nature communications

Article

https://doi.org/10.1038/s41467-024-52044-8

Artificial selection of mutations in two nearby genes gave rise to shattering resistance in soybean

Received: 28 August 2023

Accepted: 26 August 2024

Published online: 31 August 2024

Check for updates

Shuai Li[1,2,8] ✉, Weidong Wang[1,3,6,8], Lianjun Sun[1,6,8], Hong Zhu[4,8], Rui Hou[2], Huiying Zhang[2], Xuemin Tang[2], Chancelor B. Clark[1,3], Stephen A. Swarm[5,7], Randall L. Nelson[5] & Jianxin Ma[1,3] ✉

Resistance to pod shattering is a key domestication-related trait selected for seed production in many crops. Here, we show that the transition from shattering in wild soybeans to shattering resistance in cultivated soybeans resulted from selection of mutations within the coding sequences of two nearby genes - *Sh1* and *Pdh1*. *Sh1* encodes a C2H2-like zinc finger transcription factor that promotes shattering by repressing *SHAT1-5* expression, thereby reducing the secondary wall thickness of fiber cap cells in the abscission layers of pod sutures, while *Pdh1* encodes a dirigent protein that orchestrates asymmetric lignin distribution in inner sclerenchyma, creating torsion in pod walls that facilitates shattering. Integration analyses of quantitative trait locus mapping, genome-wide association studies, and allele distribution in representative soybean germplasm suggest that these two genes are primary modulators underlying this domestication trait. Our study thus provides comprehensive understanding regarding the genetic, molecular, and cellular bases of shattering resistance in soybeans.

Soybean [*Glycine max* (L.) Merr.] is an important leguminous crop domesticated from its wild relative *Glycine soja* over 6000 years ago[1]. Pod dehiscence or shattering is critical for the fitness of wild soybeans in natural environments but impedes seed harvesting in agricultural systems. Therefore, shattering resistance is one of the key traits targeted for selection during the process of soybean domestication. Currently, elite cultivars employed for crop production are generally resistant to shattering, but shattering still can occur under drought stress and account for significant yield loss[2,3]. Therefore, it is important to understand the genetic architecture of shattering resistance in soybeans. Genetic analyses using biparental populations derived from crosses between *G. soja* and *G. max* or between shattering and non-shattering soybean cultivars revealed at least 23 quantitative trait loci (QTL), distributed on chromosomes 2, 5, 10, 14, 15, 16, and 19, modulating shattering[4–9].

Two genes have been previously reported to be primary targets for selection for shattering resistance in soybeans[10,11]. Funatsuki et al. found that shattering was modulated by a gene, designated *Pdh1*, which encodes a dirigent-like protein responsible for asymmetric lignin distribution in inner sclerenchyma, producing pod wall torsion to promote shattering, and that a point mutation in the coding region of *Pdh1* resulted in the formation of a premature stop codon and thus a defective allele, *pdh1*, which reduces the pod wall tension and thus promote shattering resistance[10]. *Pdh1* was pinpointed via map-based cloning using a biparental population derived from a cross between two cultivated soybean varieties. In addition, the signal of selection for

[1]Department of Agronomy, Purdue University, West Lafayette, IN, USA. [2]College of Life Sciences, Qingdao Agricultural University, Qingdao, China. [3]Center for Plant Biology, Purdue University, West Lafayette, IN, USA. [4]College of Agronomy, Qingdao Agricultural University, Qingdao, China. [5]Department of Crop Sciences, University of Illinois, Urbana, IL, USA. [6]Present address: College of Agronomy and Biotechnology, China Agricultural University, Beijing, China. [7]Present address: Beck's Hybrids, Atlanta, IN, USA. [8]These authors contributed equally: Shuai Li, Weidong Wang, Lianjun Sun, Hong Zhu.
✉e-mail: li2014shuai@qau.edu.cn; maj@purdue.edu

the *pdh1* region was detected in landraces from non-humid region but not detected in landraces from humid region[12–15], suggesting shattering is an adaptation trait associated with varietal diversification post-domestication. Nevertheless, *Pdh1* is located within the major shattering quantitative trait locus (QTL) region (28–31 Mb) defined on chromosome 16 in our previous study using biparental populations derived from the crosses between cultivated (Williams 82) and wild soybeans (PI 468916, PI 479752)[9], and thus *pdh1* was considered as a domestication allele for shattering resistance[9,10].

Dong et al. reported that a NAM, ATAF, and CUC (NAC) domain transcription factor gene *SHATTERING1-5* (*SHAT1-5*) mediates shattering[11]. They demonstrated that the allele of this gene in a cultivated soybean cultivar HEINONG44, designated *shat1-5*, was expressed at 15-fold the level of *SHAT1-5* expression in a *G. soja* accession ZYD00755, and that the elevated expression of *shat1-5* resulted in the thickening of secondary walls of lignified fiber cap cells (FCCs) in the pod sutures, responsible for shattering resistance. *shat1-5* was considered as a domestication allele due to a line of observations[11]: (i) it is an ortholog of an Arabidopsis gene homologous to *NST1/2*, within which the ethyl methane sulfonate-induced null mutations produce non-dehiscent pods; (ii) it overlaps with one (*SHAT1-5*, peaking at ~5 Mb position on chromosome 16) of the five shattering QTL regions defined onto chromosome 16 with only 120 (*G. max* × *G. soja*) F_4 derived lines and 140 restriction fragment length polymorphism (RFLP) markers[4]; (iii) the expression of either the *SHAT1-5* allele from ZYD00755 or the *shat1-5* allele from HEINONG44 driven by the Arabidopsis *NST-1* promoter in the Arabidopsis *nst1-1;nst3-1* mutants was able to recover the wild-type (i.e., *NST-1*) phenotype; and (iv) a 20-bp deletion at ~4 kb upstream of the gene's promoter region showed an association with shattering resistance in the soybean varieties surveyed. Nevertheless, the QTL *SHAT1-5* was detected only in a single study using such a small size of mapping population and a very low density of RFLP markers[4]. Hence, the relative contribution of *shat1* to shattering resistance in cultivated soybeans was yet to be further evaluated.

To further elucidate the genetic basis and molecular mechanisms underlying shattering resistance in soybean, we conducted fine mapping and map-based cloning of the key QTL underlying the domestication transition from shattering in *G. soja* to shattering resistance in *G. max* using 3500 $F_{6:7}$ RILs[9,16,17]. In addition, we performed genome-wide association study (GWAS) on shattering resistance using phenotypic and genotypic data from 3099 *G. max* accessions randomly selected from the USDA Soybean Germplasm Collection[18]. Furthermore, the candidate genes for the domestication transition were functionally validated through genetic transformation and molecular characterization. Finally, the distribution of the wild and domestication alleles in representative soybean populations was examined.

Here, we show that shattering resistance in soybean is primarily achieved through the selection of natural mutations that occurred in two genes within a major domestication QTL in the entire genome – one is *Pdh1*, and the other is a C2H2-like zinc finger transcription factor gene, designated *Sh1*, which directly binds to the promoter region of *SHAT1-5* to repress its expression, promoting shattering in wild soybeans. We also demonstrate that selection of both *sh1* and *pdh1* mutations gave rise to shattering resistance in cultivated soybeans.

Results

QTL mapping and GWAS reveal genetic basis of shattering resistance

To understand the genetic bases of key domestication-related traits in soybean, we developed two populations comprising approximately 3500 $F_{6:7}$ RILs, as described earlier[9,16,17], by crossing a cultivar Williams 82 with each of the two highly diverged, shattering *G. soja* accessions, PI 468916 and PI 479752 (Fig. 1a, b). Among these RILs, 151 lines derived from the Williams 82 × PI 468916 cross and 510 lines derived from the Williams 82 × PI 479752 cross were genotyped through the genotyping-by-sequencing (GBS) approach and phenotyped for eleven domestication-related traits including shattering resistance in two locations and three years, which resulted in identification of a major QTL for shattering resistance at ~30 Mb position on chromosome 16 with both populations[9]. This QTL overlapped with the *Pdh1* QTL (Fig. 1c, d) and explained 23% and 53% of the phenotypic variances with the 151 (Williams 82 × PI 468916) lines and the 510 (Williams 82 × PI 479752) lines, respectively. All the F_1 plants from the crosses produced shattering pods, similar to those produced from the two *G. soja* accessions (Fig. 1b), suggesting that shattering is dominant over shattering-resistance.

Although minor QTLs for shattering resistance were also revealed (Fig. 1c, d), none of them explained a phenotypic variance of >6%, including the minor QTL on chromosome 19, which harbors the *L1/l1* locus that specifies pod colors (black in *G. soja* versus brown in *G. max*), with pleiotropic effect on shattering due likely to different efficiencies of photothermal conversion caused by different pod colors[19]. The *SHAT1-5* QTL, which corresponds to ~5 Mb position of chromosome 16, was not detected by either of our two RIL populations (Fig. 1c, d).

