Journal of Integrative Agriculture 2026, 25(8): 3139–3152

Available online at ScienceDirect

Journal of Integrative Agriculture



RESEARCH ARTICLE

A genome-wide association study revealed that *GmRGD14* positively regulates the root dry weight in soybeans

Kaili Ren[1*], Jialuo Chen[1*], Xuan Cui[1], Xiao Li[1], Dezhou Hu[1, 3], Zhongyi Yang[1], Yu'e Zhang[1], Yuming Yang[1, 4], Deyue Yu[1#], Hui Wang[1, 2#]

[1] *National Center for Soybean Improvement/National Key Laboratory of Crop Genetics and Germplasm Enhancement/Jiangsu Collaborative Innovation Center for Modern Crop Production/Nanjing Agricultural University, Nanjing 210095, China*
[2] *Zhongshan Biological Breeding Laboratory (ZSBBL), Nanjing 210095, China*
[3] *College of Life Sciences, Nanjing Agricultural University, Nanjing 210095, China*
[4] *School of Agriculture, Henan Institute of Science and Technology, Xinxiang 453003, China*

Highlights

- Seven loci containing stable SNPs significantly associated with the root dry weight (RDW) were identified.
- *GmRGD14* in *qRDW14-2* positively regulated the RDW.
- *GmRGD14* was selected during soybean domestication.

Abstract

Roots are vital for crop growth, development, yield, and tolerance to various types of environmental stress. Numerous genetic loci associated with soybean root morphological traits have been identified, but few genes associated with these traits have been reported. This study identified seven quantitative trait loci (QTLs) containing stable SNPs significantly associated with the root dry weight in soybeans through a genome-wide association study. Among these QTLs, *qRDW14-2* presented the greatest significance. In *qRDW14-2*, the gene *GmRGD14*, encoding the lysophosphatidic acid acyltransferase LPAT4, was identified as a candidate. *GmRGD14*, in block63, which contained the significant SNP S14_6521715, had the highest expression level in soybean roots, and its *Arabidopsis* homologous mutant *lpat4* presented more lateral roots than did the control Col-0. GmRGD14 was localized primarily to the cell membrane and endoplasmic reticulum. The heterologous overexpression of *GmRGD14* in *Arabidopsis* significantly increased the lateral root number, which was similar to the phenotype of *atlpat4*. Furthermore, overexpression of *GmRGD14* resulted in a greater total root length, root tip number, root surface area, and root volume in the hairy roots of transgenic soybean plants than in those of control soybean plants, whereas knockdown of the gene *via* RNA interference in soybean hairy roots resulted in the opposite phenotype. *GmRGD14*, which is highly genetically variable in wild soybean, has been gradually utilized during soybean domestication. Overall, this study revealed that *GmRGD14* is a new key gene involved in root growth, providing a promising genetic target for breeding elite soybean varieties with strong root systems.

Keywords: soybean, root dry weight, QTL, *GmRGD14*

1. Introduction

The soybean (*Glycine max* (L.) Merr.) is one of the most important grain crops and the most widely cultivated oil crop worldwide (Anderson *et al.* 2019). Soybeans have a taproot system that includes primary, lateral, and adventitious roots, which absorb water and nutrients from the soil (Fageria and Moreira 2011) and directly affect aboveground growth, ultimately impacting both yield and quality (Noh *et al.* 2022). Therefore, a robust root system plays a key role in soybean growth, development, and tolerance to various environmental

Received 9 October 2024; Received in revised form 3 January 2025; Accepted 12 February 2025; Available online 20 March 2025
[#]Correspondence Hui Wang, Tel: +86-25-84399527, Fax: +86-25-84396410, E-mail: wanghui0@njau.edu.cn; Deyue Yu, Tel/Fax: +86-25-84396410, E-mail: dyyu@njau.edu.cn
[*] These authors contributed equally to this study.

© 2026 CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer review under responsibility of Editorial Board of *Journal of Integrative Agriculture*.
doi: 10.1016/j.jia.2025.03.007

stress factors (Ye *et al.* 2018; Sun *et al.* 2022).

The growth and development of soybean roots are controlled by multiple genes. Multiple root traits, including the main/primary root length (MRL/PRL), total root length (TRL), root length (RL), lateral root number (LRN), lateral root density (LRD), root tip number (RTN), root dry weight (RDW), root fresh weight (RFW), root surface area (RSA), root volume (RV), and average root diameter (RD), have been used to evaluate the soybean root system architecture (RSA). Brensha *et al.* (2012) utilized 368 simple sequence repeat (SSR), restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP) markers to locate eight quantitative trait loci (QTLs) for RSA traits in a population composed of 94 recombinant inbred lines (RILs). Among the eight QTLs, three were for MRL, one for LRN, two for RFW, and two for RDW. Liang *et al.* (2014) employed 232 SSR markers and mapped five, four, three, and three QTLs related to the MRL, LRN, RDW, and RV, respectively, in 447 RILs. Prince *et al.* (2020) identified two, two, one, and one QTL related to MRL, TRL, RV, and RSA, respectively, using 16,469 SNPs in 184 $F_{7:8}$ RILs. Chen *et al.* (2021) mapped one QTL associated with the PRL on chromosome 16 using 103 F_7 RILs with 223 SSR markers. Wang *et al.* (2022) identified 34 QTLs associated with the RDW, LRN, RD, TRL, PRL, RV, and RSA, using over 6,300 bin maps in 404 RILs. With the development of whole-genome sequencing technology, genome-wide association studies (GWASs) have become a powerful tool for identifying RSA loci. For example, Mandozai *et al.* (2021) used SLAF-seq data from 260 spring soybean germplasm accessions, which included 179,960 SNPs, and identified two, six, four, one, and two QTLs associated with the TRL, RSA, RV, RD, and RDW, respectively. Kim *et al.* (2023) identified 29 SNPs related to the RD, TRL, and RSA *via* a GWAS by combining the RSA traits of 357 soybean landraces with 84,210 SNPs. Wang *et al.* (2023) explored four QTLs associated with the RDW using 5,341,787 SNPs from 559 soybean accessions and highlighted *Glyma.05G047900* as a strong candidate gene for improving soybean breeding for increased biomass in the future. Similarly, Li *et al.* (2024) proposed *GmGA20ox1* as a key candidate gene affecting soybean PRL traits on the basis of a GWAS, in which genotype data from 4,000,000 SNPs from 324 soybean accessions were used and eight genetic loci associated with PRL were identified. Despite the numerous loci associated with root development identified in soybeans, most of the genes related to soybean root traits have been obtained through homology cloning rather than map-based cloning (Hao *et al.* 2011; Yang *et al.* 2017, 2019; Liu *et al.* 2022; Hu *et al.* 2023). For example, *GmJAZ3* and *GmWRKY46* play crucial roles in regulating soybean root length (Liu *et al.* 2022; Hu *et al.* 2023). The overexpression of *GmJAZ3* increases the MRL (Hu *et al.* 2023). *GmWRKY46* negatively regulates root morphological traits such as the RL, RTN, and RSA by modulating the size of the meristem region in the main root apex (Liu *et al.* 2022). Additionally, *GmLBD12* and NAC-type transcription factors are essential for regulating lateral root growth (Hao *et al.* 2011; Yang *et al.* 2017, 2019). The overexpression of *GmLBD12* increases the LRN in *Arabidopsis* (Yang *et al.* 2017). The overexpression of *GmNAC20* in *Arabidopsis* significantly increases the LRN and LRD (Hao *et al.* 2011), whereas the overexpression of *GmNAC109* promotes the formation of more lateral roots in *Arabidopsis* (Yang *et al.* 2019).

Lysophosphatidic acid acyltransferase (LPAT/LPAAT) is one of the key rate-limiting enzymes in the Kennedy pathway for triacylglycerol synthesis (Misra *et al.* 2014; Körbes *et al.* 2016). LPATs have been isolated from various plant species, such as *Arabidopsis* (Kim *et al.* 2005; Angkawijaya *et al.* 2019), rice (Shaikh *et al.* 2022), soybean (Wayne *et al.* 2021), peanut (Chen *et al.* 2015), castor (Chen *et al.* 2016), and rape (Zhang K *et al.* 2022). In *Arabidopsis*, there are five LPAT isoforms: LPAT1, LPAT2, LPAT3, LPAT4, and LPAT5 (Yin *et al.* 2022). LPAT1 is a plastidial LPAT, and the absence of *AtLPAT1* leads to embryo lethality (Kim and Huang 2004; Yu *et al.* 2004). Deletion of *LPAT2* results in female gametophyte lethality (Kim *et al.* 2005). LPAT4 and LPAT5 have redundant functions under low-nitrogen conditions, as evidenced by the fact that the *lpat4-1 lpat5-1* double mutant exhibits slow shoot growth under such conditions (Angkawijaya *et al.* 2019). In rice, *LPAT2* positively regulates the response to osmotic stress by mediating abscisic acid (ABA) signalling (Shaikh *et al.* 2022). The heterologous expression of soybean *GmLPAT2* in *Arabidopsis* alters the TAG composition of seed oils (Wayne *et al.* 2021). Similarly, the overexpression of peanut *AhLPAT2* in *Arabidopsis* increases the seed oil content and weight (Chen *et al.* 2015). In castor, the overexpression of *LPAT2* increases ricinoleic acid content (Chen *et al.* 2016). The overexpression of *BnLPAT2* and *BnLPAT5* in rapeseed significantly increases the seed oil content (Zhang K *et al.* 2022). Despite these advances, importantly, to date, there have been no reports on the role of the LPAT gene family in soybean root growth and development.

In this study, RDW traits were evaluated at the seedling stage in 219 cultivated soybean accessions across five environments. Through a GWAS, a significant QTL, *qRDW14-2*, associated with RDW was identified on chromosome 14. *GmRGD14* was determined to be the primary candidate gene for *qRDW14-2 via* analysis of expression patterns and root phenotypic evaluation of *Arabidopsis* orthologous gene T-DNA insertion mutants. To further investigate the biological function of *GmRGD14* in root growth, transgenic *Arabidopsis* and soybean hairy roots were used. Additionally, the genetic diversity of *GmRGD14* was analysed using genotype data provided by a public database.

2. Materials and methods

2.1. Plant materials and related experiments

A natural population comprising 219 soybean accessions from various regions was selected for the GWAS in this study. Among these, domestic soybean accessions accounted for 83.56% and were sourced primarily from 26 provinces and municipalities, including Jiangsu (29), Jiangxi (18), Henan (17), Hubei (16), Northeast China (11), and Shandong (10). Additionally, 11 global soybean accessions were included. This population exhibited substantial genetic variation.

