Germination of soybean seed is the imminent vital process after sowing. The status of plumular axis and radicle determine whether soybean seed can emerge normally. Epicotyl, an organ between cotyledons and first functional leaves, is essential for soybean seed germination, seedling growth and early morphogenesis.

Epicotyl length (EL) is a quantitative trait controlled by multiple genes/QTLs. Here, the present study analyzes the phenotypic diversity and genetic basis of EL using 951 soybean improved cultivars and landraces from Asia, America, Europe and Africa. 3VmrMLM was used to analyze the associations between EL in 2016 and 2020 and 1,639,846 SNPs for the identification of QTNs and QTN-by-environment interactions (QEIs). A total of 180 QTNs and QEIs associated with EL were detected. Among them, 74 QTNs (ELS_Q) and 16 QEIs (ELS_QE) were identified to be associated with ELS (epicotyl length of single plant emergence), and 60 QTNs (ELT_Q) and 30 QEIs (ELT_QE) were identified to be associated with ELT (epicotyl length of three seedlings). Based on transcript abundance analysis, GO (Gene Ontology) enrichment and haplotype analysis, ten candidate genes were predicted within nine genic SNPs located in introns, upstream or downstream, which were supposed to be directly or indirectly involved in the process of seed germination and seedling development. Of 10 candidate genes, two of them (Glyma.04G122400 and Glyma.18G183600) could possibly affect epicotyl length elongation. These results indicate the genetic basis of EL and provides a valuable basis for specific functional studies of epicotyl traits.

KEYWORDS

geno-m-wide association analysis, single nucleotide polymorphism, candidate genes, 3VmrMLM, epicotyl length
Introduction

Epicotyl length (EL), an important complicated and agronomically trait, was significantly related to plant density and sowing depth of soybean (Camargos et al., 2019). EL exhibited the higher genetic variability at the early developmental stages of soybean, especially at V2 and V3 development stages (Matsuo et al., 2012). EL also affected plant height and yield of soybean (Hanyu et al., 2020). As a typical quantitative trait, EL, with relatively high heritability (more than 95%), was controlled by a few large-effect genes and a series of polygenes (Chaves et al., 2017). EL was significantly affected by environment, genotype their interactions (Chaves et al., 2017; Hanyu et al., 2020). Several studies showed that genetic and environmental variation approximately accounted for half of experimental observation. Although EL has been considered as the important feature of variety during the long-term soybean breeding, development of soybean cultivar with reasonable and stable EL through traditional selection method was still difficult (Chaves et al., 2017). It required evaluation in multiple environments over several years, and traditional selection method was expensive, time-consuming and labor-intensive (Chaves et al., 2017).

Molecular marker could effectively improve traditional selection efficiency by increasing the allele’s frequency of desirable quantitative trait loci (QTLs). Presently, linkage analysis and association analysis, were two major strategies utilized to identify QTLs of important traits in crops (Li et al., 2020; Liu et al., 2020; Wang et al., 2021). Segregating population based linkage analysis strategy is a well-known approach to obtain QTLs, followed by fine mapping using larger secondary population or other types of population with sufficient map resolution, then candidate genes could be cloned for functional characterization. (Dinka et al., 2007) mapped four additive QTLs for the length of hypocotyl in soybean. However, none of EL QTLs of soybean has been reported to date. Based on diverse germplasms, Genome-Wide Association Study (GWAS) take advantages of historical recombination events offered another strategy to effectively fine map QTL with rapid decay of linkage disequilibrium (LD) (Flint-Garcia et al., 2003). Due to the advances in next-generation sequencing (NGS) technologies or Chip with high-density SNPs, GWAS has been widely extensively used (Segura et al., 2012; Wang et al., 2016; Wen et al., 2017). As we know, there are frequently three genotypes for each marker in GWAS. Two effects should be estimated, while their polygene backgrounds should be controlled. In most GWAS methods, however, only one confound effect is estimated, while its polygene background is controlled. To solve this issue, recently, Li et al. (2022b) established a three-variance-component mixed linear model framework, 3VnrMLM, to identify QTNs, QTN-by-environment interactions (QEIs), and QTN-by-QTN interactions under controlling all the possibly polygene backgrounds.

Cytokinins and light can sometimes elicit similar morphological and biochemical responses. In the absence of light plant seedlings have long epi- or hypocotyls and appressed leaves with the plastid development blocked at the stage of etioplasts or amyloplasts. The 16 light-independent p6 homomorphogenesis (lip1) mutant of pea shows many of the characteristics normally associated with light-grown seedlings when grown in complete darkness, such as expanded leaves, a short epicotyl and partially developed chloroplast (Frances et al., 1994). It appears that the effect of cytokinin on the growth of the axis of young hypogeal (e.g., Arabidopsis) and epigeal (e.g., pea) seedlings is similar. The phenotype of wild-type Arabidopsis plants following cytokinin treatment is similar to that of the ampl1 mutant of Arabidopsis, suggesting that light and cytokinin act through a common signaling pathway (Chory et al., 1994; Seyedi et al., 2001). Genetic analysis of Arabidopsis has provided unequivocal evidence that the brassinosteroids (BRs) are essential phytohormones (He et al., 2003). Brassinolide (BL), an end product of campesterol oxidation is required for the regulation of cell elongation, stress response, male fertility, pigment biosynthesis, and numerous other developmental and physiological responses in higher plant (Grove et al., 1979), The Arabidopsis CYP90A1 (constitutive
photomorphogenesis and dwarfism, CPD) has been identified to functions as the C-23 hydroxylase in the biosynthetic pathway of brassinosteroids, and cpd mutant exhibited the most pronounced effect in dwarf phenotype than another five cytochrome P450 mutants. The biosynthetic model of BRs has been clearly identified in Arabidopsis, we supposed a similar model, It has been proved in 1998 that the transcription of Arabidopsis CYP90A1 was negatively controlled by exogenous brassinolide (Mathur et al., 1998).