To further depict the genetic architecture of shattering resistance in soybean, we conducted GWAS with the genome-wide SNP data[18] and the shattering phenotypic data from the 3099 landraces (Supplementary Data 1; www.ars-grin.gov/npgs). The shattering scores represented the levels of shattering at two weeks after harvest maturity (R8) with '1' indicating no shattering and '5' more than 50% of the pods open. This analysis revealed a major QTL peaked at the position of 29,597,918 bp on chromosome 16, approximately 340 kb downstream of the *pdh1* locus according to the soybean reference genome (Fig. 1e). In addition to this major QTL for shattering resistance, a minor QTL was detected on chromosome 7 (Fig. 1e), which appears to overlap with a minor QTL detected by QTL analysis with the 510 (Williams 82 × PI 479752) RIL lines (Fig. 1d). These observations, further suggest that the transition from shattering in *G. soja* to shattering resistance in soybean landraces was primarily modulated by the major QTL on chromosome 16.

Fine mapping of the major QTL for shattering pinpoints two candidate genes

To identify the candidate genes for the major shattering QTL on chromosome 16, we screened the 3500 RILs using two molecular markers, Satt622 and Sat_366, defining boundaries of this QTL region, and identified 30 recombinants (Fig. 2a, b and Supplementary Data 2). These recombinants were further genotyped using 10 additional markers and phenotyped in the field. Intriguingly, the genotypic and phenotypic data defined the candidate genes for the shattering QTL into two small adjacent regions. According to the Williams 82 reference genome, one region is between markers SH401K and InDel1 (18.1 kb) harboring gene Glyma.16G141100 (dubbed the *sh1* region) that encodes a C2H2-like zinc finger transcription factor, the other is flanked by markers SRM0 and SRM1 (50.8 kb) containing Glyma.16G141300 and Glyma.16G141400 (dubbed the *pdh1* region). It was reported that Glyma.16G141400 was *pdh1* for shattering resistance and that Glyma.16G141300 did not contribute to shattering resistance[10]. Thus, only Glyma.16G141100 – the candidate for *sh1*, and *pdh1* in this entire QTL region would be associated with shattering resistance.

The full-length genomic sequences of Glyma.16G141100 and Glyma.16G141400 in both *G. soja* accessions PI 468916 and PI 479752 were amplified by PCR and sequenced. At the Glyma.16G141100 locus, a small insertion/deletion (InDel) involving nine nucleotides (ACTACTACT) and a 'G to A' mutation, which is 6-nucleotides away from the InDel, were detected within the third exon of the gene between the two *G. soja* accessions (*Sh1*) and Williams 82 (*sh1*), resulting in an InDel of three amino acids (ThrThrThr) and an amino acid change from 'Ala' to 'Thr' in predicted protein sequences (Fig. 2c and Supplementary Data 3),

https://doi.org/10.1038/s41467-024-52044-8

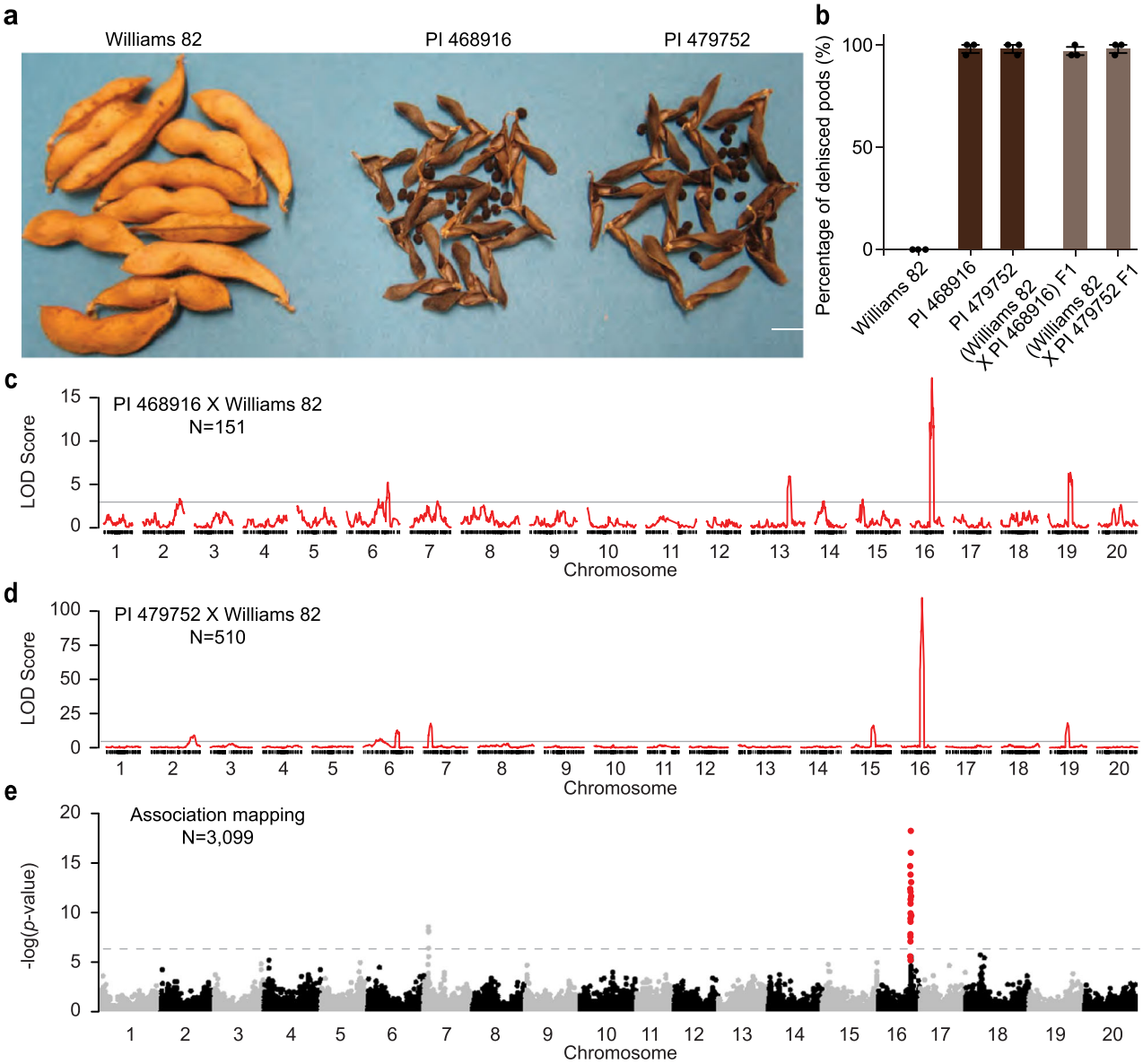

Fig. 1 | Genetic architecture of shattering resistance in soybeans. a Photographic illustration of pod dehiscence in Williams 82, PI 468916, and PI 479752. Bar = 1 cm. **b** Statistics of dehisced pods in Williams 82, PI 468916, and PI 479752, and F_1 pod derived from the crosses of Williams 82 × PI 468916 and Williams 82 × PI 4479752, respectively. **c** QTL mapping result using the RIL population derived from Williams 82 x PI 468916. The *y*-axis indicates the logarithm of the odds (LOD) score, and the *x*-axis indicates chromosome. The gray line indicates the significant threshold LOD determined by 1,000 permutations at significant level of 0.05. **d** QTL mapping result using the RIL population derived from Williams 82 x PI 479752. The *y*-axis indicates the LOD score, and the *x*-axis indicates chromosome. The gray line indicates the significant threshold LOD determined by 1,000 permutations at significant level of 0.05. **e** Genome-wide association mapping results using a natural population including ~3000 soybean landrace accessions randomly selected from the USDA Soybean Germplasm Collection. The gray dashed line indicates the significant threshold value determined by Bonferroni correction. $n = 3$ biological samples for Fig. 1b. Data are presented as mean values ± SD. Source data are provided as a Source Data file.

consistent to previously revealed by genome-resequencing[20]. At the Glyma.16G141400 locus, the two *G. soja* accessions possess *Pdh1* that is identical to the gene present in the shattering parental line, whereas Williams 82 possesses *pdh1* that is identical to the gene present in the shattering-resistant parental line described by Funatsuki et al.[10] (Fig. 2c). The genotypic and phenotypic data from the recombinants also indicated that either the *Sh1* allele or the *Pdh1* allele alone produced shattering (Fig. 2b and Supplementary Data 2).

