Plant Communications
Research article

CellPress
Partner Journal

Genome-wide scan for oil quality reveals a coregulation mechanism of tocopherols and fatty acids in soybean seeds

Danni Chu[1,3,5], Zhifang Zhang[2,5], Yang Hu[1,3], Chao Fang[2,4], Xindan Xu[1], Jia Yuan[1], Jinsong Zhang[1,3], Zhixi Tian[2,3,*] and Guodong Wang[1,3,*]

[1]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China

[2]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China

[3]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing 100039, China

[4]Present address: Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

[5]These authors contributed equally to this article.

***Correspondence: Zhixi Tian (zxtian@genetics.ac.cn), Guodong Wang (gdwang@genetics.ac.cn)**

https://doi.org/10.1016/j.xplc.2023.100598

ABSTRACT

Tocopherols (vitamin E) play essential roles in human health because of their antioxidant activity, and plant-derived oils are the richest sources of tocopherols in the human diet. Although soybean (*Glycine max*) is one of the main sources of plant-derived oil and tocopherol in the world, the relationship between tocopherol and oil in soybean seeds remains unclear. Here, we focus on dissecting tocopherol metabolism with the long-term goal of increasing α-tocopherol content and soybean oil quality. We first collected tocopherol and fatty acid profiles in a soybean population (>800 soybean accessions) and found that tocopherol content increased during soybean domestication. A strong positive correlation between tocopherol and oil content was also detected. Five tocopherol pathway–related loci were identified using a metabolite genome-wide association study strategy. Genetic variations in three tocopherol pathway genes were responsible for total tocopherol content and composition in the soybean population through effects on enzyme activity, mainly caused by non-conserved amino acid substitution or changes in gene transcription level. Moreover, the fatty acid regulatory transcription factor *GmZF351* directly activated tocopherol pathway gene expression, increasing both fatty acid and tocopherol contents in soybean seeds. Our study reveals the functional differentiation of tocopherol pathway genes in soybean populations and provides a framework for development of new soybean varieties with high α-tocopherol content and oil quality in seeds.

Key words: soybean, fatty acid, tocopherol, vitamin E, mGWAS, oil quality

Chu D., Zhang Z., Hu Y., Fang C., Xu X., Yuan J., Zhang J., Tian Z., and Wang G. (2023). Genome-wide scan for oil quality reveals a coregulation mechanism of tocopherols and fatty acids in soybean seeds. Plant Comm. **4**, 100598.

INTRODUCTION

Vitamin E is the collective name for eight fat-soluble small-molecule chemicals (α-, β-, γ-, and δ-tocopherol/tocotrienol) with distinct antioxidant activities in plants and humans. Because vitamin E is an essential dietary nutrient, vitamin E deficiency results in hemolytic anemia in premature babies and other diseases prevalent in developing countries (Dror and Allen, 2011). Among the four types of tocopherols, α-tocopherol (α-T hereafter) shows the highest antioxidant activity and is preferentially recognized by human cells (Grusak and DellaPenna, 1999; Armutcu et al., 2005; Havaux et al., 2005; Traber, 2007). Tocopherol and tocotrienol differ mainly in the polyprenyl tail: geranylgeranyl diphosphate (unsaturated C20 tail) for tocotrienol and phytyl diphosphate (saturated C20 tail) for tocopherol. As a lipophilic antioxidant, vitamin E is synthesized and stored in all plant species. The vitamin E biosynthetic

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

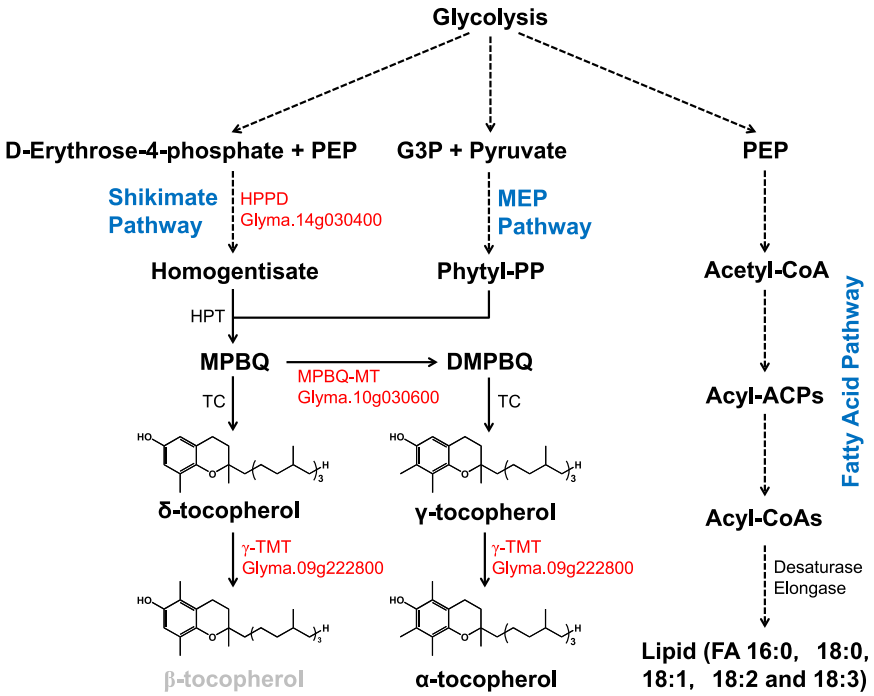


Figure 1. Simplified tocopherol and lipid pathways in soybean seeds.
DMPBQ, dimethylphytylbenzoquinol; FA, fatty acid; G3P, glyceraldehyde 3-phosphate; HPPD, 4-hydroxyphenylpyruvate dioxygenase; HPT, homogentisate phytyltransferase; MEP, methylerythritol 4-phosphate; PEP, phosphoenolpyruvate; MPBQ, 2-methyl-6- phytyl-1,4-benzoquinol; TC, tocopherol cyclase; γ-TMT, γ-tocopherol methyltransferase. Three soybean genes identified in the tocopherol GWAS are highlighted in red. β-tocopherol is shown in gray because of its low content in soybean seeds.

pathway was thoroughly investigated in *Arabidopsis thaliana* two decades ago (Figure 1; Sattler et al., 2004; DellaPenna, 2005). In general, both homogentisic acid (HGA; from the shikimate pathway) and the C20 polyprenyl tail (from the methylerythritol 4-phosphate pathway) are formed in the plastids of plant cells. These two building blocks are fused by homogentisate phytyltransferase (HPT/VTE2, vitamin E 2, the locus in Arabidopsis) to produce 2-methyl-6-phytyl-1,4-benzoquinol (MPBQ). MPBQ is subsequently acted upon by MPBQ methyltransferase (MPBQ-MT/VTE3) and tocopherol cyclase (TC/VTE1) to produce γ- and δ-tocopherol/tocotrienol. The final step in the tocopherol biosynthesis pathway is the conversion of γ- and δ-tocopherol/tocotrienol to α- and β-tocopherol/tocotrienol, respectively, by γ-tocopherol methyltransferase (γ-TMT/VTE4). Among these vitamin E enzymes, MPBQ-MT/VTE3 and γ-TMT/VTE4 play key roles in determining vitamin E composition (DellaPenna, 2005). Very recently, a new component involved in tocopherol biosynthesis, VTE7 (which belongs to the α/β hydrolase family), has been identified in plants (Albert et al., 2022). In contrast to the well-characterized tocopherol biosynthetic pathway, no transcription factor that regulates tocopherol biosynthesis has been characterized in plants (Mene-Saffrane, 2018; Munoz and Munne-Bosch, 2019).

Soybean (*Glycine max* [L.] Merr.) is one of the most important agricultural crops and a major source of oil and protein. Soybean oil has a relatively high total tocopherol (T-T hereafter) content compared with other oilseed crops; the predominant form is γ-tocopherol (γ-T; approximately 60% of T-T), followed by δ-tocopherol (δ-T; approximately 30%), α-T (approximately 10%), and β-tocopherol (less than 5%) (Wong et al., 2014; Mejean et al., 2015). No tocotrienol has been detected in soybean seeds, and the tocopherol content thus plays a key role in the storage stability of soybean oil by reducing lipid peroxidation (Kamal-Eldin, 2006). Because soybean oil accounts for 30% of global vegetable oil consumption (in million metric tons; https://www.statista.com/statistics/263937/vegetable-oils-global-consumption/) and tocopherols are good for human health, an increased content of seed tocopherols, especially α-T, is an important soybean breeding target (Van Eenennaam et al., 2003). The presence of tocopherol also protects soybean oil from double bond oxidation (to maintain oil stability), considering that unsaturated fatty acids (18:1, 18:2, and 18:3 fatty acids [FAs] contain one, two, and three double bonds, respectively) are the predominant FAs in soybean seeds (Figure 1; Ohlrogge and Browse, 1995; Sattler et al., 2004; Fang et al., 2017). Although both tocopherols and FAs play key roles in the health properties of soybean oil, little is known about the relationship between tocopherol and FAs at either the chemical or molecular level in natural soybean populations. The regulatory mechanism of lipid biosynthesis was recently revealed in soybean seeds. To date, several types of transcription factors (including zinc finger, Dof, and bZIP types), identified by gene–gene coexpression analysis, have been demonstrated to regulate lipid biosynthesis in soybean seeds (Wang et al., 2007; Song et al., 2013; Li et al., 2017; Lu et al., 2021).