To address above mentioned issues, 951 landraces and cultivars selected from Chinese primary core collection in the Chinese National Soybean GeneBank (CNSGB), were phenotyped for EL in 2016 and 2020, and genotyped by 1,639,846 SNPs in order to identify QTNs, QEIs, and their candidate genes for EL in soybean.

Materials and method

Plant materials, filed trials and epicotyl length evaluations

To construct a diversity panel of EL, a total of 951 landraces was selected from more than 20,000 samples, which delegated much of the representatives of diversity of the collection at the Chinese National Soybean GeneBank (CNSGB). These tested materials were planted with the single row plots (3-m long and 0.35-m between rows), which was performed with the completely randomized design and three replications in Sanya, Hainan China in 2016 and 2020.

A total of 3 randomly selected plants from each plot were phenotyped for EL by measuring the distance between the cotyledonary knot and the unifoliate leaves pair knot using vernier caliper.

DNA isolation and genome sequencing

The genomic DNA of each tested samples were isolated from fresh leaves of a single plant, and then resequenced. Sequencing libraries were constructed based on TrueSeqNano® DNA HT sample preparation Kit (Illumina USA), and index codes were added to attribute sequences to each accession according to the method described by (Li et al., 2020a). The illumina Hiseq X platform was used to analyze the libraries of these samples. A total of 10.58 Tb raw sequences with 150-bp read length, were obtained. After sequence quality filtering, the clean read of all tested samples, were aligned to soybean reference genome via Short Oligonucleotide Alignment Program 2 (SOAP2) software. The SNPs were calling based on MAF ≥ 0.05. The genotype was regarded as heterozygous if the depth of minor allele/the total depth of the sample was more than 1/3.

Population structure evaluation and linkage disequilibrium (LD) analysis

The population structure of GWAS panel were evaluated based on principle component analysis (PCA) programs of Software package GAPIT (Lipka et al., 2012). LD was called with SNP (MAF ≥ 0.04 and missing data ≤ 10%) based on TASSEL version 3.0 (Bradbury et al., 2007).

Association analysis of epicotyl length of soybean

A total of 1,639,846 SNPs from 951 landraces samples were utilized to detect association signals of EL in soybean. Imputed genotype of total sample panel was first transformed in to *.fam, *.bed, and *.bim format, ELS and ELT in two different environments were adopted as phenotype, evolutionary population structure encoded as B (Landrace) and C (Improved cultivar), and kinship were employed as covariates for multi-environment joint analysis with significant level of 0.01 using HHVmrMLM software of Li et al. (2022b); Li et al. (2022c). Linkage disequilibrium (LD) of 250kb up- and down-stream of significantly associated SNP were calculated by PLINK1.9, and threshold of regional average LD > 0.9 was used to define credible associated region. Functional annotation of candidate genes was performed based on annotation by phytozome (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1).

Definition and verification of candidate genes

Then SNP variations in the coding region of candidate genes were analyzed to screen candidate genes with mutation type of nonsynonymous, stoploss, stopgain, or alternative splicing. To further screen candidate genes, fixation index (FST) was calculated by published genome sequences data of 2214 soybeans (Li et al., 2022d) using vcftools (0.1.13) with window size of 100bp, and coding regions with FST ≥ 0.6 were regarded as potential domestication gene (Song et al., 2013). Subsequently, spatial and temporal expression of candidates were analyzed using publicly available soybean transcriptome integration dataset (Yu et al., 2022). Functional annotations of all candidate genes were performed based on the SoyBase database (http://www.soybase.org) and the Kyoto Encyclopedia of gene and genomes (KEGG).

Haplotype analysis

Gene region were defined using *.gff, regional genotype of hapmap diploid were extracted from imputed genotype,
then haplotypes were inferred based on regional genotype classified according to its location relative to the gene structure. Significance of traits between different haplotypes were performed by Kruskal-Wallis ($P<0.01$) (Theodorsson, 1986). Haplotype TCS network was inferred using PopART (Bandelt et al., 1999; Clement et al., 2002; French et al., 2014). Geographic mapping of different haplotypes was performed using R scripts.

**Results**

**Distribution of the landraces used in the experiment**

Globally, the improved cultivars selected for the experiment mainly comes from America and Asia, with few from Europe and Africa. Landraces were all obtained from Asia (Figure 1). To better understand the genetic architecture of these germplasms, geographic distribution and ecological types were taken into account for classification. Both domestic and foreign varieties can be divided into southern (SR), northern (NR) and central (HR) varieties, namely domestic varieties (SR, HR, NR) and foreign varieties (WDD_SR, WDD_HR, WDD_NR). Domestic NR sources are the maximum, and foreign WDD_HR varieties account for more than half of the total foreign varieties (Figure 2A and Table S1). According to ecological types, domestic cultivars can be divided into northeast spring type (NESp), northern spring type (NSp), Huang-huai spring type (HSp), Huang-huai summer type (HSu), Southern spring type (SSp), Southern summer type (SSu) and Southern autumn type (SAu), with NESp ranking the first place. The selected foreign varieties were mainly divided into spring type (WDD_Sp) and summer type (WDD_Su), and the quantity of WDD_Sp was twice as much as WDD_Su (Figure 2B and Table S2). These results demonstrated that nearly 80% of the varieties used in the experiment came from China, and 60% of the varieties obtained abroad were spring varieties in the central region.

