Chen *et al.*
Biotechnology for Biofuels and Bioproducts (2025) 18:23
https://doi.org/10.1186/s13068-025-02620-x

Biotechnology for Biofuels and Bioproducts

RESEARCH

Open Access

Seed-specific expression of phosphatidate phosphohydrolases increases soybean oil content and seed weight

Beibei Chen[1,2,3*], Jianwu Li[2,3], Shuaibing Yao[2,3], Geliang Wang[2,3] and Xuemin Wang[2,3*]

Abstract

Background Soybean is a major oil crop and a primary protein source for livestock, and soybean oil is the most common input for biodiesel. Identifying genes that enhance soybean yield and oil content is crucial for breeding programs. Phosphatidic acid (PA) phosphohydrolase (PAH), which dephosphorylates PA to diacylglycerol (DAG), plays a critical role in lipid synthesis, and yet their potential in improving agronomic traits of oil crops remains unexplored.

Results This study shows that seed-specific expression of *AtPAH1/2* enhances PA turnover into DAG and triacylglycerol (TAG) accumulation in soybean seeds. *PAH* overexpression upregulated the expression of DAG acyltransferase (*DGAT*) but suppressed phospholipid: DAG acyltransferase (*PDAT*). In addition, seed-specific expression of *AtPAH1/2* increases soybean seed size and weight. Furthermore, analysis of the variation of the soybean PAHs in 4414 soybean accessions indicated that the advantageous effects of *GmPAH*s on oil content and seed weight were selected during domestication.

Conclusion These findings suggest that targeting *PAH*s represents a promising strategy for enhancing soybean seed oil content and yield in current cultivars and landraces soybeans.

Keywords PA phosphohydrolase, Soybean oil content, Seed weight, Seed-specific expression

Background

Soybeans are a crucial global source of vegetable proteins and oils, accounting for approximately 22% of world fat and oil production. Soybean oil is the most common source of biodiesel, accounting for over 40% of the total feedstock used to make biomass-based diesel. However, soybean seeds currently contain only around 20% oil, indicating significant potential for improvement compared to other oilseed crops, such as sunflower (40–50%) and rapeseed (35–50%) [1]. Genes involved in fatty acid synthesis have been extensively investigated to enhance plant seed oil content [2, 3]. Phosphatidic acid (PA) phosphohydrolases (PAH) catalyze an Mg^{2+}-dependent dephosphorylation of PA to diacylglycerol (DAG) that can be converted to triacylglycerol (TAG) by either DAG acyltransferase (DGAT) or phospholipid: DAG acyltransferase (PDAT). DAG also serves as a precursor for the synthesis of membrane phospholipids, such as phosphatidylcholine (PC) and phosphatidylethanolamine (PE) by DAG-choline or -ethanolamine phosphotransferase. Additionally, PA contributes to the synthesis of phosphatidylglycerol (PG), phosphatidylinositols (PI), and phosphatidylserines (PS) via the cytidine diphosphate diacylglycerol (CDP-DAG) pathway [4]. Therefore,

*Correspondence:
Beibei Chen
chenbeibei2013@163.com
Xuemin Wang
wangxue@umsystem.edu
[1] State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Zhengzhou 450046, China
[2] Department of Biology, University of Missouri-St. Louis, St. Louis, MO 63121, USA
[3] Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

PAH plays a central role in regulating the levels of membrane phospholipids and storage lipids, affecting various aspects of plant growth, development, and stress response [5–8].

PAH is localized at the cytosolic and membrane fractions of cells [9]. In yeast and mammalian cells, PAH1 and PAH2 serve distinct physiological roles. PAH1, encoded by *phosphatidate phosphatase 1*, is involved in de novo lipid synthesis, whereas PAH2 participates in lipid signaling [10–14]. In Arabidopsis, PAH1 and PAH2 play redundant, important roles in phospholipid biosynthesis [7]. Studies have shown that *pah1pah2* double-mutants exhibit increased phospholipid content due to enhanced net PC biosynthesis and reduced TAG synthesis [5]. Mutants lacking *PAH* displayed elevated PA levels and retarded growth phenotypes in both plants and animals [7, 15–17]. Under phosphate-deficient conditions, *pah1/pah2* exhibited retarded plant growth and glycolipid synthesis compared to phosphate-sufficient conditions [17]. Overall, these findings suggest that PAH plays a significant role in regulating cell size and proliferation under normal and stressful conditions.

Recent findings have revealed a positive correlation between seed oil content and seed size, suggesting a shared genetic control over these traits [18–21]. The shape and size of membrane-bound organelles undergo dynamic and remarkable changes throughout the cell cycle and developmental stages, facilitated by lipid bilayer components [22, 23]. In yeast, the PAH homolog plays a regulatory role in the growth of nuclear and ER membranes during the cell cycle [16]. Additionally, PAH is implicated in cell cycle regulation through its modulation of TAG or phospholipid levels, with its activity repressed by the cell cycle regulator cyclin-dependent kinase A-1 (CDKA,1) [10, 24–27]. However, the impact of PAH on crop organ size remains poorly understood. In this study, we introduced Arabidopsis *PAH1* and *PAH*2 (*AtPAH1/2*) genes driven by a seed-specific promoter into soybean to investigate the role of PAHs in seed oil accumulation and seed development. Our results indicate that seed-specific expression of *AtPAH* genes influences lipid metabolism and seed size.

Results

Overexpression of PAHs increased TAG accumulation in soybean seeds

To investigate the role of PAH in soybean seeds, we introduced *AtPAH1* and *AtPAH2* under the control of the seed-specific promoters of *glycinin* and *β-conglycinin,* respectively (Fig. 1A). Genomic DNA analysis confirmed

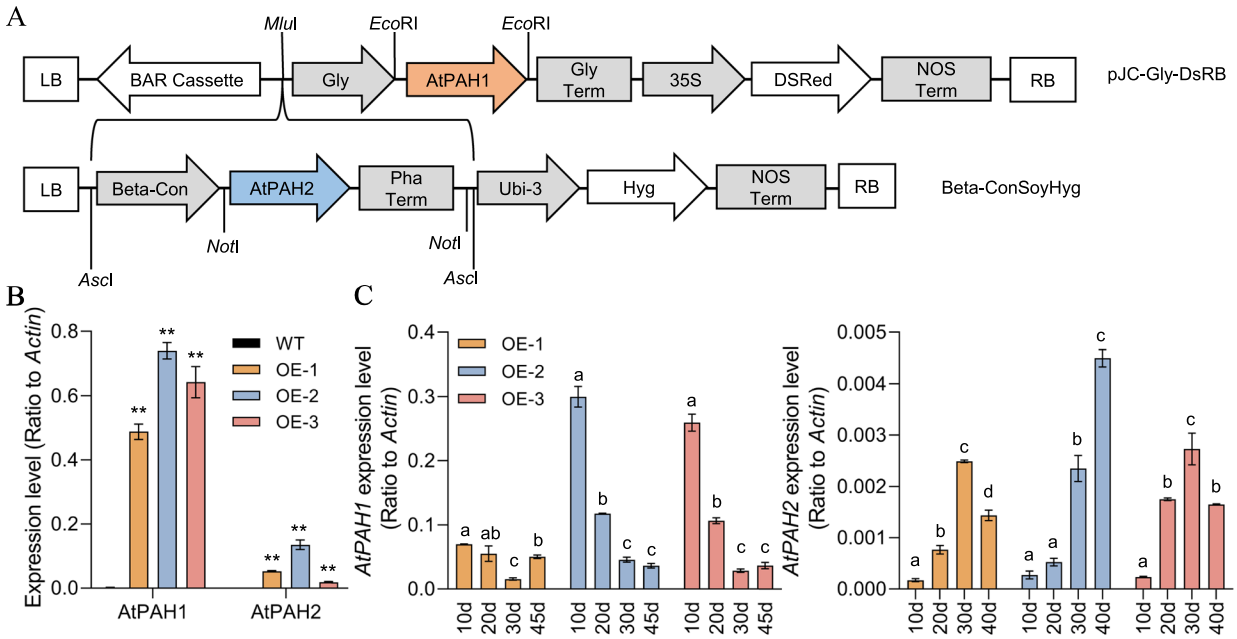

Fig. 1 Expression of *AtPAHs* in soybean with seed-specific promoters. **A** Construction of *AtPAHs* overexpression in soybean. *AtPAH1* is driven by the *glycinin* promoter and *AtPAH2* is promoted by the *β-conglycinin* promoter. **B** Transcript level of *AtPAHs* in developing seeds. Data are means ± SD ($n = 3$). Differences between WT and overexpression lines were analyzed, $^{*}P < 0.05$; $^{**}P < 0.01$ using Student's *t*-test. **C** Transcription level of *AtPAH1* and *AtPAH2* in transgenic developing soybean seeds. RNA was extracted from T5 generation developing seeds at 10, 20, 30, and 45 days after flowering. Data are mean ± SD ($n = 3$). ANOVA (analysis of variance) was employed to assess differences in *AtPAHs* gene expression across various developmental stages. A *p*-value of < 0.05 was considered statistically significant

