great amount of the identified CLES lacked predicted introns, except for 20 genes, including one to three introns (Table S2, Figure S1). Like other plants, the multiple protein sequence alignment of CLE pre-propeptides in B. napus contained a signal peptide at the N-terminus, a central variable domain, and a CLE domain at the C-terminus. The signal peptide sequences and the variable domain among distinct genes were not highly conserved, except between orthologous genes. The highly conserved CLE domain that contained approximately 12/13 aa and residues at different positions had different levels of conservation. To study the sequence of the conserved CLE domains and the degree of their conservation in different Brassicaceae species, multiple sequence alignment was used to generate the protein sequence logos in B. rapa, B. oleracea, B. napus, and Arabidopsis (Figure 3). The results showed that conserved residues patterns of the CLE domain were remarkably similar in these four plants (Figure 3). The conservation of residues 1, 4, 6, 8, 9, 11, and 12 suggested that they may be critical to the function of CLE mature peptides.

These results indicated that CLE gene family was relatively conserved, while some motif sequences changed slightly during Brassica evolution, which possibly contributed to extended special biological function.

Cis-elements in the promoters can affect gene expression (Davuluri et al., 2003; Klaas, 2009). Therefore, CLE gene promoters of these three species were investigated using PlantCARE (Magali, 2002). Four to 12 cis-elements involving development, hormone, and stress were identified in each BnCLE promoter, while 2–12 and 4–13 were found in BrCLEs and BoCLEs, respectively (Figure S2). ARE has the most elements in these three species and existed in 87.1% (61/70) BnCLE promoters, 93.1% (27/29) BrCLEs, and 96.9% (31/32) BoCLEs; this element is essential for anaerobic induction (Figure 4). Hormone-responsive elements involved in abscisic acid responsiveness, methyl jasmonate responsiveness, and ethylene responsiveness were also very common in CLE
promoters: 80% (56/70), 86.2% (25/29), and 87.5% (28/32) of 
BnCLE, BrCLE, and BoCLE promoters included ABRE, respectively. In development elements, GCN4 motif involved 
in endosperm expression (25.7%, 18/70 in BnCLEs; 37.9%, 11/29 in 
BrCLEs; and 25%, 8/32 in BoCLEs), O2 site involved in zein 
metabolism regulation (40%, 28/70 in BnCLEs; 31%, 9/29 in 
BrCLEs; and 34.4%, 11/32 in BoCLEs), CAT-box related to 
meristem expression (27.1%, 19/70 in BnCLEs; 31%, 9/29 in 
BrCLEs; and 37.5%, 12/32 in BoCLEs) and circadian (32.9%, 23/
70 in BnCLEs; 24.1%, 7/29 in BrCLEs; and 37.5%, 12/32 in 
BoCLEs) were common. According to the cluster results, some 
similar CLE promoters had similar cis-elements, like CLE4, 
CLE19, and CLE42, while other CLEs were clustered into 
different groups.

Orthologous relationship and synteny analysis of CLE genes

Syntenic genes are orthologous genes located in syntenic fragments between different species that derive from a shared 
ancestor. We obtained the syntenic genes of CLE genes of
Arabidopsis in three Brassica species by searching "syntenic gene" in BRAD (Wang et al., 2015) and showed these 
collinearity relationships using Circos (Krzywinski et al., 2009) 
software between the Aa, and Ca subgenomes of B. napus and its 
two diploid progenitors (Figure 5, Table S4). A total of 101 CLE 
genes in three Brassica species showed conserved synteny with 
those in A. thaliana and were positioned in the same conserved 
chromosomal blocks, such as A, B, D, E, F, I, J, L, R, U, and Wb.
(Schranz et al., 2006). In addition, syntenic genes in three Brassica species were divided into three fractionated subgenomes, which were specified as LF (least-fractionated), MF1 (medium-fractionated), and MF2 (most-fractionated) according to the extent of gene retention (Chalhoub et al., 2014). There were 41, 31, and 27 AtCLE genes caught in the LF, MF1, and MF2 subgenomes, respectively (Table S4). A total of 27 AtCLE genes retained corresponding syntenic genes in the three Brassica species. The existing forms of syntenic genes in the genomes of three Brassica species were different. The first type was that syntenic genes of AtCLE were completely preserved in the same block of synteny in the Ar, Co, An, and Cn subgenomes, such as AtCLV3. The second type was that AtCLE genes were retained in the Ar or/and Co genome but lost in B. napus genomes, such as AtCLE4, AtCLE5, AtCLE6, and AtCLE14. The third type was that AtCLE genes were retained in B. napus genome but lost in B. rapa or B. oleracea genomes, such as AtCLE1, AtCLE9, and AtCLE40. The results showed that the expansion of CLE gene family was also accompanied by gene loss.