**Statistical analysis for inflorescence length of the association panel**

The EL of 951 landraces in Sanya, Hainan China in 2016 and 2020, were evaluated, respectively. The skewness and kurtosis of EL the three environments were less than ±1, which exhibited a continuous variation and the near normal distribution (Table S3). Therefore, EL of the association panel in this study, were appropriate.

**Distribution of SNPs and analysis of mapping population**

Based with the frequency > 0.05 as the minor allele and the missing data less than 0.03, a total of 1,639,846 single nucleotide polymorphisms (SNPs) were unevenly distributed on 20 chromosomes of soybean genome. with a density of 578.8 bp per SNP on average, and varied from 337.3bp~1334.4bp per SNP. In detail, there were 168,498 SNPs on Chr1 with the highest density (337.3bp/SNP), 31,650 SNPs on Chr5 with lowest density (1334.4bp/SNP). (Figure 3). Based on these SNPs, principal component analysis and phylogenetic analysis were performed on the association panel. The results showed that the first PCs explained 24.52% of the genetic variation, the 951 varieties were divided into two categories with apparent discrepancy of genetic relatedness (Figure 4). For a preferably clearer study of epicotyl traits, they were also divided into two categories, ELS and ELT. Statistical methods were used to test that ELS and ELT showed normal distribution in different environments among varieties (Figure 5).
Quantitative trait nucleotide associated with epicotyl length-related traits by GWAS

QTN (Q) and QTN-by-environment interaction (QEIs) detection method in the 3VmrMLM was used to analyze SNP-trait associations in two EL two-environment datasets, ELS (2016 and 2020) and ELT (2016 and 2020). A total of 180 QTNs and QEIs associated with epicotyl length were detected. Among them, 74 QTNs (ELS_Q) and 16 QEIs (ELS_QE) were identified to be associated with ELS, and 60 QTNs (ELT_Q) and 30 QEIs (ELT_QE) were identified to be associated with ELT.

Figure 6 Of these, three sites (Gm_09_28400545, Gm_11_31100989, Gm_19_557643) could be found in all these four result datasets (Table S4).

Prediction of candidate genes for epicotyl length traits

We performed candidate gene prediction analyses with peak SNP of ±100 kb based on the physical locations of 180 SNPs associated with epicotyl length. A total of 1945 genes were included in these regions (Table S4). Functional annotation of
1945 genes were completed by using Arabidopsis annotation information. Site contribution rate, transcription abundance of candidate genes in epicotyl of two representative soybean germplasms including cultivar Williams 82 with a long epicotyl of 3.93 cm and cultivar Jack with a short epicotyl of 2.13 cm were analyzed using publicly available soybean transcriptome integration dataset (Yu et al., 2022). By comparing the epicotyl lengths of Williams 82 and Jack, a very significant difference was found (Figures 7A, B). Based on the transcriptome data of epicotyls from Williams 82 and Jack, 585 out of 1945 genes were not expressed in both epicotyls of Williams 82 and Jack, 94 genes were expressed only in the epicotyl of Jack and 60 genes were expressed only in the epicotyl of Williams 82. A total of 1206 genes were expressed in both epicotyls of Williams 82 and Jack, of them, 157 genes were significantly differentially expressed in Williams 82 and Jack. Combined with Arabidopsis annotation information, 103 genes were identified as potentially candidate genes for epicotyl length (Table S5, Figure 7C). These differentially expressed genes in long and short epicotyl cultivars might be related to the length of epicotyl of soybean.

To further elucidate whether the differentially expressed genes were related to the length of the epicotyl, GO enrichment analysis was performed (http://amigo.geneontology.org/). GO enrichment analysis showed all genes were assigned to one of three GO categories: biological process (BP), molecular function (MF), and Cellular component (CC) (Figure 8).

Further, haplotype analysis was performed for 103 potentially candidate genes screened by the above analysis. Epicotyl

In order to determine the role of the selected potential genes in soybean epicotyl growth, 22 potential candidates were screened by combining gene GO annotation and transcriptome differential expression analysis, and referring to Arabidopsis annotation information. Haplotype analysis identified 10 significantly different genes. The Hap1 and Hap2 of Glyma.01G005900 in different years of ELS (P = 0.0039) and ELT (P = 0.039) showed...
extremely significant differences (P<0.01). The Hap1 and Hap3 of Glyma.18G183600 (2016_ELS P=1.1e-09; 2020_ELS P=0.00013; 2016_ELT P=1.2e-08; 2020_ELT P=0.69), Glyma.18G185300 (2016_ELS P=0.0083; 2020_ELS P=0.00013), exhibited extremely significant differences (P<0.01), while the Hap1 and Hap3 of Glyma.18G183600 (2016_ELS P=1.1e-09; 2020_ELS P=0.00013) in different years of ELS had a very significant difference in 2016 (P<0.01), but there was no significant difference in 2020. The candidate gene Glyma.18G185300 showed a very significant difference in the two years of EL (P<0.01), and the ELT revealed as significant difference in 2016 (2016_ELT P=0.02) and showed a very significant difference in 2020 (2020_ELT P=0.0031) (Figure 9).

