

# A domestication change at *PvMYB26* in common bean sheds light on the origins of Middle American agriculture

Burcu Celebioglu<sup>1</sup>, Jayanta Roy<sup>2</sup>, Andrew Farmer<sup>3</sup>, Stephanie English<sup>3</sup>, Xingyao Yu<sup>4</sup>, Xiaosa Xu<sup>4</sup>, Phillip E. McClean<sup>2</sup>, Paul Gepts<sup>1</sup>✉, Travis Parker<sup>1</sup>✉

<sup>1</sup>Department of Plant Sciences, University of California, Davis, CA 95616, USA

<sup>2</sup>Department of Plant Sciences, North Dakota State University, Fargo, ND, USA 58108

<sup>3</sup>National Center for Genome Resources, Santa Fe, New Mexico, USA

<sup>4</sup>Department of Plant Biology, University of California, Davis, CA 95616, USA

✉ e-mail: trparker@ucdavis.edu; plgepts@ucdavis.edu

## ORCIDs

Burcu Celebioglu: 0009-0000-9456-6483

Jayanta Roy: 0000-0003-3390-958X

Andrew Farmer: 0000-0002-4224-2433

Stephanie English: 0000-0002-7659-0399

Xiaosa Xu: 0000-0002-3452-6751

Phillip E. McClean: 0000-0002-4809-0864

Paul Gepts: 0000-0002-1056-4665

Travis Parker: 0000-0002-1233-7829

## Abstract

Domestication imposed radically different selection pressures on plants, eventually transforming them into the cultivated forms that support global populations today. Here, we investigate the loss of seed dispersal via pod shattering during common bean (*Phaseolus vulgaris* L.) domestication. We identified an eight-kb deletion eliminating the transcription start site and promoter of the candidate gene *PvMYB26*. Mutants express *PvMYB26* at < 1% of the level of wild types and produce 44% less pod lignin. Whole-genome sequencing revealed that the mutation is nearly diagnostic for domestication status among Middle American common bean, indicating its importance in domestication. We also identified a high-frequency *PvMYB26* frameshift/premature stop mutation unique to Andean domesticates. Wild haplotypes most like Middle American domesticates are found in eastern Jalisco, Mexico. Our results suggest that West-Central Mexico was the site of common bean domestication and suggest that this region may have been important in the rise of Middle American agriculture.

## Introduction

Plant domestication transformed the structure and adaptation of domesticated species, leading to a revolution in human evolution and social organization<sup>1-3</sup>. Domestication imposed novel selection pressures on plants, whereby mutations that would be deleterious in the wild became advantageous in the new cultivated environments. For example, elaborate seed dispersal mechanisms are essential to the survival of wild plants, but these became disadvantageous among seed-propagated species in cultivated agroecosystems. Consequently, there was strong selection against seed dispersal during plant domestication, across diverse taxa such as cereals<sup>4</sup>, grain legumes<sup>5</sup>, *Brassicas*<sup>6</sup>, and other seed crops<sup>7</sup>.

Seed dispersal in many legumes (Fabaceae) is mediated by pod shattering, with dry pods violently dehiscing at maturity due to pod walls rapidly coiling under significant torsional forces<sup>5</sup>. This trait was problematic for harvest and yield in cultivated environments, so reduced pod shattering became a core domestication syndrome trait in the family<sup>8-10</sup>, along with loss of seed dormancy<sup>10-13</sup>, and loss of pod shattering occurred up to 40 times in parallel domestications across the family<sup>5</sup>. Among these domesticated species, common bean (*Phaseolus vulgaris* L.) is the most important for direct human nutrition<sup>14-16</sup>. *P. vulgaris* was independently domesticated in Middle America and the Andes from gene pools that had diverged more than 100,000 years ago<sup>17-20</sup>, before human habitation of the Americas. In the Andean gene pool, variation in pod dehiscence has been fine-mapped to *PvMYB26*, previously postulated to be an improvement trait in snap beans<sup>7,21,22</sup>. To date, no causal mutations have been described in this gene in common bean, and the basis of reduced pod shattering during the initial domestication of the species is unknown.

Similarly, unresolved questions remain regarding the origins of agriculture in Middle America. These agricultural systems were founded around the staple crops maize (*Zea mays*), beans (*Phaseolus* spp.), and squash (*Cucurbita* spp.), known collectively as “the three sisters”. Within common bean, contrasting hypotheses regarding the geographic domestication center have been proposed: 1) the West-Central Mexico hypothesis, particularly in the state of Jalisco<sup>23,24</sup>, and more recently, 2) the Southern Mexico hypothesis, in the state of Oaxaca<sup>25</sup>. Since *Phaseolus* beans were particularly important during agricultural intensification and are an integral component of the Middle American crop assemblage, assessing their domestication origin is valuable for understanding the rise of agriculture in Middle America. Domestication alleles are rarely exchanged between wild and domesticated gene pools<sup>26-28</sup>, thereby offering a unique opportunity to determine crop origins without the effects of genetic exchange that occurs broadly in the rest of the genome<sup>26,29-31</sup>.

Here, we investigate the domestication of this critical New World staple crop through genetic mapping of two populations, whole-genome sequencing analysis of 323 novel wild and domesticated accessions, characterization of gene expression through qPCR and RNA-seq, developmental and biochemical screens, and geographic information system analysis. Our results

shed light on the genomic basis of a vital domestication trait and the origins of agriculture in Middle America.