Here, we investigated the mechanism of tocopherol biosynthesis using a natural soybean population, focusing specifically on the relationship between tocopherol and FAs. We found an overall positive correlation between total FAs and T-T in different soybean groups. Tocopherol genome-wide association studies (GWASs) further demonstrated that transcript variation in three tocopherol pathway genes, 4-hydroxyphenylpyruvate dioxygenase 1 (HPPD1), MPBQ-MT1, and γ-TMT3, contributes to T-T and composition. A transcription factor (*GmZF351*) that regulates both FAs and tocopherols was identified and functionally characterized in soybean seeds. Collectively, these findings provide genetic information for increasing both tocopherol and FAs in soybean seeds through molecular design breeding.

RESULTS AND DISCUSSION

Tocopherol and FA profiling in natural soybean populations

Seed samples from a natural soybean population, previously used for FA and amino acid profiling (Fang et al., 2017), were used for tocopherol profiling by liquid chromatography coupled with a

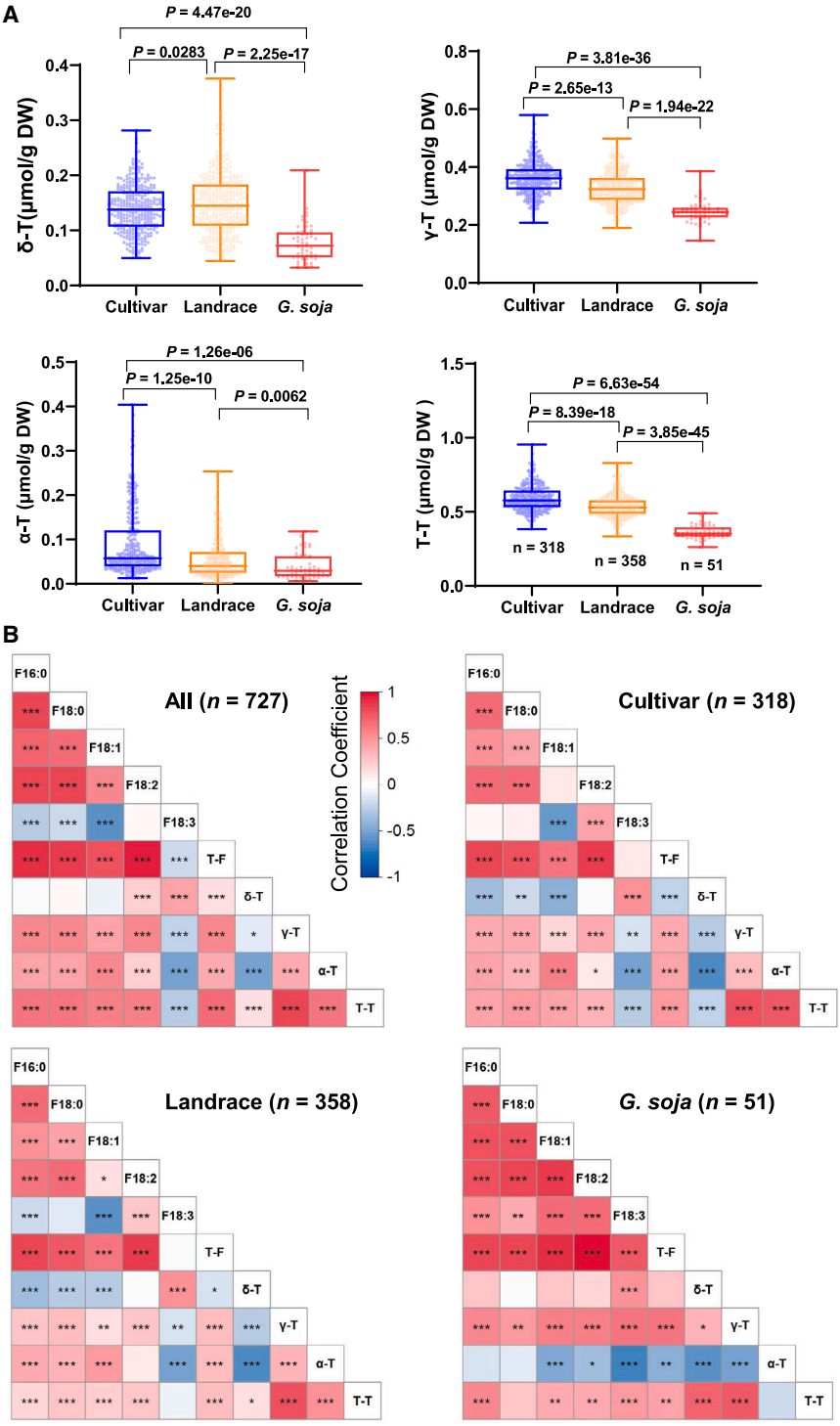


Figure 2. Tocopherol and fatty acid analysis in the soybean population planted in 2013.

(A) Tocopherol (δ-T, γ-T, α-T, and T-T) contents in mature seeds of the cultivated (n = 318), landrace (n = 358), and soja (n = 51) soybean groups. The maximum, 75% quartile, median, 25% quartile, and minimum values of the population are presented. The P values were calculated using two-tailed Student's t-test.

(B) Pearson correlation analysis between fatty acids (single fatty acids [FAs] and total FAs) and tocopherols of different soybean groups (all, cultivar, landrace, and soja). Asterisks indicate significant differences calculated using two-tailed Student's t-test. (*P < 0.05, **P < 0.01, ***P < 0.001).

diode-array detection (DAD) monitor (for detailed sample information, see Supplemental Tables 1 and 2). More than 50 *Glycine soja* (soja hereafter) accessions were included to assess changes in tocopherols during soybean domestication, a 5000-year process by which soja was transformed into landraces and then into improved cultivars (Zhou et al., 2015). There were substantial variations in seed tocopherol content among the 727 diverse soybean accessions: δ-T content ranged from 0.03 to 0.38 μmol/g dry weight (DW), γ-T ranged from 0.14 to 0.58 μmol/g DW, α-T content ranged from 0.0003 to 0.40 μmol/g DW, and T-T content ranged from 0.26 to 0.95 μmol/g DW (Figure 2A). Interestingly, the cultivars had the highest tocopherol content (0.59 ± 0.086 μmol/g DW, n = 318), followed by landraces (0.54 ± 0.075 μmol/g DW, n = 358) and soja (0.36 ± 0.044 μmol/g DW, n = 51). Individual tocopherols showed trends similar to those of T-T during soybean domestication (Figure 2; Supplemental Table 3).

Among soybean samples grown in three locations in 2014 (Heilongjiang Province, Henan Province, and Beijing), the cultivars also had the highest tocopherol content (0.55 ± 0.074 μmol/g DW, n = 341), followed by landraces (0.50 ± 0.084 μmol/g DW, n = 413) and soja (all grown in Beijing; 0.29 ± 0.044 μmol/g DW, n = 53; total FA, 345 ± 60 μmol/g DW). In addition, the tocopherol analysis showed that soybeans grown at high latitudes (Heilongjiang Province, China) had similar T-T contents (0.53 ± 0.071 μmol/g DW, total FA, 746 ± 72 μmol/g DW; n = 145), followed by those grown at low latitudes (0.52 ± 0.087 μmol/g DW, total FA, 709 ± 85 μmol/g DW; n = 603; Supplemental Figure 1; Supplemental Table 3).

Positive correlations between FAs and tocopherols in soybean

A positive correlation was found between T-T and total FA content in cultivar and landrace soybean groups that were grown in different locations. Specifically, soybeans grown in Heilongjiang (representing high latitude) and Henan (representing low latitude) differed in total FA content. (Figure 2 and Supplemental Figure 1). To investigate the relationship between FAs and tocopherols in soja, we analyzed the FA profiles of soja accessions using a method previously reported by Fang et al. (2017) and found a positive correlation between T-T and total FA similar to that observed in cultivars and landraces. These results are largely consistent

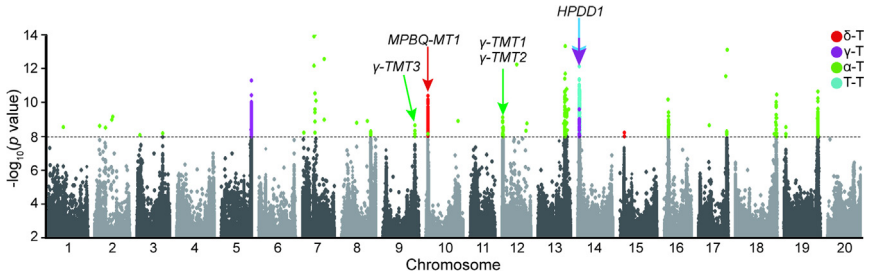


Figure 3. Manhattan plot of GWAS for tocopherol content using 2 years of best linear unbiased prediction data from the soybean population.