Meanwhile, we counted the variation sites of 10 gene haplotypes (Table S7). The results demonstrated that Glyma.04G122400, Glyma.10G031900 and Glyma.18G183600 exist in exon variation sites, of which Glyma.04G122400 and Glyma.18G183600 exist non-synonymous mutations, hence, we speculate that Glyma.04G122400 and Glyma.18G183600 are candidate genes for epicotyl differences. At the same time, we combed the geographical origin of the two gene haplotypes and the distribution of variety characteristics. From the geographical distribution, we could see that Hap1, Hap2, Hap3 and Hap4 haplotypes of the two candidate genes were absolutely dominant in the selected varieties. In terms of ecological characteristics of cultivars, Hap1 and Hap2 haplotypes of the two genes accounted for more than Landrace haplotypes in improved cultivars (Figure 10).

We predicted ten plant growth-related genes, namely Glyma.03G142200 (Ribosomal protein S10p/S20e family protein), Glyma.04G122400 (DCD domain protein), Glyma.04G145000 (nuclear factor Y, subunit B13), Glyma.10G0319000 (indole-3-acetic acid 7), Glyma.10G056000 (SAUR-like auxin-responsive protein family), Glyma.13G270800 (ubiquitin-conjugating enzyme 35), Glyma.17G005900 (Pollen Ole e 1 allergen and extensin family protein), Glyma.17G18500 (NAC domain containing protein 83), Glyma.18G183600 (far-red elongated hypocotyl 1), and Glyma.18G255300 (thioredoxin H-type 5). These results suggest that soybean epicotyl length may be regulated by multiple signaling pathways (Table 1). Additionally, none of these 10 candidates were differentiated among wild soybean, landrace and improved cultivar (Figure S1).

**Discussion**

As an important feature of soybean variety, many studies indicated that EL affected 43.12% of seeds germination and 57.12% of seedlings emergence for soybean (Hanyu et al., 2020) estimated the genotypic determination coefficient of EL was more than 80% regardless of the evaluation period. (Matsuo et al., 2012) also obtained similar results. The genotypic determination coefficient was significantly related to inheritability, thus, it made the inference about genotypes possible (Vasconcelos et al., 2012; Hanyu et al., 2020). Through screening a large enough and reasonable gene database from more than 20,000 varieties, the SNPs and potential genes related to epicotyl traits were analyzed by GWAS technology. By elucidating the epicotyl related loci, it has a potential role in the study of
FIGURE 7  
Epicotyl length of Williams82 and Jack and expression analysis of 103 candidate genes. (A) Epicotyl phenotype of W82 and Jack (B) Epicotyl Length Analysis of W82 and Jack (C) Transcriptome alignment of 103 candidate genes.

FIGURE 8  
Functional categories of the genes in 100kb flanking regions around peak SNPs.
early seed germination, seedling germination and stem strength of soybean.

To date, many seedling crop traits have been studied and elucidated, but epicotyl traits have been largely ignored and poorly studied. Four of Chr.2, Chr.4, Chr.7 and Chr.10 were identified in the F2 population of adzuki bean "Tokei1121" (T1121, long epimorph) and cultivar "Erino167" (ordinary ectomorph) with EL associated SNP (Mori et al., 2021). There are no reports on EL-related SNP sites in other plants. The genetic mechanism of the hypocotyl length trait (HL) has been extensively studied. SNP mapping of soybean root-related traits at seedling stage revealed that HL is regulated by multiple additive genes. Seven QTLs in HL associated with seedling photomorphology were identified by using recombinant inbred (RIL) populations obtained from biparental crosses between Patagonia (Pat) and Colombia (COL0) (Matsusaka et al., 2021). Compound spacer and epitaxial array localization methods were also used to identify HL loci associated with light-responsive quantitative traits (Woyn et al., 2004). To pinpoint trait-associated loci, the combination of GWAS and
transcriptome can be used to identify major genes affecting HL (Luo et al., 2017). These studies suggest that hypocotyl play a role in root growth and photomorphological responses. (Huang et al., 2006) studied the regulatory effect of brassinolide on epicotyl under low temperature conditions by proteomics. How xylan content in the gravitational bending direction of the epicotyl of adzuki bean affects its internal xylan content (Ikushima et al., 2008). Inhibitory effect of red light of the active form of phytochrome (Pfr) on epicotyl elongation in pea seedlings (Okoloko et al., 1970). These indicate that epicotyl play a non-negligible role in a variety of crops, especially dicotyledonous crops. Faced with this situation, this study used the soybean EL association panel to analyze the natural variation of epicotyl length and the related genetic structure, and analyzed the Hypothetically revealing a set of candidate genes controlling epicotyl development by GWAS analysis is undoubtedly a key step in filling in the relevant loci for epicotyl trait mapping.

**Putative genes involved in epicotyl length**

Through the Arabidopsis annotation information, candidate gene phenotype contribution rate, and combining with Yu et al. (2022) Williams 82 and Jack transcriptome results of extremely different genes, we screened 22 potential genes from 103 hypothetical genes. These genes are located in SNP peak within 100Kb.10 significantly different candidate genes were identified by haplotype analysis, these genes were genotyped significantly and distinctly of ELS and ELT. *Glyma.03G142200* is a Ribosomal protein S10p/S20e family protein, proteins involved in photosynthesis (Bah et al., 2010). Wycoff found that a lectin protein, analogous to ribosomal proteins, is detected in roots, hypocotyls and leaves and involved in soybean nodule formation (Wycoff et al., 1997).