To comprehend whether natural selection acted on the evolution of CLE gene family in B. napus, selection pressure analysis was performed on the syntenic CLE gene pairs between An and Ar, Co, and Cn. The non-synonymous rate (Ka) and synonymous rate (Ks) values were calculated. The Ka/Ks ratio > 1 represents positive selection, the Ka/Ks ratio = 1 represents neutral selection, and the Ka/Ks ratio < 1 represents purifying selection (Nekrutenko, 2002). The Ka/Ks ratios of the syntenic gene pairs are shown in Table S5. The Ka/Ks ratios for some CLE syntenic gene pairs were >1, such as BnA.CLE.1d and BnA.CLE.1e, BnA.CLE.2b and BnC.CLE.2b, BnA.CLE.10 and BraA.CLE.10, and BnA.CLE.27 and BnA.CLE.27, which indicated that these genes were subject to positive selection pressure. Many syntenic gene pairs had no Ka/Ks value in B. napus because these two genes had the same sequence or large differences. The rest were less than one, which indicated that they underwent purifying selection during the evolution process and may preferentially perform conserved functions.

### Predicted protein interactions of BnCLEs, BrCLEs, and BoCLEs

As polypeptide hormones, CLEs need to combine with receptor proteins to transmit signals between cells. To investigate the involved biological process in rapeseed and its diploid progenitors, protein–protein interaction networks were predicted based on known protein interactions in Arabidopsis. A total of 1,061 proteins Arabidopsis proteins interacted with CLEs, resulting in 3,997 proteins in rapeseed (Figure 6A). CLEs interacted with other proteins but also interacted with each other. Taking out the interacted genes for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis indicated that they played important roles in zeatin biosynthesis, amino acid metabolism (like tryptophan, arginine, and proline), polysaccharide biosynthesis, and ion channels (Figure 6B). Gene Ontology (GO) enrichment analysis showed that their molecular functions mainly were receptor serine/threonine kinase binding and phosphorelay response regulator activity; their biological processes were transmembrane receptor protein tyrosine kinase signaling pathway, polarity specification of adaxial/abaxial axis, stamen development, and lateral root development (Figure 6C).

As to CLE genes in B. rapa and B. oleracea, 1,334 and 1,356 proteins were identified as their interacted proteins, respectively. The Gene Ontology enrichment analysis (Figures S3, S4) indicated that these proteins were important in the pattern specification process, shoot system morphogenesis, meristem maintenance, root morphogenesis, and leaf development.
Expression patterns of *CLE* genes in various tissues

According to the RNA-seq data in siliques, leaves, flowers, and stems of *B. napus*, *B. rapa*, and *B. oleracea*, the expression patterns of *CLE* genes are illustrated in Figures S5-S7. The majority of *CLE* genes showed low or no expression in all four tissues. Among the expressed *CLE* genes, more *CLE* genes were highly expressed in stems in three *Brassica* species, like *BnaC.CLE42*, *BnaA.CLE46*, *BnaC.CLE46*, *BolC.CLE20.b*, *BraA.CLE6.b*, and *BraA.CLE6.c,*
whereas BnaA.CLE9.b, BolC.CLE27.a, BolC.CLE14.b, and BraA.CLE14.a were highly expressed in siliques. Only a few paralog CLE genes had a similar expression pattern, like BraA.CLE27.a/BraA.CLE27.b, BnaC.CLE42/BnaA.CLE42.b/ BnaA.CLE42.a/BnaU.CLE42, BolC.CLE9.a/BolC.CLE9.b, and BolC.CLE22.a/BolC.CLE22.b. In three Brassica species, orthologs like BnaA.CLE20.b/BraA.CLE20.b/BolC.CLE20.b and BnaC.CLE9.a/BraA.CLE9/b/BolC.CLE9 displayed a similar expression pattern. Meanwhile, orthologs in B. rapa and B. oleracea showed different expression patterns; for example, BraA.CLE4 was mainly expressed in leaves, whereas BolC.CLE4 was not. These results suggest that some CLE genes have shown functional divergence during their evolution.