## Results

### Variation in pod traits maps to *PvMYB26* and *PvPDH1*.

Both measured pod traits (number of pod valve twists and proportion of pods shattered) were significantly correlated in both the UC Canario 707 x Wild PI 638850 (CxW) and Orca x Wild PI 638850 (OxW) populations ( $P < 2 \times 10^{-16}$ ). Measurement of pod twists yielded more significant maximum LOD scores than direct measurement of pod shattering (Fig. 1, Supplementary Table 1). The most significant QTL for pod twists was mapped to Pv05 at 38,633,111 bp (LOD = 33.2) in the OxW population and 38,800,513 bp (LOD = 10.6) in the CxW population. The candidate gene *PvMYB26* (chromosome Pv05: 38.6 Mb; G19833 v2.1 gene model *Phvul.005G157600*) is located between the flanking markers of the detected QTL in both populations (Fig. 1b,e, Supplementary Table 1). A secondary locus for pod twists was identified on Pv04 in the OxW population. For pod shattering, two significant QTL were detected in the OxW population (Pv03: 49,199,252 bp, LOD=15.9; Pv05: 38,567,439 bp, LOD=10.4), with the candidate genes *PvPDH1* and *PvMYB26* falling between the peak markers of each QTL, respectively. The CxW population had three significant shattering QTL, mapping near *PvPDH1*, *PvMYB26*, and an additional locus on Pv03 (Supplementary Table 1). In both traits and both populations, no significant interactions were identified between loci, suggesting additive rather than epistatic relationships between them (fitqtl interaction  $P > 0.05$ ). RILs inheriting the wild-type allele had increased trait values in all cases. Each significant QTL explained 3.5-39.8% of phenotypic variance (Supplementary Table 1).

### A deletion eliminates 8kb of *PvMYB26*'s promoter and transcription start site in domesticated Middle American populations

Comparison of whole-genome sequencing data of the parents of the CxW and OxW populations near the pod twists and pod shattering QTL on Pv05 revealed an 8 kb deletion at *PvMYB26* in the domesticated parents Orca and UC Canario 707. The deletion, not found in the wild parent PI 638850, eliminates the transcription start site (TSS) of *PvMYB26*, including 61 bp of the 5' UTR and more than 8 kb of promoter (Fig. 1h, Supplementary Fig. 1). *PvMYB26* was therefore considered a strong candidate to account for the reduced shattering and twisting (Fig. 2a) during the domestication of Middle American beans.

To assess the functional significance of the deleted sequence through conservation analysis, we aligned *PvMYB26* and 10 kb upstream of the wild-type promoter region with 15 other legume species. Two strongly conserved noncoding sequence (CNS) motifs were identified

in the *PvMYB26* promoter across all examined legume species and in both paralogous *PvMYB26* copies of each species (Supplementary Fig. 2). An 82-bp untranslated motif at the transcription start site was found in identical form in all six evaluated species of *Phaseolus* and *Vigna*, 51 bp of which was deleted in UC Canario 707 and Orca. This strong sequence conservation suggested a functional regulatory significance, warranting further analysis.

## Gene expression and lignification are dramatically reduced in types with the *PvMYB26* TSS/promoter deletion

To characterize the effects of the 8 kb *PvMYB26* TSS/promoter deletion on gene expression, qPCR was conducted on OxW RILs at two time points. At 7 days after anthesis (DAA), *PvMYB26* expression was 18.3-fold lower in *pvmyb26* mutants than in wild-type lines (Fig. 2b,  $P=0.012$ ; heteroscedastic two-tailed *t*-test). By 20 DAA, *PvMYB26* expression was 306-fold lower in mutants ( $P=1.9 \times 10^{-8}$ ; heteroscedastic two-tailed *t*-test), indicating that the elimination of the transcription start site and promoter had a highly significant negative impact on the gene's transcription level. Additionally, analysis of published pod RNA-seq data<sup>32</sup> of four accessions across three pod sampling time points showed that *PvMYB26* expression was significantly decreased in mutant compared to wild-type individuals ( $P = 5.904 \times 10^{-5}$ , linear mixed model, Supplementary Fig. 3).

Microscopy and pod biochemical analyses were then conducted to determine the anatomical and biochemical effects of the *PvMYB26* deletion. *PvMYB26* mutants showed greatly reduced lignin deposition in the lignified wall layer (LWL) of pods, without observable differences in the bundle sheath, dehiscence zone, or other structures (Fig. 2c). Wild-type pods were composed of 6.0% lignin by mass. In contrast, mutant pods contained only 3.4% lignin, representing a statistically significant decrease of 43.5% in mutants (protected Fisher's LSD, Fig. 2d). In contrast, *PvPDH1* allele had no significant effect on pod lignin content, confirming that its effects on pod shattering are independent of lignin quantity<sup>33</sup>.

To further investigate the spatial expression pattern of *PvMYB26*, we performed mRNA in situ hybridization on pod sections sampled seven days after anthesis (DAA). Strong hybridization signals were specifically detected in the LWL cells of the pod walls (Fig. 2d), with minor expression observed in adjacent bundle sheath tissues. No visualizable hybridization signal was identified when the 8 kb deletion was present (Fig. 2d). These results indicate that *PvMYB26* expression occurs in the LWL, which is known to influence pod torsion and is differentially thickened between mutants and non-mutants<sup>5</sup>.

## A 125 kb domestication-related selective sweep occurred at *PvMYB26*

An initial assessment of the *PvMYB26* TSS/promoter deletion frequency in domesticated and wild materials was conducted by designing a high-throughput co-dominant PCR assay, which was screened on a collection of 300 *P. vulgaris* accessions. Strikingly, the deletion was

found in 99.1% of Middle American domesticates, but just 2.2% of wild beans ( $F_{ST} = 0.94$ ). The deletion, which is Middle American in origin, had also been introduced into 14% of accessions of the independently domesticated Andean Diversity Panel (Supplementary Table 2, Supplementary Fig. 4).

To better understand the haplotype diversity of *PvMYB26* and the geographic origin of the eight-kb mutation, whole-genome sequencing and analysis were then conducted on 326 lines. Only one among 55 Middle American wild lines (1.8%; PI 201016) showed the deletion. Of 138 domesticates with a Middle American *PvMYB26* haplotype, the deletion was found in 97.1%, only absent in four Mexican landraces (Fig. 3b).  $F_{ST}$  for the deletion was 0.89 in this panel.