the insertion of *AtPAH1* and *AtPAH2* into the soybean genome across three independent transgenic lines (Supplemental Fig. 1). To assess gene expression during seed development, RT-qPCR was performed on stable transgenic soybean lines (Fig. 1B). The expression levels of *AtPAH1* and *AtPAH2* varied among the lines, with the highest expression observed in transgenic line 2 (*OE-2*). During seed development, *AtPAH1,* driven by the glycinin promoter, displayed higher expression in the early stages of developing seeds, such as 10 and 20 days after flowering. In contrast, *AtPAH2,* driven by the *β-conglycinin* promoter, exhibited higher levels of expression at 30 and 45 days after flowering (Fig. 1C).

Given that PAH catalyzes a pivotal step in glycerolipid synthesis by converting PA to DAG, a precursor for TAG production, we examined whether seed-specific expression of *AtPAH1/2* influenced DAG and TAG accumulation in soybean seeds. Lipids were extracted from developmental seeds (10, 20, 30, and 45 days after flowering) and mature seeds (60 days after flowering) (Fig. 2C). TAG levels in developing (45 d) and mature *AtPHA1/2*-expressed seeds (219.8 ± 3.3 and 329.6 ± 2.7 mg per g, respectively) were 20.9% and 6.7% higher compared to wild-type (WT) seeds (181.7 ± 10.7 and 308.8 ± 4.2 mg per g, respectively) (Fig. 2A). DAG content was elevated in *AtPAH*-expressed lines at 10, 20, and 30 days, and the level went down at 45 days, which could mean efficient DAG turnover to TAG in older seeds. We assessed the transcription level of the key TAG biosynthesis genes, *GmDGAT1A (Glyma.13g106100)* and *GmPDAT1B (Glyma.13g108100),* which are known to have the highest expression in soybean seeds [28]. The results showed that the expression levels of *GmDGAT1A* began rising at 20 days and peaked at 30 days. The expression levels of *GmDGAT1A* were higher in the developing seeds of *AtPAH1/2*-expressed soybean at 30 days and 45 days compared to WT (Fig. 2D). On the other hand, *GmPDAT1B* expression was lower in *AtPAH1/2*-expressed seeds than WT (Fig. 2E). Additionally, a time-course analysis of fatty acid composition indicated no significant changes in fatty acid composition in *AtPAH1/2*-expressed seeds (Supplemental Fig. 2).

Polar lipid composition changed in developing seeds

To explore the effect of ectopic expression of *AtPAH1/2* on polar lipid metabolism in transgenic lines, lipid extraction was conducted from seeds harvested at 10, 20, 30, and 45 days after flowering, followed by mass spectrometry profiling. Significant alterations in various membrane lipid classes were detected in developing seeds compared to WT controls (Fig. 3). All three *AtPAH1/2*-expressed soybean lines displayed significant

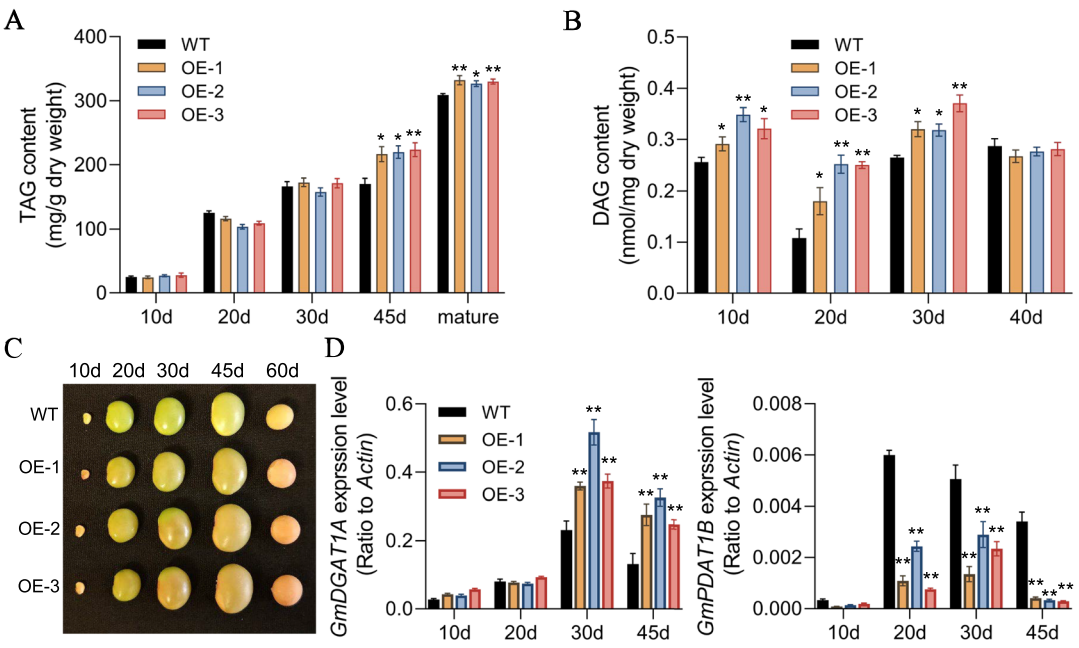

Fig. 2 Increased oil content in *AtPAHs*-expressed soybean seeds. **A** TAG contents in *AtPAHs*-expressed or WT developing and mature soybean seeds. Values are means ± SD ($n = 5$). **B** DAG content in WT and *AtPAHs* overexpression developing seeds. Data are means ± SD ($n = 5$). C Appearances of WT and *AtPAHs* overexpression developing seeds. Bar = 1 cm. D TAG synthesis genes *GmDGAT1A* and *GmPDAT1B* expression level in WT and transgenic soybean seeds. Data are means ± SD ($n = 3$). Student's *t*-test for significant difference, $^{*}p < 0.05$; $^{**}p < 0.01$