Furthermore, we performed a qPCR analysis in B. napus in various tissues—flowers, roots, stems, leaves, immature pods, immature seeds, and apical meristems. Similar to the transcriptome data, BnCLE genes had significantly different expression patterns in different tissues (Figure 5A), whereas BnaA.CLE6.a/b, BnaC.CLE6.b, BnaC.CLE7, BnaA.CLE4, BnaU.CLE4, BnaA.CLE1.c, BnaC.CLE1.a.c, BnaA.CLE9.a/b/c, and BnaC.CLE9.a/b/c were more highly expressed in immature seeds than in the other tissues examined (Figure 7A), whereas BnaA.CLE45.c, BnaA.CLE19, BnaC.CLE19, and BnaC.CLE3.a/b were more highly expressed in apical meristems than in the other tissues (Figure 7B). BnaA.CLE22.d and BnaC.CLE46 were expressed at high levels in root and stem (Figure 7D), while BnaA.CLE1.a/b, BnaA.CLE2.a, BnaC.CLE6.a, and BnaC.CLE14.a were expressed at high levels in root and immature seeds (Figure 7E). The materials and sampling time of the qPCR experiments in this study were different from the data downloaded from NCBI; therefore, the results were not completely consistent. However, there was some similarity: BnCLEs showed different expression patterns in different tissues and were modestly expressed in leaves. The expression patterns of most BnCLEs pairs were significantly different, suggesting that their roles were dissimilar. For example, BnaA.CLE1.b was highly expressed in the roots, whereas BnaC.CLE1.c was more highly expressed in immature seeds. BnaA.CLE1.d was highly expressed in immature seeds, whereas BnaA.CLE1.d was similarly expressed in the different tissues. BnaA.CLE22 was highly expressed in the stem, whereas BnaC.CLE22 was more highly expressed in apical meristems. BnaA.CLE46 was highly expressed in the stem, whereas BnaC.CLE46 was the more highly expressed root, followed by stem.

Regional association mapping for BnCLEs

A natural population including 204 accessions was selected to analyze the genetic variations of CLEs, and a total of 115 SNPs were identified; on average, 1.6 SNPs were detected for each CLE gene, which was apparently less than the whole genome level (28.9 SNPs per gene). The total and average number of SNPs in the A subgenome (74 SNPs, 2.2 SNPs/gene) were higher than in the C subgenome (41 SNPs, 1.1 SNPs/gene). Considering CLE gene length, the SNP density of the A subgenome was 5.7 SNPs/kb, whereas it was 2.2 SNPs/kb in the C subgenome. Moreover, the SNP distribution among paralogous genes was also unequal; for example, BnaA.CLE25.a/BnaA.CLE25.b/BnaC.CLE25.a/BnaC.CLE25.b, the SNP numbers were 9/1/14/0. Finally, SNP annotation implied that 73 SNPs were in exon regions and 37 SNPs could lead to missense mutations.

In this study, 12 agronomic traits (Figure S9) and SNPs within the range of 30 kb upstream and downstream of BnCLEs were used to study the impact of BnCLEs in B. napus. A total of 12 CLEs were significantly associated with at least one agronomic trait.
(p < 0.0001) (Table S6), including plant height, branch height, main inflorescence length, seed weight per silique, silique number, silique density, seed density, and seed weight of the main inflorescence. For example, there was no variation in the gene sequence of BnaC.CLV3.a, but SNPs in the upstream region (6 kb) were significantly associated with seed weight per silique; the population was clearly divided into two haplotypes based on the SNPs, and the t-test displayed the significant difference in seed weight per silique between these two groups (Figure 8A).