To investigate the possibility of a broader selective sweep, we calculated Tajima's D, nucleotide diversity, and  $F_{ST}$  over a 600 kb interval centered on *PvMYB26*. Using all three metrics, a hard selective sweep was identified in the Middle American gene pool over 125 kb (chromosome Pv05 43.977 Mb to 44.102 Mb; 5\_593 genome v1.1). The 125 kb region exhibited a mean Tajima's D of  $-2.05 \pm 0.03$  in domesticates, whereas the average value outside this sweep interval was  $1.16 \pm 0.10$  (Fig. 4a, Supplementary Table 3). Tajima's D showed no significant reduction in the region among wild lines (Fig. 4d, Supplementary Table 3). In the sweep interval, domesticates had just  $12\% \pm 0.7\%$  of the nucleotide diversity found in wild types, whereas in the surrounding region they had  $53.4\% \pm 1.7\%$  of the diversity found in wild types (Fig. 4b,c). An increased mean  $F_{ST}$  of  $0.49 \pm 0.01$  was also observed between wild and domesticated lines, compared to a mean  $F_{ST}$  of  $0.25 \pm 0.01$  in the surrounding region (Fig. 4d). All of these statistics indicate that a hard sweep occurred across the region among domesticated Middle American populations, differentiating it from wild types.

## A 1 bp *PvMYB26* frameshift deletion exists solely in Andean domesticates

The whole-genome sequencing data were also analyzed for variation that might be responsible for reduced shattering in the independently domesticated Andean gene pool of common bean. We identified a one-bp frameshift deletion in exon 3 of *PvMYB26* (38,639,348 bp), leading to a complete disruption of the remaining 79 amino acid sequence and introducing a premature stop codon, which truncates its length by 45 amino acids. The deletion is found in 78.9% of Andean domesticates ( $n=76$ ) but is absent among wild beans ( $F_{ST} = 0.66$ ).

Seven wild lines (PI 638850, PI 390770, PI 638864, PI 638872, W6 18809, W6 18810, W6 18811) lacking the 1 bp deletion had otherwise identical haplotypes to this mutant sequence. These wild types spanned three countries (southern Peru, Bolivia, and Argentina), and no selection sweep was identified in the genome area around *PvMYB26* (Supplementary Table 4, Supplementary Fig. 5). The 1 bp deletion was therefore not considered informative in identifying the specific geography of domestication in Andean beans.

## West-Central Mexico as a center of origin for Middle American agriculture

The extreme  $F_{ST}$  and near total lack of gene flow between wilds and domesticates at *PvMYB26* offered a unique opportunity to identify the geographic origins of Middle American beans without the effects of historical admixture. To this end, consensus FASTA sequences within 2 kb of the *PvMYB26* promoter deletion were extracted and aligned with the predominant domesticated haplotype. The wild haplotypes most similar to the domesticated haplotypes were found in eastern Jalisco (PI 417707, PI 417708). The entire coding DNA sequence of *PvMYB26* is identical to that of domesticates in these types, with only a one-bp promoter indel distinguishing them from the predominant domesticated haplotype. Accessions from northern Michoacan (G12894, PI 417621) also had *PvMYB26* transcribed sequences identical to domesticates. Three wild accessions from Morelos (PI 325679, PI 535430, PI 535444) and one from southern Guanajuato (G23433) had promoter sequences identical to domesticates but differences in the transcribed region. The identification of wild haplotypes most similar to domesticates in this region suggests that this area was the site of origin for the *pvmbyb26* mutation that strongly distinguishes wild and domesticated Middle American populations, and indicates the region was likely the Middle American center of domestication of common bean (Fig. 5).

## Discussion

In grain-propagated crops, including cereals, pulses, and *Brassica* oil crops, the absence of grain release at maturity is of utmost importance to maximize yields under cultivation. Preventing grain shedding has taken multiple anatomical appearances, including acquiring a stiff rachis through the disruption of abscission layer formation in Fertile Crescent cereals<sup>4,34</sup> and losing explosive pod shattering in grain legumes, due to reduced pod torsion force or changes in dehiscence zones<sup>5,9</sup>. Identifying the molecular basis of crucial domestication traits sheds light on the entire domestication process and more broadly on the origins of agriculture<sup>2,35</sup>. Here, we document research results leading us to the identification of a single major gene – *PvMYB26* – which likely reduced pod shattering in the initial stages of the two domestications of common bean, in Middle America and the southern Andes<sup>17,36</sup>. Furthermore, a comparison of haplotypes in and around the locus suggests a domestication area for the Middle American common bean and, more broadly, an initiation of Mesoamerican agriculture in west-central Mexico.

The evidence supporting a key role of the *PvMYB26* gene on chromosome Pv03 in pod shattering includes genetic linkage mapping in two recombinant inbred populations between a domesticated and a wild bean parent and the identification of an eight-kb deletion affecting the upstream promoter-containing sequence and the transcription start site in Middle American domesticates. This deletion mutation strongly reduces the gene's expression, is correlated with reduced lignin deposition in the pod walls and is accompanied by a substantial reduction in

genetic diversity in nearly every Middle American domesticated type tested ( $n = 134$ ) compared to the corresponding wild progenitor, characterizing a marked selective sweep of some 125 kb in the genome region in and around the gene. These sequence features indicate a Middle American origin in west-central Mexico for the mutation, with subsequent introgression in a minority of the Andean domesticated gene pool. In contrast, most of the Andean domesticated gene pool shows a one-bp frameshift deletion in *PvMYB26*, causing a truncation and loss-of-function of the gene product. All these results strongly suggest a primordial role for *PvMYB26* in pod dehiscence and bean domestication.