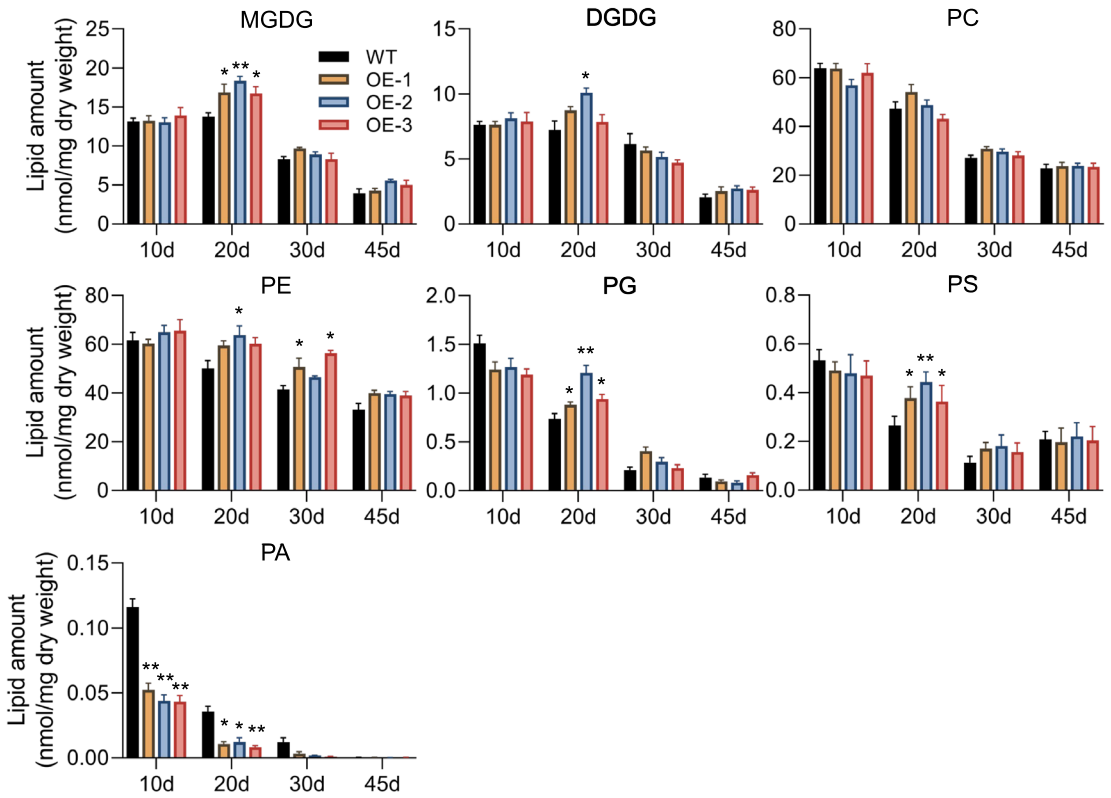

Fig. 3 Effect of *AtPAHs* expression on membrane glycerolipid levels in developing seeds. Total lipids were extracted and analyzed using ESI–MS/MS. Data are means ± SD ($n = 5$). Significance compared with WT at 10, 20, 30, and 45 days after flowering. *$p < 0.05$; **$p < 0.01$ (Student's *t*-test)

increases in monogalactosyldiacylglycerol (MGDG), PG, and PS in 20-day-old developing seeds compared to WT. At the 20-day-old seeds, the level of digalactosyldiacylglycerol (DGDG) and PE tended to increase in all three *AtPAH1/2*-expressed lines, but only the line *OE-2* that had the highest *AtPAH1/2* expression showed a significant increase. In 30-day-old seeds, the PE level increased in *OE-1* and *OE-3* (Fig. 3). By comparison, the PC level showed no significant difference between WT and *AtPAH1/2*-expressed soybean lines. The PC level decreased as seeds matured (Fig. 3). In contrast, all three *AtPAH1/2*-expressed soybean lines showed a marked decrease in PA levels, the substrate for PAH, at the early stages of seed development compared to WT (Fig. 3). These results suggest that the ectopic expression of *AtPAH1/2* in soybean seeds promotes the conversion of PA into polar lipid and DAG.

Seeds expressing *AtPAH1/2* enhanced polar lipid turnover to TAG

To further investigate the impact of ectopic expression of *AtPAH1/2* on lipid metabolism in seeds, we analyzed lipid biosynthesis in developing seeds using [^{3}H]-acetate pulse-chase labeling. Developing seeds at 19–20 days after flowering from WT, *OE-1*, and *OE-2* lines were labeled with [^{3}H]-acetate for one hour and then chased for 0, 1, 6, and 12 h. The results revealed a progressive decrease in the concentration of ^{3}H in DAG and PC throughout the chase period. Conversely, levels of TAG and PE, which both utilize DAG as substrate, showed an increase (Fig. 4). Concurrently, the radioactivity percentage in MGDG and DGDG exhibited a gradual rise, peaking at the 6-h chase period (Fig. 4).

In transgenic lines, ^{3}H concentration in PA initially rose but declined by the 12-h mark. By comparison, the PA level remained elevated in WT and higher than that in *AtPAH1/2*-expressed lines. Notably, transgenic plants exhibited a higher DAG content than WT before the chase, and the labeled DAG level became similar at 6 h after the chase (Fig. 4). TAG content in transgenic seeds remained higher than in WT. Compared to WT seeds, *AtPAH1/2*-OE seeds had a lower level of PA but a higher level of DAG at earlier stages, consistent with the effect of PAH conversion to DAG. The lack of difference of DAG at later stages between *AtPAH1/2*-OE and WT plants could result from the increased conversion of DAG to TAG (Fig. 4). The lower PA labeling than DAG could result from the lower PA than DAG levels in seeds.

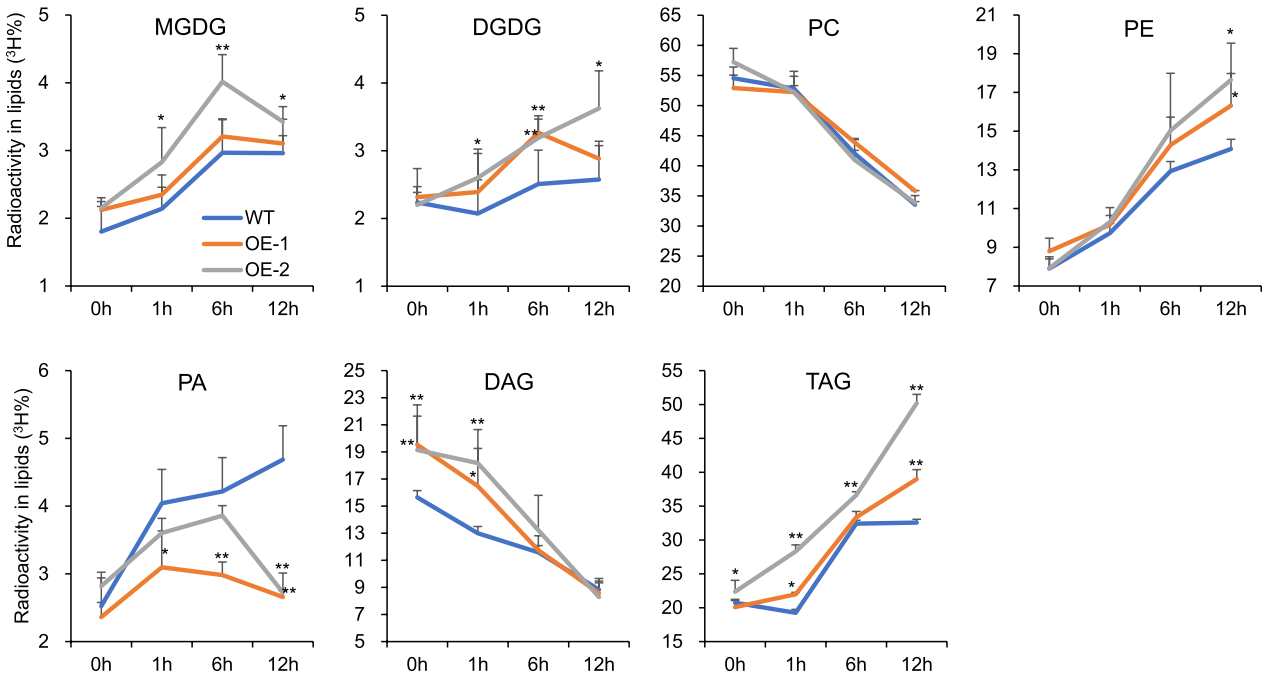

Fig. 4 In vivo labeling lipids of developing seeds with [^{3}H]-acetic acid. Developing seeds at 19 to 20 days after flowering were incubated with [^{3}H]-acetic acid for 1 h and subjected to lipid extraction at 0, 1, 6, and 12 h, respectively. Values are means ± SD ($n = 5$). Significance compared with WT, *$p < 0.05$; ** $p < 0.01$ (Student's *t*-test)

Overall, these labeling patterns support that *AtPAH1/2*-expressed soybean seeds increase the conversion of PA to DAG and subsequently to TAG compared to WT seeds (Fig. 4).