*Glyma.04G122400* DCD (Development and Cell Death) domain protein, thought to be involved in the hypersensitive response and programmed (Ludwig and Tenhaken, 2001, Enhaken et al., 2005). In previous studies, DCD domain proteins was believed to be involved in extracellular matrix or cytoskeleton proteins involved in growth and differentiation processes (Ichinose et al., 1990, Massimiliano et al., 2007).

| Chr. | Physical position (bp) | Gene model | Trait | $R^2$ (contribution rate) | Pvalue | Functional annotation |
|------|------------------------|------------|-------|--------------------------|--------|-----------------------|
| 3    | 35863419               | Glyma.03G142200 | ELT_Q | 0.5768                   | 6.09167E-21 | Ribosomal protein S10p/S20e family protein |
| 4    | 15439303               | Glyma.04G122400 | ELT_Q | 0.4336                   | 8.01377E-07 | DCD (Development and Cell Death) domain protein |
| 4    | 26351924               | Glyma.04G145000 | ELS_Q | 0.2279                   | 4.20387E-22 | nuclear factor Y, subunit B13 |
| 10   | 2738580                | Glyma.10G031900 | ELS_Q | 0.5234                   | 1.14306E-11 | indole-3-acetic acid 7 |
| 10   | 5143580                | Glyma.10G056000 | ELS_Q | 0.5294                   | 5.84009E-32 | SAUR-like auxin-responsive protein family |
| 13    | 37248488              | Glyma.13G270800 | ELS_Q | 1.5012                   | 7.02497E-35 | ubiquitin-conjugating enzyme 35 |
| 17    | 637613                | Glyma.17G055900 | ELS_Q | 0.5942                   | 5.10164E-10 | pollen Ole e 1 allergen and extensin family protein |
| 17    | 23689587              | Glyma.17G185000 | ELS_Q | 0.7863                   | 4.96905E-13 | NAC domain containing protein 83 |
| 18    | 44381201              | Glyma.18G136300 | ELS_Q | 2.1064                   | 1.12718E-32 | far-red elongated hypocotyl 1 |
| 18    | 44381201              | Glyma.18G185300 | ELS_Q | 2.1064                   | 1.12718E-32 | one helix protein |

TABLE 1 Gene based association of candidate genes.
principal component of the ABA-and auxin dependent reactions during post-germination seed growth (Belin et al., 2009). Glyma.13G270800 ubiquitin-conjugating enzyme 35. Previous studies have shown that ubiquitination plays important roles in plant abiotic stress responses, Protein ubiquitinations play crucial roles for numerous cellular processes such as cell growth, development, and response to diverse biotic and abiotic stresses. (Takahashi et al., 2009; Zhou et al., 2010). The ubiquitin-depen-dent protein degradation pathway is involved in photo-morphogenesis, hormone regulation, familial homeosis, senescence, and pathogen defense (Suzuki et al., 2002; Devoto et al., 2003).

Glyma.17G185000 NAC domain containing protein 83, The NAC (for NAM-ATAF1/2-CUC2) transcription factors constitute one of the largest transcription factor families in plant genomes (Ooka et al., 2004; Olsen et al., 2005b). Roles of many NAC transcription factors have been demonstrated in diverse developmental processes and plant responses to biotic and abiotic stresses, such apical meristem formation (Hibara et al., 2003), cell cycle control (Kim et al., 2006), AtNAC2 functioning in root development (He et al., 2005), cell divi-sion (Riechmann et al., 2000; Kim et al., 2006), NTHM2 inte-grates auxin and salt signals in regulating Arabidopsis seed germination (Park et al., 2011), In Arabidopsis thaliana, 105 genes are predicted to encode NAC proteins (Ooka et al., 2004). Song et al. study found The highly homologous NAC transcription factors ANAC060, ANAC040 and ANAC089 regulate important transitions in the early phases of plant development. All three genes play a role in the interplay between the environment and the developmental switch that results in germination and/or seedling development (Song et al., 2022). For germination and seedling development to occur, the protein has to be released from the membrane, which for ANAC089 was shown to be directly affected by changes in the cellular redox status (Albertos et al., 2021).

Glyma.18g183600 far-red elongated hypocotyl 1, Phytochrome A (phyA) is the primary photoreceptor for mediating the far-red high irradiance response in Arabidopsis thaliana. FAR-RED ELONGATED HYCOTYL1 (FHY1) and its homolog FHY1-LIKE (FHL) define two positive regulators in the phyA signaling pathway (Shen et al., 2009). Most abundant in young seedlings in the dark,encodes FHY1 protein that mediates the transfer of phytochrome A (phyA) to the nucleus. Phytochrome A (phyA) acts as red and far red (FR) sensing photoreceptors to regulate plant growth and development (Helizon et al., 2018). Multiple metabolic pathways are required to regulate the length of soybean epicotyl (Clouse et al., 1992; Hao et al., 2014).