To evaluate root traits, we determined the RDW of soybean seedlings grown in five environments using either hydroponic or sand culture methods. The hydroponic experiments were conducted in March 2016 (Environment 1, E1), September 2016 (E2), and March 2017 (E3), as previously described (Hu *et al.* 2024). Specifically, soybean accessions were grown in an artificial climate chamber under consistent conditions: an average temperature of 28°C, an average relative humidity of 48%, and a photoperiod of 16 h with a canopy light intensity of 30,000 lx. For this process, plump, disease-free, and uniform-sized seeds were selected and germinated in sterile vermiculite. Once the soybean cotyledons slightly opened and the apical meristem emerged from between the cotyledons, four uniformly growing seedlings were selected and transferred to Hoagland nutrient solution (pH=5.8). The plants were then fixed in plastic planting baskets *via* sponge blocks as fixing devices and arranged in a 60-well hydroponic box (65 cm×42 cm×14 cm) with a 6 cm distance between holes. The nutrient mixture was replaced every 3 d, and the hydroponic boxes were cleaned regularly to reduce the growth of moulds and mosses. The experiment followed a completely randomized design with four biological replicates. The roots below the cotyledonary nodes of the seedlings were sampled after 15 d of cultivation in Hoagland nutrient solution. All the roots were thoroughly cleaned with water, placed in kraft paper bags, and oven-dried at 105°C for 30 min and then at 75°C to a constant weight, after which the RDW (g) was measured. Sand culture experiments were conducted using potted plants in 2016 (E4) and 2017 (E5), as previously described (Yang Y M *et al.* 2024). During the seedling stage, Hoagland nutrient solution (pH=5.8) was applied every 4 d. The experiment also followed a completely randomized design but with three biological replicates. The sampling was performed at 30 d after sowing, and the detailed method was similar to that used for cultivation in the Hoagland nutrient solution.

2.2. Phenotypic data analysis

Descriptive statistical analysis was performed, and frequency distribution plots were generated using IBM SPSS Statistics 26 (Cronk 2017). Analysis of variance (ANOVA) was conducted on the basis of the general line model $Y_{ijk}=\mu+G_i+E_j+(GE)_{ij}+\varepsilon_{ij}$ in R, where Y_{ijk} represents the phenotypic value of the ith genotype in the jth environment and the kth replicate, μ is the overall mean, G_i is the genotype effect, E_j is the environment effect, $(GE)_{ij}$ is the genotype-by-environment interaction effect, and ε_{ij} is the residual error. The variance components (V_g, V_{ge}, and V_e) of the RDW were obtained *via* PROC GLM in SAS 9.2. The broad-sense heritability (H^2) was calculated according to the formula $H^2=V_g/(V_g+V_{ge}/n+V\varepsilon/nr)$, where V_g is the genotypic variance, V_{ge} is the genotype-by-environment interaction variance, $V\varepsilon$ is the residual variance, n is the number of environments, and r is the number of replicates. The analysis of the phenotypic data from the two planting methods was conducted separately.

2.3. GWAS

Genomic resequencing of the 219 soybean materials yielded a total of 5,410,112 SNPs (Lu *et al.* 2020; Li *et al.* 2022). Rare allelic variants were removed using a minor allele frequency (MAF) threshold<0.05. The GWAS between phenotypes and genotypes was conducted *via* the multi-locus mixed model (MLMM) in the GAPIT package (Wang and Zhang 2021; Hyten 2022) and the general linear model (GLM) and the fixed and random model circulating probability unification (FarmCPU) models in the rMVP package (Yin *et al.* 2021). Owing to the significant influence of the environment on RDW, a threshold of $-\log_{10}P\geq4.00$ was selected to identify significant SNPs. The Manhattan plot was redrawn using the qqman package (Turner 2018).

2.4. LD analysis

Linkage disequilibrium (LD) decay for the resequenced genotypic data of the natural soybean population was calculated using PopLDdecay (Zhang *et al.* 2019). The average LD decay distance for the population was determined as the distance at which the LD coefficient decreased to 0.2 (Hu *et al.* 2024). LDBlockShow (Dong *et al.* 2021) was used to calculate D' values for the 2,126 SNPs within *qRDW14-2* and to visualize LD levels. The theoretical range of D' is 0 to 1. The default cut-off for strong linkage in LDBlockShow is D'>0.8 (Marsh 2022).

2.5. Homozygosity identification for *Arabidopsis thaliana* mutants

The *A. thaliana* T-DNA insertion mutants *lpat4* (SALK_115626), *prp1* (SALK_089145) and *prp2* (SALK_070905) were obtained from the Arabidopsis Biological Resource Center (AraShare, https://www.arashare.cn/index/Product/index.html). The seeds were first sterilized with chlorine gas for 3 h and then evenly sown on solid MS media. After vernalization in the dark at 4°C for 2 d, the plates were transferred to a growth chamber, where they were incubated under a 16 h/8 h light/dark photoperiod at 22°C and a relative humidity of 60–70%. Once the seedlings developed two to four true leaves, they were transplanted into a substrate consisting of a 1:1 mixture of nutrient soil and vermiculite. The transplants were then covered with plastic wrap to maintain humidity, after which they were grown under the same temperature and light conditions. The plastic wrap was removed once new leaves emerged, allowing the plants to acclimate. After 2 wk of growth, leaf samples were collected, and genomic DNA was extracted using the Plant Genomic DNA Extraction Kit (Proteinssci, Shanghai, China).

To identify the homozygosity of the *Arabidopsis* mutants, we employed the three-primer method (O'Malley *et al.* 2015). The primer sequences were obtained from the T-DNA Primer Design website (http://signal.salk.edu/tdnaprimers.2.html) and are listed in Appendix A. All primer sequences were synthesized by Nanjing GenScript. PCR amplification was performed using *Taq* DNA polymerase (Vazyme, Nanjing, China). The amplification program included initial denaturation at 95°C for 3 min, followed by 35 cycles of 95°C for 15 s, 58°C for 15 s, and 72°C for 20 s, with a

final extension at 72°C for 5 min. The PCR products were subsequently analysed *via* 1% agarose gel electrophoresis and imaged *via* a gel documentation system.

2.6. Cloning and bioinformatics analysis of *GmRGD14*

Primers for amplifying the full-length CDS of *GmRGD14* were designed based on the *Glyma.14G077500.2* transcript sequence, and their primer sequences are provided in Appendix A. The full-length CDS of *GmRGD14* was amplified from the root cDNA of two materials, NJAU_C032 (Hap6) and NJAU_C081 (Hap3), at the vegetative stage three (V3). After amplification, the product was cloned and inserted into the entry vector pCE2-TA/Blunt-Zero (Vazyme, Nanjing, China). Next, sequencing verification was carried out by Sangon Biotech (China), after which the construct was named pT-*GmRGD14*. Nucleotide sequence alignment was performed using MEGA_X_10.1.7 (Kumar *et al.* 2018). The amino acid count, molecular weight, and isoelectric point of GmRGD14 were predicted online using the ExPASy ProtParam tool (https://web.expasy.org/protparam/). Domain analysis was conducted online using the SMART tool (http://smart.embl-heidelberg.de/). Phylogenetic analysis of LPATs was performed using the neighbor-joining method with default parameters in the MEGA_X_10.1.7 software.

2.7. Tissue expression analysis

The expression patterns of *Glyma.14G077300*, *Glyma.14G077400*, and *Glyma.14G077500* in various tissues of Williams 82 (W82) were analysed *via* the RNA-Seq database (https://soyatlas.venanciogroup.uenf.br/) and RT-qPCR. For RT-qPCR analysis, roots, shoots, and leaves at the V3 stage, flowers at the reproductive stage two (R2), and pods and seeds at the reproductive stage four (R4) of W82 were used. To analyse the differences in *GmRGD14* expression among different tissues of different haplotypes, the soybean materials NJAU_C123 (Hap1), NJAU_C201 (Hap2), NJAU_C081 (Hap3), NJAU_C170 (Hap4), NJAU_C054 (Hap5), NJAU_C032 (Hap6), and NJAU_C116 (Hap7) were selected. The cultivation was performed according to the hydroponic method described in Section 2.1. After 15 d, the roots, stems, and leaves were rapidly frozen in liquid nitrogen and stored at –80°C for subsequent detection.

Total RNA was extracted from the collected tissues using a Total RNA Extraction Kit (AF503A, Proteinssci, Shanghai, China). The extracted RNA was subsequently reverse-transcribed to cDNA *via* the HiScript® II Q RT SuperMix for qPCR (+gDNA wiper) Kit (R223, Vazyme, Nanjing, China). For RT-qPCR, the ChamQ SYBR qPCR Master Mix (without ROX) Kit (Q321-02, Vazyme, Nanjing, China) was used, and the reactions were carried out on a CFX96 Real-Time PCR System (Bio-Rad, USA). The data were subsequently analysed using the ABI 7500 Sequence Detection System (SDS) v.1.4. The relative expression levels of the target genes were calculated *via* the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen 2001). To ensure accuracy, *GmTubulin* (*Glyma.03G124400*) was used as the internal reference gene for expression normalization. The experiment was conducted with three biological replicates and three technical replicates. The expression patterns of *Glyma.14G077300* and *Glyma.14G077400* in various tissues of W82 were normalized using the ΔC_T value of the root tissue as the reference, whereas *Glyma.14G077500* was normalized with the seed tissue as the reference. For *GmRGD14* expression in different haplotype tissues, the ΔC_T value of stem tissue from NJAU_C116 was used as a reference to normalize other tissues. The sequences of the primers used are listed in Appendix A.

2.8. Subcellular localization

The pFGC5941 vector was digested with the restriction enzymes *Xho*I and *Bam*HI, and primers containing *Xho*I and *Bam*HI restriction sites were used to amplify the *GmRGD14* CDS from the pT-*GmRGD14* vector. The *GmRGD14* sequence was subsequently cloned and inserted into the pFGC5941 vector, thereby constructing a *GmRGD14-GFP* fusion expression vector *via* homologous recombination. This fusion expression vector (35S:*GmRGD14*-GFP), along with an empty control vector (35S:GFP), was then transformed into *Agrobacterium tumefaciens* strain EHA105 and injected into tobacco mesophyll cells for transient expression. After 8 h of dark treatment, followed by 48 h of light incubation, the images were observed using a Leica SP8 confocal laser scanning microscope. mCherry-Marker was used for nuclear membrane labelling, and ER-Marker was employed for endoplasmic reticulum labelling.

2.9. *Arabidopsis* transformation

The pBA002 vector was first digested with the restriction enzymes *Mlu*I and *Sac*I, and primers containing *Mlu*I and *Sac*I restriction sites were subsequently used to amplify the *GmRGD14* CDS from the pT-*GmRGD14* vector. Using homologous recombination, a 35S-driven constitutive *GmRGD14* overexpression vector, pBA002-*GmRGD14*, was successfully constructed. The recombinant plasmid pBA002-*GmRGD14* was then introduced into *A. tumefaciens* strain EHA105. To prepare the infection solution for generating transgenic *Arabidopsis* plants, a sucrose (5%) solution containing 0.03% surfactant (Silwet L-77) was used (Clough and Bent 1998). For transformation, flowers and siliques of wild-type *Arabidopsis* (WT, Col-0) and mutant *Arabidopsis* (*lpat4*) at the flowering stage were removed, and the inflorescence axes were completely immersed in infection medium for 3–5 s. After infection, the plants were kept in the dark overnight, followed by normal light incubation. This infection process was repeated 3–5 times throughout the flowering period (Clough and Bent 1998). Following the infection and seed harvest from infected *Arabidopsis* plants, the seeds were sown on glufosinate-ammonium selection media. *Arabidopsis* plants that survived for two weeks were then transferred to the substrate for further positive identification through PCR and RT-qPCR. The transgenic plants obtained from WT were classified as *GmRGD14* overexpression (OE) lines, whereas those obtained from mutant *Arabidopsis* were classified as *GmRGD14* complementation (COM) lines. PCR amplification for positive identification was performed in a 50 μL reaction system using

phanta max super-fidelity DNA polymerase (P505, Vazyme, Nanjing, China). The RT-qPCR analysis was conducted as described in Section 2.7, using Arabidopsis *AtTubulin* (*At5g62690*) as the reference gene. The sequences of the primers used are listed in Appendix A.