We then analyzed the expression of the four genes between markers SH401K and SRM1 in flowers and/or developing pods (excluding seeds) in the three parental lines from the developmental stage R1 (beginning bloom stage) to R6 (full seed stage) (Fig. 2b, d). Glyma.16G141100 in the two *G. soja* accessions (i.e., the *Sh1* candidate) was expressed at the highest level in developing pods at R3 (beginning pod stage) and exhibited differential expression in developing pods at R3 and R4 (full pod stage) between the *G. soja* accessions and Williams 82 (i.e., the *sh1* candidate). *Pdh1* in the two *G. soja* accessions was expressed at the highest level in developing pods also at R3, but *Pdh1* in the *G. soja* accessions and *pdh1* in Williams 82 showed differential expression across the developmental stages from R1 to R6, suggesting that Glyma.16G141100 and *Pdh1* may be involved in different molecular mechanisms modulating shattering. The expression patterns of Glyma.16G141200 and Glyma.16G141300 were distinct from those of Glyma.16G141100 and *Pdh1*. For example, in the R3 stage when

https://doi.org/10.1038/s41467-024-52044-8

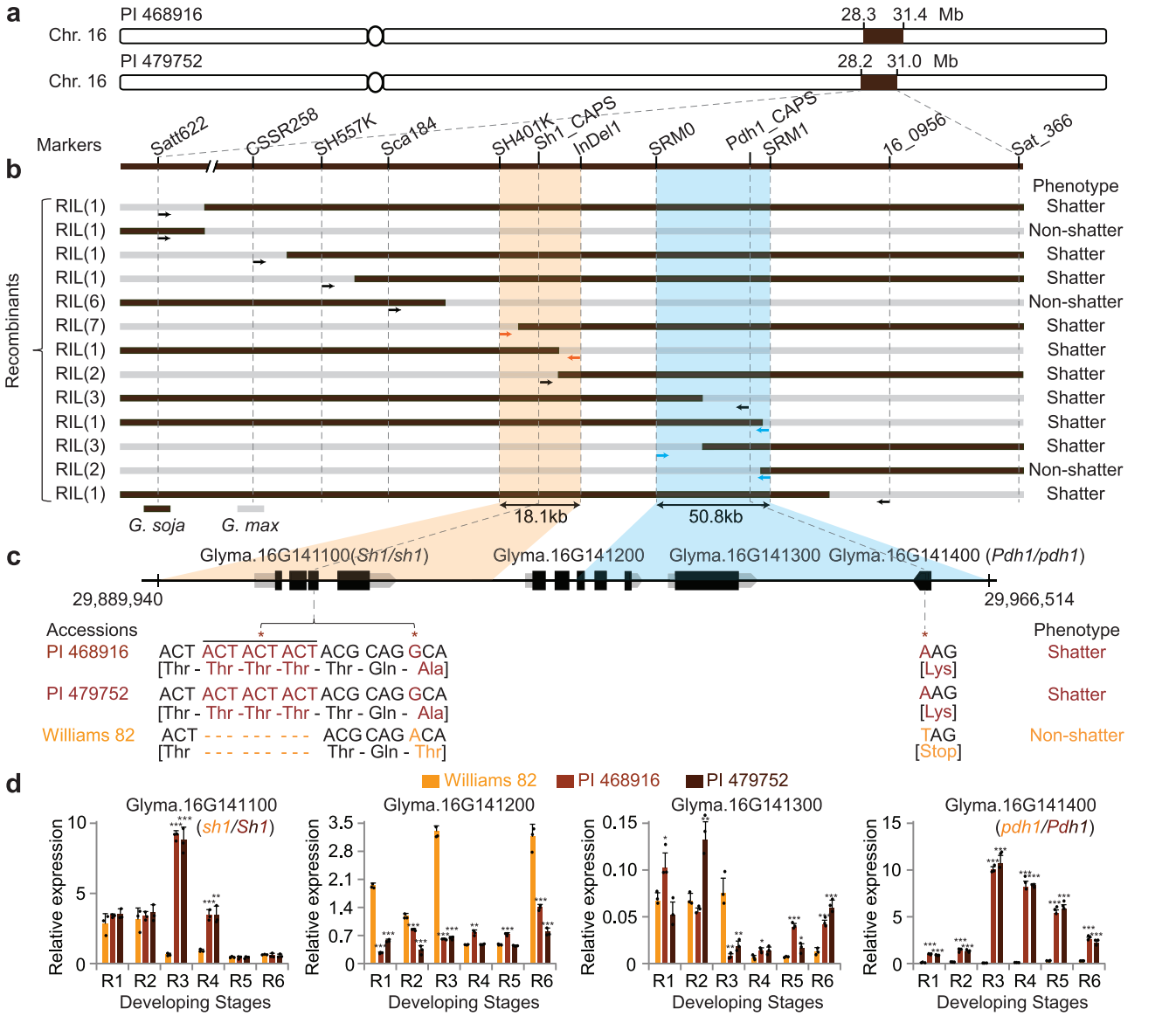

Fig. 2 | Map-based cloning of soybean shattering resistance genes. a Physical positions of shattering QTL region identified using the two RIL populations from Williams 82 × PI 468916 and Williams 82 × PI 479752, respectively. The black boxes indicate the shattering QTL region defined on chromosome 16 according to the Williams 82 reference genome assembly 2.0. **b** Fine mapping of shattering resistance genes. The physical positions of molecular markers used for fine mapping are shown. Recombinants carrying crossovers are identified using molecular markers and phenotypes from the two RIL populations derived from Williams 82 × PI 468916 and Williams 82 × PI 479752. Black bars indicate *G. soja* genotype, and gray bars indicate Williams 82 genotype. The annotated genes in the mapped region are shown. Arrows indicate the deduced direction of causal genes. **c** Polymorphisms in the coding sequence of Glyma.16g141100 and *Pdh1* that result in amino acid changes between shattering resistant line Williams 82 and susceptible lines PI 468916 and PI 479752. **d** The expression levels of Glyma.16g141100, Glyma.16g141200, Glyma.16g141300 and *Pdh1* in developing soybean pods in the parental line Williams 82, PI 468916, and PI 479752. The expression levels of these genes relative to a *GmCons4* gene were analyzed by qRT-PCR. R1: beginning bloom, R2: full bloom, R3: beginning pod, R4: full pod, R5: beginning seed, R6: full seed. $n = 3$ biological samples for **d**. Data are presented as mean values ± SD. The statistical significance is determined by a two-sided *t* test, and ***, **, * indicate $P < 0.001$, 0.01, and 0.05, respectively. Source data are provided as a Source Data file.

Glyma.16G141100 and *Pdh1* were most highly expressed in the two *G. soja* accessions but barely expressed in Williams 82, the expression levels of Glyma.16G141200 and Glyma.16G141300 in the two *G. soja* accessions are lower than those of their respective allelic copies in Williams 82.

Genetic transformation validates that both genes modulate shattering

To validate the scenario of 'two genes in one QTL modulating shattering' as deduced above, as well as the causal mutations for shattering resistance in cultivated soybean, we fused the cauliflower mosaic virus 35S promoter with the coding sequence (CDS) of the *Sh1* candidate, the CDS of the *sh1* candidate, the CDS of *Pdh1*, and the CDS of *pdh1*, respectively. We introduced the four fusion constructs (p35S:CDS-Sh1, p35S:CDS-sh1, p35S:CDS-Pdh1, p35S:CDS-pdh1), separately, into Williams 82, and obtained nine, six, seven, and four independent transformation events, respectively. For each construct, the T_2 progeny derived from three transformation events were further analyzed. In each of the chosen T_2 progeny, the transgene was expressed in the developing pods at the R3 stage at a much higher level than the native *sh1* or *pdh1* gene in Williams 82 (Fig. 3a, b). As expected, the p35S:CDS-Sh1 transgenic lines and the p35S:CDS-Pdh1 transgenic lines both produced more shattered pods than Williams 82 (Fig. 3c–f). By contrast, neither the p35S:CDS-sh1 transgenic lines nor the p35S:CDS-pdh1 transgenic lines produced more