Glyma.18G185300 one helix protein, The cellular functions of two Arabidopsis (Arabidopsis thaliana) one-helix proteins, OHP1 and OHP2 (also named LIGHTHARVESTING-LIKE2 [LIL2] and LIL6, respectively, because they have sequence similarity to light-harvesting chlorophyll a/b-binding proteins), OHP1 and OHP2 play an essential role in chloroplast development as well as in vegetative growth, The photosynthetic capacity of ohp1-1 and ohp1-2 mutants also was decreased significantly (Myouga et al., 2018). The protein is localized to the thylakoid membrane and its transcript is transiently induced by exposure to high light conditions. increased expression of OHP1 is observed under light stress (Jansson et al., 2000), may constitute a novel mechanism of photoprotection in the plant photosynthetic apparatus (Psencik et al., 2020).

We speculate that traits during soybean domestication are gradually selected, and the priority traits are yield-related traits, such as seed size, oil content, and protein content (Wang et al., 2020). The epicotyl length involved in this study is not a major direct yield trait and therefore demonstrated weak signal of domestication selection.

In general, It is certain that most of the above candidate genes are related to the regulation of light and temperature. For example, the candidate gene Glyma.18G183600 is a phytochrome A (phyA) gene, which is the main photoreceptor mediating the far-red high-irradiation response in Arabidopsis. Cellular function of Glyma.18G185300 with sequence similarity to light-harvesting chlorophyll a/b binding protein, Glyma.03G142200 is a protein involved in photosynthesis, and the analysis results show that they are all involved in the growth and development of soybean epicotyl. This is consistent with the results that soybean epicotyl length is greatly affected by different environments. These results can be reflected from the haplotype analysis of ten candidate genes, which can be reflected in the significant differences in different environments (Figure 9).epicotyl However, further functional verification is needed to clarify the whole mechanism of action.

More importantly, since the epicotyl is located in the country of cotyledons and true leaves, it is not only involved in seed germination and seedling growth, but also affects early morphogenesis of seedlings. Understanding and regulating the molecular regulatory network of epicotyl length has important guiding significance for crop breeding.

**Data availability statement**

All whole genome sequencing data in this study have been deposited in the NCBI Sequence Read Archive under accession number PRJNA681974.

**Author contributions**

HH, and ML conceived the study and contributed to population development. YC, HW, JW, and BG contributed to phenotypic evaluation. HG, and HR analyzed the data. MY, and HG contributed to genotyping. LQ, and YH contributed to experimental design and writing the paper. All authors contributed to the article and approved the submitted version.
Funding

This study was financially supported by the National Natural Science Foundation of China (32172005), Agricultural Science and Technology Innovation Program (ASTIP) of Chinese Academy of Agricultural Sciences.

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.

References

Albertos, P., Tatematsu, K., Mateos, I., Sánchez-Vicente, I., Fernández-Arbazaar, A., Nakabayashi, K., et al. (2021). Redox feedback regulation of ANAC089 signaling alters seed germination and stress response. Cell Rep. 35 (11), 109263. doi: 10.1016/j.celrep.2021.109263

Albertos, P., Tatematsu, K., Mateos, I., Sánchez-Vicente, I., and Lorenzo, O. (2021). Redox feedback regulation of ANAC089 signaling alters seed germination and stress response. Cell Rep. 35 (11), 109263. doi: 10.1016/j.celrep.2021.109263

Bab, A. M., Sun, H., Fei, C., Zhou, J., Dai, H., Zhang, G., et al. (2010). Comparative proteomic analysis of typha angustifolia leaf under chromium, cadmium and lead stress. J. Hazard Mater. 184 (1-3), 191–203. doi: 10.1016/j.jhazmat.2010.08.023

Ballif, J., Endo, S., Kotani, M., Macadam, J., and Wu, Y. (2011). Over-expression of HAP3b enhances primary root elongation in arabidopsis. Plant Physiol. Bioch. 49 (6), 579–583. doi: 10.1016/j.plaphy.2011.01.013

Bandelt, H., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16 (1), 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Belin, C., Megies, C., Hauersona, E., and Lopez-Molina, L. (2009). Abscisic acid represses growth of the arabidopsis embryonic axis after germination by enhancing auxin signaling. Plant Cell. 21 (8), 2253–2268. doi: 10.1105/tpc.109.077702

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23 (19), 2633–2635. doi: 10.1093/bioinformatics/btm308

Camargos, T., Campos, N., Alves, G., Ferreira, S., and Matsuo, D. (2019). The effect of soil volume, plant density and sowing depth on soybean seedlings characteristics. Agron. Sci. Biotech. 5 (2), 47. doi: 10.33158/ASB.2019%52p47

Chaves, M. V. A., Silva, N. S., Silva, R. H. O., Jorge, G. L., Silveira, I. C., Medetos, L. A., et al. (2017). Genotype x environment interaction and stability of soybean cultivars for vegetative-stage characters. Genetics Mol. Biol. 40 (3). doi: 10.1590/1415-4762-gmb.2016-2976

Chory, J., Reinecke, D., Sim, S., Washburn, T., and Brenner, M. (1994). A role for ERF1 in etiolation in arabidopsis. Plant Physiol. 104 (2), 339–347. doi: 10.1104/pp.104.2.339

Clément, M., Snell, Q., Walker, P., Posada, D., and Crandall, K. (2002). TCS: Estimating gene genealogies. parallel distributed Process. symposium Int. Proc. 2, 184.