Seed-specific expression of *AtPAH1/2* increased seed weight in soybean

The Arabidopsis *pah1/pha2* double-mutants displayed an abnormal growth phenotype with slower growth rates than WT [7]. To investigate the impact of seed-specific expression of *AtPAH1/2*, we evaluated seed size and weight. Results showed that *AtPAH1/2*-expressed soybean seeds were larger in ten-seed width (7.3 ± 0.05 cm) and ten-seed length (8.4 ± 0.07 cm) by 5.0% and 6.7%, respectively, compared to WT soybean seeds (6.9 ± 0.12 width and 7.86 ± 0.14 length) (Fig. 5A, B). Consistent with the increase in seed size, the hundred-seed weight of the *AtPAH1/2* lines was 11.4% heavier than that of WT (Fig. 5C). Moreover, the transgenic plants displayed a marked increase in the number of seeds per plant (Fig. 5D). To probe the mechanisms underlying seed size effect by *AtPAH1/2*, we examined the expression level of *BIG SEEDS1* (*BS1*), a conserved gene known to regulate seed size by inhibiting primary cell proliferation [29]. Transcriptional analysis revealed a notable decrease in *GmBS1* expression in 20-day-old seeds of three independent transgenic soybean plants (Fig. 5E). Additionally, genes specific to the S-phase of cell division, *cyclin D3* (*GmCYCD3)* and *GmHISTONE4*, were significantly upregulated in these transgenic plants (Fig. 5E). These data could mean that enhanced cell proliferation during seed development could play a role in the larger seed size, but the effect of the *AtPHA1/2* overexpression on cell size and proliferation requires further investigation.

GmPAH1 effects on seed size and oil content have been selected through domestication

The functionality of the lipid biosynthesis gene tends to be conserved across different species. In soybean, there are three *PAH1* homologous genes: *Glyma.10g046400, Glyma.13g134500*, and *Glyma.19g175600* (Supplemental Fig. 3). Protein sequence alignment revealed that *Glyma.10g046400* and *Glyma.13g134500* exhibit a high degree of similarity (94% identity) with each other, while their similarity with *Glyma.19g175600* is lower, at 73% and 74% identity, respectively. To investigate the impact of PAH homologous on seed size (100-seed weight) and seed oil content in soybean, genetic variation was explored using the soybean multi-omics database (soyMD) [30]. Analysis of nucleotide polymorphisms in the *PAH1* homolog *Glyma.10g046400* revealed nine haplotypes (hap). The oil content and 100-seed weight of the hap0 to hap7 genotypes were considerably higher than those of the hap8 genotype, with 139 SNPs displayed

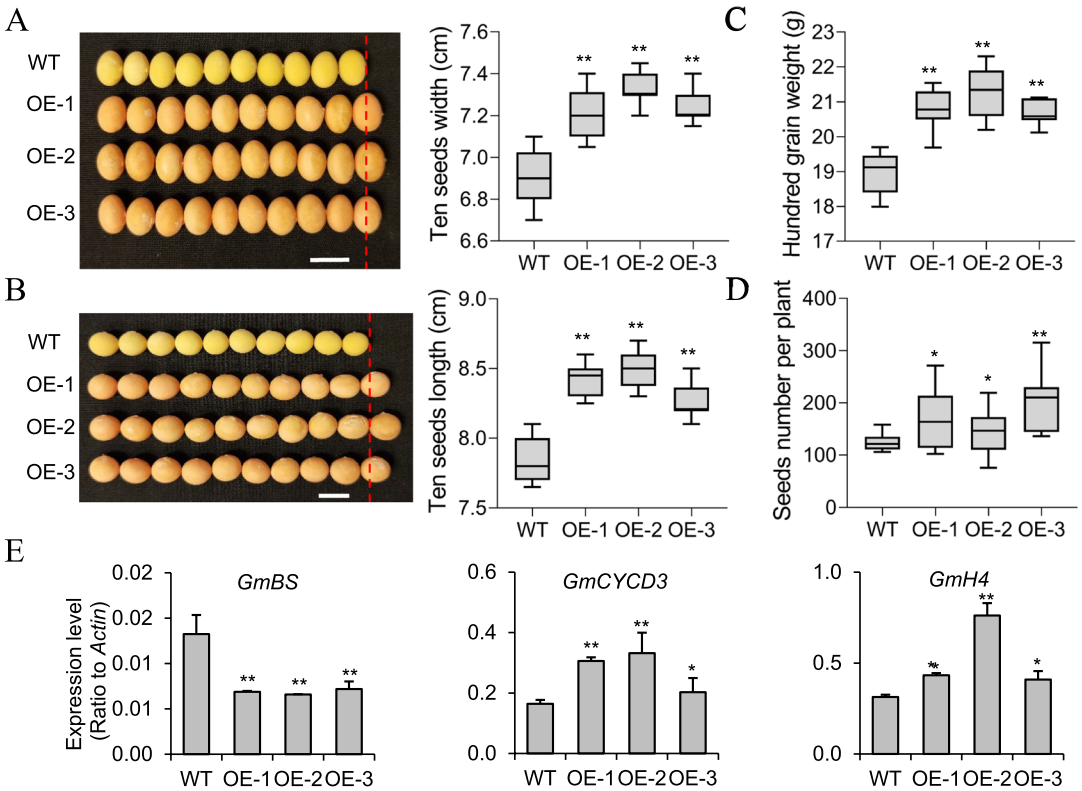

Fig. 5 Enhanced seed size and weight in *AtPAHs* expressing soybean. **A** Seed width is measured from the left to the right, with the seed umbilicus being oriented to the right. B Seed length is measured from the left to right, with the seed umbilicus facing upwards. Bars = 1cm. Comparisons of seed width (**A**, right), seed length (**B**, right), seed weight (**C**), and seed number per plant **D** of WT and *AtPAHs* overexpression line. Values are means ± SD ($n = 10$). **E** Expression level of seed size-related genes in 20d soybean seeds. Data are means ± SD ($n = 3$). $P < 0.05$ is indicated by *, and $P < 0.01$ is indicated by **, determined by Student's *t*-tests

p-values of $9.78e^{-20}$ and $7.735e^{-11}$, respectively (Fig. 6A, B). We studied the distribution of various haplotypes in cultivars, landraces, and wild populations since domestication is the process by which humans apply artificial selection to a wild soybean (*Glycine soja*). Allele-frequency analysis of 4,414 accessions from various subpopulations showed that both cultivars and landraces have hap0 through hap7. On the other hand, hap8 was only present in wild soybeans and not in cultivars or landraces (Fig. 6C). Further geographical distribution analysis of the nine haplotypes revealed that hap8 is exclusively distributed in Asia, the origin of soybean (Fig. 6D).

Haplotype analysis of another homolog, Glyma.13g134500, showed that hap0 and hap1 have higher oil content and 100-seed weight than hap2, with 97 SNPs showing p-values of $9.629e^{-22}$ and $3.837e^{-4}$, respectively (Fig. 7A, B). Hap2 was only found in wild soybeans (Fig. 7C) and is distributed in Asia (Fig. 7D). Meanwhile, Glyma.19g175600 exhibited haplotype differences in 100-seed weight (p-value = $6.404e^{-3}$) but not in total oil content (p-value = 0.1075) (Fig. 7E, F). The receptor soybean cultivar Jack used in this study contained the beneficial *PAH* haplotypes of both *Glyma.10g046400* and *Glyma.13g134500*. These results suggest that *Glyma.10g046400* and *Glyma.13g134500* affect seed oil content and yield traits simultaneously and have been selected during the domestication and breeding of wild soybeans to cultivar or landrace varieties. In other words, the *GmPAH* genotype found in existing cultivars and landraces is a superior variant characterized by elevated oil content and increased seed weight. This implies that increasing oil content and seed weight in current cultivars and landraces through *PAH* genotype screening may not be feasible. Instead, using seed-specific promoters to increase the expression of *PAH* genes in soybean, as shown in the present study, presents a potential avenue to enhance oil content and seed weight in soybean cultivars.