The *PvMYB26* hard sweep, extending over 125 kb, is significantly larger than the 66 kb selective sweep found at the *sh4* shattering locus in rice<sup>37</sup>, or the 65.6 kb sweep at the *tb1*<sup>38</sup> in maize, both of which are key domestication genes in their respective species. The size difference may be attributed to stronger linkage disequilibrium in common bean compared to the two cereal species and the limited gene exchange between wild and domesticated beans, especially around domestication genes<sup>26-28</sup>. The hard sweep around *PvMYB26* contrasts with the multiple mutations and soft sweeps that have been commonly observed among domestication syndrome traits across species<sup>39,40</sup>, including common bean<sup>41-45</sup>. In the Andean gene pool, where the selective sweep at *PvMYB26* was weaker (Fig. 4C, Supplementary Table S4, Supplementary Fig. S5), a greater number of secondary loci have been described<sup>33</sup>. There has also been a significant introgression of the Middle American mutant haplotype into the Andean gene pool (Supplementary Table S2, Supplementary Fig. S4).

The expression pattern of *PvMYB26*, which was limited to the LWL fibers, correlates with the observed lignin deficiency phenotype in *pvmyb26* mutants and supports *PvMYB26*'s direct role in regulating lignification within this specialized cell layer. The connection between MYB transcription factors and lignin biosynthesis is well-established across plant species<sup>46-52</sup>. In *Arabidopsis*, *MYB58* and *MYB63* directly activate lignin biosynthetic genes like *PHENYLALANINE AMMONIA LYASE 1 (PAL1)*, *CINNAMATE 4-HYDROXYLASE (C4H)*, and *4-COUMARATE COA LIGASE 1 (4CL1)*, by binding AC-rich promoter elements<sup>47,53</sup>. Similarly, *MYB26* regulates secondary cell wall thickening in anthers by controlling *NAC SECONDARY WALL THICKENING PROMOTING FACTOR1 (NST1)* and *NST2* expression, while *MYB46/MYB83* orchestrate a hierarchical network activating both lignin biosynthesis and secondary wall formation<sup>46,47</sup>. These findings align with our observations that *PvMYB26* disruption specifically impacts LWL lignification, highlighting the precision of MYB-mediated transcriptional control in spatial regulation of lignification. The conserved role of MYB transcription factors in activating lignin pathway genes provides a mechanistic context for the severe lignin reduction caused by the *PvMYB26* promoter deletion.

While *PvMYB26* may have been sufficient for reduced pod shattering during the initial domestication phase of common bean, it was not sufficient for control of pod shattering in arid environments, which exacerbate pod shattering. In these environments, additional mutations in the *PvPDH1* dirigent gene on chromosome Pv03<sup>5,33,54</sup> and other genes, selected after domestication, provided additional resistance to pod shattering and local ecogeographic

adaptation (Fig. 6). *PvMYB26* had previously been described as an improvement gene selected subsequently in the development of modern snap beans<sup>7,21,22</sup>. However, the near-uniform presence of the 8-kb mutation in the domesticated Middle American gene pool strongly suggests that *PvMYB26* is a foundational component of the initial domestication process of Middle American common beans. Identifying the one-bp frameshift mutation in 79% of dry beans with Andean haplotypes also indicates its importance in the initial domestication of that gene pool.

Our results further suggest that, at least in the case of pod shattering, orthologous genes may have been involved across grain legume domestications. In both common bean domestications – Middle American and southern Andean – mutations in the same gene that reduced pod shattering were selected. This observation raises the question of whether other, more distantly related grain legumes show a similar convergence. The legume family was domesticated approximately 40 times independently, more than any other plant family<sup>5</sup>, creating an unequaled natural experiment in parallel evolution. Growing evidence indicates that *MYB26* homologs are critical regulators of seed dispersal and carpel development across legumes, contrasting their anther-specific role in *Arabidopsis*<sup>46,55</sup>. In *Vigna* sp., loss of pod shattering has been linked to loss-of-function mutations in *MYB26* orthologs in five independently domesticated lineages<sup>56,57</sup>. These *Vigna MYB26* mutants also show reduced fiber deposition in the pod LWL, consistent with our anatomical and biochemical results and other studies in common bean<sup>21,58</sup>. Unlike some *Vigna* mutants, however, the TSS/promoter deletion of *PvMYB26* does not lead to a complete elimination of pod lignin in the LWL, which may be advantageous for threshing during harvest as it facilitates the recovery of grains with limited damage.

Similarities in protein or nucleic acid sequences<sup>2,35,59-61</sup> between wild and domesticated types of crop species can provide information about the geography of domestication, including the number and location of putative domestication regions, with the caveat that introgression can have a confounding effect in this geographic assessment. In this case, the almost exclusive presence of the mutant *PvMYB26* haplotype in the domesticated Middle American gene pool and its near absence in the sympatric wild gene pool confirmed the restricted gene flow around domestication genes observed by Papa et al.<sup>27,28</sup> and limited any confounding effects due to admixture. The comparison of the two-kb haplotype surrounding the *PvMYB26* deletion showed that wild haplotypes most like the domesticated ones were situated in a region encompassing eastern Jalisco and northern Michoacán (Fig. 5). This region is largely congruent with a region identified by Gepts<sup>23</sup> and Kwak et al.<sup>24</sup>, using different genome-wide molecular markers. This region overlaps the current area and includes areas to the east in Guanajuato and western Jalisco in the Rio Lerma-Rio Santiago basin.

It is striking that this area overlaps with a region in which the presumed ancestors of the three major crops of the principal cropping system (milpa system) are distributed: beans (*Phaseolus* spp.), maize (*Zea mays* L.), and squash (*Cucurbita* spp., and more specifically *C. argyrosperma*) (Zizumbo-Villarreal and Colunga García-Marín 2010). A putative domestication region for maize and *C. argyrosperma* has been widely hypothesized to be in the lower Balsas



River basin<sup>62-66</sup>, leading Zizumbo-Villarreal and Colunga García-Marín<sup>67</sup> to suggest that the northwestern part of the Balsas-Jalisco biogeographic region, which encompasses the two largest river basins of Mexico, could have witnessed the initiation of the cultivation and domestication of the three major component crops of the milpa agroecological food system and, hence, agriculture in Mesoamerica. Further cultural developments led to the expansion of the basic milpa cropping system by integrating additional crops and animals and developing more complex food systems supporting the diverse pre-Hispanic civilizations<sup>68,69</sup>.