Negative $\log_{10}$-transformed *P* values are plotted against the position of SNPs on 20 chromosomes. The significant tocopherol-associated SNPs ($P < 1 \times 10^{-8}$, corresponding to the dashed threshold line) are highlighted in different colors (δ-T in red, γ-T in purple, α-T in green, and T-T in blue). Candidate tocopherol biosynthetic genes are indicated with arrows.

with a previous report (Ghosh et al., 2021). A similar correlation between tocopherol and FAs was also observed in *Brassica* oilseeds and maize germplasm collections (Goffman and Bohme, 2001; Li et al., 2013). These results suggested that a possible coregulation mechanism for FA and tocopherol biosynthesis is conserved in seed plants. Notably, α-T always showed a negative correlation with C18:3 FA in our analysis, although α-T has the highest ability to protect unsaturated FAs from oxidation (Kamal-Eldin, 2006). A recent study also found that C18:3 FA content was decreased, alongside increased total FAs, in metabolically engineered soybean seeds with enhanced vitamin E content (Konda et al., 2020). One possible explanation for this phenomenon is that a high content of α-T might inhibit the activity of FA desaturase 3, which is responsible for conversion of C18:2 FA to C18:3 FA. The underlying mechanism merits further investigation.

GWAS for tocopherol content in soybean seeds

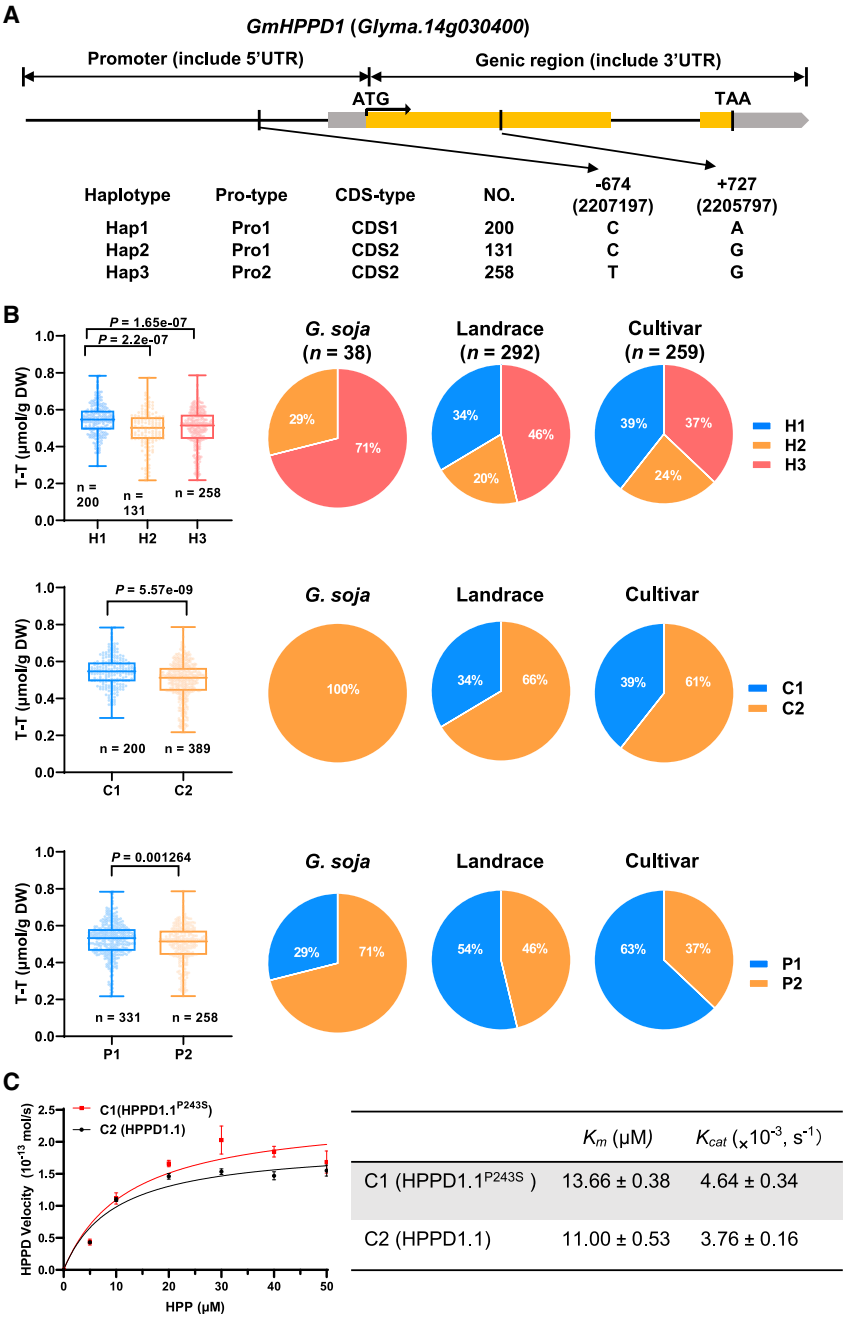
A GWAS of tocopherol traits identified 18 SNP loci associated with individual tocopherols or T-T of soybean seeds. The SNP loci were distributed on 11 different chromosomes: Chr.08, Chr.09, Chr.10, Chr.12, Chr.13, Chr.14, Chr.15, Chr.16, Chr.17, Chr.18, and Chr.19 (Figure 3; Supplemental Tables 4 and 5). According to the linkage disequilibrium threshold of the soybean genome, the region 100-kb upstream and downstream of each peak SNP locus was used to screen candidate genes. As shown in Figure 3, three known tocopherol biosynthetic genes were identified by the metabolic GWAS strategy: *GmHPPD1* was associated with T-T content, and *GmMPBQ-MT1* and *Gmγ-TMT*s were associated with tocopherol composition and/or ratios. In addition to these known tocopherol pathway genes, we also found several transcription factors that may regulate tocopherol biosynthesis and that merit further investigation (Supplemental Table 5).

Genetic variations in *GmHPPD1* and T-T in the soybean population

As shown in Supplemental Table 5, the strongest association signal for T-T was detected on Chr.14, with the lead SNP locus (2 241 904 bp; $P = 6.29 \times 10^{-13}$) located around *GmHPPD1* (4-hydroxyphenylpyruvate dioxygenase, *Glyma.14g030400* from 2 204 142 to 2 206 726 bp). There are two splicing variants of *GmHPPD1*: *GmHPPD1.1* undergoes normal intron splicing and encodes a 490-amino-acid peptide, whereas *GmHPPD1.2* does not undergo intron splicing, resulting in a truncated peptide with 409 amino acids. Biochemical characterization revealed that GmHPPD1.1, rather than GmHPPD1.2, acts as a bona fide HPPD enzyme *in vitro* (Supplemental Figure 2A). The α-T content was significantly increased in *GmHPPD1.1*-overexpressing (OE) hairy roots, although endogenous HGA was undetectable (Supplemental Figure 2B and 2C). Sequence variation in *GmHPPD1.1* was then investigated in a representative population of 589 soybean accessions (259 cultivars, 292 landraces, and 38 soja). There were three representative haplotypes for the *GmHPPD1.1* region based on a combination of one SNP in the promoter region (including the 5′ untranslated region in this study; −674 position, C and T) and one SNP in the genic region (+727 position in the coding region, A and G), which led to substitution of Pro^{243} by Ser^{243} (Figure 4A). The 243rd amino acid is not located in the catalytic cavity of HPPD enzymes (Supplemental Figure 3). The T-T content of the Hap1 type ($n = 200$) was higher than that of the Hap2 type ($n = 131$) and the Hap3 type ($n = 258$). Moreover, the Pro1 and CDS1 (coding sequence 1) types had significantly higher T-T content than the Pro2 and CDS2 types, respectively (Figure 4B). There were no potential superior alleles (Hap1) in the wild soybean population, whereas 34% and 39% were detected in landraces and improved cultivars (Figure 4B). These results suggested that *GmHPPD1.1* was probably selected during soybean domestication. Notably, although there was no significant difference in HPPD catalytic efficiency (K_{cat}/K_m) between the CDS1 (Pro^{243}) and CDS2 (Ser^{243}) proteins, the K_{cat} value of CDS1 was approximately 20% higher than that of CDS2 (46.36 ± 3.41 vs. 37.63 ± 1.64 s^{-1}, $n = 3$; Figure 4C).

Genetic variations in *GmMPBQ-MT1* and γ-T/δ-T in the soybean population

The strongest association signal for γ-T/δ-T was detected on Chr.10, with the lead SNP locus (2 597 916 bp; $P = 1.62 \times 10^{-18}$) located near *GmMPBQ-MT1* (*Glyma.10g030600* from 2 658 064 to 2 661 302 bp). In Arabidopsis, *MPBQ-MT* (*At3g63410*, *VTE3*) was demonstrated to catalyze the conversion of δ-T to γ-T (Cheng et al., 2003). *GmMPBQ-MT1* contains two introns and encodes a 342-amino-acid peptide. Sequence variation in *GmMPBQ-MT1* was investigated in a representative population of 631 soybean accessions (275 cultivars, 320 landraces, 36 soja). There were two representative haplotypes for the *GmMPBQ-MT1* region determined by one SNP (+904 position in the coding region, A and G), which led to the substitution of Ser^{302} by Pro^{302} (Supplemental Figures 4 and 5). The 302nd amino acid (Ser or Pro) of GmMPBQ-MT1 was far from conserved regions such as the *S*-adenosyl-l-methionine binding domain (Supplemental Figure 5B; Cheng et al., 2003). The T-T and δ-T contents in



	K_m (μM)	K_{cat} ($\times 10^{-3}$, s^{-1})
C1 (HPPD1.1^{P243S})	13.66 ± 0.38	4.64 ± 0.34
C2 (HPPD1.1)	11.00 ± 0.53	3.76 ± 0.16

Figure 4. Haplotype effects of two *GmHPPD1.1* SNPs on T-T content.