2.10. Root and shoot phenotype identification in *Arabidopsis*

Arabidopsis homozygous mutants (*lpat4*, *prp1*, and *prp2*), *GmRGD14* T_2 generation *Arabidopsis* transgenic lines (OE and COM), and the WT were utilized for root phenotype identification. For the three homozygous mutants, the seeds were first sterilized with chlorine gas and then carefully sown on solid MS media *via* sterile toothpicks. The seeds were evenly distributed in a 10 cm×10 cm square plastic Petri dish with grid lines at 20 seeds per dish, with 10 WT seeds on the left and 10 mutant seeds (*lpat4*, *prp1*, or *prp2*) on the right. Each dish served as one replicate, and three replicates were established. After 2 d of vernalization at 4°C, the dishes were transferred to a growth chamber, where they were positioned vertically and cultivated under a 16 h light/8 h dark cycle at 22°C. After 15 d of growth, the PRL and LRN of the *Arabidopsis* plants were measured. For the *GmRGD14* transgenic lines (OE and COM), seeds were similarly sterilized with chlorine gas. WT and mutant (*lpat4*) seeds were sown on MS media, whereas seeds of the OE and COM lines were sown on MS media containing glufosinate-ammonium. For root phenotype identification, five seedlings of similar growth were selected from each of the WT, *lpat4*, and OE lines and transferred to the same MS media for continued vertical cultivation after 5 d of vertical cultivation in a 22°C growth chamber. Three replicates were established for each group. After 10 d, both the PRL and LRN of *Arabidopsis* were measured, and samples were collected for positive identification at the DNA and RNA levels. For shoot phenotype identification, 15 seedlings of similar growth were selected from each of the WT, *lpat4*, and OE lines and transferred into pure vermiculite for further cultivation after 5 d of cultivation in a 22°C growth chamber. After 18 d, the maximum rosette radius, rosette leaf number, and shoot fresh weight of the *Arabidopsis* plants were measured, and samples were collected for positive identification at the DNA and RNA levels. The method for measuring the maximum rosette radius, rosette leaf number, and shoot fresh weight is based on that of Boyes *et al.* (2001).

2.11. Construction of a transgenic *GmRGD14* soybean hairy root system

The pMDC83 vector, which contains a 35S promoter-driven green fluorescent protein (GFP) gene, was used as a plant overexpression vector. The *GmRGD14* CDS was amplified from the pT-*GmRGD14* vector and cloned and inserted into the pMDC83 vector to obtain the *GmRGD14*-GFP fusion-overexpressing homologous recombination vector OE-*GmRGD14*. The pTCK303 vector, which contains a 35S promoter-driven enhanced green fluorescent protein (eGFP) gene, was used as a plant RNA interference (RNAi) vector. A 150–250 bp fragment from the *GmRGD14* CDS region was selected, cloned, and inserted into the pTCK303 vector *via* two pairs of primers to obtain the RNAi homologous recombination vector RNAi-*GmRGD14*. The primer sequences are listed in Appendix A. The OE-*GmRGD14* and RNAi-*GmRGD14* plasmids were transformed into the *Agrobacterium rhizogenes* strain K599, which was then used to infect the cotyledons of the soybean cultivar Jack, following a modified version of a previously described method (Huang *et al.* 2022). After infection, the plants were cultured in the dark at 25°C. Since both the pMDC83 and pTCK303 vectors contain independently expressed GFP genes, positive hairy roots could be screened using a portable fluorescence lamp (LUYOR-3415, LUYOR, USA) that emits dual fluorescence excitation light. After the hairy roots had grown, GFP fluorescence was observed to identify transformed roots (those with green fluorescence), while untransformed roots (without green fluorescence) were discarded. For each vector, 24 GFP-positive hairy roots were selected for root phenotype assessment and RT-qPCR analysis. The root phenotypic data were analysed with an automatic scanning apparatus (Epson Perfection V850 Pro, Epson, China) equipped with WinRHIZO 2022a software (Zealquest, Shanghai). RT-qPCR analysis was conducted as described in Section 2.7.

2.12. Haplotypic and domestication analysis

The physical location of *GmRGD14* in the *Glycine max* Wm82.a2.v1 reference genome was determined using the Phytozome website (https://phytozome-next.jgi.doe.gov/). Haplotype analysis was conducted using the *GmRGD14* CDS region, UTR, and 2 kb promoter region (Chr14: 6,529,905...6,536,064), encompassing a total range of 6.16 kb. The polymorphic sites for haplotype analysis were sourced from both the cultivated soybean population used in the 2.3 GWAS and a natural population comprising 4,414 soybean accessions provided by the SoyMD database (Yang Z Q *et al.* 2024). The *t*-test in SPSS was used to analyse the significance of differences between different haplotypes. The 4,414 soybean accessions included 588 wild soybeans (*Glycine soja*), 1,253 landraces (*Glycine max*), and 2,573 improved cultivars (*Glycine max*). Among these, 3,811 were from Asia, 316 from North America, 35 from Europe, 5 from Africa, 68 from South America, 3 from Oceania, and 176 of unknown origin. The genetic differentiation index (F_{ST}), Tajima's *D*, and nucleotide diversity (π) among wild soybean, landraces, and improved cultivars were calculated using VCFtools (Cadzow *et al.* 2014) with a sliding window set to 10k–10k.

3. Results

3.1. Phenotypic characteristics of the RDW trait in 219 soybean accessions

To evaluate the phenotypic variation in RDW in soybean, we measured RDW at the seedling stage in a natural population of 219 cultivated soybean accessions (Table 1). In the five environments, the RDW ranged from 0.04 g in E1 to 3.69 g in E5. The mean RDW values varied between (0.21±0.07) g and (1.48±0.63) g, with a coefficient of variation ranging from

Table 1 Descriptive statistical analysis, analysis of variance, and generalized heritability of the root dry weight (RDW)[1)]

Env.[2)]	Mean	SD	Min.	Max.	CV (%)	G	G×E	H^2 (%)
E1	0.30	0.12	0.04	0.67	40	**	**	83
E2	0.21	0.07	0.05	0.43	33			
E3	0.33	0.14	0.08	1.22	42			
E4	0.40	0.16	0.10	0.94	40	**	**	68
E5	1.48	0.63	0.16	3.69	43			

[1)] Env., environment; SD, standard deviation; Min., minimum; Max., maximum; CV, coefficient of variation; G, genotype; G×E, genotype-by-environment interaction; H^2, heritability.
[2)] E1–E5, the five environments from Environment 1 to 5.
**, P<0.01.

33% in E2 to 43% in E5. The soybean genotype (G) and the genotype×environment interaction (G×E) had significant effects (P<0.01) on the RDW under the two planting methods (Table 1). The broad-sense heritability was 83% for E1–E3 and 68% for E4–E5, indicating substantial environmental influence on phenotypic data collection methods. The frequency distribution histogram revealed that the RDW trait approximated a normal distribution (Appendix B). These results suggest considerable variation in RDW within this natural soybean population.

3.2. GWAS

After quality control, 5,352,391 high-quality SNPs were retained (Appendix C). We conducted a genome-wide association analysis by combining the RDW phenotype with these SNPs. Across the five environments, a total of 3,619, 6,579, and 3,089 significant SNPs were identified *via* the MLMM, GLM, and FarmCPU models, respectively, and were distributed across all 20 chromosomes (Fig. 1-A and B; Appendices D–H). The LD decay distance between collinear SNPs, with r^2=0.2, was approximately 180 kb (Appendix D). On the

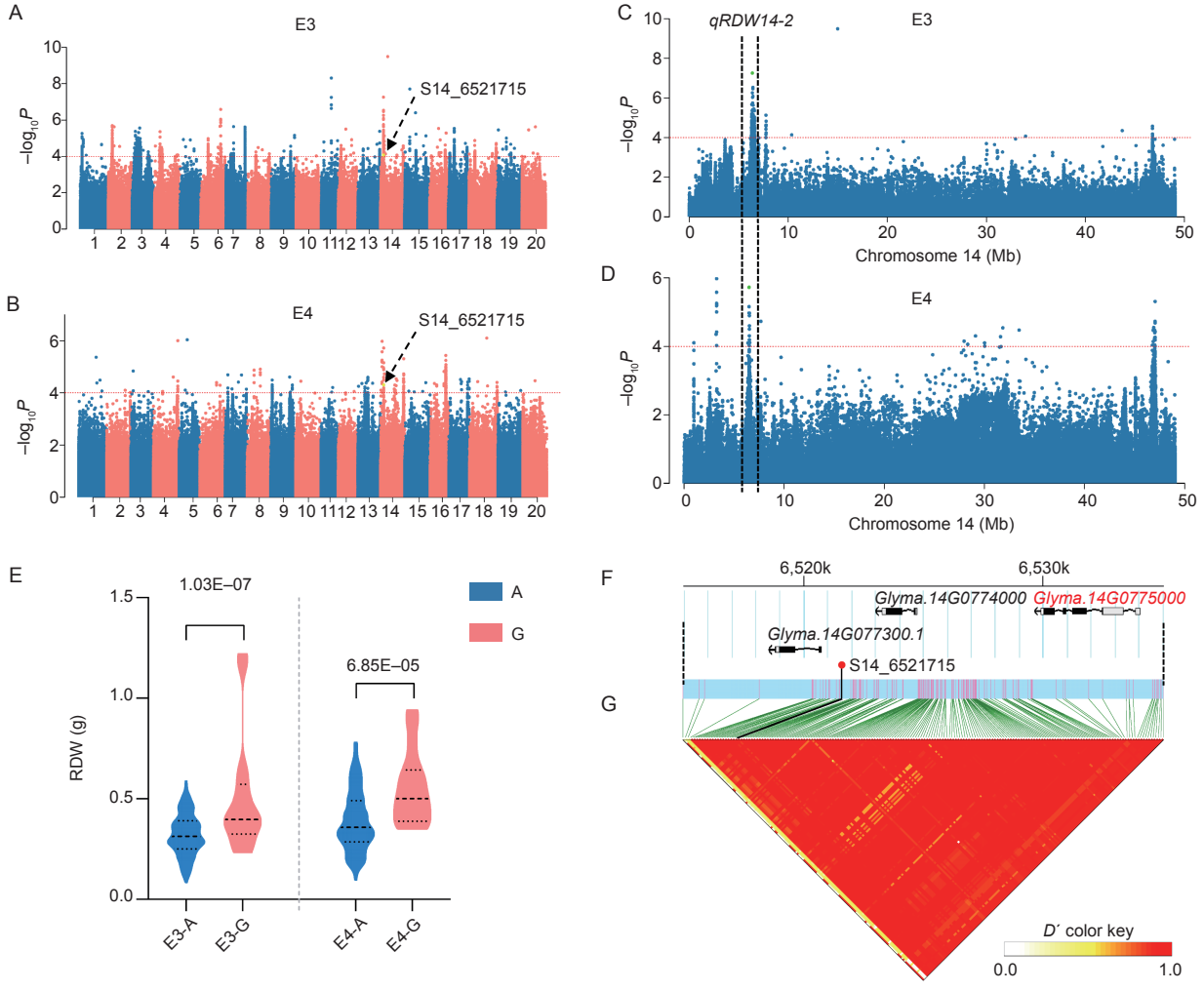


Fig. 1 Genome-wide association study (GWAS) results for the root dry weight (RDW) in Environment 3 (E3) and E4 and a linkage disequilibrium plot of block63. A and B, Manhattan plots of the RDW *via* the MLMM model in E3 and E4. The same significantly associated SNPs located in E3 and E4 are displayed as yellow dots. Plots above the red dotted line show the genome-wide significance with a threshold of P=1.00×10^{-4}. C and D, Manhattan plots of chromosome 14 in E3 and E4. The leading SNPs located in E3 and E4 are displayed as green dots. Plots above the red dotted line show the genome-wide significance with a threshold of P=1.00×10^{-4}. E, phenotypic comparison of RDW in E3 and E4 between the two groups divided by the significant SNP S14_6521715 (A/G). A two-tailed t-test was used for difference analyses (P<0.05). F, genes within the 6.515–6.535 Mb region (block63) on chromosome 14 of the soybean Wm82.a2 genome, according to https://www.soybase.org/. G, linkage disequilibrium plot of block63.