Discussion

Enhancing crop productivity and ensuring global food security are critical goals for sustaining adequate human nutrition. Transgenic technologies offer an efficient approach to boost crop productivity, protection, and quality. Numerous genes from various plant

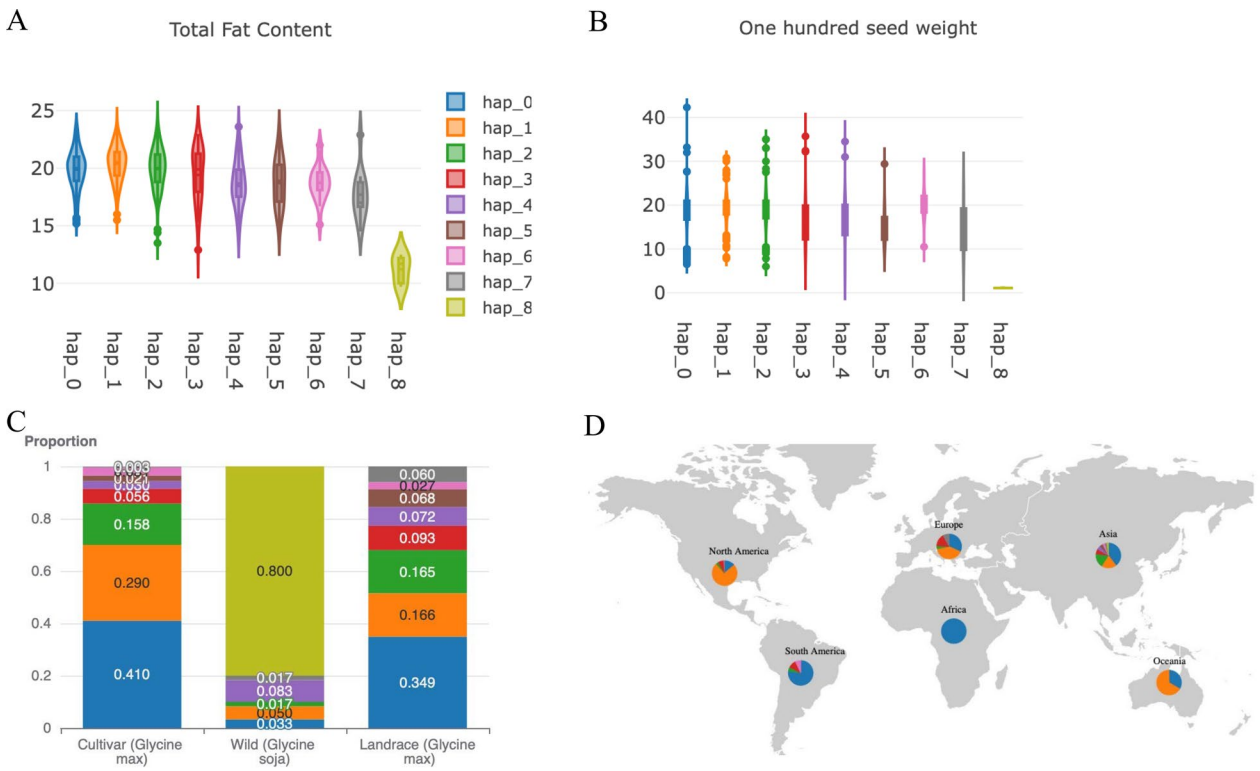

Fig. 6 Haplotype analysis of *Glyma.10g046400* in the soybean natural variation population. **A** Total oil content (%) of each soybean accession across nine haplotypes of *Glyma.10g046400* on Chr10:4155215.4164211 ± 2 kb. **B** One hundred seed weight (g) of each soybean accession across the nine haplotypes of *Glyma.10g046400*. C Allele frequencies of the nine haplotypes of *Glyma.10g046400* in cultivar, wild, and landrace sub-populations. Numbers represent the percentage of accessions of each haplotype in the three sub-populations. **D** Geographical distribution of *Glyma.10g046400* gene across different haplotypes. The pie chart areas represent the proportion of each haplotype in various regions. The colors corresponding to the haplotypes are shown to the right of panel **A**

species have proven invaluable for crop enhancement. However, the use of constitutive expression promoters in transgenic plants can lead to abnormal morphology, including retarded development and sterility. [31, 32]. In response to these challenges, seed-specific promoters have been utilized to enhance seed oil accumulation and improve fatty acid composition. The promoters of *Napins*, which are vital seed storage proteins in *Brassica* species, have been utilized to drive the expression of *leafy cotyledon1* (*BnLEC1*) and *LEC1-like* (*BnL1L*) genes, resulting in a significant increase of up to 20% in seed oil content without adverse effects on major agronomic traits [33]. Another study compared the effect of the *β-conglycinin* promoter-driven seed-specific vs the Cauliflower Mosaic Virus (CaMV) 35S promoter-driven constitutive expression of fatty acid desaturase 3 (*FAD3*) and found a 1.6-fold higher increase in fatty acid content with the seed-specific promoter compared to the CaMV 35S promoter [34]. Moreover, overexpression target genes in transgenic plants can introduce genetic diversity beyond a single species or genus, opening avenues for introducing specific traits from other species.

Cultivated soybeans, including landraces and cultivars, were domesticated from wild soybeans in China approximately 5,000 years ago [35]. Traits associated with advantageous genes and loci were selected during the transition from wild types to landraces, and these genes may not be used for further improvement in molecular design breeding. In this study, we identified favorable *PAH* genotypes selected through domestication and breeding and specifically expressed the Arabidopsis *PAH* gene in soybean seeds, resulting in increased seed oil content and seed weight (Fig. 2A, Fig. 5C). This suggests that tissue-specific expression of exogenous genes can further improve agronomic traits of crops based on previously selected beneficial genotypes. Therefore, leveraging transgenic approaches, alongside the identification and incorporation of beneficial genes from diverse organisms, holds substantial promise for significantly improving crop traits, particularly seed yield and quality in soybeans.

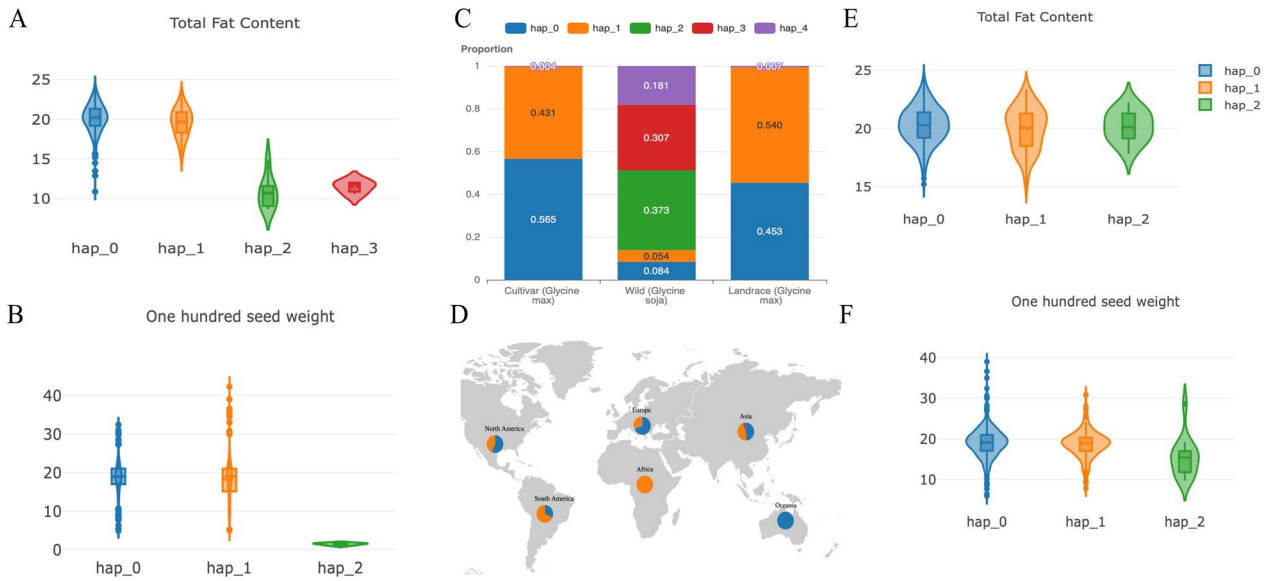

Fig. 7 Haplotype analysis of *Glyma.13g134500* and *Glyma.19g175600* in soybean population. **A** Total fat content (%) of each soybean accession across four haplotypes of *Glyma.13g134500* (Chr13:24696109.24705068 ± 2 kb). **B** One hundred seed weight (g) of each soybean accession across the four haplotypes of Glyma.13g134500. **C** Allele frequencies of the five haplotypes of *Glyma.13g134500* in cultivar, wild, and landrace sub-populations. Numbers represent the percentage of accessions of each haplotype in the three sub-populations. **D** Geographical distribution of *Glyma.13g134500* gene across different haplotypes. The pie chart areas represent the proportion of each haplotype in various regions. The color of *Glyma.13g134500* associated with the haplotypes is shown in panel **C** of the upper section. E Total fat content (%) of each soybean accession across three haplotypes of *Glyma.19g175600* (Chr19:43561808.43569019 ± 2 kb). **F** One hundred seed weight (g) for each soybean accession in the three haplotypes of *Glyma.19g175600*. The colors of *Glyma.19g175600* corresponding to the haplotypes are displayed to the right of panel **E**