(A) Haplotypes detected in the 4.6-kb genomic region of *GmHPPD1.1*. Yellow and gray represent exons and untranslated regions, respectively.

(B) T-T content in seeds of different haplotype populations; the frequencies of different haplotypes in wild soybean, landraces, and improved cultivars are shown on the right. The maximum, 75% quartile, median, 25% quartile, and minimum values are represented. The *P* values were calculated using two-tailed Student's *t*-test.

(C) Enzymatic characterization of GmHPPD1.1 and GmHPPD1.1^{P243S}. HPP, 4-hydroxyphenylpyruvate. All data points represent the average and SD from triplicate experiments.

Hap1 accessions (n = 426) were significantly higher than those in Hap2 (n = 205; Supplemental Figure 4B). This result suggested that Pro^{302} is likely to play an important role in MPBQ-MT activity. It has probably been selected during soybean domestication and improvement, given the frequencies of Hap1 (superior allele) in wild soybeans (17%), landraces (62%), and improved cultivars (80%). Notably, there was no significant difference in α-T between the Hap1 and Hap2 populations, which suggested that *GmMPBQ-MT1* haplotypes have little effect on variation in α-T content within the soybean population. We could not compare the enzymatic activity of *GmMPBQ-MT1* Hap1 and Hap2 because MPBQ is not commercially available. However, it is reasonable to propose that Hap1 has higher MPBQ-MT activity than Hap2.

Genetic variations in *Gmγ-TMT3* and α-T in the soybean population

Because α-T has much higher vitamin E activity than δ-T and γ-T (KamalEldin and Appelqvist, 1996) and is especially relevant to human health, we focused on GWAS signals associated with α-T content in this study. The strongest association signal for α-T was detected on Chr.9, with the lead SNP locus (44 341 827 bp; $P = 4.35 \times 10^{-13}$) located near *Gmγ-TMT3* (*Glyma.09g222800* from 44 341 974 to 44 346 311 bp). Another two duplicated γ-*TMT*s (*Gmγ-TMT1*, *Glyma.12g014200* from 1 025 584 to 1 029 095 bp and *Gmγ-TMT2*, *Glyma.12g014300* from 1 033 151 to 1 037 054 bp; Supplemental Table 5) were also associated with α-T content in the soybean population. Tocopherol profiling clearly showed that α-T accumulated to high levels in green tissues, including cotyledons and leaves (Supplemental Figure 6A). We first characterized the three γ-TMTs *in vitro* and *in planta*. *Gmγ-TMT2* was predominately expressed in the leaf and cotyledon, whereas *Gmγ-TMT2* and *Gmγ-TMT3* had comparable expression levels in developing seeds (Supplemental Figure 6B). Sequence analysis showed that only Gmγ-TMT2 has a clear plastid signal peptide, and this was experimentally verified by subcellular assays. Gmγ-TMT1 and Gmγ-TMT3 are cytosol proteins (Supplemental Figure 6C). *In vitro* enzymatic assays showed that Gmγ-TMT3 had three-fold higher activity toward γ-T than Gmγ-TMT1 and Gmγ-TMT2 (Supplemental Figure 6D). All these data suggested that Gmγ-TMT3 plays an important role in α-T production in soybean seeds, whereas Gmγ-TMT2 is responsible for α-T production in the green tissues of soybean plants.

We next analyzed *Gmγ-TMT3* in detail because of its high GWAS signal associated with α-T content and relatively higher expression level in developing seeds. *Gmγ-TMT3* contains six exons and five introns, encoding a 303-amino-acid peptide. Sequence variation in *Gmγ-TMT3* was investigated in a representative population of

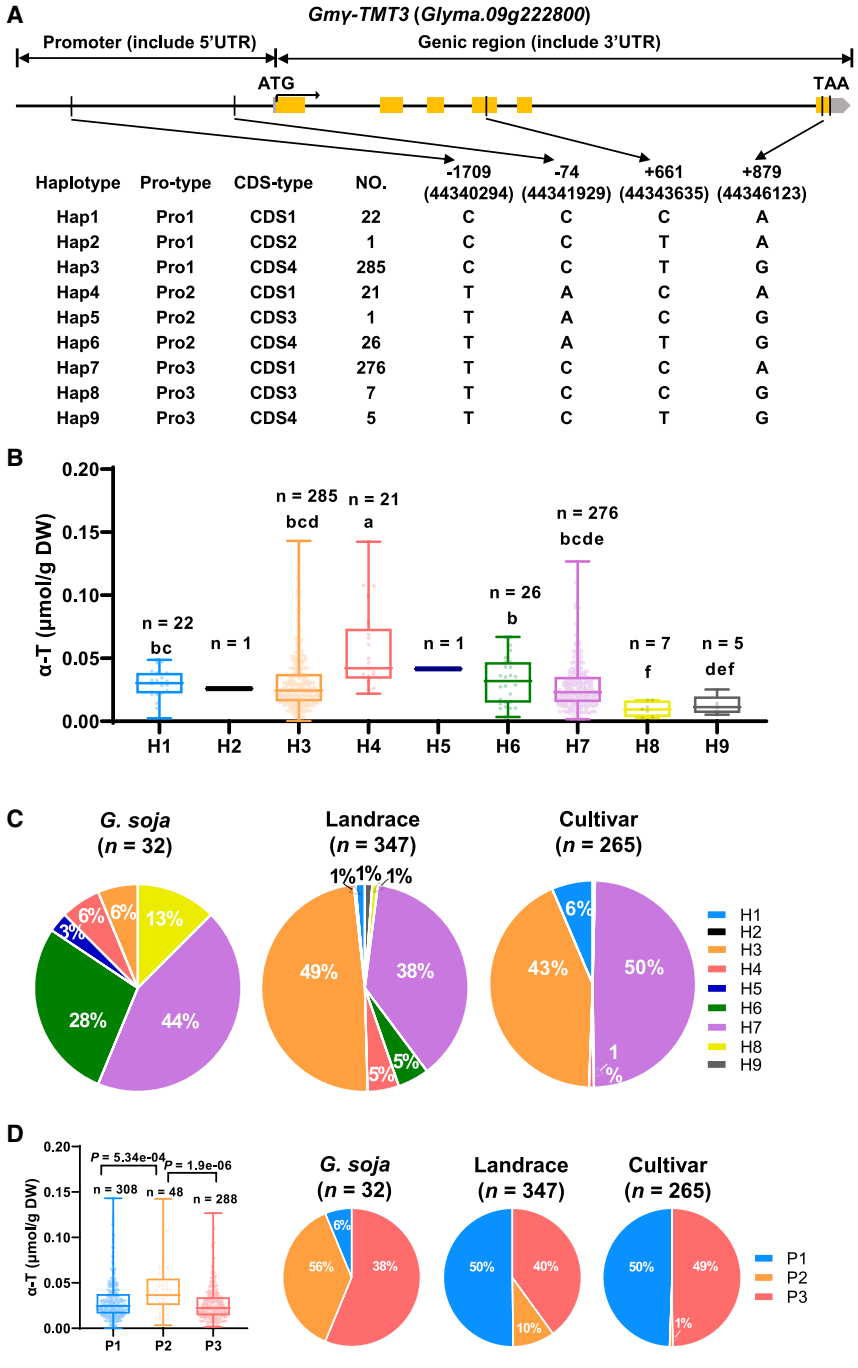


Figure 5. Sequence and allelic variation in *Gmγ-TMT3* among soybean populations.

(A) Haplotypes detected in the 6.3-kb genomic region of *Gmγ-TMT3*. Yellow and gray represent exons and untranslated regions, respectively.

(B) α-T in the seeds of nine different haplotype populations. The maximum, 75% quartile, median, 25% quartile, and minimum values of different soybean populations are presented. Different lowercase letters indicate statistically significant differences at the $P < 0.05$ level determined by Tukey's multiple comparisons test.

(C) Frequencies of different haplotypes in wild soybeans, landraces, and improved cultivars are shown on the left.

(D) α-T in seeds of different haplotype populations; frequencies of different haplotypes in wild soybeans, landraces, and improved cultivars are shown on the left. The maximum, 75% quartile, median, 25% quartile, and minimum values of different soybean populations are presented. The *P* values were calculated using two-tailed Student's *t*-test.