basis of this decay distance, a QTL was defined as the region extending 180 kb upstream and downstream of the significant SNPs. As a result, 298, 473, and 384 QTLs were identified *via* the MLMM, GLM, and FarmCPU models, respectively, and QTLs containing clustered significant SNPs were identified on each chromosome (Appendices F–H). Sixty-two stable QTLs had overlapping intervals across two or more models and environments (Appendix I). Notably, the QTLs *qRDW6-2*, *qRDW6-3*, *qRDW7-1*, *qRDW13-4*, *qRDW14-2*, *qRDW16-2*, and *qRDW18-3* each contained one to six stable SNPs significantly associated with RDW in the two environments. Among the seven stable loci, the leading SNPs of *qRDW14-2* presented the greatest significance, including one stable SNP S14_6521715 (Fig. 1-C and D; Appendix J). The soybean variety carrying the G allele of the SNP had a greater RDW than did those carrying the A allele, by approximately 1.4- to 1.6-fold (P=6.85×10^{-5} to 1.03×10^{-7}) (Fig. 1-E). Consequently, we selected the *qRDW14-2* locus for further investigation.

3.3. *Glyma.14G077500* was the candidate gene of *qRDW14-2*

To identify the candidate gene, we calculated the *D´* values between collinear SNPs within *qRDW14-2*. The results indicated that this segment was divided into 100 blocks (Appendix K). Block63, which contained the significant SNP S14_6521715, included three genes: *Glyma.14G077300*, *Glyma.14G077400*, and *Glyma.14G077500* (Fig. 1-F and G). On the basis of the RNA-Seq data (Almeida-Silva *et al.* 2023) and RT-qPCR results, *Glyma.14G077300* and *Glyma.14G077400* presented the highest expression levels in flowers, whereas their expression levels were relatively lower in the roots (Appendix L). In contrast, *Glyma.14G077500* presented relatively high expression levels in the roots compared with other tissues of W82 (Appendix L). Both *Glyma.14G077300* and *Glyma.14G077400* are functionally annotated as encoding disease resistance-related proteins and are highly homologous to *AtPRP* (*At1g75030*) in *Arabidopsis*, with amino acid sequence similarities of 73.81 and 72.91%, respectively (Appendix M). *Glyma.14G077500* encodes the lysophosphatidic acid acyltransferase LPAT4, which is homologous to *AtLPAT4* (*At1g75020*) in *Arabidopsis*, with an amino acid sequence similarity of 77.28% (Appendix M). A comparison of the root phenotypes among two T-DNA insertion mutants of *AtPRP* (*prp1*, *prp2*), a T-DNA insertion mutant of *AtLPAT4* (*lpat4*), and the WT revealed that the LRNs of *lpat4*, *prp1*, and *prp2* were considerably greater than those of the WT. Of the three mutants, *lpat4* had the greatest number of lateral roots, which was approximately 1.8 times greater than that of the WT, whereas *prp1* and *prp2* presented approximately 1.2- and 1.3-fold greater lateral roots, respectively (Appendix N). Given that the *lpat4* mutant presented more significant root system alterations, we repeated the same experiment twice, and similar results were obtained each time (Appendix O). Based on these findings, *Glyma.14G077500* was considered a candidate gene for *qRDW14-2* and named *GmRGD14* (*root growth and development 14*) based on its associated traits.

3.4. Natural variation in *GmRGD14* and its association with the RDW

To investigate the relationship between *GmRGD14* and RDW in soybeans, we further analysed the genetic diversity and haplotypes of the gene. Within the approximately 6.16 kb region covering the gene and 2,000 bp upstream of the 5´ UTR, we identified 26 SNPs (Fig. 2-A). Among these, 18 SNPs were located in the promoter region of *GmRGD14*, three were in the 5´ UTR, and five were in the intron region. Based on these 26 SNPs, the 219 soybean accessions were grouped into seven haplotypes, with Hap1, Hap3, and Hap6 being the major haplotypes (Fig. 2-A). Compared with the other six haplotypes, Hap6 presented a significantly greater RDW in E3 and E4, indicating that it is a superior haplotype (Fig. 2-B). These results also suggest that *GmRGD14* is associated with an increased RDW.

3.5. Tissue expression pattern and subcellular localization of *GmRGD14*

GmRGD14 has seven alternative transcripts without nucleotide differences in CDS region in the soybean reference genome Wm82.a6.v1 on the Phytozome website (Appendix M; https://phytozome-next.jgi.doe.gov/jbrowse/index.html?data=genomes%2FGmax_Wm82_a6_v1&tracks=Transcripts%2CAlt_Transcripts&loc=Gm14%3A6474895..6481821&highlight=Gm14%3A6475691..6481019). To further verify the differentiation of *GmRGD14* between superior and inferior haplotypes, we cloned the *GmRGD14* CDS from the cDNA of NJAU_C032 (superior haplotype Hap6) and NJAU_C081 (inferior haplotype Hap3) using the mRNA of *Glyma.14G077500.2* as a cloning template. The sequences showed no nucleotide differences; both were 1,152 bp in length and encoded 383 amino acids. The predicted molecular weight of the protein is 44.1 kDa, with an isoelectric point of 9.12. The protein contains a phosphatidyltransferase PlsC domain with a transmembrane region on either side, classifying it into the lysophosphatidic acid acyltransferase (LPAT) subfamily (Appendix M). Nine LPAT genes have been found in the soybean genome (Wayne *et al.* 2021). *GmRGD14* shares the highest amino acid sequence similarity with *Glyma.17G248200*, at 97.4% (Appendix M). Tissue expression analysis revealed that *GmRGD14* was expressed in roots, stems, and leaves, with significantly greater expression in the roots of Hap6 than in those of the other six haplotypes (Fig. 2-C). Subcellular localization analysis revealed that in tobacco mesophyll cells expressing the fusion protein (35S:*GmRGD14*-GFP), the fluorescence signals overlapped with those of the endoplasmic reticulum marker and the nuclear membrane marker. In contrast, the empty vector (35S:GFP), expressing only GFP, displayed fluorescence signals distributed throughout the cell (Fig. 2-D and E). These findings suggest that GmRGD14 is primarily localized to the endomembrane system, including the endoplasmic reticulum and cell membrane.

3.6. *GmRGD14* positively regulates lateral root growth in *Arabidopsis*

To explore the biological function of *GmRGD14*, we generated

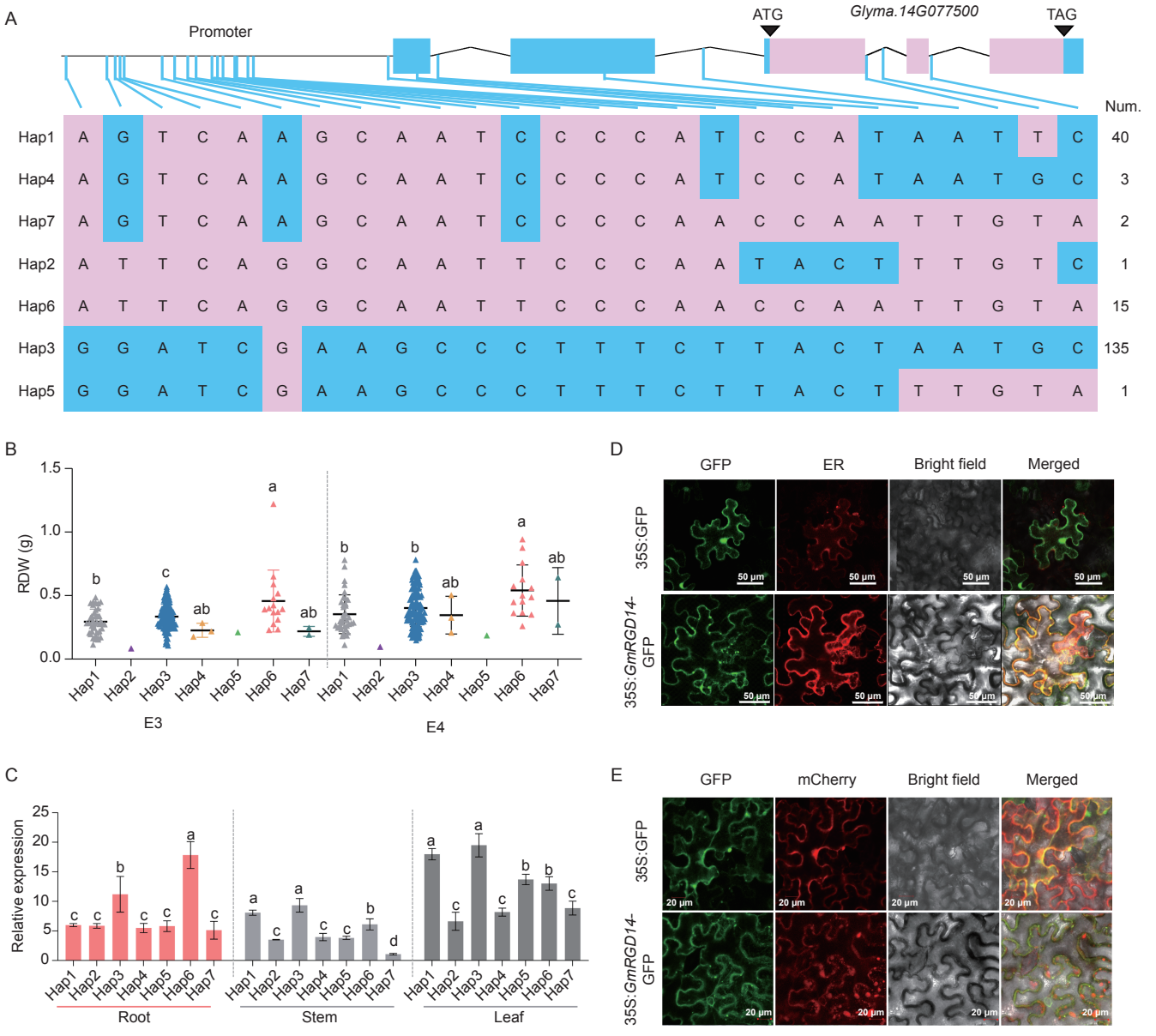


Fig. 2 Haplotype analysis, tissue expression pattern, and subcellular localization. A, haplotype division of 219 soybean accessions on the basis of the genotype of *Glyma.14G077500*. The gene model of *Glyma.14G077500* is depicted above the plot. Num., the number of cultivated soybeans contained in each haplotype. B, root dry weights (RDWs) of seven haplotypes (Hap1, n=40; Hap2, n=1; Hap3, n=135; Hap4, n=3; Hap5, n=1; Hap6, n=15; Hap7, n=2). The data are presented as the mean±SD (P<0.05; two-tailed t-test). C, tissue expression pattern comparison of *GmRGD14* in different haplotypes. The data are presented as the mean±SD (n=3, P<0.05; two-tailed t-test). D and E, subcellular localization of GmRGD14. The ER is used as a marker for endoplasmic reticulum localization. Scale bar, 50 μm. mCherry is used as a marker for nuclear and cytoplasmic localization.