According to JASPAR, the SNP region (TCCGTACA) was predicted as the binding site of C2H2 zinc finger factors (SPL3). Moreover, there were another four genes also strongly associated with yield traits, like BnaA.CLE20.b and BnaC.CLE20.b; SNPs in the upstream region (1.4 and 3.4 kb) were associated with main inflorescence seed density (Figures 8B, C). Their interacted proteins were not only enriched in receptor serine/threonine kinase binding (GO:0033612), signaling receptor binding (GO:0005102), and cell–cell signaling involved in cell fate commitment (GO:0045168) but also enriched in GO terms including regulation of meristem structural organization (GO:0009934), maintenance of meristem identity (GO:0010074), and maintenance of root meristem identity (GO:0010078). Therefore, it was speculated that these CLEs could affect the development of rapeseed and result in phenotypic variations.
Discussion

CLE peptides are plant-specific peptide hormones that act as mediators of cell-to-cell communication (Fletcher, 2020). Genome-wide studies on CLE gene family have been performed in tomato, soybean, wheat, and populus (Zhang et al., 2014; Hastwell et al., 2015; Han et al., 2016; Li Z. et al., 2019; Han et al., 2020). In this study, systematic identification, classification, evolution, expression, and association mapping analysis were performed in *B. napus*. In total, we identified 70, 32, and 29 CLEs in *B. napus*, *B. oleracea*, and *B. rapa*, respectively, which were less than those in the previous work (Han et al., 2020), and the possible reason was more strict parameters used in this work. These CLEs were divided into seven subfamilies according to the phylogenetic tree. Furthermore, this work analyzed the relationship of CLEs among three Brassica species and displayed the similarities and differences of cis-acting elements, interacted proteins, and expression patterns. With the aid of a natural population of *B. napus*, the genetic variations of CLE genes were uncovered, and several CLEs were recognized as candidate genes for important agronomic traits in rapeseed by regional association mapping analysis.

Polyplody, an important force in the evolution of species (especially magnoliophyte), played an important role in plant adaptation to new environments (Schemske, 1998; Zhang et al., 2020). Each Brassicaceae genome underwent a WGD event ~35 MYA ago (Bowers, 2003; Yuannian, 2011). Comparative genomic research showed that Brassica species experienced triploidy at the genomic level after they diverged from the Arabidopsis lineage approximately 20 MYA ago (Chalhoub et al., 2014; Liu et al., 2014). Therefore, it is apparent that the Brassica genome underwent paleopolyploidyization (Bowers, 2003; Lysak et al., 2005). The differentiation of *B. rapa* and *B. oleracea* occurred approximately 4.6 MYA ago (Liu et al., 2014), and their natural hybridization formed the *B. napus* about 7,500 years ago (Chalhoub et al., 2014). Segmental duplication also leads to increased gene numbers (Flagel and Wendel, 2009). Most plants have undergone polyploidyization events and thus retain a large number of duplicated chromosomal blocks at the genome level (Cannon et al., 2004). For example, in *B. napus*, segmental duplication/WGT is the primary force for WOX expansion (Li M. et al., 2019). In this study, nearly 80% of CLEs were distributed in syntenic blocks, suggesting that segmental duplication/WGT might contribute significantly to CLE expansion in Brassica. Tandem duplicated genes were defined as an array of at least two homologous genes within 50 kb (Cannon et al., 2004). However, we only found one pair of tandem duplicated genes in *B. rapa* and *B. oleracea*, suggesting that it is not a major factor for CLE expansion in Brassica.