Haplotype	Pro-type	CDS-type	NO.	-1709 (44340294)	-74 (44341929)	+661 (44343635)	+879 (44346123)
Hap1	Pro1	CDS1	22	C	C	C	A
Hap2	Pro1	CDS2	1	C	C	T	A
Hap3	Pro1	CDS4	285	C	C	T	G
Hap4	Pro2	CDS1	21	T	A	C	A
Hap5	Pro2	CDS3	1	T	A	C	G
Hap6	Pro2	CDS4	26	T	A	T	G
Hap7	Pro3	CDS1	276	T	C	C	A
Hap8	Pro3	CDS3	7	T	C	C	G
Hap9	Pro3	CDS4	5	T	C	T	G

644 soybean accessions (265 cultivars, 347 landraces, and 32 soja). A total of 4 SNPs were found in the 6.4-kb genomic region of *Gmγ-TMT3*, including two SNPs in the promoter region (−1709 [P1, C or T] and −74 bp [P2, C or A]) and two SNPs in the genic region (+661 bp in the coding region [C1, C or T (T221 or I221)] and +879 bp in the coding region [C2, A or G (S294 or G294)]), for a total of nine haplotypes (Figure 5A). The average seed α-T content was generally higher in Hap4 (P2C1) accessions ($n = 22$) than in other haplotypes. For the Pro and CDS types, the average seed α-T content was significantly higher in P2-type accessions than in P1 and P3 (Figure 5B–5D). These results suggest that Hap4 and Pro2 are the potential superior alleles of *Gmγ-TMT3* that may improve soybean seed α-T content. Haplotype frequency analysis revealed that soybean cultivars have a lower proportion of potentially superior *Gmγ-TMT3* alleles than landraces and wild soybean. The frequencies of Hap4 in wild soybean, landraces, and improved cultivars were 5%, 6%, and 1%, respectively (Figure 5C), which suggests that the *Gmγ-TMT3* gene was not selected during soybean domestication. Therefore, these superior alleles could be used to qualitatively increase the α-T level in soybean cultivars by marker-assisted selection. We next investigated differences in enzymatic activity between four CDS types (C1 T221/S294, C2 I221/S294, C3 T221/G294, and C4 I221/G294), among which C1 and C4 have been reported previously (Dwiyanti et al., 2011). C2 I221/S294 and C3 T221/G294, the rare haplotypes in the soybean population, had significantly lower *γ-TMT* activity than C1 and C4, and these results were validated by mutagenesis and enzymatic assays (Supplemental Figure 7C and 7D). These data support the previous claim that C1 and C4 are not responsible for the difference in α-T content between a high α-T variety (Keszthelyi Aproszemu Sarga) and a low α-T variety (Ichihime) (Dwiyanti et al., 2011).

The haplotype effects of three tocopherol pathway genes (*GmHPPD1*, *GmMPBQ-MT1*, and *Gmγ-TMT3*) on tocopherol in improved cultivars are summarized in Supplemental Table 6.

Coregulation of tocopherol and FA by *GmZF351* in soybean seeds

The overall positive correlation between T-T and total FA in the soybean population suggested a shared regulatory mechanism for

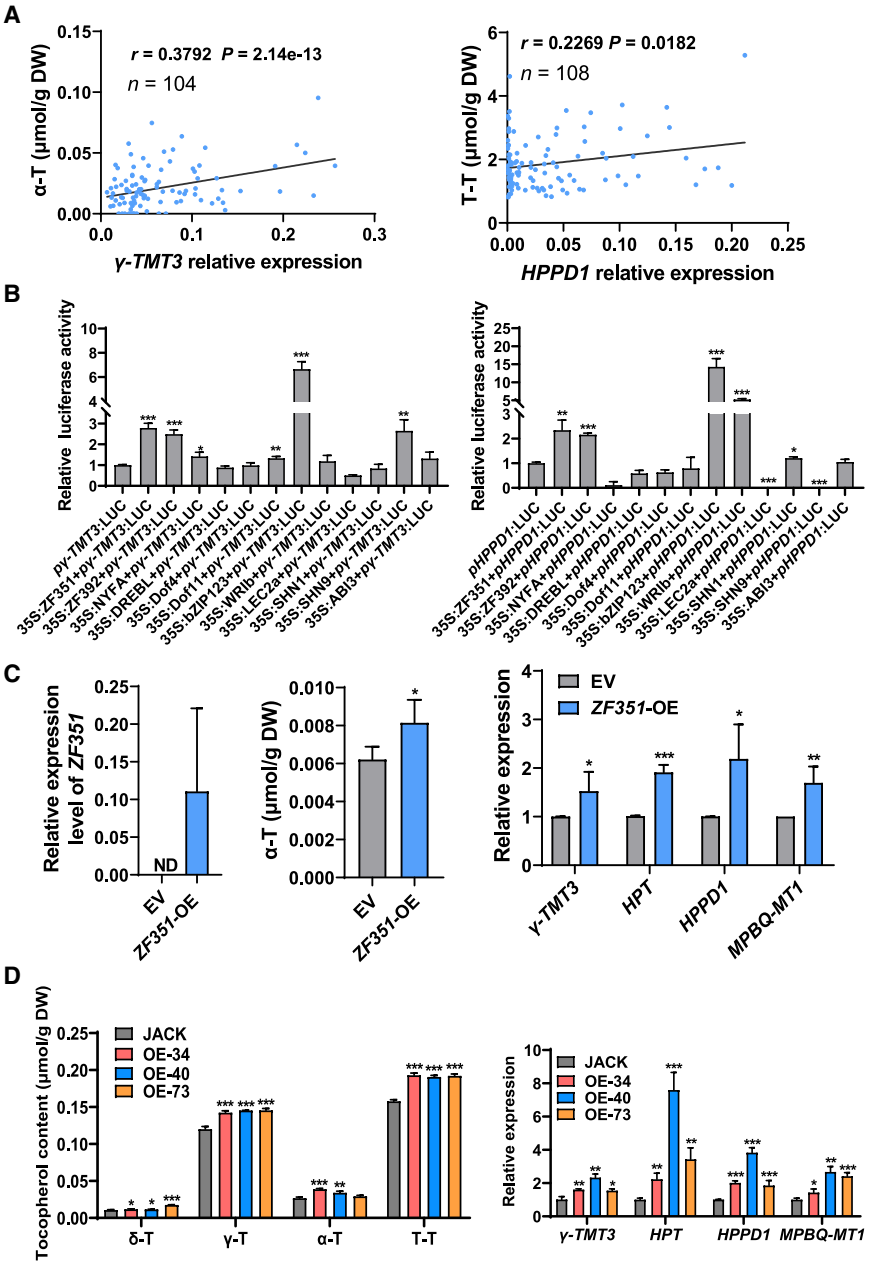


Figure 6. GmZF351 activates tocopherol biosynthesis in soybean seeds.

(A) Correlation analysis of *Gmγ-TMT3* (left panel, data collected from 104 accessions; for details, see Supplemental Table 7) and *GmHPPD1* (right panel, data collected from 108 accessions; for details, see Supplemental Table 8) expression levels and tocopherol content among soybean accessions. DW, dry weight.

(B) Activation assays of *Gmγ-TMT3* (left panel) and *GmHPPD1* (right panel) promoters driven by 12 known FA TFs in an *N. benthamiana* system. Data are presented as mean ± SD (n = 3). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's t-test).

(C) Analysis of α-T (middle panel) and four tocopherol genes (*Gmγ-TMT3*, *GmHPT*, *GmHPPD1*, and *GmMPBQ-MT1*, the right panel) in *GmZF351*-OE soybean hairy roots. DW, dry weight; EV, empty vector, used as control; ND, not detectable. Data are presented as mean ± SD (n = 4). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's t-test).

(D) Analysis of tocopherol contents (δ-T, γ-T, α-T, and T-T in the middle panel) and expression of four tocopherol genes (bottom panel) in green seeds of three independent *GmZF351*-OE soybean lines. Data are presented as mean ± SD (n = 3). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's t-test).

both tocopherol and FA biosynthesis (Figure 2 and Supplemental Figure 1). Variation in the promoter regions of *GmHPPD1* and *Gmγ-TMT3* prompted us to test whether there was a correlation between the expression level of tocopherol pathway genes and tocopherol content in the soybean population. We tested this proposal in a small, randomly selected soybean population that included landraces and improved cultivars (for detailed information, see Supplemental Tables 7 and 8). Positive correlations were detected between *GmHPPD1* and T-T and between *Gmγ-TMT3* and α-T (Figure 6A), suggesting that activation transcription factor(s) are involved in tocopherol biosynthesis in soybean. To date, no tocopherol transcription factor has been functionally identified in plants, although several putative transcription factors have been proposed (Park et al., 2019). By contrast, several types of transcription factors (NYFA, DREBL, zinc finger type, Dof type, and bZIP type) involved in FA biosynthesis have been identified in soybean (Song et al., 2013; Lu et al., 2016, 2021; Li et al., 2017). We first screened the binding and activation ability of 12 known FA transcription factors (TFs) to the promoters of *GmHPPD1* and *Gmγ-TMT3*. At least three TFs, GmZF351 (*Glyma06g290100*), GmZF392 (*Glyma.12G205700*), and GmbZIP123 (*Glyma.06G010200*), clearly activated the luciferase (LUC) gene fused with the *GmHPPD1* or *Gmγ-TMT3* promoter region in an *N. benthamiana* system (Figure 6B). Among these three FA TFs, only *GmZF351* and *GmZF392* have been functionally validated in transgenic soybean plants, and both genes were selected during soybean domestication (Li et al., 2017; Lu et al., 2021). Unfortunately, transcript levels of *GmZF392* and the FA marker gene (*GmTAG1-1*, *Glyma.19g063400*) in four independent *ZF392*-OE lines were comparable to those of control plants, probably because of gene silencing after several generations (Supplemental Figure 8). We therefore did not include *GmZF392* in further analyses.