Arabidopsis transgenic lines in which *GmRGD14* was overexpressed in WT (OE#1 and OE#2) and in mutant *lpat4* to complement *AtLPAT4* (COM#4 and COM#8). Owing to the significantly delayed germination of the complementation lines (COM#4 and COM#8), compared with the *lpat4*, WT, OE#1, and OE#2, the phenotypes of COM#4 and COM#8 were not evaluated (Appendix P). Compared with those in WT plants, the number of lateral roots in *lpat4* plants increased by approximately 64.6%, whereas the numbers in OE#1 and OE#2 plants significantly increased by approximately 52.4 and 50.9%, respectively (Fig. 3-A and E). The RLs of *lpat4* and the OE#1 and OE#2 lines did not significantly differ from those of the WT (Fig. 3-A and D). Therefore, it is speculated that *GmRGD14* positively regulates lateral root growth and development in *Arabidopsis*, which contrasts with the function of its homologous gene *AtLPAT4* in *Arabidopsis*.

In addition, compared with those of the WT, the maximum rosette radius, rosette leaf number, and shoot fresh weight of the *lpat4* plants significantly increased by 15.0, 27.8, and 53.8%, respectively (Fig. 3-F–I). These traits increased by approximately 16.3, 18.3, 38.7, 10.0, 20.0, and 36.8%, respectively, in OE#1 and OE#2 plants (Fig. 3-F–I). These findings suggest that *GmRGD14* also promotes shoot growth in *Arabidopsis*.

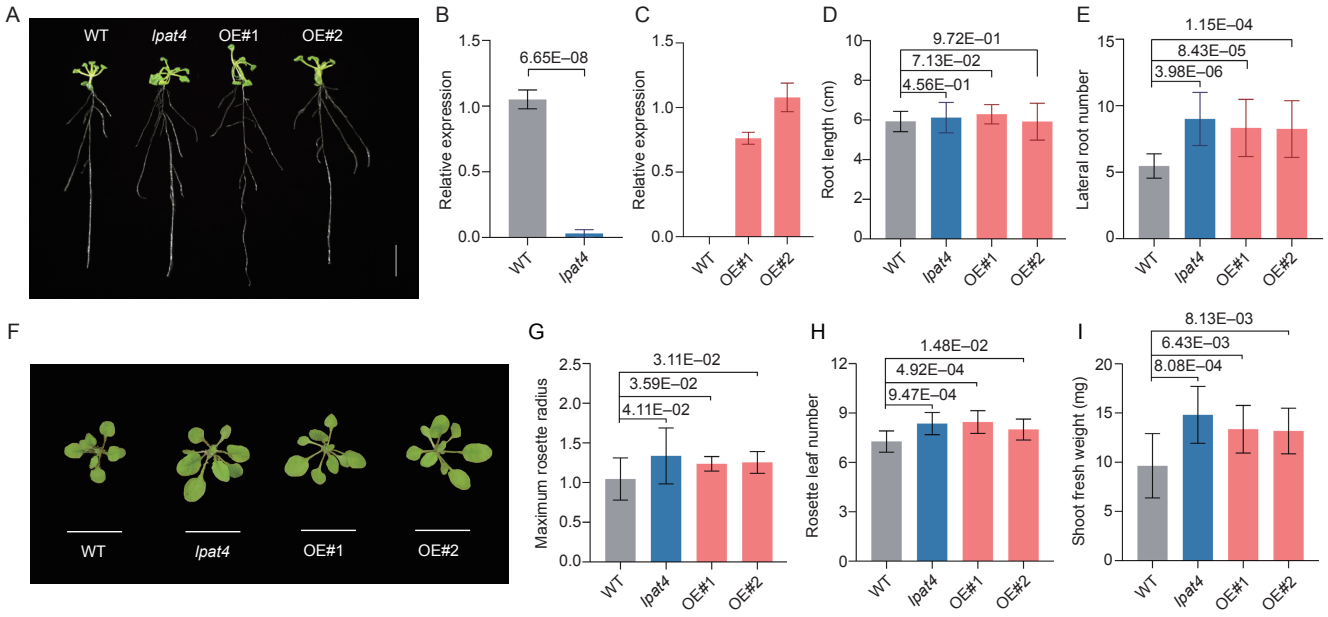


Fig. 3 Overexpression of *GmRGD14* enhanced root and shoot growth in *Arabidopsis*. A, image of the root phenotypes of wild type (WT), *AtLPAT4* mutant (*lpat4*), and *GmRGD14*-overexpressing plants (OE#1 and OE#2). Scale bar=1 cm. B, relative expression levels of *AtLPAT4* in WT and *lpat4* plants subjected to root trait analysis. The data are presented as the mean±SD (n≥4, P<0.05; two-tailed t-test). C, relative expression level of *GmRGD14* in OE#1, and OE#2 for root trait analysis. The data are presented as the mean±SD (n=4). D and E, root length and lateral root number of WT, *lpat4*, OE#1, and OE#2. The data are presented as the mean±SD (n≥10, P<0.05; two-tailed t-test). F, image of the aboveground part phenotypes of WT, *AtLPAT4* mutant (*lpat4*), and *GmRGD14*-overexpressing plants (OE#1 and OE#2). Scale bar=0.5 cm. G, H, and I, maximum rosette radius, rosette leaf number, and shoot fresh weight of WT, *lpat4*, OE#1, and OE#2. The data are presented as the mean±SD (n≥10, P<0.05; two-tailed t-test).

3.7. Overexpression of *GmRGD14* increases soybean hairy root biomass

To further validate the function of *GmRGD14* in regulating soybean root growth and development, we generated transgenic soybean hairy roots overexpressing *GmRGD14* (OE-*GmRGD14*) and those subjected to RNA interference with *GmRGD14* (RNAi-*GmRGD14*) (Appendix Q). The RT-qPCR results indicated that the expression level of *GmRGD14* was approximately 24.0 times greater in OE-*GmRGD14* than in the overexpression empty vector control (OE-EV), whereas the expression level in RNAi-*GmRGD14* was reduced by approximately 53.9% compared with that in the RNA interference empty vector control (RNAi-EV) (Appendix Q). The TRL, RSA, RTN, and RV in OE-*GmRGD14* were significantly greater than those in the control (OE-EV) at approximately 2.1, 2.2, 1.5, and 2.2 times greater than those in the control, respectively (Fig. 4-A, C–F). The RFW and RDW of the transgenic hairy roots overexpressing the gene were also significantly greater than those of the control, with the RFW increasing by approximately 1.9-fold and the RDW increasing by approximately 1.9-fold (Fig. 4-G and H). These results demonstrate that overexpression of *GmRGD14* promotes root growth in soybeans, consistent with findings in *Arabidopsis*. In contrast, the TRL, RSA, RTN, and RV of the RNAi-*GmRGD14* hairy roots were significantly lower than those of the control RNAi-EV hairy roots at approximately 81.1, 75.7, 59.3, and 70.2% of those of the control, respectively (Fig. 4-B–F). These results indicate that decreased *GmRGD14* expression inhibited the growth of hairy roots in transgenic plants. In terms of biomass, there was no significant difference between RNAi-*GmRGD14* and RNAi-EV hairy roots (Fig. 4-G and H). This may be related to the reduced *GmRGD14* expression. The experiments yielded similar results (Appendices R and S). Collectively, these findings indicate that *GmRGD14* positively regulates soybean root growth and development.

3.8. *GmRGD14* was selected during soybean domestication

To investigate whether *GmRGD14* has been subject to selection during population evolution, we calculated nucleotide diversity within a 0.24 Mb interval encompassing the *GmRGD14* gene region and the upstream 2k promoter region across wild soybean, landraces, and improved cultivars. The F_{ST} index indicated the greatest differentiation between wild soybean and improved cultivars, followed by wild soybean and landraces, and the lowest differentiation was observed between soybean landraces and improved cultivars. Both the π values and Tajima's D values were greater for wild soybean than for cultivated soybean (Fig. 5-A–C). These results indicate that *GmRGD14* is being gradually utilized during soybean domestication, but its application in improved cultivars remains limited.

We further analysed the frequency and geographical distribution of the superior haplotype Hap6 in the three subgroups. In the data for the 4,414 soybean genotypes, 47 SNPs and 11 Indels were identified (Appendix T). These 58 polymorphic sites classified the soybean accessions into 96