Finally, the following evolutionary pod domestication scenarios can be proposed. In the Middle American gene pool (Fig. 6), more specifically in what is now west-central Mexico (Jalisco, northern Michoacan, and western Guanajuato), selection for reduced pod shattering during the initial phase of domestication may have led to the *de novo* selection and near-fixation of an 8-kb deletion affecting the *PvMYB26* locus on chromosome Pv05 (current results). In a subsequent phase, differentiation of the incipient MA gene pool into ecogeographic races and dissemination across the diverse ecosystems of the Middle American area<sup>36,70</sup>, selection for nearly complete pod indehiscence in the arid northern Mexican highlands led to the preponderance of a *PvPDHI* (Pv03) loss-of-function, non-synonymous substitution in race Durango, which includes bean types such as pintos, pinks, and great northern<sup>54</sup>. In contrast, a separate migration of Middle American domesticates towards the Middle American lowlands (race Mesoamerica, represented by smaller-seeded navy and black beans<sup>70</sup>) did not result in selection for additional pod shattering tolerance<sup>54</sup>, given the more humid nature of the habitat. A similar scenario, involving the same genes, can be developed for the Andean domestication. Reductions in pod shattering can be traced to either the one-bp mutation in *PvMYB26*, to the introgression of the 8-kb mutation from Middle American domesticates, or selection of hitherto undetermined pod shattering genes in the Andean gene pool. As observed earlier, however, the reduced diversity of the latter gene pool prevents us from identifying a more specific location of the Andean domestication anywhere between southern Peru and northwestern Argentina.

Understanding the genetic basis of domestication enhances the ability to identify critical genetic diversity in landraces and crop wild relatives maintained in gene banks to develop more resilient crops and improve global food security.

## Online methods

### Plant Materials and Phenotyping

‘Orca’ (domesticated Middle American common beans<sup>71</sup>) and ‘PI 638850’ (wild Andean common bean) were used as parents to develop a 292-member “OxW” RIL population segregating for domestication-related traits, which were phenotyped at the F<sub>4</sub> and F<sub>5</sub> generations. The domesticated Andean bean ‘UC Canario 707’<sup>72</sup> and wild PI 638850 were used as parents for a second RIL population, (“CxW”), consisting of 269 members phenotyped at the F<sub>5</sub> and F<sub>6</sub> generations. While UC Canario 707 is primarily Andean in ancestry, it contains a Middle

American haplotype at *PvMYB26*. Orca and UC Canario 707 are highly resistant to shattering, while PI638850 is highly prone to shattering, typical of wild beans. RIL populations were planted in the University of California, Davis greenhouse during the fall and winter of 2022-2023 and 2023-2024, to allow for the long nights required for flowering in photoperiod-sensitive wild materials. Temperatures were maintained between 15°C at night and 26°C during the day. Dry pods were harvested before shattering and allowed to dry for at least 30 days at room temperature in laboratory conditions to express pod shattering. The proportion of pods shattering was then quantified. To test pod twists, non-shattered pods were then artificially split and then desiccated through incubation at 40°C for two weeks, with an additional two weeks of re-equilibration at room temperature. The number of pod twists (rotations) was then taken from three representative pods of each RIL.

Initial marker screens were run on a panel of 300 lines: 90 wild types, 110 Middle American Diversity Panel (MDP)<sup>73</sup> domesticates, and 100 Andean Diversity Panel (ADP)<sup>74</sup> domesticates. This panel of 300 lines included all wild accessions in the NPGS gene bank with seed sizes smaller than 15 g/100. Domesticates were chosen randomly among the diversity panels.

Whole-genome shotgun sequencing was conducted on a panel of 289 novel lines, analyzed along with 37 recently published genomes<sup>45</sup>. In total, the collection included 326 lines, of which 225 were domesticated, 89 were wild, and 12 were admixed. According to genome-wide sequencing data, accessions were classified as domesticated, wild, or admixed.

## Genetic mapping

Genotyping of the CxW and OxW populations was conducted according to the Khufu method<sup>75</sup> by the HudsonAlpha Institute (Huntsville, AL, USA), yielding 48,197 initial SNPs in the CxW population and 189,595 initial SNPs in the OxW population. SNP data were cleaned using the methods of Xie et al.<sup>76</sup>. After filtering co-segregating markers, 2,113 SNPs were retained in the CxW population, and 2,156 were retained in the OxW population. The ASMAP and R/QTL packages were used for linkage and QTL mapping<sup>77-79</sup> using the Haley-Knott algorithm<sup>80</sup>. The fitqtl() function was used to assess allelic interactions and effects. A LOD significance threshold of 3.21 was used based on the 95% threshold of 1000 randomized permutations of the data. Allele effect plots were visualized in ggplot2<sup>81</sup>.

## Sequencing and analysis

Illumina 150 bp paired-end sequencing was conducted to a depth of 12-16x was conducted on the 289 novel accessions at the UC Davis Genome Center's DNA Technologies Core and the HudsonAlpha Institute for Biotechnology (Huntsville, AL, USA). DNA for sequencing was isolated using the Qiagen DNeasy plant mini kit, verified for quality by Nanodrop and gel. Raw sequences were quality-checked with FASTQC<sup>82</sup> and trimmed using Trimmomatic v0.39's ILLUMINACLIP function<sup>83</sup>. The Burrow-Wheeler Aligner (BWA-

MEM<sup>84</sup>) was used with default settings to align trimmed reads to the reference genomes G19833 v2.1<sup>19</sup> ([https://phytozome-next.jgi.doe.gov/info/Pvulgaris\\_v2\\_1](https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1)) and 5-593 ([https://phytozome-next.jgi.doe.gov/info/Pvulgaris5\\_593\\_v1\\_1](https://phytozome-next.jgi.doe.gov/info/Pvulgaris5_593_v1_1)). Consensus sequence was extracted using samtools consensus<sup>85</sup>. Alignments were visualized in the Integrative Genomics Viewer (IGV)<sup>86</sup>. Protein models were generated using the SWISS-MODEL Expasy web server [SWISS-MODEL (expasy.org)].