PAH is an evolutionarily conserved enzyme found in both plants and animals. PAH plays a crucial role in maintaining lipid homeostasis by regulating cellular levels of PA to DAG. Mutation in *PAH* retarded growth in Arabidopsis and mice [7, 53]. So, this study chose seed-specific expression of *PAH1/2* to minimize *PAH* effects on other tissues. In soybean seeds, ectopic expression of *AtPAHs* not only increased TAG content, but also elevated glycolipid levels. This suggests that increased cellular pools of DAG can be utilized for the synthesis of both storage and membrane glycerolipids, consistent with similar findings in Arabidopsis [7]. While *AtPAH* expression increased TAG content in developing and mature seeds, there were no significant differences observed in the fatty acid composition of the TAGs.

The level of *AtPAH1* ectopic expression was way higher than that of *AtPAH2*. In addition, *AtPAH1* and 2 displayed different patterns of expression during the seed development, with *AtPAH1* expressed highly in the early, whereas *AtPAH2* highly in later stages (Fig. 1C). These disparities in expression could result from the use of different promoters as *AtPAH1* was driven by the *glycinin* promoter whereas AtPAH2 was under the control of the *β-conglycinin* promoter. The current experimental design involved co-expression of both genes simultaneously and thus could not attribute the observed seed phenotype to the OE of single *AtPAH1* or *AtPAH2*.

Genes involved in oil synthesis also have been reported to influence seed size and weight [36–40]. [19, 21]), potentially due to an increased supply of the source repository. Ectopic expression of *AtPAH1/2* in soybean seeds upregulated *GmDGAT1A* RNA levels, enhancing DAG synthesis and subsequent TAG production by boosting GmDGAT1A activity. Previous Arabidopsis studies have shown that regulating carbon flux into TAGs can increase seed oil content and seed weight [39]. Analysis of dividing cells has revealed that lipids, including PA and TAG, contribute to the structural integrity during cell division. Knockdown of *DGAT* and other lipid metabolic enzymes results in cell division defects [41, 42]. In our study, ectopic expression of *AtPAH1/2* in soybean seeds increased seed size, indicating an impact on cell proliferation during seed development. On the other hand, phospholipids also play a crucial role as cell membrane components during cell division, supporting changes in the plasma membrane. Notably, around 20 days after flowering, corresponding to rapid seed enlargement, the disparity in polar lipids content between soybeans expressing *AtPAH1/2* and WT soybeans becomes most prominent. Moreover, both the substrate and the product involved in the PAH-catalyzed reaction serve as signaling

lipids. Lipids can act directly as ligands or assist in assembling proteins into signaling platforms, which is crucial for cellular signaling processes. Given that PAH affects levels of PA and DAG in developing soybean seeds, investigating the potential contribution of these lipid signaling molecules to cell size regulation would be valuable.

Conclusion

Here, we identified that the superior *GmPAH* genotype, which enhances oil content and 100-seed weight, has been selected during the domestication of soybeans. We further showed that seed-specific expression of *AtPAHs* in cultivated soybeans enhanced seed oil content and seed weight. The ectopic expression of *AtPAHs* increased the conversion of PA to DAG that is further acylated to TAG via DGAT. Additionally, the expression of *AtPAH*s promoted the synthesis of phospholipids and glycolipids during seed development. These results indicate the potential application to use the seed-specific expression of *PAHs* to improve seed oil and crop yield even in the present superior genotypes.

Materials and methods

Plant materials and growth conditions

Soybean transformation was performed using an improved Agrobacterium-mediated method from the Plant Transformation Core (PTC) at the University of Missouri [43–45]. Agrobacterium-mediated soybean plants were selected based on glufosinate ammonium (Basta) resistance and the DsRed marker for selection. Positive plants were subsequently confirmed using gene-specific primers via PCR (Supplemental Fig. 1). Three independent lines from the T5 generation were used for subsequent experiments. Soybean wild-type Jack ecotype and three transgenic lines were grown in 2.5-gallon pots filled with Berger 7–35% soil mix in a greenhouse. The plants were cultivated under controlled conditions with 14 h of light and 10 h of darkness, maintaining temperatures at 26 °C during the day and 24 °C at night. Supplemental lighting was used when sunlight levels fell below 400W/m^2 between 6 and 8 am, with humidity maintained at 40%. Additionally, a shade curtain automatically closed to 50% when sunlight exceeded 900W/m^2, and fully closed at 100% when levels surpassed 1000W/m^2.

Construction of plant transformation vectors

The genomic sequences of *PAH1* and *PAH2* from Arabidopsis were amplified using the following primer pairs: AtPAH1F (5'-GCGGAATTCATGAGTTTGGTTGGAAG-3') and AtPAH1R (5'-GCGGAATTCTCATTCAACCTCTTCTAT-3') for *AtPHA1* (At3g09560), and AtPAH2F (5'-GCGGCGGCCGCATGAATGCCGTCGG-3') and AtPAH2R (5'-GCGGCGGCCGCTCACATAAGCGATG-3') for *AtPAH2* (At5g42870). The *AtPAH1* PCR product was then ligated into the pJC-Gly-DsRed binary vector at the *E*coRI site, which includes the seed-specific *glycinin* promoter, the Basta resistance gene, and the DsRed marker. The *AtPAH2* DNA fragment was cloned into the *Not*I site of the Beta-ConSoyHyg plasmid, containing the β-conglycinin promoter. Subsequently, the DNA segment comprising the β-conglycinin promoter, *AtPAH2* genomic sequence, and terminator was excised from the Beta-ConSoyHyg plasmid using *Asc*I, which has a compatible sticky end with *Mlu*I, and then inserted into the *Mlu*I site of the pJC-Gly-DsRed vector containing *AtPAH1*. The resultant constructs, named pJC-Gly-DsRed-AtPAH1/2, contained the genomic sequence of *AtPAH1* under the control of the Gly-promoter, while *AtPAH2* was driven by the β-conglycinin promoter. These constructs were then transformed into the GV3101 strain of *Agrobacterium tumefaciens* for soybean plant transformation.

RNA extraction and real-time PCR analyses

Total RNA was extracted from soybean seeds at various developmental stages using an RNA isolation kit (Qiagen) according to the manufacturer's instructions. Genomic DNA contamination was eliminated using RNase-free RQ1 DNase (Promega), and RNA concentration was assessed with a NanoPhotometer (Implen). Subsequently, cDNA synthesis was performed using a qScript cDNA Synthesis Kit (Quanta Biosciences) starting with 1 µg of total RNA. Transcript amplification utilized gene-specific primers, as detailed in Supplemental Table 1. PCR conditions involved initial denaturation at 94 °C for 3 min, followed by 40 cycles of 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 30 s, employing the iQ5 Real-time PCR system (BioRad) with SYBR Green. We processed the RT-qPCR data using the 2(-Delta Delta C(T)) method [46]. Expression levels of target genes were normalized to the internal control, soybean *ACTIN*.