https://doi.org/10.1038/s41467-024-52044-8

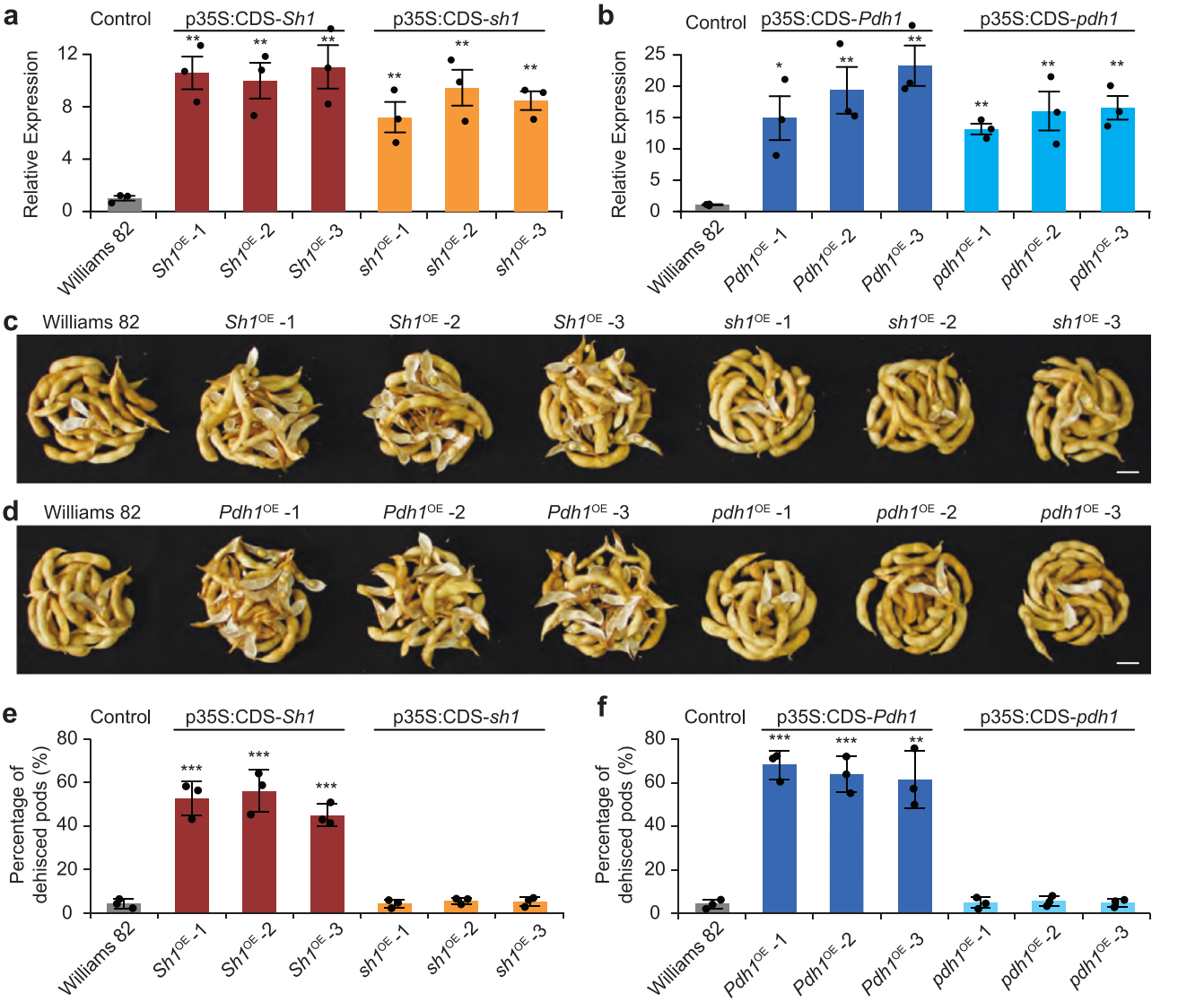

Fig. 3 | Validation of *Sh1* and *Pdh1* functions by genetic transformation. a The relative expression of *Sh1/sh1* in Williams 82 and three independent *Sh1/sh1* transgenic lines. $Sh1^{OE}$/$sh1^{OE}$ indicates *Sh1* and *sh1* transgenic lines, respectively. The expression level of genes in Williams 82 was set as 1, and the others were adjusted accordingly. **b** The relative expression of *Pdh1/pdh1* in Williams 82 and three independent *Pdh1/pdh1* transgenic lines. $Pdh1^{OE}$/$pdh1^{OE}$ indicates *Pdh1* and *pdh1* transgenic lines, respectively. The expression level of genes in Williams 82 was set as 1, and the others were adjusted accordingly. **c** Photographic illustration of pod dehiscence in *Sh1* and *sh1* transgenic line compared to Williams 82. Bar = 1 cm. **d** Photographic illustration of pod in *Pdh1* and *pdh1* transgenic line compared to Williams 82. Bar = 1 cm. **e** Statistics of pod dehiscence in *Sh1* and *sh1* transgenic line compared to Williams 82. **f** Statistics of pod dehiscence in *Pdh1* and *pdh1* transgenic line compared to Williams 82. $n = 3$ biological samples for **a**, **b**, and $n = 3$ biological replications (with six individual transgenic plants for each replication) for **e**, **f**. Data are presented as mean values ± SD. The statistical significance is determined by a two-sided *t* test, and ***, **, and * indicate $P < 0.001$, $P < 0.01$, and $P < 0.05$, respectively. Source data are provided as a Source Data file.

shattered pods than Williams 82 (Fig. 3c–f). These observations indicate: (i) either Glyma.16G141100 (now designated *Sh1)* or *Pdh1* in the *G. soja* accessions promotes shattering; (ii) the 9-bp deletion, the nearby G/A mutation, or the combination of both as a whole in the coding region of *Sh1*, is the causal mutations that formed *sh1* (Fig. 2c); (iii) consistent with the study by Funatsuki et al.[10], the (A/T) mutation is the causal mutation that formed *pdh1*; and (iv) the domestication transition from shattering to shattering resistance resulted from selection of both the *sh1* and *pdh1* mutations.

Allelic distribution reveals selection signals and dynamic processes

To understand the selection process for shattering resistance during soybean domestication, we investigated the genetic diversity of an ~2 Mb region surrounding this major shattering QTL region in 2898 re-sequenced soybean accessions, including 103 *G. soja*, 1048 landraces and 1747 improved cultivars[21,22]. The genomic sequence analysis revealed that the genetic diversity of landraces with the *sh1* allele (π^{sh1}) or the *pdh1* allele (π^{pdh1}) is lower than that in *G. soja* ($\pi^{G.soja}$) at the position from 30.2 to 30.8 Mb (based on ZH13.v2.0), where *Sh1* and *Pdh1* were located respectively (Fig. 4a). The ratios of the nucleotide diversity among the landraces carrying *sh1* allele or *pdh1* allele to the nucleotide diversity among the *G. soja* revealed selective sweeps of approximately 560 kb and 670 kb (Fig. 4b, c, Supplementary data 4), respectively. This sweep region was also detected when the ratios of nucleotide diversities among the landraces to the nucleotide diversities among the *G. soja* accessions, without sorting the *Pdh1* and *pdh1* alleles or the *Sh1* and *sh1* alleles into distinct groups[21,22], demonstrating that this region was targeted for selection in the process of soybean domestication.

https://doi.org/10.1038/s41467-024-52044-8

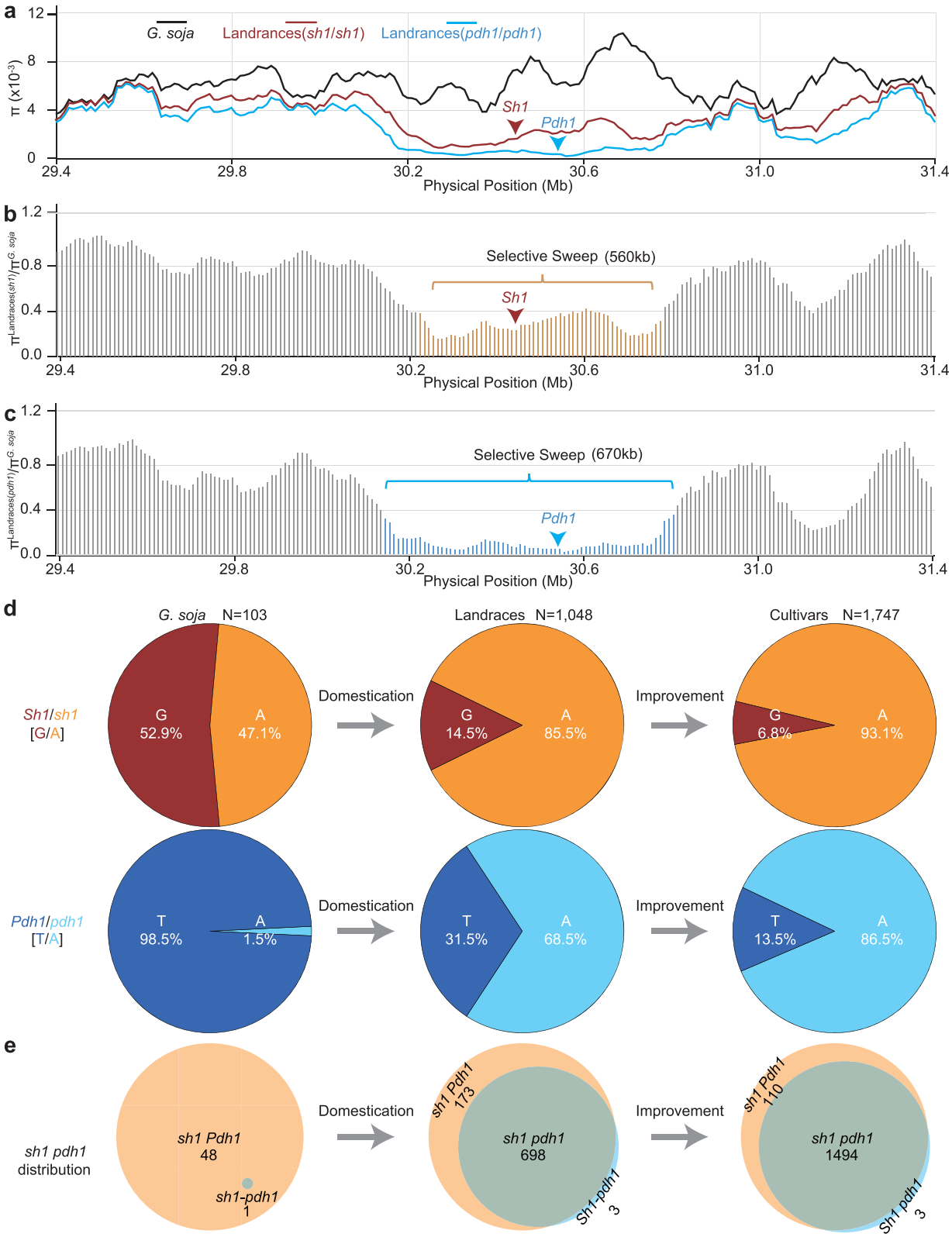

Fig. 4 | Analyses of genetic diversity in the *Sh1* and *Pdh1* region and allele distribution in soybean populations. a Genetic diversity (π) of *G. soja* (black), landrace with *sh1* allele (brown) and landraces with *pdh1* allele (blue) at the *Sh1* and *Pdh1* region on chromosome 16 (Zhonghuang 13 version 2) using 2,898 re-sequenced soybean accessions, including 103 *G. soja*, 1,048 landraces and 1,747 improved cultivars. 100-kb sliding windows with 10-kb steps were used. The selective sweep surrounding *Sh1* (**b**) and *Pdh1* (**c**). Individual vertical bars indicate 100-kb sliding windows with 10-kb steps. The selective sweep was identified by nucleotide diversity ratios of landrace with *sh1* or *pdh1* over the *G. soja* populations. **d** The frequencies of the causal mutations at *sh1* and *pdh1* positions in *G. soja*, landraces and cultivars. 'G' and 'A' indicate shattering susceptible and resistant alleles at *Sh1*, respectively. 'T' and 'A' indicated shattering susceptible and resistant alleles at *Pdh1*, respectively. **e** Co-selection of *sh1* and *pdh1* during domestication and improvements.