Not all duplicated genes are retained in plants; gene loss always occurred due to the genomic sequence rearrangement after hybridization or chromosome doubling (Paterson et al., 2004; Ye et al., 2020). Each *AtCLE* should have three syntenic orthologs in *B. rapa* and *B. oleracea* based on the triploidy hypothesis. However, only *AtCLE5/6/45* had such a pattern, and the other *AtCLEs* had
The presence of TEs in the genome can cause the rearrangement of chromosomal sequences, which affects the genomic structure, including deletion, inversion, and translocation. Second, the C subgenome underwent more active homologous exchanges than the A subgenome during polyploidization (Chalhoub et al., 2014).

Only a few CLE peptides are studied in Brassica; however, the functions of most AtCLE peptides are characterized. Therefore, Brassica CLE functions can be predicted through sequence similarity and phylogenetic analysis. Seven distinct CLE subfamilies were defined using Arabidopsis and three Brassica species. For example, all CLV3 genes are clustered into subfamily III, while Brclv3 and Bnclv3 cause enlarged meristems and lead to extra organs such as multilocular siliques, similar to the phenotype of Atclv3 (Fletcher et al., 1999; Fan et al., 2014; Yang et al., 2018). This approach could elucidate the function of unknown CLE peptides. However, many genes contain more than one copy in polyploid species (Chalhoub et al., 2014; Li Z. et al., 2019). These duplicated genes are redundant, and variation in promoters could affect their gene expression (Davuluri et al., 2003; Klaas, 2009; Arsovski et al., 2015). For example, BrCLE45 have three highly similar copies, but cis-elements have significant differences, so their expression patterns were different. Distinct expression patterns are a direct sign that duplicated genes may diverge in different directions, indicating the occurrence of non-functionalization, neo-functionalization, or sub-functionalization at the transcriptional level (Gallagher et al., 2016; Cheng et al., 2018). BrCLEs appear to be highly expressed in immature seeds, roots, and stems based on their expression levels in previous studies. Moreover, their interacting proteins are enriched in the development of these organs. Therefore, it is concluded that CLEs perform important functions in the biological process of these organs. However, the majority of CLEs hardly expressed in the tissues analyzed in this study. Therefore, they were either expressed in other developmental stages or conditions, or they were suppressed after duplication. Promoters that regulate the CLEs expression may also play a role. Abundant cis-elements involved in hormone response, development, and stress were identified, and their existence made the regulation more flexible, which helps plants adapt to complex environments.

Genetic variations in CLEs were investigated in the associated B. napus population; the smaller number of SNPs in CLEs indicated that this gene family sequence was very conserved. SNP density was
higher in the A subgenome compared with the C subgenome, which is consistent with other published gene families (Zhu et al., 2020; Wahid et al., 2022; Xie et al., 2022). Twelve CLEs were found to be significantly associated with important agronomic traits using regional association mapping analysis. Among them, the variations in the upstream region of BnaC.CLV3.a (the orthologs of the famous gene CLV3) were significantly associated with seed weight per siliqua. These variations was predicted as the binding sites of transcription factor SPL3, which can directly bind to the promoters of many genes to affect flower development (Jung et al., 2016). In Brassica, BrCLV3 and BnCLV3 control multilocular siliqua traits and increase seed production (Fan et al., 2014; Yang et al., 2018). Although no variation was detected in the genetic region of BnaC.CLV3.a in this natural population, the upstream variations that occurred in the binding sites of transcription factor SPL3 may influence the binding ability, which would possibly regulate the expression of BnaC.CLV3.a and affect the yield trait. In summary, the association mapping analysis conducted in this study could provide a better way to explore the significance of CLEs in phenotypic variation and offer candidate genes for further genetic improvement breeding in B. napus.

**Conclusion**

A total of 29, 32, and 70 CLEs were identified in B. rapa, B. oleracea, and B. napus, respectively, and divided into seven subfamilies in the phylogenetic tree. The conservation of the CLE domain suggested that the CLE family is relatively conserved in its biological function. WGT and segmental duplication were the major contributors to the expansion of CLEs in Brassica species. Transcriptome and qPCR analyses indicated that BnCLEs were highly expressed in immature seeds, roots, and stems. Some CLE pairs exhibited different expression patterns in the same tissue, indicating duplicated CLE differentiation. Moreover, genetic variations and regional association mapping analysis indicated that 12 CLE genes were potential genes for regulating important agronomic traits. In summary, this study was helpful in understanding the molecular evolution and biological function of CLEs in rapeseed and its diploid progenitors, which would aid future genetic improvement for high-yield breeding in B. napus.