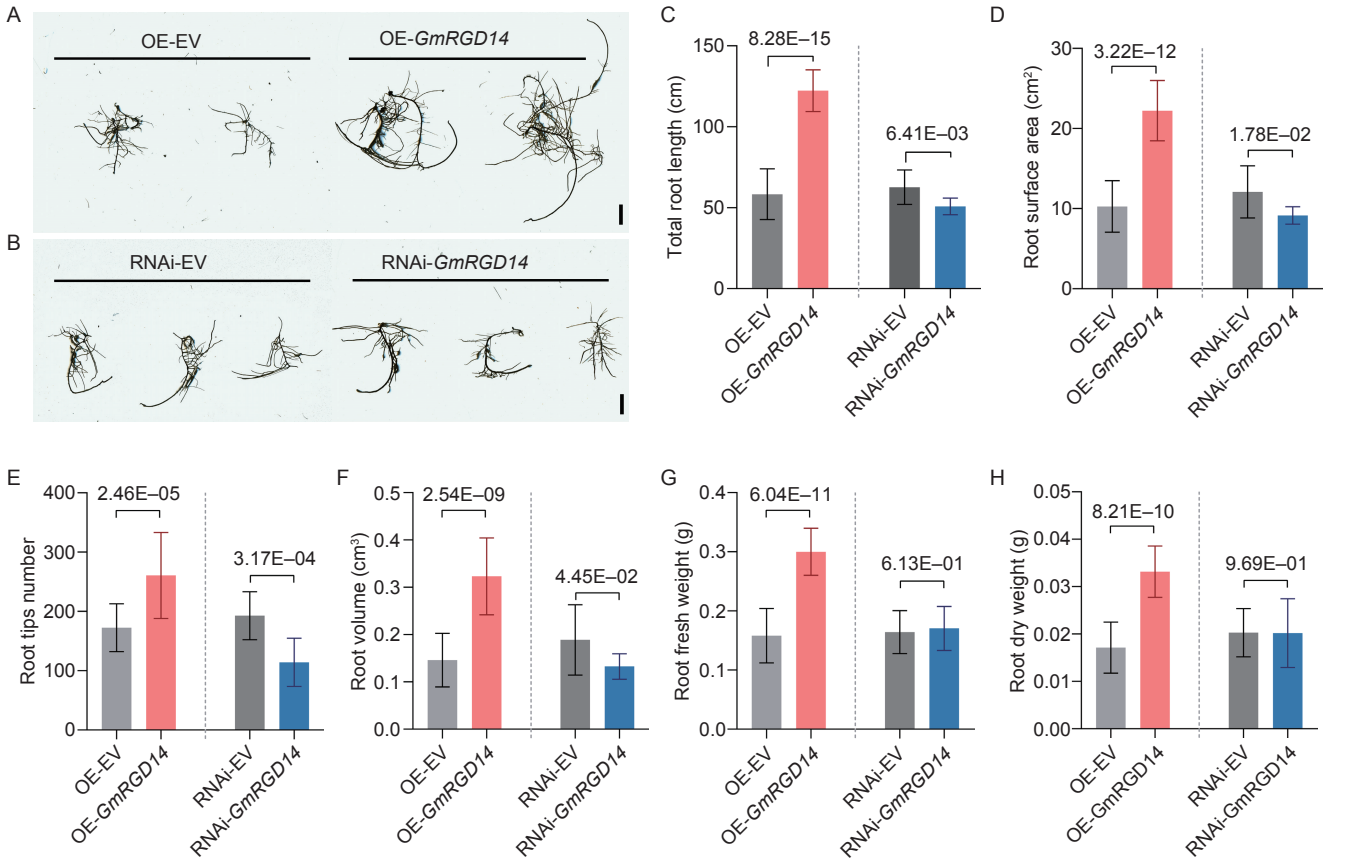


Fig. 4 *GmRGD14* positively regulates root traits. A, image of soybean hairy roots of the overexpression empty vector (OE-EV) and *GmRGD14*-overexpressing (OE-*GmRGD14*) lines. Scale bar=1 cm. B, image of soybean hairy roots from the RNA interference empty vector (RNAi-EV) and *GmRGD14*-knockdown (RNAi-*GmRGD14*) lines. Scale bar=1 cm. C–H, comparison of the total root length, root surface area, number of root tips, root volume, root fresh weight, and root dry weight of *GmRGD14* transgenic hairy roots and the corresponding empty vector controls. The data are presented as the mean±SD (n≥9, P<0.05; two-tailed t-test).

haplotypes. A total of 77, 24, and 16 haplotypes of *GmRGD14* were detected in wild soybean, landraces, and improved cultivars, respectively. Sixty-seven haplotypes of *GmRGD14* were present only in wild soybean, indicating that wild soybean may contain a relatively high genetic diversity of *GmRGD14* (Appendix U). H13, H29, H39, H57, and H59 were the major haplotypes, each including more than 70 materials (Appendix T). In our population, Hap1 corresponded to H13 in the public population. In our population, Hap3 was divided into two haplotypes: H29 and H39. Additionally, in our population, Hap6 was also divided into two haplotypes: H57 and H59. The superior haplotype Hap6 (including H57 and H59) accounted for only 1% of the haplotype in wild soybean, increasing to 8%/4% (landrace/improved cultivars) in cultivated soybeans (Fig. 5-D; Appendix U), indicating that Hap6 was selectively retained during domestication from wild to cultivated soybeans. Hap6 (including H57 and H59) was distributed across Africa, Asia, Europe, North America, and South America, with the highest proportion found in Africa (Fig. 5-E; Appendix V).

4. Discussion

4.1. Six novel loci associated with RDW were discovered

RSA is one of the most critical morphological traits in plants; it is a complex quantitative trait characterized by high plasticity. According to data from the SoyBase website, over 80 QTLs associated with soybean root traits have been identified and are distributed across chromosomes 2, 3, 5, 6, 7, 8, 9, 10, 11, 13, 14, 18, 19, and 20 (Brensha *et al.* 2012; Liang *et al.* 2014; Prince *et al.* 2020). Notably, a QTL on chromosome 3 simultaneously controls MRL and RDW (Brensha *et al.* 2012), whereas a QTL on chromosome 20 controls LRN, RV, RSA, TRL, and RDW (Wang *et al.* 2022). In this study, we identified 62 stable QTLs significantly associated with RDW through a GWAS (Appendix I). Among these, seven stable QTLs, *qRDW6-2*, *qRDW6-3*, *qRDW7-1*, *qRDW13-4*, *qRDW14-2*, *qRDW16-2*, and *qRDW18-3*, each contained stable SNPs that were repeatedly located in multiple environments (Appendices I and J). On chromosome 14, *qRDW14-2* was found to colocalize with the QTL *Qhti-14-3*, which is related to RD (van Nguyen *et al.* 2017) (Appendix J). The other six QTLs identified in this study were novel loci associated with RDW. Further identification confirmed that the LPAT gene *GmRGD14* is a candidate gene for *qRDW14-2* and positively regulates RDW in soybeans (Fig. 4; Appendices N and S).

4.2. *GmRGD14* plays a positive role in regulating root growth, which is different from *AtLPAT4*

In *Arabidopsis*, LPATs play a significant role in root growth.

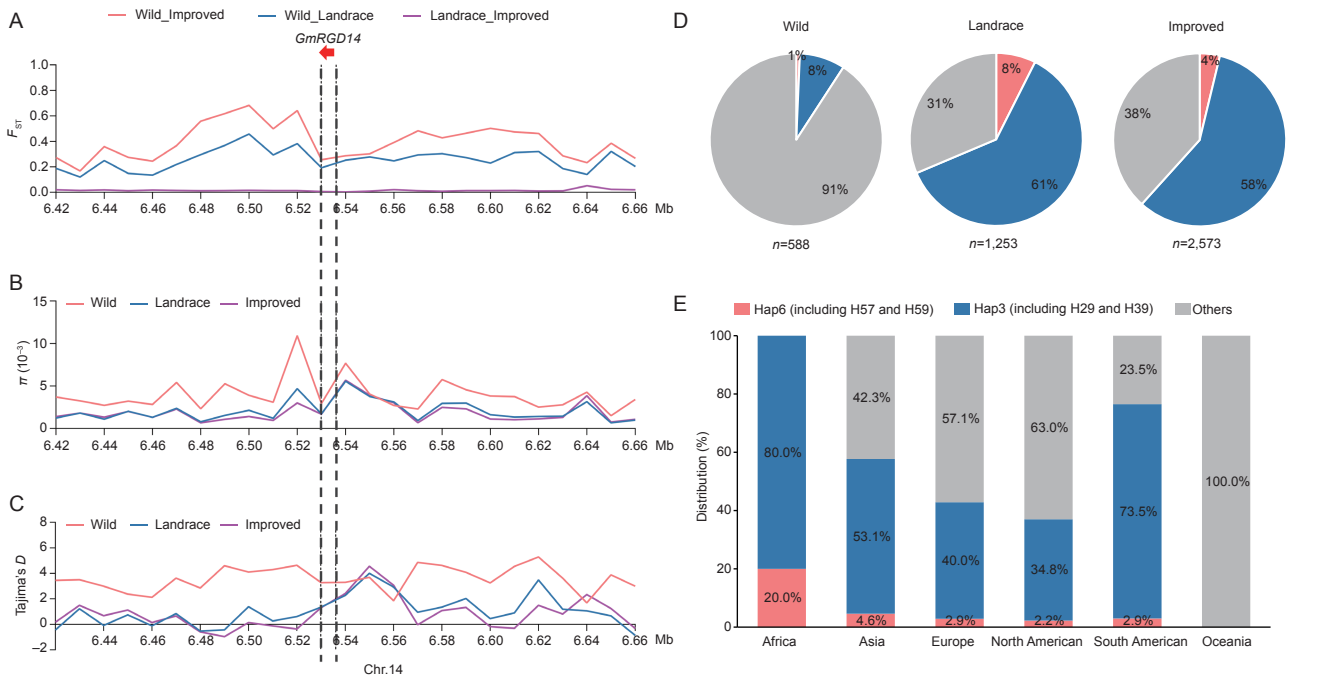


Fig. 5 Natural allelic variation of *GmRGD14* in wild and cultivated soybean lines. A–C, F_{ST}, π, and Tajima's *D* in the 6.42–6.66 Mb region on Chr.14 among the three groups. The *GmRGD14* gene region and 2k upstream promoter region are located between the two black dashed lines. D, proportion of different *GmRGD14* haplotypes among the three groups. E, distribution of different *GmRGD14* haplotypes across the six continents.

LPAT2 overexpression promotes primary root growth under low-phosphorus stress (Angkawijaya *et al.* 2017). Double mutants of *LPAT3*, *LPAT4*, and *LPAT5*, such as *lpaat3-1 lpaat4-2*, *lpaat3-1 lpaat5-2*, and *lpaat4-2 lpaat5-2*, exhibit reduced PRLs (Wattelet-Boyer *et al.* 2022). LPAT members are clearly directly or indirectly involved in regulating primary root growth. In soybeans, there are nine LPAT homologous genes (Wayne *et al.* 2021), but none have been reported to be associated with root traits. In this study, *GmRGD14* and *AtLPAT4* were both found to be involved in regulating root growth. However, these genes exerted opposite effects on LRN regulation, despite sharing 77.28% sequence similarity and containing the same phosphatidyltransferase domain, PlsC (Appendix M). The *Arabidopsis* mutant *lpat4* presented a significant increase in the number of lateral roots (Fig. 3-E; Appendices N and O), whereas overexpression of *GmRGD14* in both *Arabidopsis* and soybean hairy roots increased the number of lateral roots or hairy roots (Figs. 3-E and 4; Appendix S). The divergence among paralogues has also been reported in other plants. In *Arabidopsis* and wheat, *AtCOL5* and *TaCol-B5* overexpression lines flower early, in contrast to the late heading observed in response to *OsCOL5* overexpression in rice (Hassidim *et al.* 2009; Zhang X Y *et al.* 2022; Wen *et al.* 2024). Similarly, the *DRO1* (DEEPER ROOTING 1) gene in *Arabidopsis* and peach plays different roles in regulating root architecture. The *Arabidopsis dro1* mutant presented relatively large lateral root angles and reduced PRL, with no significant change or a slight decrease in the RL observed upon *AtDRO1* overexpression. In contrast, the overexpression of peach *PpeDRO1* significantly increases the PRL (Guseman *et al.* 2017). These results suggest that the divergence between paralogues may be species-specific during species evolution. Soybean is a short-day plant, whereas *Arabidopsis* is a long-day plant. Light and temperature conditions not only influence flowering time but also play a critical role in root development (van Gelderen *et al.* 2018; Cao *et al.* 2021). The distinct light and temperature requirements between soybean and *Arabidopsis* may be one of the reasons for the observed difference in root development mediated by *AtLPAT4* and *GmRGD14*. Additionally, soybean is an ancient tetraploid, and the soybean LPAT family contains more genes than the diploid *Arabidopsis*. Apart from *GmRGD14* (*Glyma.14G077500*), *Glyma.17G248200* was clustered in the same clade as *AtLPAT4* (Appendix M), consistent with the findings reported by Wayne *et al.* (2021). The function of *Glyma.17G248200* is unknown, and whether it shares a similar role with *AtLPAT4* in root development remains to be verified. Abundant genetic variation is important for plant adaptation to the environment and for crop improvement. Therefore, it is necessary to carry out functional studies on genes across different species to provide valuable genetic resources for crop improvement.