We further analyzed the frequencies of the causal mutations of *sh1* (G/A, which is co-segregating with the deletion of 9-bp, or 6-bp in some *G. max* varieties) and *phd1* (A/T) in the *G. soja*, landraces and elite cultivars. It was found that 47.1% of the *G. soja*, 85.5% of landraces and 93.1% elite cultivars carry *sh1*, and 1.5% of the *G. soja*, 68.5% of landraces and 86.5% elite cultivars carry *pdh1* (Fig. 4d). Although the *sh1-Pdh1* and *Sh1-pdh1* genotypes were also seen in both landraces and elite cultivars, the majority of the landraces (79.9%) and elite cultivars (93.0%) carry both *sh1* and *phd1* (Fig. 4e). These results, together with the detected selective sweep harboring the two genes, suggest that both the mutant alleles were targeted for selection during soybean domestication and improvement.

We also analyzed the distribution of the ZYD00755 (*G. soja*)-type allele *SHAT1-5* and the HEINONG44 (*G. max*)-type allele *shat1-5*, based on the 20-bp InDel at ~4 kb upstream of their promoter region[11], in the same set of *G. soja*, landraces and elite cultivars[21,22]. It was found that 13.5% of the *G. soja* accessions, 34.4% of landraces and 40.9% elite cultivars carry *shat1-5* (Supplementary Fig. 1a). Of the landraces and elite cultivars possessing *pdh1*, 68.8% and 58.1% carry *SHAT1-5*, respectively (Supplementary Fig 1b). Contrastingly, only 0.4% and 0.2% of the landraces and elite cultivars possessing *pdh1* carry *Sh1*, respectively (Fig. 4e). These observations suggest that the 20-bp deletion is not associated with shattering resistance and that the *shat-1-5* is unlikely to be a direct target for selection during soybean domestication and improvement.

Molecular and cellular assays reveal interaction between Sh1 and *SHAT1-5*

Then, how does *Sh1* promote shattering in *G. soja*? Since Sh1 is a transcription factor, we wondered whether it directly binds to the promoter regions of *Pdh1* and *SHAT1-5* to regulate their expression. To explore these possibilities, we fragmented the putative promoter regions (2-kb segments) upstream of the first exons of *Pdh1* and *SHAT1-5* each into five 420-bp segments with a 20-bp overlap between adjacent segments, and then examined the binding capability of Sh1 to these segments using yeast one-hybrid (Y1H) assays (Fig. 5a, b, Supplementary Fig. 2). Our data showed that Sh1 was able to bind to the S2 segment of the *SHAT1-5* promoter region (Fig. 5a, b). Subsequently, the S2 segment was further fragmented into 13 contiguous/overlapping mini-segments (S2-1 to S2-13), which were screened for Sh1-binding using Y1H assays (Fig. 5c, d). The data revealed that Sh1 was able to bind to S2-5, S2-6, S2-11, and S2-12, which cover a 180 bp contiguous sequence (Fig. 5d). The binding of Sh1 to these mini segments was further validated using electrophoretic mobility shift assays (EMSAs) (Fig. 5e).

Subsequently, we employed dual-luciferase transient expression assays to determine the action that Sh1 exerts on *SHAT1-5*. Briefly, the CDS of *Sh1* was cloned into the pGreen II 0029 62-SK vector (62-SK) as the Sh1 effector construct and expressed by the CaMV 35S promoter, and the S2 segment from the *SHAT1-5* promoter region was cloned into the pGreen II 0800-LUC vector (LUC), which also contains the REN gene, as the *SHAT1-5* reporter construct. The Sh1 effector construct or the 62-Sk control was combined with the *SHAT1-5* reporter construct and co-injected into tobacco leaves, separately, to monitor the transactivity of the reporter by the effector. We found that the LUC/REN ratio was reduced in leaves co-transformed with the Sh1 effector and the *SHAT1-5* reporter constructs compared to the LUC/REN ratio in leaves co-transformed with the 62-SK vector and the *SHAT1-5* reporter construct (Fig. 5f), indicating that Sh1 is a repressor of *SHAT1-5*. This explains why *SHAT1-5* in *G. soja* was expressed at such a low level compared with *shat1-5* in *G. max*[11].

Given that the mini-segments in the *SHAT1-5* promoter region bound by Sh1 are shared by the *shat1-5* allele, and that the *Sh1* transgenic lines of Williams 82 showed increased shattering compared with Williams 82, we speculated that the Sh1 protein produced by the transgene *Sh1* would bind to the promoter region of *shat1-5* to downregulate its expression. As expected, reduced levels of *shat1-5* expression were observed in all three independent *Sh1* transgenic lines of Williams 82, compared with Williams 82 (Fig. 5g).

The zinc finger domain in a typical C2H2 zinc finger protein is the core DNA-binding domain[23]. As the putative zinc finger domains of Sh1 and sh1 are identical, it is, as expected, that sh1, like Sh1, was detected to be able to bind to the S2 segment in the promoter of *SHAT1-5* (Supplementary Figs. 3 and 4). To determine whether sh1 was able to repress the expression of *SHAT1-5*, we conducted the dual-luciferase transient expression assays with sh1 as did with Sh1. Different from that observed in the Sh1 assays, the LUC/REN ratio in the tobacco leaves co-transformed with the sh1 effector and the *SHAT1-5* reporter constructs was similar to that in tobacco leaves co-transformed with the 62-SK vector and the *SHAT1-5* reporter construct (Supplementary Fig. 5), suggesting that sh1 had lost the ability to repress the expression of *SHAT1-5*. Consistent to this observation, similar levels of *shat1-5* expression were detected in the three *sh1* transgenic lines and in Williams 82 (Supplementary Fig. 6). These results suggest that the loss of the Sh1 function as a repressor of *SHAT1-5*/*shat1-5* in sh1 was not related to the DNA-binding ability of the zinc finger domain but the detected casual mutations that changed amino acids in other domain of the gene.

To determine whether the sequence polymorphisms between the promoter regions of Sh1/sh1 are associated with their ability/inability to repress the expression of *SHAT1-5*, we conducted the dual-luciferase transient expression assays using the CDSs of *Sh1* and *sh1*, separately, whose expressions were driven by either their respective promoters or with swapped promoters of the two alleles, rather than the 35S promoter. We found that the expression of *Sh1* under the control of either the *Sh1* or *sh1* promoter repressed the expression of *SHAT1-5*, while the expression of *sh1* under the control of either the *Sh1* or *sh1* promoter did not affect the expression of *SHAT1-5* (Supplementary Fig. 7). These observations further suggest that the detected mutations in the CDS of *sh1* was responsible for the loss of the allele's ability to repress the expression of *SHAT1-5*/*shat1-5*.

According to Dong et al., the higher expression level of *shat1-5* in soybean cultivar HEINONG44, relative to *SHAT1-5* in *G. soja* accession ZYD00755, was associated with the thickening of secondary walls of lignified FCCs in the ventral suture underlying shattering resistance[11]. As Sh1 promotes shattering through downregulating *SHAT1-5*, we deduced that the *Sh1* transgenic lines would exhibit reduced thickness of lignified FCC secondary walls compared with Williams 82; whereas the *sh1* transgenic lines and Williams 82 would show similar thickness of lignified FCC secondary walls. As exemplified in Fig. 5h, the observed data were consistent with our deduction.