Clouse, S. D., Zurek, D. M., McMannis, T. C., and Baker, M. E. (1992). Effect of brassinolide on gene expression in elongating soybean epicycles. Plant Physiol. 100 (3), 1377–1383. doi: 10.1104/pp.100.3.1377

Contreras-Soto, R. I., Mora, F., Lazzari, F., de Oliveira, M. A. B., Scapim, C. A., and Schuster, I. (2017). Genome-wide association mapping for flowering and maturity in tropical soybean: implications for breeding strategies. Breed Sci. 67 (5). doi: 10.3839/fpls.2017.01222

Depuydt, T., and Vandepoele, K. (2021). Multi-omics network-based functional annotation of unknown arabidopsis genes. Plant J. 108 (4), 1193–1212. doi: 10.1111/tpj.15507

Devoto, A., Muskett, P. R., and Shirrau, K. (2003). Role of ubiquitination in the regulation of plant defense against pathogens. Curr. Opin. Plant Biol. 6 (4), 307–311. doi: 10.1016/s1369-5266(03)00060-8

Dinka, S. J., Campbell, M. A., Demers, T., and Raizada, M. N. (2007). Predicting the size of the progeny mapping population required to positionally clone a gene[J]. Genetics 176 (4), 2035–2054. doi: 10.1534/genetics.107.074377

Enshaker, R., Doerk, T., and Bork, P. (2005). DCD – a novel plant specific domain in proteins involved in development and programmed cell death. BMC Biol. 5, 169. doi: 10.1186/1471-2105-5-169

Flint-Garcia, S. A., Thornberry, J. M., and Th, B. E. (2003). Structure of linkagedisequilibrium in plants. Annu. Rev. Plant Biol. 54 (4), 357–374. doi: 10.1146/annurev.arplant.54.031902

French, S., White, M. J., Edgerton, M. D., Jones, A. M., Elliott, R. C., and Thompson, W. F. (1992). Initial characterization of a pea mutant with light-independent photomorphogenesis. Plant Cell. 4 (12), 1519–1530. doi: 10.1105/tpc.4.12.1519

French, N., Yu, S., Biggs, P., Holland, B., Fearnhead, P., Binney, B., et al. (2014). “Evolution of campylobacter species in new Zealand,” in Campylobacter ecology and evolution. Eds. S. K. Sheppard and G. Meric., 221–240. ISBN-978-1-908230-36-2.

Grove, M. D., Spencer, G. F., and Rohwereder, W. K. (1979). Brassinolide, a plant growth-promoting steroid isolated from brassica napa pollen[J]. Nature 281. doi: 10.1038/281216a0

Hanyu, J., Costa, S., Concon, P., and Matsu, D. (2020). Genetic parameters estimate and characters analysis in phenotypic phase of soybean during two evaluation periods. Agron. Sci. Biotech. 6, 1–12. doi: 10.33158/ASB.r104.v6.2020

Hao, H. P., He, Z., Li, H., Shi, L., and Tang, Y. D. (2014). Effect of root length on epistatic dormancy release in seeds of paeonia ludlowii. Tibetan peony. Ann. Bot. 113 (3), 443–452. doi: 10.1093/arb/mct273

He, J. X., Fujukawa, S., Li, T. C., Kang, S. G., Seto, H., Takatsuto, S., et al. (2003). Sterols regulate development and gene expression in Arabidopsis. Plant Physiol. 131 (3), 1258–1269. doi: 10.1104/pp.104.14605

Helimon, H., Roeder-Dalton, J., Gasch, P., von Horsten, S., Essel, L. O., and Zeidler, M. (2018). Arabidopsis phytochrome a nuclear translocation is mediated by a far-red elongated hypocotyl 1-importin complex. Plant J. 96 (6), 1255–1268. doi: 10.1111/tpj.14107

He, X. J., Mu, R. L., Cao, W. H., Zhang, Z. G., Zhang, J. S., and Chen, S. Y. (2005). AINAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. Plant J. 44 (6), 903–916.

Hibara, K., Takada, S., and Tasaka, M. (2003). CUC1 gene activates the expression of SAM-related genes to induce adventitious shoot formation. Plant J. 36 (5), 687–696.

Huang, B., Chiu, C. H., Chen, S. L., Juan, H. F., and Chen, Y. M. (2006). A sterols regulate development and gene expression in Arabidopsis. Annu. Rev. Plant Biol. 57 (1), 293–328. doi: 10.1146/annurev.arplant.57.032905.105519

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.2022.1033120/full#supplementary-material
multiple categories for soybean genotypes. Acta Sci. Agron. 34 (3), 265–273. doi: 10.4025/acsagron.v34i3.14015

Matsuoka, D., Fujita, D., Sanchez, D. H., and Botto, J. F. (2021). Ultra-High-Density QTL marker mapping for seedling photomorphogenesis mediating Arabidopsis establishment in southern patagonia. Front. Plant Sci. 12. doi: 10.3389/fpls.2021.77732

Mori, M., Maki, K., Kawahata, T., Kawahara, D., Kato, Y., Yoshida, T., et al. (2021). Mapping of QTLs controlling epicotyl length in adzuki bean (Vigna angularis). Breed Sci. 71 (2), 208–216. doi: 10.1270/jsbbs.2009.3

Myouga, F., Takashashi, K., Tanaka, R., Nagata, N., and Shinozaki, K. (2018). Stable accumulation of photosystem II requires ONE HELIX PROTEIN (OHP1) of the light harvesting-like family. Plant Physiol. 176 (3), 1672–1686. doi: 10.1104/pp.17.01782