The functional differences between *GmRGD14* and *AtLPAT4* prompted us to further compare other results for these two genes. We found that *GmRGD14* and *AtLPAT4* differ in their predicted protein domains, expression patterns, and subcellular localization. Specifically, *GmRGD14* has two transmembrane regions flanking the PlsC domain (Appendix M), whereas *AtLPAT4* has three such regions (Körbes *et al.* 2016). In terms of expression patterns, *GmRGD14* was highly expressed in soybean roots, followed by leaves, with the lowest expression observed in shoots (Appendix L).

In contrast, *AtLPAT4* is most highly expressed in siliques, followed by leaves and roots, and is expressed at the lowest levels in flowers (Kim *et al*. 2005). Additionally, GmRGD14 is primarily localized in the endoplasmic reticulum and cell membrane (Fig. 2-D and E), whereas AtLPAT4 has been reported to localize in the endoplasmic reticulum (Angkawijaya *et al*. 2019). These differences may influence how *GmRGD14* and *AtLPAT4* regulate lateral root growth.

4.3. *GmRGD14* may play a role in soybean tolerance to abiotic stresses

Numerous studies have shown that some genes that regulate root traits also play crucial roles in plant tolerance to abiotic stresses. For example, *GmNTL1* and *GmMDH2* promote root growth and increase salt stress tolerance in plants (Zhang *et al.* 2024; Zhou *et al.* 2025). Field experiments have demonstrated that, compared with control plants, transgenic soybean plants overexpressing *GmNTL1* exhibit greater yields under both low- and high-salinity conditions (Zhang *et al.* 2024). *GmGDPD2* enhances low phosphorus tolerance and increases yield by increasing both RL and RSA (Hu *et al.* 2024). *GmRR1* improves yield-related traits by negatively regulating RL, RSA, and RDW in response to low-phosphorus stress (Yang Y M *et al.* 2024). The *OsLPAT2* mutation reduces rice tolerance to salt and drought stress (Shaikh *et al.* 2022). A mutation in *AtLPAT1* can suppress the sensitivity of the fatty acid biosynthesis mutant *fab1* to low temperatures (Kim *et al.* 2010). In this study, *GmRGD14*, which encodes the phosphatidyltransferase LPAT4 and is associated with root traits such as TRL, RSA, RTN, and RV, may also play a role in soybean tolerance to abiotic stresses, including drought, salinity, low temperatures, and nutrient starvation. This topic warrants further investigation in future studies.

5. Conclusion

Seven stable QTLs significantly associated with RDW were identified across chromosomes 6, 7, 13, 14, 16, and 18. Among these QTLs, *qRDW14-2*, located on chromosome 14, presented the highest significance level. *GmRGD14*, encoding the phosphatidyltransferase LPAT4, was the candidate gene for *qRDW14-2*. Overexpression of *GmRGD14* in *Arabidopsis* increased LRN. In soybean hairy roots, overexpression of *GmRGD14* improved TRL, RSA, RTN, and RV, whereas RNA interference-mediated silencing of *GmRGD14* produced phenotypes opposite to those of the respective controls. Additionally, *GmRGD14* has been gradually utilized during domestication. Hap6 is the elite haplotype. This study offers new insights into the genetic basis of soybean root growth.

Acknowledgements

The funding for this research was provided by the Biological Breeding-National Science and Technology Major Project, China (2023ZD04070), the Zhongshan Biological Breeding Laboratory, China (ZSBBL-KY2023-03), and the National Natural Science Foundation of China (32101742, 32072080, and 32090065).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that they did not use AI in the preparation and writing of this manuscript.

Appendices associated with this paper are available at https://doi.org/10.1016/j.jia.2025.03.007

References

Almeida-Silva F, Pedrosa-Silva F, Venancio T M. 2023. The Soybean Expression Atlas v2: A comprehensive database of over 5000 RNA-seq samples. *The Plant Journal*, **116**, 1041–1051.

Anderson E J, Ali M L, Beavis W D, Chen P Y, Clemente T E, Diers B W, Graef G L, Grassini P, Hyten D L, McHale L K, Nelson R L, Parrott W A, Patil G B, Stupar R M, Tilmon K J. 2019. Soybean [*Glycine max* (L.) Merr.] breeding: History, improvement, production and future opportunities. In: Al-Khayri J M, Jain S M, Johnson D V, ed., *Advances in Plant Breeding Strategies*: *Legumes*. Springer Publishing, Cham, Switzerland. pp. 431–516.

Angkawijaya A E, Nguyen V C, Nakamura Y. 2017. Enhanced root growth in phosphate-starved *Arabidopsis* by stimulating *de novo* phospholipid biosynthesis through the overexpression of *LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE 2* (*LPAT2*). *Plant Cell and Environment*, **40**, 1807–1818.

Angkawijaya A E, Nguyen V C, Nakamura Y. 2019. LYSOPHOSPHATIDIC ACID ACYLTRANSFERASES 4 and 5 are involved in glycerolipid metabolism and nitrogen starvation response in *Arabidopsis*. *New Phytologist*, **224**, 336–351.

Boyes D C, Zayed A M, Ascenzi R, McCaskill A J, Hoffman N E, Davis K R, Görlach J. 2001. Growth stage-based phenotypic analysis of *Arabidopsis*: A model for high throughput functional genomics in plants. *The Plant Cell*, **13**, 1499–1510.

Brensha W, Kantartzi S K, Meksem K, Grier R L, Barakat A, Lightfoot D A, Kassem M A. 2012. Genetic analysis of root and shoot traits in the 'Essex' by 'Forrest' recombinant inbred line (RIL) population of soybean [*Glycine max* (L.) Merr.]. *Journal of Plant Genome Sciences*, **1**, 1–9.

Cadzow M, Boocock J, Nguyen H T, Wilcox P, Merriman T R, Black M A. 2014. A bioinformatics workflow for detecting signatures of selection in genomic data. *Frontiers in Genetics*, **5**, 293.

Cao S H, Luo X M, Xu D G, Tian X L, Song J, Xia X C, Chu C C, He Z H. 2021. Genetic architecture underlying light and temperature mediated flowering in *Arabidopsis*, rice, and temperate cereals. *New Phytologist*, **230**, 1731–1745.

Chen G Q, van Erp H, Martin-Moreno J, Johnson K, Morales E, Browse J, Eastmond P J, Lin J T. 2016. Expression of castor LPAT2 enhances ricinoleic acid content at the *sn*-2 position of triacylglycerols in lesquerella seed. *International Journal of Molecular Sciences*, **17**, 507.

Chen H T, Kumawat G, Yan Y L, Fan B J, Xu D H. 2021. Mapping and validation of a major QTL for primary root length of soybean seedlings grown in hydroponic conditions. *BMC Genomics*, **22**, 132.

Chen S L, Lei Y, Xu X, Huang J Q, Jiang H F, Wang J, Cheng Z S, Zhang J N, Song Y H, Liao B S, Li Y R. 2015. The peanut (*Arachis hypogaea* L.) gene *AhLPAT2* increases the lipid content of transgenic *Arabidopsis* seeds. *PLoS ONE*, **10**, e0136170.

Clough S J, Bent A F. 1998. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal*, **16**, 735–743.

Cronk B C. 2017. *How to Use IBM SPSS Statistics*: *A Step-by-Step Guide to Analysis and Interpretation*. 8th ed. Routledge Publishing, New York, USA.

Dong S S, He W M, Ji J J, Zhang C, Guo Y, Yang T L. 2021. LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics*, **22**, bbaa227.