Lipid extraction and profiling

Total lipids were extracted from developing seeds following a previously established protocol with minor adjustments [47]. Briefly, seeds at various developmental stages were isolated from pods and immediately immersed in 3 mL of preheated (75 °C) isopropanol containing 0.01% butylated hydroxytoluene (BHT), then incubated for 15 min. Next, 1.5 mL of chloroform and 0.6 mL of water were added, and the mixture was shaken for 1.5 h at 220 rpm. The resulting lipid extract was transferred to a new glass tube. The seeds were sliced into approximately 1-mm-thin pieces and placed in another glass tube. These slices were incubated with 3 mL of chloroform/methanol (2/1, v/v) in a shaking

incubator for one hour. This chloroform/methanol extraction step was repeated five times. The combined lipid extracts were washed with 1 mL of 1 M KCl solution and 2 mL of water and then dried under a stream of nitrogen gas. Chloroform was added to dissolve the dried lipids, with the amount calculated based on the dry weight of the tissue. Phospholipids, galactolipids, and diacylglycerol were analyzed using an electrospray ionization triple quadrupole mass spectrometer, and lipid molecular species were quantified by comparison with internal standards, as previously described [48].

For fatty acid analysis, approximately 30 mg of seed powder was accurately weighed and placed into glass tubes with Teflon cover. For each tube, 1.5 mL of 2.5% H_2SO_4 in methanol with 0.01% BHT, 400 µL of toluene, and 200 µL of a 2 mg/mL internal standard of 17:0 fatty acid were added. The tubes were then heated at 90 °C for 1 h to esterify the fatty acids into methyl esters. The mixture was subsequently extracted with 1.8 mL of water and 1 mL of hexane. The upper hexane phase containing the fatty acid methyl esters was collected for analysis using gas chromatography (Shimadzu GC-17A, Kyoto, Japan), as described [49].

[^{3}H] Acetate chase labeling

Soybean pods were collected 19 or 20 days after flowering. Seeds extracted from these pods were directly placed onto the culture medium composed of 5 mM MES, 0.5% sucrose, 1/2 Murashige and Skoog salts, pH5.8, kept on ice. Thirty embryos from each line were dissected and pre-incubated at room temperature for 40 min under a light intensity of 40 µmole m^{-2} s^{-1}, with continuous shaking. Radiolabeling was initiated by adding 10 mL of culture medium containing 20 µCi/mL [^{3}H] -acetate (5 mCi/ml in ethanol, American Radiolabeled Chemicals). The samples were then incubated for 1 h at room temperature with gentle agitation. After labeling, the embryos were washed twice with fresh culture medium, and the chase period was started. At specified time points, embryos were collected and transferred to preheated (75 °C) isopropanol containing 0.01% butylated hydroxytoluene for lipid extraction. Lipid components were extracted and separated using thin-layer chromatography (TLC), and the levels of ^{3}H in glycolipid (MGDG, DGDG), phospholipid (PA, PC, PE), and glycerolipid (DAG, TAG) were quantified using a scintillation counter (Perkin Elmer). Phospholipids and glycolipids were separated on TLC plates using the following developing solvents: chloroform/methanol/water (65/25/4, v/v/v) and hexane/diethyl ether/acetic acid (75/25/1, v/v/v), respectively [50–52].

Soybean multi-omics database (soyMD) analysis of *GmPAHs*

The variation of *GmPAH* in 4414 resequenced soybean accessions was analyzed using the soybean multi-omics database (SoyMD) [30]. The single-locus model was employed to analyze the haplotypes of three *GmPHA* genes and their 2-kb flanking regions in relation to total oil content and one-hundred seed weight. The module on allele frequency in sub-population, geographical distribution, and phenotypic value was plotted based on the soyMD website to illustrate the functional role of *GmPHAs*.

Abbreviations

PA	Phosphatidic acid
DAG	Diacylglycerol
TAG	Triacylglycerol
DGAT	DAG acyltransferase
PDAT	Phospholipid: DAG acyltransferase
PAH	Phosphatidic acid phosphohydrolases
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositols
PS	Phosphatidylserines
CDP-DAG	Cytidine diphosphate diacylglycerol
TLC	Thin-layer chromatography
MGDG	Monogalactosyldiacylglycerol
DGDG	Digalactosyldiacylglycerol
hap	Haplotype

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13068-025-02620-x.

Supplementary Material 1.

Acknowledgements

The authors gratefully acknowledge the support from Des Lee Collaborative Vision and Agriculture and Food Research Initiative (AFRI) award no. [2016-67013-24429/project accession number 1007600 and 2020-67013-30908/project accession number 1022148] from the USDA National Institute of Food and Agriculture. B.C. was supported in part by the National Natural Science Foundation of China (32200200), China Postdoctoral Science Foundation (GZC20240402).

Author contributions

B.C. designed and performed most of the experiments and wrote and revised the manuscript. J.L. did transgenic soybean screening. S.Y. helped analyze lipid data and edit the manuscript. G.W. performed AtPAH vector construction. X.W. proposed, designed, and supervised the study and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from Agriculture and Food Research Initiative (AFRI) award no. [2016-67013-24429/project accession number 1007600 and 2020-67013-30908/project accession number 1022148] from the USDA National Institute of Food and Agriculture. B.C. was supported in part by the National Natural Science Foundation of China (32200200), China Postdoctoral Science Foundation (GZC20240402).

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 8 September 2024 Accepted: 11 February 2025
Published online: 24 February 2025