We first tested the effect of *GmZF351* on tocopherol biosynthesis in soybean hairy roots. Both expression levels of tocopherol biosynthetic genes (*Gmγ-TMT3*, *GmHPPD1*, *GmHPT* [*Glyma.17G061900*], and *GmMPBQ-MT1*) and α-T content were significantly increased in the *GmZF351*-OE hairy roots (Figure 6C). As expected, *GmTAG1-1* expression was also upregulated in the *GmZF351*-OE hairy roots (Supplemental Figure 9). We next analyzed gene expression and

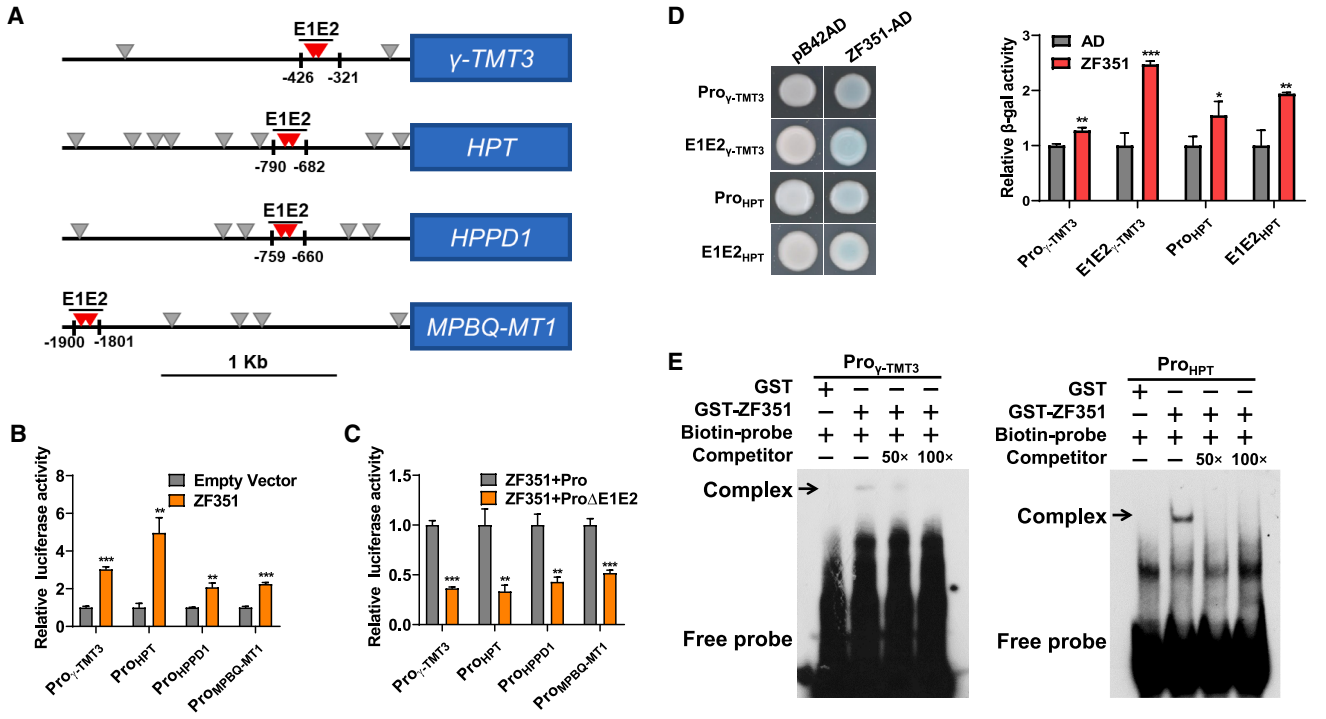


Figure 7. GmZF351 binds and activates the promoters of tocopherol biosynthetic genes.

(A) Distribution of putative E1E2 (elements 1 and 2 marked in red) GmZF351-binding sites in the 2-kb promoter regions of four tocopherol biosynthetic genes. Gray triangles in the promoter region represent single CT(G/C) (T/A)AA elements.

(B) Activation assays of four tocopherol gene promoters by GmZF351 in an *N. benthamiana* system. The relative LUC activities for the empty vector were set to 1. Data are presented as mean ± SD ($n = 3$). $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's *t*-test).

(C) Activation assays of four tocopherol gene promoters with E1E2 deletions (positions marked in the promoters shown in A). The relative LUC activities of full-length promoters were set to 1 in this experiment. Data are presented as mean ± SD ($n = 3$). $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's *t*-test).

(D) Quantitative measurement of the transcription activation ability of GmZF351 on full-length promoters and the E1E2 fragments of *Gmγ-TMT3* and *GmHPT*. Data are presented as mean ± SD ($n = 3$). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's *t*-test).

(E) An EMSA was used to test the physical interactions of E1E2 fragments from *Gmγ-TMT3* and *GmHPT* with GmZF351.

tocopherol content in green seeds (10 days after flowering) of three independent *GmZF351*-OE stable transgenic lines. Both *GmZF351* and *GmTAG1-1* were highly upregulated in seeds of *GmZF351*-OE stable transgenic lines, consistent with a previous report (Li et al., 2017; Supplemental Figure 10A and 10B). Contents of γ-T, δ-T, α-T, and T-T were significantly increased to different extents (approximately 20%–30%) in seeds of *GmZF351*-OE soybean plants (Figure 6D), and expression levels of *Gmγ-TMT3*, *GmHPT*, *GmHPPD1,* and *GmMPBQ-MT1* were upregulated accordingly (Figure 6D). Expression levels of other tocopherol-related genes, including *GmTC* (*Glyma.04G082300*), *Gmγ-TMT1*, and *Gmγ-TMT2,* were also upregulated in *GmZF351*-OE seeds (Supplemental Figure 10C). These results suggested that increased tocopherol content in seeds of *GmZF351*-OE soybean plants was caused by increased expression of tocopherol biosynthetic genes.

Next, we examined the mechanism by which GmZF351 activates the expression of tocopherol biosynthetic genes. To determine the specific binding element for GmZF351, we used an online tool for motif-base sequence analysis (multiple expectation maximization for motif elicitation, http://meme-suite.org) to analyze the GmZF351-bound promoter fragments. In addition, a recent study identified a GmZF351-binding sequence that consisted of two CT(G/C) (T/A)AA elements (designated E1E2), which were usually separated by approximately 100 bp (Wei et al., 2023). As shown in Figure 7A and Supplemental Table 9, putative GmZF351-binding elements (E1E2) were found in the 2-kb promoter regions of *Gmγ-TMT3*, *GmHPT*, *GmMPBQ-MT1,* and *GmHPPD1.* We then tested whether GmZF351 directly binds to the promoter regions of tocopherol genes using the *N. benthamiana* system. GmZF351 bound directly to the *Gmγ-TMT3*, *GmHPT*, *GmMPBQ-MT1,* and *GmHPPD1* promoters, as reflected by two-fold to five-fold upregulation of relative luciferase activity (Figure 7B). Activation by GmZF351 was significantly decreased by approximately 50%–60% when the E1E2 sequences were deleted from the promoters of the four tested tocopherol biosynthetic genes (Figure 7C). The importance of E1E2 sequences for GmZF351 binding and activation was also confirmed using the yeast one-hybrid (Y1H) system: the E1E2 fragment showed higher β-gal activity than the full-length promoters of *Gmγ-TMT3* and *GmHPT* (Figure 7D). Positive signals for *GmHPPD1* and *GmMPBQ-MT1* promoters were not detected in Y1H assays for an unknown reason. An electrophoretic mobility shift assay (EMSA) clearly showed that GmZF351 could physically interact with E1E2 probes from promoters of *Gmγ-TMT3* and *GmHPT* (Figure 7E; for probe sequences, see Supplemental Table 9). Together, these experiments demonstrate that GmZF351 can bind to the promoters of key tocopherol-related genes and activate their expression *in vitro* and *in planta*.

Although we did not detect a strong GWAS signal on Chr.6 ($P > 1 \times 10^{-8}$) where *GmZF351* is located, we investigated the haplotype of *GmZF351* in a representative population of 585 soybean accessions (253 cultivars, 291 landraces, and 41 soja). Three representative haplotypes for the *GmZF351* region are determined by two SNPs (+155 position in the coding region, A and G; +340 position in the coding region, A and G), which lead to missense mutations (Ile52Thr and Ala114Cys; Supplemental Figure 11A). The T-T and individual tocopherol contents in Hap1 accessions (n = 43) did not differ from those in Hap2 (n = 200), but T-T and individual tocopherols were significantly lower in Hap2 accessions than in Hap3 accessions by approximately 7% (Supplemental Figure 11B). The frequency of superior alleles (Hap3) was increased during soybean domestication from soja (15%) to landraces (59%) and improved cultivars (65%), whereas Hap2 showed the opposite trend, decreasing from soja (83%) to landraces (34%) and improved cultivars (27%).

In summary, total vitamin E levels in common plant seed oils are species dependent, ranging from dozens to several thousand μg/g oil (DellaPenna, 2005). Tocopherol is an essential nutrient in the human diet, and improvement of tocopherol content, especially α-T content, is an important breeding goal for increasing the nutritional quality of soybean oil. Many studies have demonstrated that γ-*TMT* genes play a key role in α-T in various plant species (Grusak and DellaPenna, 1999). Moreover, overexpression of γ-*TMT*, especially together with upstream genes, leads to enhanced T-T and α-T contents in soybean seeds (Van Eenennaam et al., 2003; Tavva et al., 2007; Konda et al., 2020). However, traditional breeding remains a more publicly acceptable strategy for development of soybean cultivars with high vitamin E levels. In this study, we identified three tocopherol biosynthetic genes (*GmHPPD1*, *GmMPBQ-MT1*, and *Gmγ-TMT3*) and one TF gene (*GmZF351*) that are mainly responsible for tocopherol variation in a natural soybean population. This study provides foundational information for breeding new soybean varieties with high α-T and high-quality oils. In addition to *GmZF351*, several more TFs were screened out from the tocopherol GWAS analysis (shown in Supplemental Table 5) and merit testing to determine whether they also regulate tocopherol biosynthesis in soybean plants.