Discussion

The exploration of the genetic foundation underlying shattering resistance carries significant scientific importance, not only in understanding the process of soybean domestication but also in addressing the persistent issue of yield loss caused by shattering. We demonstrate that shattering in wild soybeans is predominantly governed by the *Sh1-Pdh1* QTL, and that artificial selection of the *sh1-pdh1* double mutants was primarily responsible for the rise of shattering resistance in cultivated soybeans. Our study also suggest that Sh1 is a repressor of *SHAT1-5* and modulates shattering through downregulating *SHAT1-5* expression, whereas *sh1* is a recessive allele, which has lost the ability to repress the expression of *shat1-5*, giving rise to shattering resistance in the *pdh1* background in cultivated soybeans. While it has been clear that the small deletion and the nearby point mutation mark the *sh1* allele, whether the deletion, the point mutation, or both were causal for the loss-of-function of *sh1* has yet to be investigated. As the zinc finger binding domains of the *Sh1* and *sh1* are identical and both were able to bind to the promoters of *SHAT1-5*, the mechanism by which the

https://doi.org/10.1038/s41467-024-52044-8

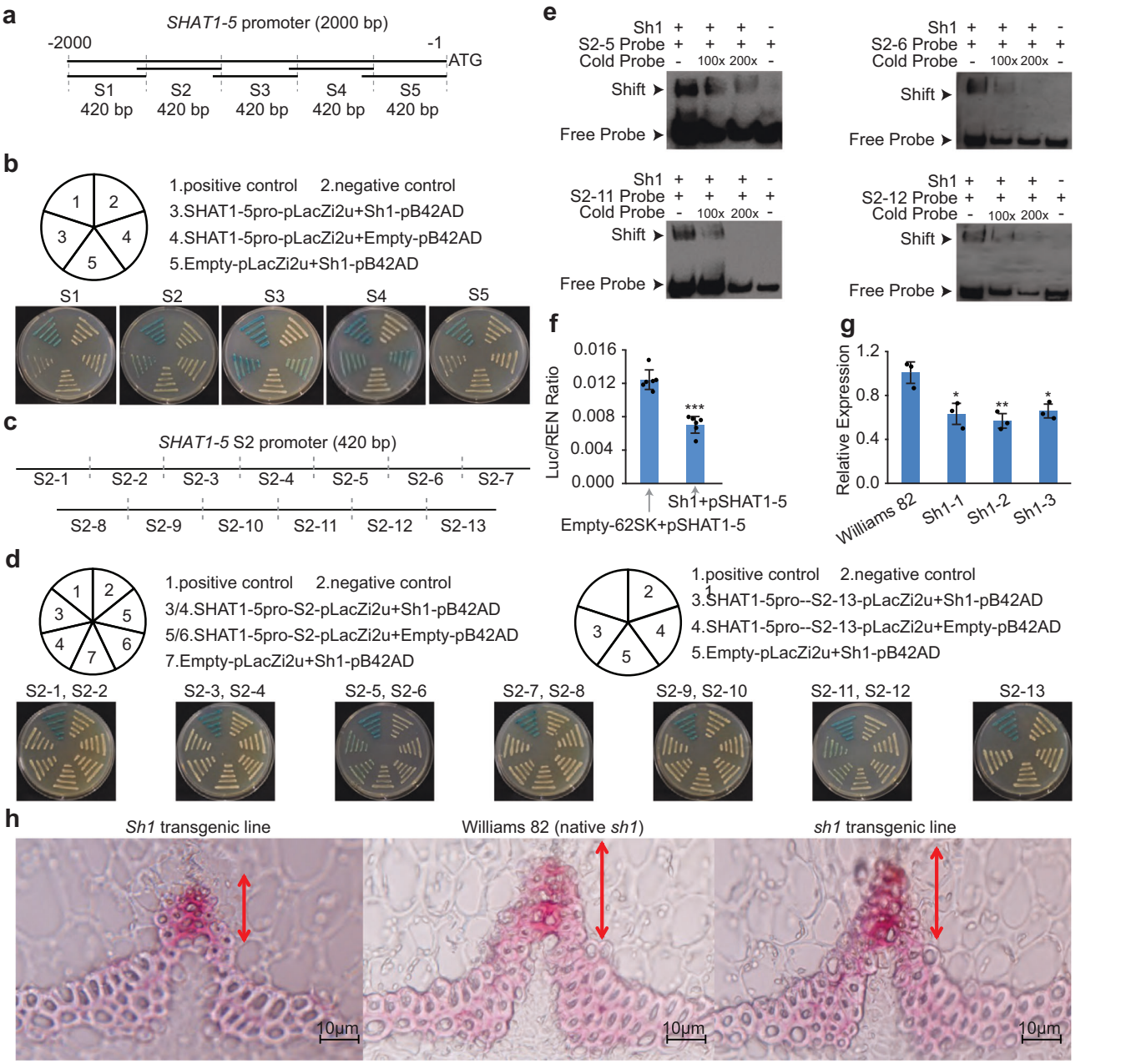

Fig. 5 | Molecular interaction between Sh1 and *SHAT1-5* and cellular mechanism for shattering. a Schematic diagram of *SHAT1-5* promoter segments used in Y1H assay. **b** Results of Y1H assays with *SHAT1-5* promoter segments. Diagram illustrating the results of Y1H assays. From left to right, S1 to S5. **c** Schematic diagram of the *SHAT1-5* promoter S2 segments fragmentation (60 bp for each). **d** Results of Y1H assays with *SHAT1-5* promoter S2 segments fragmentation. Diagram illustrating the results of Y1H assays. For the plate have seven groups, both 3 and 5 indicate S2-1, S2-3, S2-5, S2-7, S2-9, and S2-11 segments, and both 4 and 6 indicate S2-2, S2-4, S2-6, S2-8, and S2-12 segments, respectively. **e** EMSA detection of Sh1 binding to S2-5, S2-6, S2-11 and S2-12 segments of *SHAT1-5* promoter. EMSA of 3′-biotin-labeled dsDNA probes with the purified Sh1 protein. The presence (+) or absence (−) of specific probes is marked. The concentration of the cold probe was 1 μM (100x), 2 μM (200x), while that of the biotinylated probe was 10 nM. Water was added in place of Sh1 protein as a control. **f** Regulation of Sh1 to *SHAT1-5* using the dual luciferase assay. Relative LUC/REN activities after infiltration for 3 days. **g** The expression of *SHAT1-5* in R3 pods of *Sh1* transgenic plants. **h** Cross section analysis of ventral suture of *Sh1/sh1* and Williams 82. The lignified fiber cap cell is indicated by arrows. $n = 6$ biological samples for **f**, and $n = 3$ biological samples for **g**, respectively. For **f**, **g**, data are presented as mean values ± SD, and the statistical significance is determined by a two-sided t test. ***, **, * indicate $P < 0.001$, 0.01, and 0.05, respectively. For **e**, **h**, three independent replicates are performed, and a representative result is shown. Source data are provided as a Source Data file.

causal mutations resulted in the loss of the gene's ability to repress the expression of *SHAT1-5* remains unclear. Overall, the structures and functions of C2H2 zinc finger proteins are poorly characterized in plants; nevertheless, a few non-zinc finger domains have been annotated to enable various functions such as protein-protein interactions, transcription repression, and oligomerization in animals. Thus, although only the zinc finger domain in Sh1 is detectable, our study suggests that the domain of Sh1, where the amino acids were altered in sh1, is critical for the *SHAT1-5* expression, exemplifying the role of an unknown domain of a plant C2H2-like zinc finger protein in transcription repression.

In addition to the enhanced understanding of the genetic architecture and molecular mechanisms underlying this domestication trait, our study would facilitate the utilization of wild soybeans for crop improvement. For example, our observations from the RILs for QTL mapping, genetic transformation, and molecular and cellular assays

suggest that both *Sh1* and *Pdh1* need to be knocked out for re-domestication or de novo domestication of soybean from its wild relatives for shattering resistance. This suggestion would be easily taken given that the great majority of landraces and elite cultivars are the *sh1-pdh1* mutants. Nevertheless, 13% of the 1048 landraces and 6% of the 1747 elite soybean cultivars[21,22] harbor both *Sh1* and *Pdh1* (the latter was defined based on the absence of the causal mutation for *pdh1* defined by Funatsuki et al.[10]), and one of those cultivars showed improved shattering resistance after its *Pdh1* was knocked out by gene-editing[14]. As those elite cultivars were developed for soybean production, they should carry some degrees of shattering resistance. Such resistance could have been obtained by uncharacterized mutations in the *Pdh1* and *Sh1* alleles, such as a few rare recently found loss-of-function *pdh1* alleles[15], which, prior to their discovery, must have been defined as *Pdh1*. Of course, such resistance could also be attributed to mutations in uncharacterized genes involved in the *Pdh1*- and *Sh1*-mediated regulatory pathways, or other unknown genetic factors. Alternatively, some elite cultivars possessing *Sh1* and *Pdh1* may have been adapted to specific environments such as humid eco-regions where shattering generally does not occur regardless of their genotypes[12,13]. More comprehensive investigations such as phenotyping and expression analysis of the soybean cultivars with *Sh1* and *Pdh1* under diverse environments, creation of double mutations within these two alleles or expression of either of these two alleles with their native promoters in the *sh1* and *pdh1* background, identification of genes targeted by Sh1 or its interacting proteins, genetics and functional genomics studies would further validate the functions of the two genes and/or lead to discovery of uncharacterized alleles, genes, and/or genetic pathways underlying this important trait in soybean and, probably, in many other plant species producing seeds in pods as well.