References

1. Woodfield HK, Harwood J K. Oilseed crops: linseed, rapeseed, soybean, and sunflower, reference module in life sciences. Encyclopedia of Applied Plant Sciences (Second Edition); 2017. p. 34–38.
2. Manan S, Chen B, She G, Wan X, Zhao J. Transport and transcriptional regulation of oil production in plants. Crit Rev Biotechnol. 2016;23:1–15.
3. Bates PD, Stymne S, Ohlrogge J. Biochemical pathways in seed oil synthesis. Curr Opin Plant Biol. 2013;16:358–64.
4. Jennings W, Epand RM. CDP-diacylglycerol, a critical intermediate in lipid metabolism. Chem Phys Lipids. 2020;230: 104914.
5. Craddock CP, Adams N, Bryant FM, Kurup S, Eastmond PJ. PHOSPHATIDIC ACID PHOSPHOHYDROLASE regulates phosphatidylcholine biosynthesis in arabidopsis by phosphatidic acid-mediated activation of CTP:PHOSPHOCHOLINE CYTIDYLYLTRANSFERASE activity. Plant Cell. 2015;27:1251–64.
6. Craddock CP, Adams N, Kroon JT, Bryant FM, Hussey PJ, Kurup S, et al. Cyclin-dependent kinase activity enhances phosphatidylcholine biosynthesis in Arabidopsis by repressing phosphatidic acid phosphohydrolase activity. Plant J. 2017;89:3–14.
7. Eastmond PJ, Quettier AL, Kroon JT, Craddock C, Adams N, Slabas AR. Phosphatidic acid phosphohydrolase 1 and 2 regulate phospholipid synthesis at the endoplasmic reticulum in Arabidopsis. Plant Cell. 2010;22:2796–811.
8. Wang L, Kazachkov M, Shen W, Bai M, Wu H, Zou J. Deciphering the roles of Arabidopsis LPCAT and PAH in phosphatidylcholine homeostasis and pathway coordination for chloroplast lipid synthesis. Plant J. 2014;80:965–76.
9. O'Hara L, Han GS, Peak-Chew S, Grimsey N, Carman GM, Sinios-soglou S. Control of phospholipid synthesis by phosphorylation of the yeast lipin Pah1p/Smp2p Mg^{2+}-dependent phosphatidate phosphatase. J Biol Chem. 2006;281:34537–48.
10. Brindley DN. Lipid phosphate phosphatases and related proteins: signaling functions in development, cell division, and cancer. J Cell Biochem. 2004;92:900–12.
11. Carman GM, Han G-S. Roles of phosphatidate phosphatase enzymes in lipid metabolism. Trends Biochem Sci. 2006;31:694–9.
12. Carman GM. Phosphatidate phosphatases and diacylglycerol pyrophosphate phosphatases in saccharomyces cerevisiae and *Escherichia coli*. Biochim Biophys Acta. 1997;1348:45–55.
13. Nanjundan M, Possmayer F. Pulmonary phosphatidic acid phosphatase and lipid phosphate phosphohydrolase. Am J Physiol Lung Cell Mol Physiol. 2003;284:L1-23.
14. Pyne S, Long JS, Ktistakis NT, Pyne NJ. Lipid phosphate phosphatases and lipid phosphate signalling. Biochem Soc Trans. 2005;33:1370–4.
15. Han GS, Siniossoglou S, Carman GM. The cellular functions of the yeast lipin homolog PAH1p are dependent on its phosphatidate phosphatase activity. J Biol Chem. 2007;282:37026–35.
16. Santos-Rosa H, Leung J, Grimsey N, Peak-Chew S, Siniossoglou S. The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth. EMBO J. 2005;24:1931–41.
17. Nakamura Y, Koizumi R, Shui G, Shimojima M, Wenk MR, Ito T, et al. Arabidopsis lipins mediate eukaryotic pathway of lipid metabolism and cope critically with phosphate starvation. Proc Natl Acad Sci U S A. 2009;106:20978–83.
18. Goettel W, Zhang H, Li Y, Qiao Z, Jiang H, Hou D, et al. POWR1 is a domestication gene pleiotropically regulating seed quality and yield in soybean. Nat Commun. 2022;13:3051.
19. Liu JY, Zhang YW, Han X, Zuo JF, Zhang Z, Shang H, et al. An evolutionary population structure model reveals pleiotropic effects of GmPDAT for traits related to seed size and oil content in soybean. J Exp Bot. 2020;71:6988–7002.
20. Miao L, Yang S, Zhang K, He J, Wu C, Ren Y, et al. Natural variation and selection in GmSWEET39 affect soybean seed oil content. New Phytol. 2020;225:1651–66.
21. Wang S, Liu S, Wang J, Yokosho K, Zhou B, Yu YC, et al. Simultaneous changes in seed size, oil content and protein content driven by selection of *SWEET* homologues during soybean domestication. Natl Sci Rev. 2020;7:1776–86.
22. Burke B, Ellenberg J. Remodelling the walls of the nucleus. Nat Rev Mol Cell Biol. 2002;3:487–97.
23. Nunnari J, Walter P. Regulation of organelle biogenesis. Cell. 1996;84:389–94.
24. Kurat CF, Wolinski H, Petschnigg J, Kaluarachchi S, Andrews B, Natter K, Kohlwein SD. Cdk1/Cdc28-dependent activation of the major triacylglycerol lipase Tgl4 in yeast links lipolysis to cell-cycle progression. Mol Cell. 2009;33:53–63.
25. Sciorra VA, Morris AJ. Roles for lipid phosphate phosphatases in regulation of cellular signaling. Biochim Biophys Acta. 2002;1582:45–51.
26. Waggoner DW, Xu J, Singh I, Jasinska R, Zhang QX, Brindley DN. Structural organization of mammalian lipid phosphate phosphatases: implications for signal transduction. Biochim Biophys Acta. 1999;1439:299–316.
27. Wang X, Devaiah SP, Zhang W, Welti R. Signaling functions of phosphatidic acid. Prog Lipid Res. 2006;45:250–78.
28. Zhang G, Bahn SC, Wang G, Zhang Y, Chen B, Zhang Y, et al. PLDα1-knockdown soybean seeds display higher unsaturated glycerolipid contents and seed vigor in high temperature and humidity environments. Biotechnol Biofuels. 2019;12:9.
29. Ge L, Yu J, Wang H, Luth D, Bai G, Wang K, Chen R. Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. Proc Natl Acad Sci U S A. 2016;113:12414–9.
30. Yang Z, Luo C, Pei X, Wang S, Huang Y, Li J, et al. SoyMD: a platform combining multi-omics data with various tools for soybean research and breeding. Nucleic Acids Res. 2024;52:D1639–50.
31. Chen B, Zhang G, Wang J, Li P, Yang J, Benning C, et al. Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (Glycine max). Plant Biotechnol J. 2020;18:155–71.
32. Yang Y, Munz J, Cass C, Zienkiewicz A, Kong Q, Ma W, et al. Ectopic expression of WRINKLED1 affects fatty acid homeostasis in brachypodium distachyon vegetative tissues. Plant Physiol. 2015;169:1836–47.
33. Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, et al. Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. Plant Physiol. 2011;156:1577–88.
34. Yeom WW, Kim HJ, Lee KR, Cho HS, Kim JY, Jung HW, et al. Increased production of α-linolenic acid in soybean seeds by overexpression of lesquerella FAD3-1. Front Plant Sci. 2020;10:1812.
35. Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, et al. Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci U S A. 2006;103:16666–71.
36. Ding J, Ruan C, Guan Y, Krishna P. Identification of microRNAs involved in lipid biosynthesis and seed size in developing sea buckthorn seeds using high-throughput sequencing. Sci Rep. 2018;8:4022.
37. Fatihi A, Zbierzak AM, Dörmann P. Alterations in seed development gene expression affect size and oil content of Arabidopsis seeds. Plant Physiol. 2013;163:973–85.
38. Guo ZH, Haslam RP, Michaelson LV, Yeung EC, Lung SC, Napier JA, et al. The overexpression of rice ACYL-CoA-BINDING PROTEIN2 increases grain size and bran oil content in transgenic rice. Plant J. 2019;100:1132–47.
39. Jako C, Kumar A, Wei Y, Zou J, Barton DL, Giblin EM, et al. Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol. 2001;126:861–74.

40. Zhang J, Martin JM, Beecher B, Lu C, Hannah LC, Wall ML, et al. The ectopic expression of the wheat Puroindoline genes increase germ size and seed oil content in transgenic corn. Plant Mol Biol. 2010;74:353–65.
41. Atilla-Gokcumen GE, Muro E, Relat-Goberna J, Sasse S, Bedigian A, Coughlin ML, et al. Dividing cells regulate their lipid composition and localization. Cell. 2014;156:428–39.
42. Storck EM, Özbalci C, Eggert US. Lipid cell biology: a focus on lipids in cell division. Annu Rev Biochem. 2018;87:839–69.
43. Zeng P, Vadnais DA, Zhang Z, Polacco JC. Refined glufosinate selection in Agrobacterium-mediated transformation of soybean [*Glycine max* (L.) Merrill]. Plant Cell Rep. 2004;22:478–82.
44. Zhang G, Yang J, Chen X, Zhao D, Zhou X, Zhang Y, et al. Phospholipase D- and phosphatidic acid-mediated phospholipid metabolism and signaling modulate symbiotic interaction and nodulation in soybean (Glycine max). Plant J. 2021;106:142–58.
45. Zhang Z, Xing A, Staswick P, Clemente TE. The use of glufosinate as a selective agent in Agrobacterium-mediated transformation of soybean. Plant Cell Tissue Organ Cult. 1999;56:37–46.
46. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Method. 2002;25:402–8.
47. Welti R, Li W, Li M, Sang Y, Biesiada H, Zhou H, Rajashekar CB, Williams TD, Wang X. Profiling membrane lipids in plant stress responses role of phospholipase Dα in freezing-induced lipid changes in Arabidopsis. J Biol Chem. 2002;277:31994–2002.
48. Narasimhan R, Wang G, Li M, Roth M, Welti R, Wang X. Differential changes in galactolipid and phospholipid species in soybean leaves and roots undernitrogen deficiency and after nodulation. Phytochemistry. 2013;96:81–91.
49. Chen B, Wang J, Zhang G, Liu J, Manan S, Hu H, et al. Two types of soybean diacylglycerol acyltransferases are differentially involved in triacylglycerol biosynthesis and response to environmental stresses and hormones. Sci Rep. 2016;6:28541.
50. Cai G, Fan C, Liu S, Yang Q, Liu D, Wu J, et al. Nonspecific phospholipase C6 increases seed oil production in oilseed Brassicaceae plants. New Phytol. 2020;226:1055–73.
51. Dutton HJ, Jones EP, Scholfield CR, Chorney W, Scully NJ. Countercurrent distribution of soybean fatty acid methyl esters biosynthetically labeled with H3 and C14. J Lipid Res. 1960;2:63–7.
52. Yang W, Wang G, Li J, Bates PD, Wang X, Allen DK. Phospholipase Dζ enhances diacylglycerol flux into triacylglycerol. Plant Physiol. 2017;174:110–23.
53. Phan J, Reue K. Lipin, a lipodystrophy and obesity gene. Cell Metab. 2005;1:73–83.

Publisher's Note