**Data availability statement**

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

**Author contributions**

YX, MX, and CZ designed the research, analyzed the data, and wrote the manuscript. MS analyzed the data. YX and CT reviewed the manuscript. All authors contributed to the article and approved the submitted version.

**Funding**

This study was supported by National Natural Science Foundation of China (32070217), Precursor projects of Guizhou province for biological breeding supporting by science and technology in 2022 (Fine identification and evaluation of crop germplasm resources), Subsidy project from NSFC of Guizhou Academy of Agricultural Sciences (No. [2021] 50), Science and technology project of Shandong Education Department to MS (Grant no. J15LE02), China Postdoctoral Science Foundation funded project to MS (Grant no. 2018M632646), the Young Top-notch Talent Cultivation Program of Hubei Province for CT, and the National Natural Science Foundation of China (31770250).

**Acknowledgments**

We thank Zhixian Qiao of the Analysis and Testing Center at IHB for the technical support in RNA-seq analysis.

**Conflict of interest**

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

**Publisher’s note**

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

**Supplementary material**

The Supplementary Material for this article can be found online at: [https://www.frontiersin.org/articles/10.3389/fpls.2022.998082/full#supplementary-material](https://www.frontiersin.org/articles/10.3389/fpls.2022.998082/full#supplementary-material)
Lu, S., Wang, J., Chitnaz, F., Derbyshire, M., Geer, R., Gonzales, N., et al. (2020). CD/DSPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 48, D265–D268. doi: 10.1093/nar/gkaa991

Lysak, M. A., Koch, M. A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe brassicaeae. Genome Res. 15 (4), 516–525. doi: 10.1101/gr.3531105

Magali, L. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 30 (1), 325–327. doi: 10.1093/nar/gai3125

Nekrutkena, A. (2002). The K A/K s ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. Genome Res. 12 (1), 198–202. doi: 10.1101/gr.200901

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. (2013). Root-derived CLE glycopeptides control nodulation by direct binding to HARK receptor kinases. Nat. Commun. 4, 2191. doi: 10.1038/ncomms3191

Ostergaard, L., and King, G. J. (2008). Standardized gene nomenclature for the brassica genus. Plant Methods 4 (1), 1–10. doi: 10.1186/1746-4811-4-10

Patterson, A., Bowers, J., and Chapman, B. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc. Natl. Acad. Sci. U. S. A. 101 (26), 9903–9908. doi: 10.1073/pnas.0307901101

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33 (3), 290–295. doi: 10.1038/nbt.3122

Rojo, E., Sharma, V., Kovalova, V., Raikhel, N., and Fletcher, J. (2002). CLV3 is localized to the extracellular space, where it activates the arabidopsis CLAVATA receptor kinase. J. Biol. Chemistry 277, 327–327. doi: 10.1074/jbc.M110097200

Song, X., Guo, P., Ren, S., Xu, T., and Liu, C. (2013). Antagonistic peptide regulation of seed oil accumulation in Brassica napus. BMC Genomics 14 (5), 969–977. doi: 10.1186/1471-2164-14-5

Scheule, R. D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Systematics 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Schoof, H., Lenhard, M., Haeccker, A., Mayer, K. F. X., Jurgens, G., and Laut, T. (2000). The stem cell population of arabidopsis shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. Plant Cell 12 (6), 635–644. doi: 10.1105/tpc.002747

Schranz, M. E., Lysak, M. A., and Mitchell-Olds, T. (2006). The ABCs of comparative genomics in the brassicaceae: building blocks of crucifer genomes. Trends Plant Sci. 11 (11), 535–542. doi: 10.1016/j.tplants.2006.09.002