METHODS

Soybean materials

Soybean seeds were selected and sown in deeply plowed fields with proper moisture content (15%–20%). The seeds were planted in three-row plots in a randomized complete block design with three replications for each environment. Only one accession was planted in each plot, and the plots were 5 m in length with a row spacing of 0.4 m. The space between two plots was 0.4 m. After 3 weeks, the seedlings were manually thinned to achieve an equal density of 120 000 individuals per hectare.

A total of 727 accessions used for tocopherol profiling (318 cultivars, 358 landraces, and 51 soja; for detailed information, see Supplemental Table 1) were planted at the Experimental Station of the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing (40°22′N and 116°23′E) during the summer season in 2013.

A total of 807 soybean accessions (341 cultivars, 413 landraces, and 53 soja; for detailed information, see Supplemental Table 2) were used for tocopherol profiling in 2014. Among them, 145 accessions collected from northern areas were planted in Mudanjiang (44°58′N and 129°60′E), Heilongjiang Province, during the summer season. A total of 603 accessions collected from Huang Huai Hai and southern areas were planted in Zhoukou (33°62′N and 114°65′E), Henan Province, during the summer season. Fifty-three soja were planted at the Experimental Station of the Institute of Genetics and Developmental Biology, Beijing. Information on growth location was missing for six soybean accessions.

The dry soybean seeds were ground into a fine powder, then stored at −80°C until needed for tocopherol measurement. The samples were stored at −80°C for less than three months before liquid chromatography (LC) analysis.

Tocopherol quantification in soybean seeds

In brief, 1 ml 95% acetone was added to approximately 200 mg DW seed powder in a 2-ml Eppendorf tube. The tocopherols were extracted by rotation for 30 min at 4°C. The supernatant (2 μl injection volume for each sample) was collected after centrifugation at 13 000 × g for 10 min and analyzed by LC (Agilent 1290 Infinity, Santa Clara, CA, USA) with a DAD detector (Agilent G4212A) at 295 nm. Tocopherols were resolved on a ZORBAX Eclipse Plus C18 column (2.1 mm × 50 mm × 1.8 μm; Agilent), and the column temperature was maintained at 35°C during the analytical process. The mobile phase comprised 80% acetonitrile and 20% isopropanol, and the flow rate was 0.3 ml/min. Standard curves were generated with each tocopherol (1, 10, 20, 50, 80, and 100 μM) using the same analytical method. All data for each soybean accession were the average value of two duplicate measurements.

LC–mass spectrometry (MS) analysis for HPP and HGA

HPP and HGA were measured using an ultrahigh performance LC quadrupole time-of-flight MS analytical platform consisting of an Agilent 1290 Infinity LC pump and a 6550 single quadrupole mass spectrometer (Agilent). A ZORBAX Extend C18 column (100 × 2.1 mm internal diameter, 1.8 μm; Agilent) was used for HGA measurement, and a solvent mixture of 0.1% formic acid in water (solution A) and acetonitrile (solution B) was used as the mobile phase with a linear gradient. The step gradient was as follows: 1:99 v/v (0 min), 1:99 v/v (1 min), 99.05:0.5 v/v (15.5 min), 99.05:0.5 v/v (17 min), 0.4 ml/min, mobile phase temperature 35°C. The capillary voltage was 4000 V, the carrier gas temperature was 225°C, the dry gas flow rate was 13 l/min, and the sheath gas flow rate was 12 l/min. Negative ion mode was used for data acquisition. The HGA content was calculated using a standard curve.

Tocopherol GWAS analysis

SNP data from our previous study (Fang et al., 2017) were used to analyze the genetic diversity of tocopherol-related genes in soybean. Resequencing data from 53 soja accessions (approximately 1× coverage) were also included in the GWAS. A total of 6 291 929 SNPs with a minor allele frequency >1% and a missing rate <10% were used for association analysis. The soybean accessions were divided into three populations: soja,

landrace, and cultivar. The overall chemical contents of the soybean accessions (2 years of data) were also calculated using the best linear unbiased prediction for GWAS analysis. GWAS was carried out based on a mixed linear model using the EMMAX software package (Kang et al., 2010). The phenotypic variance explained (PVE) of the lead SNP was calculated using Genome Association and Prediction Integrated Tool (https://zzlab.net/GAPIT/).

Haplotype analyses

SNPs with a minimum minor allele frequency of 0.05 were selected for haplotype analysis, as were those located in the exon of a CDS that resulted in a nonsynonymous mutation that could potentially affect enzyme activity. In addition, SNPs in the 2-kb promoter region including the 5′ untranslated region were selected based on the presence of a predicted TF binding motif from the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Nucleic acid isolation and quantitative reverse transcription PCR

Genomic DNA was isolated from fresh tender leaves by the CTAB method (Murray and Thompson, 1980). To analyze the expression patterns of tocopherol biosynthetic genes at the S5 stage, seeds of more than 100 soybean accessions were collected from the Experimental Station of IGDB, CAS, in 2020 (detailed in Supplemental Tables 7 and 8) and ground to a fine powder in liquid nitrogen using a tissue homogenizer. RNA extraction, reverse transcription, and quantitative reverse transcription PCR were performed as described previously (He et al., 2019), and *GmUBC* (encoding putative ubiquitin-conjugating enzyme E2; *Glyma.18G051400*) was used as an internal control gene. All primers used in the study are listed in Supplemental Table 10.

Protein expression, purification, and enzymatic assays

The targeted tocopherol gene was ligated into a pMAL-c2x expression vector and expressed in *E. coli*, and recombinant Gmγ-TMT and GmHPPD proteins were purified by dextrin Sepharose affinity chromatography. Detailed primer information is provided in Supplemental Table 10.

To verify the activity of Gmγ-TMT, a 100-μl reaction containing 10 μl 1 mM γ-T, 10 μl 2 mM *S-adenosyl-l-methionine*, 10 μl 0.5 M Tris (pH 7.5), and 10 μg purified γ-TMT protein was incubated at 28°C for 2 h. Then, 100 μl 95% acetone was added to terminate the reaction. After centrifugation for 5 min, the product was detected by the LC-DAD method described above.

GmHPPD1.1 assays were carried out as described previously (Garcia et al., 1999). In brief, a 100-μl reaction containing 2 μl 1 mM HPP, 10 μl 500 mM sodium ascorbate, 10 μl 1 M Tris (pH 7.5), and 5 μg purified GmHPPD1.1 protein was incubated at 30°C for 30 min. Then, 100 μl 80% methanol was added to terminate the reaction. The product was analyzed by LC-MS as described above after centrifugation at top speed for 3 min. To determine the kinetic parameters of GmHPPDs, an appropriate concentration of purified protein was incubated with various concentrations of HPP and saturating concentrations of sodium ascorbate for 30 min at 30°C in a 100-μl volume. All kinetic parameters of GmHPPDs in this study were calculated from a Hanes Plot (Hyper 32 software, v.1.0.0).

Soybean hairy root system

Constructs with PROKII as the backbone vector were transferred into *Agrobacterium rhizogenes* strain K599 to infect 5-day-old soybean (Kefeng-1) seedlings and obtain transgenic soybean hairy roots as described previously (Cao et al., 2008).

Promoter activity assays in *N. benthamiana*

An approximately 2-kb promoter fragment upstream of the tocopherol-related genes translation start site was amplified from genomic DNA. The amplified fragments were cloned into the multiple cloning sites of pGreenII 0800-LUC to produce the target constructs. The effector construct was the above-mentioned 35S:GmZF351. All constructs were sequenced and introduced into Agrobacterium tumefaciens GV3101 (with pSoup-p19 plasmid). Firefly and Renilla luciferase activities were quantified using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA) as described previously (Hellens et al., 2005). The relative LUC activity levels were quantified as the ratio of LUC/Renilla enzyme activities. Agrobacteria harboring empty effector vector were coinfiltrated with reporters as a negative control.

Y1H assays

Y1H assays were performed as described previously (Xu et al., 2022). In brief, pLacZi-*tocopherol related gene-pro* (reporter) and pB42AD-*ZF351* (effector) were cotransformed into yeast strain EGY48, and the transformants were plated on minimal synthetic defined medium with −Ura/−Trp dropout mix and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) for blue color development. Quantification of β-galactosidase activity was performed as described in the Yeast Protocols Handbook (BD Clontech, Palo Alto, CA, USA).

EMSA

Single-stranded complementary fragments containing the E1E2 motif were synthesized and labeled with biotin at their 5′ ends (Invitrogen; for primers, see Supplemental Table 10). Double-stranded DNA probes were obtained by heating complementary fragments at 75°C for 30 min and then slowly cooling at room temperature. An EMSA was performed as described by Bian et al. (2020) using the LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific, Waltham, MA, USA).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was financially supported by the National Key Research and Development Program of China (2021YFF1001200 and 2018YFA0900600), the Chinese Academy of Science (XDA24040202), and the State Key Laboratory of Plant Genomics of China (grant no. SKLPG2016A-13).

AUTHOR CONTRIBUTIONS

G.W. and D.C. designed the research. D.C. carried out most experiments in this study. Z.Z., C.F., and Z.T. prepared the soybean samples and carried out the GWAS analysis. D.C. and J.Y. carried out the tocopherol and FA analysis. X.X. prepared the figures. Y.H. carried out the EMSA

experiments. J.Z. generated the transgenic soybean plants. D.C., Z.T., and G.W. analyzed the data and wrote the manuscript with input from all coauthors.