To identify conserved non-coding sequence, *PvMYB26* and 10 kb of its promoter sequence were aligned with 15 other legume species extracted from Phytozome 13<sup>87</sup> (accessed 14 September 2024), as well as the *Pisum sativum* assembly cultivar ‘Zhewan’ from NCBI (<https://www.ncbi.nlm.nih.gov/datasets/taxonomy/3888/>, accessed 24 January, 2025), with sequence alignment conducted using NCBI BLASTn.

## Microscopy

Pods from the OxW population were collected at stage R8 (full-seeded green pod) from lines characterized at allele for both *PvPdh1* and *PvMYB26* genes based on selectable markers for each gene (Supplementary Table 5). Sections 100 µm thick were made using a vibratome and treated 0.01% Auramine O and 0.007% Calcofluor solution for 20 minutes<sup>88</sup>. These two stains enable the secondary cell wall components in plant tissues to be distinguished from each other under the microscope, with Calcofluor dyeing cellulose blue and Auramine O staining lignin and other hydrophobic tissues (e.g., cuticle) green. Images were visualized using a fluorescence microscope (Olympus BH2-RFL, Waltham, MA, USA) with ultraviolet filter set (UG-1 and DM-400 + L-420).

## mRNA *in situ* hybridization

mRNA *in situ* hybridizations and probe synthesis were performed as previously described<sup>89,90</sup>. Briefly, bean pods of six OxW RILs were collected at seven days after anthesis (DAA). Pods were hand-sectioned transversely and fixed in 4% paraformaldehyde (Electron Microscopy Sciences). Fixed tissue was dehydrated through a graded ethanol series (50%, 70%, 85%, 95%, and 100%), followed by a Histoclear series, and embedded in paraffin (epredia Precision Cut, B1002490). Sections of 10 µm thickness were cut using a Leica microtome and mounted on ProbeOn Plus Slides (Fisher Scientific, 22-230-900). Primer sequences used to amplify *PvMYB26* for probe synthesis were TACACACCAGGCATTCGTGC and GAACTAGCATGGCTCTGAGGT. Images were acquired using an OLYMPUS BX50 DIC microscope.

## Expression analyses

Pod samples were harvested at two growth stages, 7 and 20 days after anthesis (DAA). Pod sections 1 cm thick were taken with a razor blade, and seeds were removed from the 20 DAA samples sections. Samples were flash frozen in liquid N<sub>2</sub> and stored frozen at -80°C. Pod

sections were ground in liquid N<sub>2</sub> using a mortar and pestle, and total RNA was extracted using the RNeasy Plant Mini kit (Qiagen, Hilden, Germany). Concentration and quality were subsequently confirmed by NanoDrop and 1% bleach gel. Then 3 µl RNA per sample was converted to cDNA using SuperScript IV VILO Master Mix with ezDNase. RT-qPCR primers were designed using NCBI primer blast. *Act11* (*Phvul.002G269900*) was used as a control gene. Primers spanning intron-exon junctions were used for both genes (Supplementary Table 1). qPCR primer efficiency was checked on pooled cDNA from each sample. For 7 DAA,  $n=11$  for *PvMYB26/PvMYB26*,  $n=7$  for *pvmby26/pvmby26*; for 20 DAA  $n=11$  *PvMYB26/PvMYB26*,  $n=5$  for *pvmby26/pvmby26*. For  $\Delta C_T$  calculation, the  $C_T$  values of the reference gene were subtracted from the  $C_T$  values of the candidate gene. These values were then converted to  $2^{-\Delta C_T}$  values. The mean, standard deviation, and standard errors of  $2^{-\Delta C_T}$  data for both *PvMYB26* wild and domesticated allele lines were calculated. Expression differences between these two alleles were compared using heteroscedastic two-tailed  $t$ -test.

Pod RNA-seq data from Parker et al.<sup>32</sup> were analyzed for four genotypes: Midas, G12873, ICA Bunsu, and SXB 405; at three different stages (R7, pod formation; R8, pod fill; and R9, pod maturation<sup>91</sup>). Three replicates were conducted for each stage and genotype. The expression patterns of ICA Bunsu and SXB 405 were compared with the expression patterns of Midas and G12873 to identify transcripts differentially expressed between *PvMYB26* mutants and wild types. The expression of candidate loci was compared by mixed linear model with mutation status as a fixed effect and maturity stage as a random effect.

## Lignin content

OxW RILs were classified according to genotype at both *PvMYB26* and *PvPDH1* using codominant PCR markers, with four total allelic categories. For each RIL and year, three pods were assigned to one of four technical replicate bins for each of the four corresponding allelic categories. These dry pods were then submitted to Cumberland Valley Analytical Services for lignin quantification. Lignin mass was determined using the Goering and Van Soest<sup>92</sup> method with modifications. The fiber residue from the acid-detergent fiber (ADF) extraction was recovered on a Whatman Glass Fiber Filter in a California Buchner Funnel. Approximately 45 ml of 72% Sulfuric Acid was added to the fiber residue, followed by shaking for two hours to ensure that fibers were washed with acid. After the acid-washed fiber residue was filtered, rinsed, dried and weighed, lignin residue was washed for two hours to remove lignin organic matter. The ash residue was weighed again, and the amount of lignin was determined by subtracting it from the original weight. Results were analyzed using an ANOVA-protected Fisher's LSD test based on mutation status at *PvMYB26* and *PvPDH1*.