Song, X., Guo, P., Ren, S., Xu, T., and Liu, C. (2013). Antagonistic peptide technology for functional dissection of CLV3/ESR genes in arabidopsis. Plant Physiol. 161 (3), 1076–1085. doi: 10.1104/pp.111.211029

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., et al. (2018). A small peptide modulates stomatal control via abscisic acid in long-distance signalling. Nature 556 (7700), 235–238. doi: 10.1038/s41586-018-0099-2

Tatsuiki, K. (2006). A plant peptide encoded by CLV3 identified by in situ MALDI-TOF MS analysis. Sc. (New York N.Y.) 5788, 845–848. doi: 10.1126/science.1128439

Wahid, S., Xie, M., Sarfarz, S., Liu, J., Zhao, C., Bai, Z., et al. (2022). Genome-wide identification and analysis of Ariadne gene family reveal its genetic effects on agronomic traits of brassica napus. Int. J. Mol. Sci. 23, 6265. doi: 10.3390/ijms23112605

Wang, J., Replogle, A., Hussey, R., and Baum, T. (2011). Identification of potential host plant mimics of CLAVATAESR (CLE)-like peptides from the plant-parasitic nematode Heteroder a schachtii. Mol. Plant Pathol. 12 (2), 177–186. doi: 10.1111/j.1364-3703.2010.00660.x

Wang, X., Wu, J., Liang, J., Cheng, F., and Wang, X. (2015). Brassica database (BRAD) version 2.0: Integrating and mining brassicaceae species genomic resources. Database J. Biol. Database Curation 2015, bav093. doi: 10.1093/database/bav093

Xie, M., Zuo, R., Bai, Z., Yang, L., Zhao, C., Gao, F., et al. (2022). Genome-wide characterization of Serine/Arginine rich gene family and its genetic effects on agronomic traits of brassica napus. Front. Plant Sci. 13. doi: 10.3389/fpls.2022.829668

Yang, Y., Zhu, K., Li, H., Han, S., Meng, Q., Khan, S., et al. (2018). Precise editing of CLAVATA genes in brassica napus I. regulates multilocular silique development. Plant Biotechnol. J. 16 (7), 1322–1335. doi: 10.1111/pbi.12872

Ye, C., Wu, D., Mao, L., Jia, L., Qu, J., Lao, S., et al. (2020). The genomes of the allopolyhaploid chinochloa crus-galli and its progenitors provide insights into polyploidization-driven adaptation. Mol Plant 13 (9), 1298–1310. doi: 10.1016/j.molp.2020.07.001

Yuanmian, J. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature 474, 97–100. doi: 10.1038/nature09916

Zhang, L., Wu, S., Chang, X., Wang, X., Zhao, Y., Xia, Y., et al. (2020). The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress. Plant Cell Environ. 43 (12), 2847–2856. doi: 10.1111/pce.13898

Zhang, Y., Yang, S., Song, Y., and Wang, J. (2014). Genome-wide characterization, expression and functional analysis of CLV3/ESR gene family in tomato. BMC Genomics 15 (1), 827. doi: 10.1186/1471-2164-15-827

Zhao, C., Safdar, I. B., Xie, M., Shi, M., Dong, Z., Yang, L., et al. (2021). Mutation analysis of PHYTOENES DESATURASE 3 gene causes yellowish-white petals in Brassica napus. Crop J. 9, 1124–1134. doi: 10.1016/j.cj.2020.10.012

Zhao, C., Xie, M., Liang, L., Yang, L., Han, H., Qin, X., et al. (2022). Genome-wide association analysis combined with quantitative trait loci mapping and dynamic transcriptome unveil the genetic control of seed oil content in brassica napus I. Front. Plant Sci. 13. doi: 10.3389/fpls.2022.929197

Zhu, W., Guo, Y., Chen, Y., Wu, D., and Jiang, L. (2020). Genome-wide identification, phylogenetic and expression pattern analysis of GATA family genes in brassica napus. BMC Plant Biol. 20 (1), 543. doi: 10.1186/s12870-020-02572-2