ACKNOWLEDGMENTS

No conflict of interest is declared.

Received: October 31, 2022
Revised: March 24, 2023
Accepted: April 4, 2023
Published: April 7, 2023

REFERENCES

Albert, E., Kim, S., Magallanes-Lundback, M., Bao, Y., Deason, N., Danilo, B., Wu, D., Li, X., Wood, J.C., Bornowski, N., et al. (2022). Genome-wide association identifies a missing hydrolase for tocopherol synthesis in plants. Proc. Natl. Acad. Sci. USA **119**. e2113488119.

Armutcu, F., Coskun, O., Gurel, A., Sahin, S., Kanter, M., Cihan, A., Numanoglu, K.V., and Altinyazar, C. (2005). Vitamin E protects against acetone-induced oxidative stress in rat red blood cells. Cell Biol. Toxicol. **21**:53–60.

Bian, X.H., Li, W., Niu, C.F., Wei, W., Hu, Y., Han, J.Q., Lu, X., Tao, J.J., Jin, M., Qin, H., et al. (2020). A class B heat shock factor selected for during soybean domestication contributes to salt tolerance by promoting flavonoid biosynthesis. New Phytol. **225**:268–283.

Cao, D., Hou, W., Song, S., Sun, H., Wu, C., Gao, Y., and Han, T. (2008). Assessment of conditions affecting Agrobacterium rhizogenes-mediated transformation of soybean. Plant Cell Tiss. Organ Cult. **96**:45–52.

Cheng, Z., Sattler, S., Maeda, H., Sakuragi, Y., Bryant, D.A., and DellaPenna, D. (2003). Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. Plant Cell **15**:2343–2356.

DellaPenna, D. (2005). A decade of progress in understanding vitamin E synthesis in plants. J. Plant Physiol. **162**:729–737.

Dror, D.K., and Allen, L.H. (2011). Vitamin E deficiency in developing countries. Food Nutr. Bull. **32**:124–143.

Dwiyanti, M.S., Yamada, T., Sato, M., Abe, J., and Kitamura, K. (2011). Genetic variation of gamma-tocopherol methyltransferase gene contributes to elevated alpha-tocopherol content in soybean seeds. BMC Plant Biol. **11**:152.

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G., Zhou, Z., Yu, H., Zhang, M., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol. **18**:161.

Garcia, I., Rodgers, M., Pepin, R., Hsieh, T.F., and Matringe, M. (1999). Characterization and subcellular compartmentation of recombinant 4-hydroxyphenylpyruvate dioxygenase from Arabidopsis in transgenic tobacco. Plant Physiol. **119**:1507–1516.

Ghosh, S., Zhang, S., Azam, M., Qi, J., Abdelghany, A.M., Shaibu, A.S., Gebregziabher, B.S., Feng, Y., Huai, Y., Htway, H.T.P., et al. (2021). Seed tocopherol assessment and geographical distribution of 1151 Chinese soybean accessions from diverse ecoregions. J. Food Compost. Anal. **100**:103932.

Goffman, F.D., and Böhme, T. (2001). Relationship between fatty acid profile and vitamin E content in maize hybrids (Zea mays L.). J. Agric. Food Chem. **49**:4990–4994.

Grusak, M.A., and DellaPenna, D. (1999). Improving the nutrient composition of plants to enhance human nutrition and health. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50**:133–161.

Havaux, M., Eymery, F., Porfirova, S., Rey, P., and Dörmann, P. (2005). Vitamin E protects against photoinhibition and photooxidative stress in *Arabidopsis thaliana*. Plant Cell **17**:3451–3469.

He, J., Chen, Q., Xin, P., Yuan, J., Ma, Y., Wang, X., Xu, M., Chu, J., Peters, R.J., and Wang, G. (2019). CYP72A enzymes catalyse 13-hydrolyzation of gibberellins. Nat. Plants **5**:1057–1065.

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods **1**:13.

Kamal-Eldin, A. (2006). Effect of fatty acids and tocopherols on the oxidative stability of vegetable oils. Eur. J. Lipid Sci. Technol. **108**:1051–1061.

KamalEldin, A., and Appelqvist, L.A. (1996). The chemistry and antioxidant properties of tocopherols and tocotrienols. Lipids **31**:671–701.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42**:348–354.

Konda, A.R., Nazarenus, T.J., Nguyen, H., Yang, J., Gelli, M., Swenson, S., Shipp, J.M., Schmidt, M.A., Cahoon, R.E., Ciftci, O.N., et al. (2020). Metabolic engineering of soybean seeds for enhanced vitamin E tocochromanol content and effects on oil antioxidant properties in polyunsaturated fatty acid-rich germplasm. Metab. Eng. **57**:63–73.

Li, Q.T., Lu, X., Song, Q.X., Chen, H.W., Wei, W., Tao, J.J., Bian, X.H., Shen, M., Ma, B., Zhang, W.K., et al. (2017). Selection for a zinc-finger protein contributes to seed oil increase during soybean domestication. Plant Physiol. **173**:2208–2224.

Li, Y., Hussain, N., Zhang, L., Chen, X., Ali, E., and Jiang, L. (2013). Correlations between tocopherol and fatty acid components in germplasm collections of Brassica oilseeds. J. Agric. Food Chem. **61**:34–40.

Lu, L., Wei, W., Li, Q.T., Bian, X.H., Lu, X., Hu, Y., Cheng, T., Wang, Z.Y., Jin, M., Tao, J.J., et al. (2021). A transcriptional regulatory module controls lipid accumulation in soybean. New Phytol. **231**:661–678.

Lu, X., Li, Q.T., Xiong, Q., Li, W., Bi, Y.D., Lai, Y.C., Liu, X.L., Man, W.Q., Zhang, W.K., Ma, B., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. Plant J. **86**:530–544.

Méjean, M., Brunelle, A., and Touboul, D. (2015). Quantification of tocopherols and tocotrienols in soybean oil by supercritical-fluid chromatography coupled to high-resolution mass spectrometry. Anal. Bioanal. Chem. **407**:5133–5142.

Mene-Saffrane, L. (2018). Vitamin E biosynthesis and its regulation in plants. Antioxidants **7**:2.

Muñoz, P., and Munné-Bosch, S. (2019). Vitamin E in plants: biosynthesis, transport, and function. Trends Plant Sci. **24**:1040–1051.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Res. **8**:4321–4325.

Ohlrogge, J., and Browse, J. (1995). Lipid biosynthesis. Plant Cell **7**:957–970.

Park, C., Dwiyanti, M.S., Nagano, A.J., Liu, B., Yamada, T., and Abe, J. (2019). Identification of quantitative trait loci for increased alpha-tocopherol biosynthesis in wild soybean using a high-density genetic map. BMC Plant Biol. **19**:510.

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M., and DellaPenna, D. (2004). Vitamin E is essential for seed longevity,

and for preventing lipid peroxidation during germination. Plant Cell **16**:1419–1432.

Song, Q.X., Li, Q.T., Liu, Y.F., Zhang, F.X., Ma, B., Zhang, W.K., Man, W.Q., Du, W.G., Wang, G.D., Chen, S.Y., et al. (2013). Soybean *GmbZIP123* gene enhances lipid content in the seeds of transgenic Arabidopsis plants. J. Exp. Bot. **64**:4329–4341.

Tavva, V.S., Kim, Y.H., Kagan, I.A., Dinkins, R.D., Kim, K.H., and Collins, G.B. (2007). Increased alpha-tocopherol content in soybean seed overexpressing the Perilla frutescens gamma-tocopherol methyltransferase gene. Plant Cell Rep. **26**:61–70.

Traber, M.G. (2007). Vitamin E regulatory mechanisms. Annu. Rev. Nutr. **27**:347–362.

Van Eenennaam, A.L., Lincoln, K., Durrett, T.P., Valentin, H.E., Shewmaker, C.K., Thorne, G.M., Jiang, J., Baszis, S.R., Levering, C.K., Aasen, E.D., et al. (2003). Engineering vitamin E content: from Arabidopsis mutant to soy oil. Plant Cell **15**:3007–3019.

Wang, H.W., Zhang, B., Hao, Y.J., Huang, J., Tian, A.G., Liao, Y., Zhang, J.S., and Chen, S.Y. (2007). The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants. Plant J. **52**:716–729.

Wei, W., Lu, L., Bian, X.H., Li, Q.T., Han, J.Q., Tao, J.J., Yin, C.C., Lai, Y.C., Li, W., Bi, Y.D., et al. (2023). Zinc-finger protein GmZF351 improves both salt and drought stress tolerance in soybean. J. Integr. Plant Biol. https://doi.org/10.1111/jipb.13474.

Wong, Y.F., Makahleh, A., Saad, B., Ibrahim, M.N.M., Rahim, A.A., and Brosse, N. (2014). UPLC method for the determination of vitamin E homologues and derivatives in vegetable oils, margarines and supplement capsules using pentafluorophenyl column. Talanta **130**:299–306.

Xu, M., Wang, X., Liu, J., Jia, A., Xu, C., Deng, X.W., and He, G. (2022). Natural variation in the transcription factor REPLUMLESS contributes to both disease resistance and plant growth in Arabidopsis. Plant Commun. **3**:100351.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. **33**:408–414.