## Haplotype diversity

To generate haplotype networks and gene trees, consensus FASTA within 2 kb of the *PvMYB26* TSS/promoter deletion was extracted from the reference genome aligned whole

genome sequencing lines using samtools consensus function<sup>85</sup>. This 4 kb region included the full transcribed region of the gene and approximately 2 kb of promoter. Alignments were built with MAFFT (version 7)<sup>93</sup>. A single gene tree was estimated with IQ-TREE (v.1.6.12)<sup>94</sup> under the GTR+G model<sup>95</sup> with 1,000 bootstrap replicates. The phylogenetic tree was visualized by the iTOL online server<sup>96</sup>. To find out the haplotype network, the NeighborNet algorithm was implemented in the SplitsTree program (version 6.4.11)<sup>97</sup>.

Population genetics statistics were calculated on biallelic variants identified within a 600 kb region centered on *PvMYB26*. A vcf was generated from .bam files using bcftools mpileup and bcftools call, based on the 5\_593 v1.1 reference genome. Principal component analysis was performed using PLINK 1.9<sup>98</sup> and used to classify materials as having Andean, Middle American, or admixed haplotypes at the region. Wild, domesticated, and admixed status were classified based on genome-wide SNP data, identifying several misclassified lines in the NPGS-GRIN database. The biologically distinct and independently domesticated Middle American (wild  $n=55$ , domesticated  $n=139$ ) and Andean (wild  $n=31$ , domesticated  $n=76$ ) gene pool materials were then processed and handled separately for subsequent steps. The R package PopGenome<sup>99</sup> was used to calculate nucleotide diversity,  $F_{st}$ , and Tajima's D. Variants were loaded using the option "include.unknown=TRUE" after determining that the level of missing data was not substantial enough to bias the analyses when run in nucleotide mode. A sliding window transform was used to split the variants into 25 kb windows with a 5 kb step. Statistics were plotted using the R package ggplot2<sup>81</sup> with the position on the x-axis representing the position at the center of each window.

## Geographic Analysis

Heatmaps of similarity between wild lines and the core domesticated haplotype were generated using the consensus sequences within 2 kb of the TSS/promoter deletion used for haplotype network analysis. Consensus sequences were aligned to the predominant Middle American domesticated haplotype of accession 5-593 using NCBI BLASTn. E scores and geographic origins of wild Middle American accessions were then loaded into QGIS v3.22.6. Heatmaps were generated using the IDW interpolation function with a distance coefficient set to  $P=3$ . National, state, and oceanic boundaries were downloaded from Natural Earth data (Kelso and Patterson 2010).

## References

- 1 Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700-707 (2002).
- 2 Larson, G. *et al.* Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* **111**, 6139-6146 (2014).
- 3 Purugganan, M. D. What is domestication? *Trends in ecology & evolution* **37**, 663-671 (2022).

489 4 Yu, Y. & Kellogg, E. A. Multifaceted mechanisms controlling grain disarticulation in the  
490 Poaceae. *Current Opinion in Plant Biology* **81**, 102564 (2024).  
491 5 Parker, T. A., Lo, S. & Gepts, P. Pod shattering in grain legumes: emerging genetic and  
492 environment-related patterns. *Plant Cell* **33**, 179-199 (2021).  
493 <https://doi.org/10.1093/plcell/koaa025>  
494 6 Ogutcen, E. *et al.* Pod shattering: a homologous series of variation underlying domestication and  
495 an avenue for crop improvement. *Agronomy* **8**, 137 (2018).  
496 7 Di Vittori, V. *et al.* Convergent Evolution of the Seed Shattering Trait. *Genes (Basel)* **10** (2019).  
497 <https://doi.org/10.3390/genes10010068>  
498 8 Hammer, K. Das Domestikationssyndrom. *Die Kulturpflanze* **32**, 11-34 (1985).  
499 <https://doi.org/10.1007/bf02098682>  
500 9 Smartt, J. *Grain legumes: evolution and genetic resources*. (Cambridge university press, 1990).  
501 10 Zhang, H., Mascher, M., Abbo, S. & Jayakodi, M. Advancing grain legumes domestication and  
502 evolution studies with genomics. *Plant and Cell Physiology* **63**, 1540-1553 (2022).  
503 11 Koinange, E. M., Singh, S. P. & Gepts, P. Genetic control of the domestication syndrome in  
504 common bean. *Crop Science* **36**, 1037-1045 (1996).  
505 12 Soltani, A. *et al.* The genetics and physiology of seed dormancy, a crucial trait in common bean  
506 domestication. *BMC plant biology* **21**, 1-17 (2021).  
507 13 Palmer, J. P. *et al.* Pectin acetyltransferase 8 influences pectin acetylation in the seed coat, seed  
508 imbibition, and dormancy in common bean (*Phaseolus vulgaris* L.). *Legume Science* **4**, e130  
509 (2022).  
510 14 Broughton, W. J. *et al.* Beans (*Phaseolus* spp.)—model food legumes. *Plant and soil* **252**, 55-128  
511 (2003).  
512 15 Bitocchi, E. *et al.* Beans (*Phaseolus* spp.) as a model for understanding crop evolution. *Frontiers*  
513 *in plant science* **8**, 722 (2017).  
514 16 Parker, T. A. *et al.* Genetic resources and breeding priorities in *Phaseolus* beans: Vulnerability,  
515 resilience, and future challenges. *Plant breeding reviews* **46**, 289-420 (2022).  
516 17 Gepts, P., Osborn, T., Rashka, K. & Bliss, F. Phaseolin-protein variability in wild forms and  
517 landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of  
518 domestication. *Economic botany* **40**, 451-468 (1986).  
519 18 Mamidi, S. *et al.* Demographic factors shaped diversity in the two gene pools of wild common  
520 bean *Phaseolus vulgaris* L. *Heredity* **110**, 267-276 (2013).  
521 19 Schmutz, J. *et al.* A reference genome for common bean and genome-wide analysis of dual  
522 domestications. *Nature genetics* **46**, 707-713 (2014).  
523 20 Ariani, A., Berny Mier y Teran, J. C. & Gepts, P. Spatial and temporal scales of range expansion  
524 in wild *Phaseolus vulgaris*. *Molecular biology and evolution* **35**, 119-131 (2018).  
525 21 Rau, D. *et al.* Genomic dissection of pod shattering in common bean: mutations at non-  
526 orthologous loci at the basis of convergent phenotypic evolution under domestication of  
527 leguminous species. *The Plant Journal* **97**, 693-714 (2019).  
528 22 Di Vittori, V. *et al.* Pod indehiscence in common bean is associated with the fine regulation of  
529 PvMYB26. *J Exp Bot* **72**, 1617-1633 (2021). <https://doi.org/10.1093/jxb/eraa553>  
530 23 Gepts, P. in *Genetic Resources of Phaseolus Beans: Their maintenance, domestication, evolution*  
531 *and utilization* 215-241 (Springer, 1988).  
532 24 Kwak, M., Kami, J. A. & Gepts, P. The putative Mesoamerican domestication center of  
533 *Phaseolus vulgaris* is located in the Lerma–Santiago Basin of Mexico. *Crop Science* **49**, 554-563  
534 (2009).  
535 25 Bitocchi, E. *et al.* Molecular analysis of the parallel domestication of the common bean (*P*  
536 *haseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist* **197**, 300-313 (2013).  
537 26 Papa, R. & Gepts, P. Asymmetry of gene flow and differential geographical structure of  
538 molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from  
539 Mesoamerica. *Theoretical and applied genetics* **106**, 239-250 (2003).