Fageria N K, Moreira A. 2011. The role of mineral nutrition on root

growth of crop plants. In: Sparks D L, ed., *Advances in Agronomy*. Academic Press Publishing, New York, USA. pp. 251–331.
van Gelderen K, Kang C K, Pierik R. 2018. Light signaling, root development, and plasticity. *Plant Physiology*, **176**, 1049–1060.
Guseman J M, Webb K, Srinivasan C, Dardick C. 2017. *DRO1* influences root system architecture in *Arabidopsis* and *Prunus* species. *The Plant Journal*, **89**, 1093–1105.
Hao Y J, Wei W, Song Q X, Chen H W, Zhang Y Q, Wang F, Zou H F, Lei G, Tian A G, Zhang W K, Ma B, Zhang J S, Chen S Y. 2011. Soybean NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants. *The Plant Journal*, **68**, 302–313.
Hassidim M, Harir Y, Yakir E, Kron I, Green R M. 2009. Over-expression of *CONSTANS-LIKE 5* can induce flowering in short-day grown *Arabidopsis*. *Planta*, **230**, 481–491.
Hu D D, Cui R F, Wang K, Yang Y M, Wang R Y, Zhu H Q, He M S, Fan Y K, Wang L, Wang L, Chu S S, Zhang J Y, Zhang S S, Yang Y F, Zhai X H, Lü H Y, Zhang D D, Wang J S, Kong F J, Yu D Y, *et al.* 2024. The Myb73-GDPD2-GA2ox1 transcriptional regulatory module confers phosphate deficiency tolerance in soybean. *The Plant Cell*, **36**, 2176–2200.
Hu Y, Liu Y, Tao J J, Lu L, Jiang Z H, Wei J J, Wu C M, Yin C C, Li W, Bi Y D, Lai Y C, Wei W, Zhang W K, Chen S Y, Zhang J S. 2023. GmJAZ3 interacts with GmRR18a and GmMYC2a to regulate seed traits in soybean. *Journal of Integrative Plant Biology*, **65**, 1983–2000.
Huang P H, Lu M Y, Li X B, Sun H Y, Cheng Z Y, Miao Y C, Fu Y F, Zhang X M. 2022. An efficient *Agrobacterium rhizogenes*-mediated hairy root transformation method in a soybean root biology study. *International Journal of Molecular Sciences*, **23**, 12261.
Hyten D L. 2022. Genotyping platforms for genome-wide association studies: Options and practical considerations. In: Torkamaneh D, Belzile F, eds., *Genome-Wide Association Studies*. Humana Publishing, New York, USA. pp. 29–42.
Kim H U, Huang A H C. 2004. Plastid lysophosphatidyl acyltransferase is essential for embryo development in *Arabidopsis*. *Plant Physiology*, **134**, 1206–1216.
Kim H U, Li Y B, Huang A H C. 2005. Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in *Arabidopsis*. *The Plant Cell*, **17**, 1073–89.
Kim H U, Vijayan P, Carlsson A S, Barkan L, Browse J. 2010. A mutation in the *LPAT1* gene suppresses the sensitivity of *fab1* plants to low temperature. *Plant Physiology*, **153**, 1135–1143.
Kim S H, Tayade R, Kang B H, Hahn B S, Ha B K, Kim Y H. 2023. Genome-wide association studies of seven root traits in soybean (*Glycine max* L.) landraces. *International Journal of Molecular Sciences*, **24**, 873.
Körbes A P, Kulcheski F R, Margis R, Margis-Pinheiro M, Turchetto-Zolet A C. 2016. Molecular evolution of the lysophosphatidic acid acyltransferase (LPAAT) gene family. *Molecular Phylogenetics and Evolution*, **96**, 55–69.
Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, **35**, 1547–1549.
Li X, Hu D Z, Cai L Y, Wang H Q, Liu X Y, Du H P, Yang Z Y, Zhang H R, Hu Z B, Huang F, Kan G Z, Kong F J, Liu B H, Yu D Y, Wang H. 2022. *CALCIUM-DEPENDENT PROTEIN KINASE38* regulates flowering time and common cutworm resistance in soybean. *Plant Physiology*, **190**, 480–499.
Li Y, Gu J B, Zhao B Y, Yuan J B, Li C, Lin Y, Chen Y H, Yang X L, Li Y, Wang Z Y. 2024. Identification and confirmation of novel genetic loci and domestication gene *GmGA20ox1* regulating primary root length in soybean seedling stage. *Industrial Crops and Products*, **217**, 118814.
Liang H Z, Yu Y L, Yang H Q, Xu L J, Dong W, Du H, Cui W W, Zhang H Y. 2014. Inheritance and QTL mapping of related root traits in soybean at the seedling stage. *Theoretical and Applied Genetics*, **127**, 2127–2137.
Liu X Q, Yang Y M, Wang R Y, Cui R F, Xu H Q, Sun C Y, Wang J S, Zhang H Y, Chen H T, Zhang D. 2022. *GmWRKY46*, a WRKY transcription factor, negatively regulates phosphorus tolerance primarily through modifying root morphology in soybean. *Plant Science*, **315**, 111148.
Livak K J, Schmittgen T D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, **25**, 402–408.
Lu S J, Dong L D, Fang C, Liu S L, Kong L P, Cheng Q, Chen L Y, Su T, Nan H Y, Zhang D, Zhang L, Wang Z J, Yang Y Q, Yu D Y, Liu X L, Yang Q Y, Lin X Y, Tang Y, Zhao X H, Yang X Q, *et al.* 2020. Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nature Genetics*, **52**, 428–436.
Mandozai A, Moussa A A, Zhang Q, Qu J, Du Y Y, Anwari G, Amin N A, Wang P W. 2021. Genome-wide association study of root and shoot related traits in spring soybean (*Glycine max* L.) at seedling stages using SLAF-Seq. *Frontiers in Plant Science*, **12**, 568995.
Marsh J. 2022. Linkage disequilibrium statistics and block visualization. In: Edwards D, ed., *Plant Bioinformatics*: *Methods and Protocols*. Humana Publishing, New York, USA. pp. 483–496.
Misra N, Panda P K, Parida B K. 2014. Genome-wide identification and evolutionary analysis of algal LPAT genes involved in TAG biosynthesis using bioinformatic approaches. *Molecular Biology Reports*, **41**, 8319–8332.
van Nguyen L, Takahashi R, Githiri S M, Rodriguez T O, Tsutsumi N, Kajihara S, Sayama T, Ishimoto M, Harada K, Suematsu K, Abiko T, Mochizuki T. 2017. Mapping quantitative trait loci for root development under hypoxia conditions in soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics*, **130**, 743–755.
Noh E, Fallen B, Payero J, Narayanan S. 2022. Parsimonious root systems and better root distribution can improve biomass production and yield of soybean. *PLoS ONE*, **17**, e0270109.
O'Malley R C, Barragan C C, Ecker J R. 2015. A user's guide to the *Arabidopsis* T-DNA insertion mutant collections. In: Ionso J, Stepanova A, eds., *Plant Functional Genomics*: *Methods and Protocols*. Humana Publishing, New York, USA. pp. 323–342.
Prince S J, Vuong T D, Wu X L, Bai Y H, Lu F, Kumpatla S P, Valliyodan B, Shannon J G, Nguyen H T. 2020. Mapping quantitative trait loci for soybean seedling shoot and root architecture traits in an inter-specific genetic population. *Frontiers in Plant Science*, **11**, 1284.
Shaikh A A, Alamin A, Jia C X, Gong W, Deng X J, Shen Q W, Hong Y Y. 2022. The examination of the role of rice lysophosphatidic acid acyltransferase 2 in response to salt and drought stresses. *International Journal of Molecular Sciences*, **23**, 9796.
Sun M L, Li Y, Zheng J Q, Wu D P, Li C X, Li Z Y, Zang Z W, Zhang Y Z, Fang Q W, Li W B, Han Y P, Zhao X, Li Y G. 2022. A nuclear factor Y-B transcription factor, *GmNFYB17*, regulates resistance to drought stress in soybean. *International Journal of Molecular Sciences*, **23**, 7242.
Turner S D. 2018. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *The Journal of Open Source Software*, **3**, 731.
Wang J B, Zhang Z W. 2021. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics*, **19**, 629–640.
Wang X, Zhou S D, Wang J, Lin W X, Yao X L, Su J Q, Li H Y, Fang C, Kong F J, Guan Y F. 2023. Genome-wide association study for biomass accumulation traits in soybean. *Molecular Breeding*, **43**, 33.
Wang Z L, Huang C, Niu Y C, Yung W S, Xiao Z X, Wong F L, Huang M K, Wang X, Man C K, Sze C C, Liu A L, Wang Q W, Chen Y L, Liu S, Wu C X, Liu L F, Hou W S, Han T F, Li M W, Lam H M. 2022. QTL analyses of soybean root system architecture revealed genetic relationships with shoot-related traits. *Theoretical and Applied Genetics*, **135**, 4507–4522.
Wattelet-Boyer V, Guédard M L, DIttrich-Domergue F, Maneta-Peyret L, Kriechbaumer V, Boutté Y, Bessoule J J, Moreau P. 2022. Lysophosphatidic acid acyltransferases: A link with intracellular protein trafficking in *Arabidopsis* root cells? *Journal of Experimental Botany*, **73**, 1327–1343.
Wayne L L, Gachotte D J, Graupner P R, Adelfinskaya Y, McCaskill D G, Metz J G, Zirkle R, Walsh T A. 2021. Plant and algal lysophosphatidic acid acyltransferases increase docosahexaenoic acid accumulation at the *sn*-2 position of triacylglycerol in transgenic *Arabidopsis* seed oil. *PLoS ONE*, **16**, e0256625.
Wen X X, Zhong Z Z, Xu P, Yang Q Q, Wang Y P, Liu L, Wu Z Z, Wu Y W, Zhang Y X, Liu Q N, Zhou Z P, Peng Z Q, He Y Q, Cheng S H, Cao L Y, Zhan X D, Wu W X. 2024. *OsCOL5* suppresses heading through modulation of *Ghd7* and *Ehd2*, enhancing rice

yield. *Theoretical and Applied Genetics*, **137**, 162.
Yang H, Shi G X, Du H Y, Wang H, Zhang Z Z, Hu D Z, Wang J, Huang F, Yu D Y. 2017. Genome-wide analysis of soybean *LATERAL ORGAN BOUNDARIES Domain*-containing genes: A functional investigation of *GmLBD12*. *The Plant Genome*, **10**, 1–19.
Yang X F, Kim M Y, Ha J M, Lee S H. 2019. Overexpression of the soybean NAC gene *GmNAC109* increases lateral root formation and abiotic stress tolerance in transgenic *Arabidopsis* plants. *Frontiers in Plant Science*, **10**, 1036.
Yang Y M, Wang L, Zhang D, Che Z J, Wang Q, Cui R F, Zhao W, Huang F, Zhang H Y, Cheng H, Yu D Y. 2024. Soybean type-B response regulator GmRR1 mediates phosphorus uptake and yield by modifying root architecture. *Plant Physiology*, **194**, 1527–1544.
Yang Z Q, Luo C F, Pei X X, Wang S B, Huang Y M, Li J W, Liu B H, Kong F J, Yang Q Y, Fang C. 2024. SoyMD: A platform combining multi-omics data with various tools for soybean research and breeding. *Nucleic Acids Research*, **52**, D1639–D1650.
Ye H, Song L, Chen H T, Valliyodan B, Cheng P, Ali L, Vuong T, Wu C J, Orlowski J, Buckley B, Chen P Y, Shannon J G, Nguyen H T. 2018. A major natural genetic variation associated with root system architecture and plasticity improves waterlogging tolerance and yield in soybean. *Plant Cell and Environment*, **41**, 2169–2182.
Yin L L, Zhang H H, Tang Z S, Xu J Y, Yin D, Zhang Z W, Yuan X H, Zhu M J, Zhao S H, Li X Y, Liu X L. 2021. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinformatics*, **19**, 619–628.
Yin Y T, Raboanatahiry N, Chen K, Chen X F, Tian T, Jia J, He H S, He J J, Guo Z Y, Yu L J, Li M T. 2022. Class A lysophosphatidic acid acyltransferase 2 from *Camelina sativa* promotes very long-chain fatty acids accumulation in phospholipid and triacylglycerol. *The Plant Journal*, **112**, 1141–1158.
Yu B, Wakao S, Fan J L, Benning C. 2004. Loss of plastidic lysophosphatidic acid acyltransferase causes embryo-lethality in *Arabidopsis*. *Plant and Cell Physiology*, **45**, 503–510.
Zhang C, Dong S S, Xu J Y, He W M, Yang T L. 2019. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**, 1786–1788.
Zhang K, He J J, Yin Y T, Chen K, Deng X, Yu P, Li H X, Zhao W G, Yan S X, Li M T. 2022. Lysophosphatidic acid acyltransferase 2 and 5 commonly, but differently, promote seed oil accumulation in *Brassica napus*. *Biotechnology for Biofuels and Bioproducts*, **15**, 83.
Zhang W X, Zhi W J, Qiao H, Huang J J, Li S, Lu Q, Wang N, Li Q, Zhou Q, Sun J Q, Bai Y T, Zheng X J, Bai M Y, Breusegem F V, Xiang F N. 2024. H_2O_2-dependent oxidation of the transcription factor GmNTL1 promotes salt tolerance in soybean. *The Plant Cell*, **36**, 112–135.
Zhang X Y, Jia H Y, Li T, Wu J Z, Nagarajan R, Lei L, Powers C, Kan C C, Hua W, Liu Z Y, Chen C, Carver B F, Yan L L. 2022. *TaCol-B5* modifies spike architecture and enhances grain yield in wheat. *Plant Science*, **376**, 180–183.
Zhou R N, Wang S H, Liu P Y, Cui Y F, Hu Z B, Liu C Y, Zhang Z G, Yang M L, Li X, Wu X X, Chen Q S, Zhao Y. 2025. Genome-wide characterization of soybean malate dehydrogenase genes reveals a positive role for *GmMDH2* in the salt stress response. *Journal of Integrative Agriculture*, **24**, 2492–2510.

Executive Editor-in-Chief Xueyong Zhang
Managing Editor Ning Wang