Shattering resistance is one of the quintessential domestication syndrome traits strongly favored by early farmers across leguminous and other crop species[24]. The dramatic differences in allelic frequencies of *Sh1* and *Pdh1* (Fig. 4d) between wild and domesticated soybeans reiterate this point. Despite the strong selection pressure for the recessive shattering resistance alleles of these two genes, the dominant wild-type shattering alleles do persist at low levels in landraces and even among cultivars. This may be a result of differential selection pressures in different environments. Bandillo et al. reported an environmental association analysis of soybean landraces and concluded that selection pressure for the non-shattering *pdh1* allele was stronger in drier environments conducive to shattering compared to more humid environments where rates of shattering may be lower[25]. Additionally, in some areas of smallholder soybean production, whole soybean plants are harvested before full maturity and stacked in piles to dry, and in these systems, shattering may be neutral or even favorable for efficient recovery of seeds.

It remains unclear how *sh1* and *pdh1* were selected and fixed in the majority of cultivated soybeans. One possible scenario is hitchhiking, which suggests that as the frequency of the beneficial variant increases through artificial selection, other genetic variants linked to it can also rise in frequency even if they do not provide any direct benefit[26]. Hitchhiking effects play crucial roles in the process of plant domestication, yielding both detrimental and advantageous outcomes[27,28]. Since *sh1* and *pdh1* are so closely linked and both are critical for shattering resistance, the high proportion of the *sh1-pdh1* genotype in cultivated soybeans would be easily explained by mutual hitchhiking (Fig. 6). It is also possible that the two mutant alleles were initially selected separately, and even within distinct timeframes, and were later merged through recombination between the two loci from different genotypes (Fig. 6). Extensive and frequent genomic introgression between *G. soja* and *G. max* and within each of the two subspecies had occurred during domestication and during the radiation of landraces to various eco-regions post domestication[29,30], which may have led to re-selection or recurrent selection for the *sh1-pdh1* mutants, re-

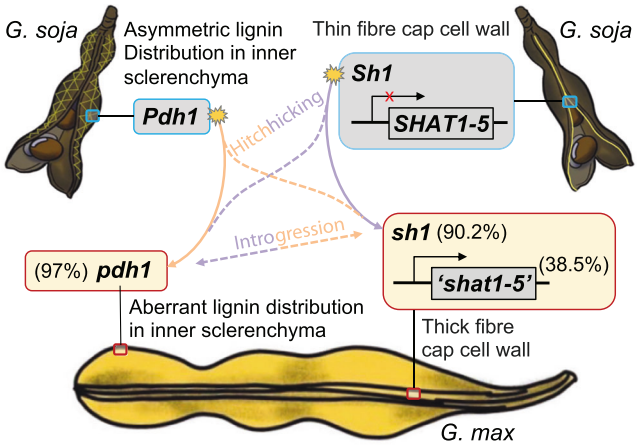

Fig. 6 | Schematic demonstration of genetic mechanisms and selection process for shattering. Allele *Pdh1* produces pod wall torsion, whereas *Sh1* weakens pod ventral suture through repressing *SHAT1-5* expression. *pdh1* and *sh1* are recessive alleles selected for shattering resistance. The frequencies of *pdh1*, *sh1*, and *shat1-5* in cultivated soybeans are shown in parentheses.

shaping the frequencies of the alleles underlying this domestication trait.

Our previous work revealed nearly perfect genic collinearity between the *Sh1-Pdh1* regions of soybean and common bean[31], which were estimated to have diverged from a common ancestor for ~18 million years[32]. As many wild legumes show pod shattering, it would be interesting to investigate whether the orthologs of these two genes in these legumes retain similar functions and functionality, and whether mutations, if occurred, in the two orthologous genes were also targeted for selection for shattering resistance in common bean and other leguminous crops.

Methods

Plant materials

The $F_{6:7}$ RIL populations used in this study were developed using a shattering-resistance cultivar Williams 82 and two shattering-susceptibility accessions, PI 468916 and PI 479752[9,16,17]. The genotypes were analyzed using Genotyping-by-Sequencing (GBS) and the phenotypes of 661 RIL lines were used for the QTL mapping for Williams 82 × PI 468916 and Williams 82 × PI 479752 populations. The shatter phenotypes of the parental lines Williams 82, PI 468916, PI 479752 grown in the field in West Lafayette, IN, were analyzed 10 days after full maturity (R8 stage). For gene expression analysis in different developing pods, the parental lines Williams 82, PI 468916 and PI 479752 were grown in the greenhouse and used to collect different growth stages of pods (R1: beginning bloom, R2: full bloom, R3: beginning pod, R4: full pod, R5: beginning seed, R6: full seed). For the Genome wide association study (GWAS) analysis, 3,099 soybean landraces were randomly selected from USDA Soybean Germplasm Collection (Supplementary Data 1). The phenotypic data for GWAS was downloaded from the USDA National Plant Germplasm System (https://npgsweb.ars-grin.gov/) and the SoySNP50K data was obtained from a previous study[18].

Phenotyping

The shattering phenotypes of parental lines and RIL populations were analyzed in the field, and the phenotypes of the parental lines and key recombinants were confirmed in both the field and the greenhouse. The pod dehiscence percentage >=30% was designated as shattering, and the pod dehiscence percentage <30% was designated as shattering resistant for the RIL populations. The transgenic soybean plants and control Williams 82 were grown in the greenhouse and >100 full

maturity pods (R8 stage) from the top part of plants of each line were collected and transferred into 37 °C oven for 50 days to analyze the percentage of pod dehiscence. Soybean pods from the six individual transgenic plants were harvested together as one biological replication, and three biological replications were analyzed in this study.

QTL mapping, fine mapping and GWAS analysis

The GBS of the 661 RIL population and QTL mapping were performed using composite interval mapping (CIM) model incorporated in the r/qtl package[9,33,34]. After the identification of shattering QTL on chromosome 16, we used the simple sequence repeat (SSR) markers and additional SNP, Indel markers to identify crossovers between individual markers, and fine mapped the QTL region. Genome-wide association analyses were performed using TASSEL5 with a mixed linear model[35,36]. All the primers used in this study are listed in Supplementary Data 5.

Plasmid construction and soybean transformation

The coding sequence of *Sh1* and *Pdh1* were amplified from the pod indehiscence variety Williams 82 and shattering genotype PI 468916, respectively, using specific primers with *Xho*I and *Xba*I enzyme site sequences. The resulting PCR products were digested by *Xho*I and *Xba*I to produce cohesive ends, respectively. The plasmid pPTN1171 was also digested using *Xho*I and *Xba*I to produce a linear plasmid[37]. The PCR fragment and the linear pPTN1171 were ligated using T4 ligase. The constructed plasmids were introduced into *Agrobacterium tumefaciens* LBA4404 to transform into soybean Williams 82, respectively. The soybean transformation was performed using *Agrobacterium*-mediated cotyledonary-node method[38]. T_2 transgenic soybean lines were used for shattering phenotype analysis. The soybean transgenic lines were confirmed using PCR, and the PCR fragments were sequenced to confirm the target gene. The expression levels of *Sh1/sh1* or *Pdh1/pdh1* in transgenic lines were analyzed using R3 stage pods, respectively. All the primers used in this study are listed in Supplementary Data 5.

RNA isolation, cDNA synthesis and gene expression analysis

The RNA isolation of soybean samples was performed using TRIzol reagent (Invitrogen, USA). The cDNA synthesis was conducted using SuperScript II reverse transcriptase (Promega, USA). Quantitative reverse transcription PCR (qRT-PCR) were performed using an ABI QuantStudio®5 (ABI, USA) machine with the set as follows: 95 °C for 5 s, 60 °C for 30 s, 40 cycles. Each sample was analyzed using three biological replicates. All the primers used in this study are listed in Supplementary Data 5.

Genetic diversity analysis

The genetic diversity (π) of the *G. soja*, landrace and cultivar populations were calculated in 100 kb windows with 10 kb steps across the *Sh1-Pdh1* region (ZH13.v2.0[39]) using VCFTools[40].

Y1H, dual luciferase, and EMSA assays

The CDS of *Sh1* was introduced into the yeast expression vector pB42AD. Different truncated *SHAT1-5* and *Pdh1* promoter regions from PI 468916 were cloned into yeast expression vector pLaczi2u, respectively. These fusion constructs were then co-transformed into the yeast strain EGY48. The transformed strains were incubated on SD-Ura/-Trp agar at 28 °C for 3–5 d, following which their transcriptional activities were assessed on SD-Ura/-Trp/Gal/Raf agar supplemented with X-gal. For dual-luciferase system analysis, the promoter fragment of *SHAT1-5*, spanning from 1201 bp to 1620 bp upstream of the ATG (S2 segment), was cloned into the pGreen II 0800-LUC vector (LUC), and the CDS of *Sh1/sh1* was cloned into the pGreen II 0029 62-SK vector (SK). The promoter sequences of *Sh1/sh1*, approximately 2.5 kb upstream of the ATG, and the CDS of *Sh1/sh1* were obtained from PI 468916 and Williams 82 by PCR, respectively. Different combinations of *Sh1/sh1* promoters and CDS were cloned into pZY101.2 construct digested with *XhoI/XbaI*[47].