Okoloko, G. E., Lewis, L. N., and Reid, B. R. (1970). Changes in nucleic acids in phytochrome-dependent elongation of the alaska pea epicotyl. Plant Physiol. 46 (5), 660–665. doi: 10.1104/pp.46.3.660

Olsen, A. N., Ernst, H. A., Leggio, L. L., and Skriver, K. (2005). DNA-Binding specificity and molecular functions of NAC transcription factors. Plant Cell 16 (9), 785–797. doi: 10.1105/tpc.105.03160

Okita, H., Toh, K., Doki, K., Nagaia, T. S., and Kikuchi, S. (2004). Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. DNA Res. 10 (6), 239–247. doi: 10.1093/dnares/10.6.239

Park, J., Kim, Y. S., Kim, S., Jung, J. H., Woo, J. C., and Park, C. M. (2011). Integration of auxin and salt signals by the NAC transcription factor NTM2 during seed germination in Arabidopsis. Plant Physiol. 156 (2), 537–549. doi: 10.1104/pp.111.177071

Pienck, J., Hey, D., Grimm, B., and Lokstein, H. (2020). Photoprotection of photosynthetic pigments in plant one-helix protein 1/2 heterodimers. J. Phys. Chem. 111 (23), 9387–9392. doi: 10.1021/acs.jpclett.0c00266

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., et al. (2000). Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science 290 (5499), 2105–2110. doi: 10.1126/science.290.5499.2105

Segura, V., Viljhummsdottir, B. J., Platt, A., Korte, A., Seren, U., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44 (7), 825–830. doi: 10.1038/ng.2314

Seyedi, M., Selastam, E., Timko, M. P., and Sundquist, C. (2001). The cytokinin 2-spermine synthetase causes partial reversal to skotomorphogenesis and induces formation of prolamellar bodies and prochlorophyllide57 in the hpi mutant of pea. Physiol. Plant 112 (2), 261–272. doi: 10.1034/j.1399-3054.2001.00013.x

Shen, Y., Zhou, S., Feng, S., Li, J., Tan-Wilson, A., Qu, L. J., et al. (2009). Phytochrome a mediates rapid red light-induced phosphorylation of arabidopsis FAR-RED ELONGATED HYPOCOTYL1 in a low fluence response. Plant Cell 21 (2), 494–506. doi: 10.1105/tpc.108.061259

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Ficks, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8 (1), e54985. doi: 10.1371/journal.pone.0054985

Song, S., Willems, L., Jiao, A., Zhao, T., Eric, S. M., and Leinie, B. (2022). The membrane associated NAC transcription factors ANAC060 and ANAC040 are functionally redundant in the inhibition of seed dormancy in arabidopsis thaliana. J. Exp. Bot. 73 (16), 5514–5528. doi: 10.1093/jxb/erc232

Sui, M., Jia, J., Li, H., Zhan, Y., Luo, J., Teng, W., et al. (2020). Identification of loci and candidate genes analyses for tocopherol concentration of soybean seed. Front. Plant Sci. 11. doi: 10.3389/fpls.2020.539460

Suzuki, G., Yanagawa, Y., Krok, S. F., Matsu, M., and Deng, X. W. (2002). Arabidopsis COP10 is a ubiquitin-conjugating enzyme variant that acts together with COP1 and the COP9 signalosome in repressing photomorphogenesis. Genes Dev. 16 (5), 554–559. doi: 10.1101/gad.964602

Takahashi, H., Noroaka, S., Seki, M., Shinozaki, K., Endo, Y., and Sasaki, T. (2009). A simple and high-sensitivity method for analysis of ubiquitination and polyubiquitination based on wheat cell-free protein synthesis. BMC Plant Biol. 9 (1), 39. doi: 10.1186/1471-2229-9-39

Theodorsson, N. E. (1986). Kruskal-Wallis test: BASIC computer program to perform non-parametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. Comput. Math. Prog. Bioc. 21 (1), 57–62. doi: 10.1007/BF02670807

Wang, S. B., Feng, J. Y., Ren, W. L., Huang, B., Zhou, L., Wen, J. Y., et al. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci. Rep. 6, 24444. doi: 10.1038/srep24444

Wang, X., Gao, P., Xin, M., Wang, Y., Chen, X., Zhao, A., et al. (2021). Genome-wide association study identifies QTL for thousand grain weight in winter wheat under normal- and late-sown stressed environments. Theor. Appl. Genet. 134 (1), 143–157. doi: 10.1007/s00122-020-03687-w
Overexpression of the poplar NF-YB7 transcription factor confers drought tolerance and improves water-use efficiency in *Arabidopsis thaliana*.

Yu, J., Pressoir, G., Briggs, W. H., Yroh, B. I., Yamashita, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38 (2), 203–208. doi: 10.1038/ng1702

Yu, Y., Zhang, H., Long, Y. P., Shu, Y., and Zhai, J. X. (2022). PPRD: a comprehensive online database for expression analysis of ~45,000 plant public RNA-seq libraries. *Plant Biotechnol. J.* 20 (5), 806–808. doi: 10.1111/pbi.13798

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42 (4), 355–360. doi: 10.1038/ng.456

Zheng, Y. M., Mao, Y., Xie, C., Smith, H., Lus, L., and Xu, S. (2005). Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). *Genetics* 169 (4), 2267–2275. doi: 10.1534/genetics.104.033217