The recombinant plasmids were transformed into *Agrobacterium tumefaciens* (GV3101), and dual-luciferase assay was performed for enzyme activity determination in *Nicotiana benthamiana* leaves 3 days after injection[41]. For EMSA, *Sh1* was cloned into pGEX-6P-1 vector and transformed into *Escherichia coli* strain BL21 codon plus (DE3). The transformed cells were induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by incubation at 37 °C for 4 h. The cell pellet was suspended in GST lysis buffer (1 mM PMSF, 1 mM DTT and binding buffer) and then subjected to sonication on ice at 22 W with a 4-s/4-s on/off cycle for 20 min and centrifuged at 8000 g for 30 min at 4 °C. The supernatant was then purified using a GST Sefinose™ Resin (Sangon Biotech) according to the manufacturer's instructions. For EMSA analysis, a Chemiluminescent EMSA Kit (Beyotime) was employed as per the manufacturer's instructions. Based on the *SHAT1-5* promoter sequence, four pairs of 60 bp double-stranded probes were designed, with biotinylation at the 3' end to function as Probe-Biotin, while the unlabeled probes served as Probe-Cold for competition. Documentation of the EMSA results was carried out using the plant in vitro fluorescence detector (Newton7.0, Vilber). The probes used for EMSA and primers are listed in Supplementary Data 5.

Analysis of fiber cap cells in the pod ventral suture

Mid-sections of soybean pods at the R7 stage were selected, which were fixed in 4% (w/v) paraformaldehyde and embedded in paraffin, and cut using a Leica RM 2235 microtome (Leica, Wetzlar, Germany)[42]. The 8 μm thick paraffin sections were stained with phloroglucinol for 5 min using a lignin staining solution from the SAINT-bio lignin staining kit (R23301, SAINT-bio, Shanghai, China), and the fiber cap cells within ventral suture were observed and imaged using a Carl Zeiss AG Axio Scope A1 microscope (Model 0340108).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. Source data are provided with the paper. Source data are provided with this paper.

References

1. Carter, T., Hymowitz, T. & Nelson, R. Biogeography, local adaptation, Vavilov, and genetic diversity in soybean. In: *Biological resources and migration*) (Springer, 2004).
2. Kataliko, R. K. et al. Resistance and correlation of pod shattering and selected agronomic traits in soybeans. *J. Plant Stud.* **8**, 1–39 (2021).
3. Tukamuhabwa, P., Dashiell, K. E., Rubaihayo, P. & Nabasirye, M. Determination of field yield loss and effect of environment on pod shattering in soybean. *Afr. Crop Sci. J.* **10**, 203–209 (2002).
4. Bailey, M., Mian, M., Carter, T. Jr, Ashley, D. & Boerma, H. Pod dehiscence of soybean: identification of quantitative trait loci. *J. Hered* **88**, 152–154 (1997).
5. Funatsuki, H. et al. Simple sequence repeat markers linked to a major QTL controlling pod shattering in soybean. *Plant Breed.* **125**, 195–197 (2006).
6. Liu, B. et al. QTL mapping of domestication-related traits in soybean (*Glycine max*). *Ann. Bot.* **100**, 1027–1038 (2007).
7. Kang, S.-T. et al. Population-specific QTLs and their different epistatic interactions for pod dehiscence in soybean [*Glycine max* (L.) Merr.]. *Euphytica* **166**, 15–24 (2009).
8. Suzuki, M., Fujino, K., Nakamoto, Y., Ishimoto, M. & Funatsuki, H. Fine mapping and development of DNA markers for the qPDH1 locus associated with pod dehiscence in soybean. *Mol. Breed.* **25**, 407–418 (2010).

9. Swarm, S. A. et al. Genetic dissection of domestication-related traits in soybean through genotyping-by-sequencing of two interspecific mapping populations. *Theor. Appl. Genet.* **132**, 1195–1209 (2019).
10. Funatsuki, H. et al. Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc. Natl. Acad. Sci. USA* **111**, 17797–17802 (2014).
11. Dong, Y. et al. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nat. Commun.* **5**, 1–11 (2014).
12. Hu, D. et al. Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments. *Front. Plant Sci.* **10**, 811 (2019).
13. Zhang, J. & Singh, A. K. Genetic control and geo-climate adaptation of pod dehiscence provide novel insights into soybean domestication. *Genes Genom Genet.* **10**, 545–554 (2020).
14. Zhang, Z. et al. Elimination of an unfavorable allele conferring pod shattering in an elite soybean cultivar by CRISPR/Cas9. *aBIOTECH* **3**, 110–114 (2022).
15. Yong, B. et al. Parallel selection of loss-of-function alleles of *Pdh1* orthologous genes in warm-season legumes for pod indehiscence and plasticity is related to precipitation. *New Phytol.* **240**, 863–879 (2023).
16. Sun, L. et al. *GmHs1-1*, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nat. Genet.* **47**, 939–943 (2015).
17. Zhang, D. et al. Elevation of soybean seed oil content through selection for seed coat shininess. *Nat. Plants* **4**, 30–35 (2018).
18. Song, Q. et al. Fingerprinting soybean germplasm and its utility in genomic research. *Genes Genom Genet.* **5**, 1999–2006 (2015).
19. Lyu, X. et al. The domestication-associated *L1* gene encodes a eucomic acid synthase pleiotropically modulating pod pigmentation and shattering in soybean. *Mol. Plant* **16**, 1178–1191 (2023).
20. Bayer, P. E. et al. Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. *Plant Genome.* **15**, e20109 (2022).
21. Zhou, Z. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**, 408–414 (2015).
22. Liu, Y. et al. Pan-genome of wild and cultivated soybeans. *Cell* **182**, 162–176.e113 (2020).
23. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
24. Parker, T. A., Lo, S. & Gepts, P. Pod shattering in grain legumes: emerging genetic and environment-related patterns. *Plant Cell* **33**, 179–199 (2021).
25. Bandillo, N. B. et al. Dissecting the genetic basis of local adaptation in soybean. *Sci. Rep.* **7**, 17195 (2017).
26. Futuyma, D. J. *Evolution* 3rd edn. (Sinauer Associates, Inc, 2013).
27. Lu, S. et al. Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* **52**, 428–436 (2020).
28. Huang, X., Huang, S., Han, B. & Li, J. The integrated genomics of crop domestication and breeding. *Cell* **185**, 2828–2839 (2022).
29. Wang, X., Chen, L. & Ma, J. Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication. *Genome Biol.* **20**, 1–5 (2019).
30. Wang, W. et al. A transposon-mediated reciprocal translocation promotes environmental adaptation but compromises domesticability of wild soybeans. *New Phytol.* **232**, 1765–1777 (2021).
31. Zhao, M., Zhang, B., Lisch, D. & Ma, J. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *Plant Cell* **29**, 2974–2994 (2017).
32. McClean, P. E., Mamidi, S., McConnell, M., Chikara, S. & Lee, R. Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genom.* **11**, 184 (2010).
33. Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
34. Broman, K. W., Wu, H., Sen, Ś. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
35. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
36. Zhang, Z. et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).
37. Ping, J. et al. *Dt2* is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. *Plant Cell* **26**, 2831–2842 (2014).
38. Song, Z.-y. et al. Screening Chinese soybean genotypes for *Agrobacterium*-mediated genetic transformation suitability. *J. Zhejiang Univ. Sci. B* **14**, 289–298 (2013).
39. Shen, Y. et al. Update soybean Zhonghuang 13 genome to a golden reference. *Sci. China Life Sci.* **62**, 1257–1260 (2019).
40. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
41. Zhang, Y. et al. Dual functions of PsmiR172b-PsTOE3 module in dormancy release and flowering in tree peony (*Paeonia suffruticosa*). *Hortic. Res.* **10**, uhad033 (2023).
42. Tang, X. et al. Ubiquitinated DA1 negatively regulates vascular cambium activity through modulating the stability of *WOX4* in Populus. *Plant Cell* **34**, 3364–3382 (2022).

Acknowledgements

We thank Wenying Xu and Xiaolin Yao for technical support and Chuanlong Lu for drawing pods in Fig. 6. This work was partially supported by the Shandong Province Taishan Young Scholar Program (tsqn202211190 to S.L.) and the Shandong Provincial Science Fund for Distinguished Young Scholars (ZR2022YQ21 to S.L.), Indiana Soybean Alliance (Soybean genetic diversity projects to J.M. and R.L.N.), and the Agriculture and Food Research Initiative of the U.S. Department of Agriculture National Institute of Food and Agriculture (2018-67013-27425 and 2021-67013-33722 to J.M.).

Author contributions

J.M., R.L.N., S.L., and W.W. designed the research; S.L., W.W., L.S., H.Z., R.H., H.Z., X.T., C.B.C., and S.A.S. performed the research and analyzed the data; J.M., S.L., and W.W. drafted the manuscript; J.M., R.L.N., C.B.C., and S.A.S. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52044-8.

Correspondence and requests for materials should be addressed to Shuai Li or Jianxin Ma.

Peer review information *Nature Communications* thanks Fanjiang Kong, Bin Liu, Petr Smýkal and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.