- 27 Papa, R., Acosta, J., Delgado-Salinas, A. & Gepts, P. A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theoretical and Applied Genetics* **111**, 1147-1158 (2005).
- 28 Papa, R. *et al.* Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Annals of Botany* **100**, 1039-1051 (2007).
- 29 Hufford, M. B. *et al.* The genomic signature of crop-wild introgression in maize. *PLoS genetics* **9**, e1003477 (2013).
- 30 Janzen, G. M., Wang, L. & Hufford, M. B. The extent of adaptive wild introgression in crops. *New Phytologist* **221**, 1279-1288 (2019).
- 31 Yang, N. *et al.* Two teosintes made modern maize. *Science* **382**, eadg8940 (2023).
- 32 Parker, T. A. *et al.* Loss of pod strings in common bean is associated with gene duplication, retrotransposon insertion and overexpression of PvIND. *New Phytologist* **235**, 2454-2465 (2022).
- 33 Parker, T. A., Berny Mier y Teran, J. C., Palkovic, A., Jernstedt, J. & Gepts, P. Pod indehiscence is a domestication and aridity resilience trait in common bean. *New Phytologist* **225**, 558-570 (2020).
- 34 Tanno, K.-i. & Willcox, G. How fast was wild wheat domesticated? *Science* **311**, 1886-1886 (2006).
- 35 Gepts, P. The contribution of genetic and genomic approaches to plant domestication studies. *Current opinion in plant biology* **18**, 51-59 (2014).
- 36 Kwak, M. & Gepts, P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theoretical and Applied Genetics* **118**, 979-992 (2009).
- 37 Zhang, L. B. *et al.* Selection on grain shattering genes and rates of rice domestication. *New Phytologist* **184**, 708-720 (2009).
- 38 Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature genetics* **43**, 1160-1163 (2011).
- 39 Turner-Hissong, S. D., Mabry, M. E., Beissinger, T. M., Ross-Ibarra, J. & Pires, J. C. Evolutionary insights into plant breeding. *Current Opinion in Plant Biology* **54**, 93-100 (2020).
- 40 Zhong, L., Zhu, Y. & Olsen, K. M. Hard versus soft selective sweeps during domestication and improvement in soybean. *Molecular Ecology* **31**, 3137-3153 (2022).
- 41 Kwak, M., Toro, O., Debouck, D. G. & Gepts, P. Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). *Annals of Botany* **110**, 1573-1580 (2012).
- 42 McClean, P. E. *et al.* White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytologist* **219**, 1112-1123 (2018).
- 43 McClean, P. E. *et al.* T and Z, partial seed coat patterning genes in common bean, provide insight into the structure and protein interactions of a plant MBW complex. *G3: Genes, Genomes, Genetics* **14**, jkae184 (2024).
- 44 Weller, J. L. *et al.* Parallel origins of photoperiod adaptation following dual domestications of common bean. *Journal of Experimental Botany* **70**, 1209-1219 (2019).
- 45 Parker, T. *et al.* Seed color patterns in domesticated common bean are regulated by MYB-bHLH-WD40 transcription factors and temperature. *The Plant Journal* **119**, 2765-2781 (2024).
- 46 Yang, C. *et al.* Transcription factor MYB26 is key to spatial specificity in anther secondary thickening formation. *Plant physiology* **175**, 333-350 (2017).
- 47 Xie, M. *et al.* Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Frontiers in plant science* **9**, 1427 (2018).
- 48 Kim, S. H. *et al.* The Arabidopsis R2R3 MYB transcription factor MYB15 is a key regulator of lignin biosynthesis in effector-triggered immunity. *Frontiers in Plant Science* **11**, 583153 (2020).
- 49 Miyamoto, T., Tobimatsu, Y. & Umezawa, T. MYB-mediated regulation of lignin biosynthesis in grasses. *Current Plant Biology* **24**, 100174 (2020).
- 50 Xiao, R., Zhang, C., Guo, X., Li, H. & Lu, H. MYB transcription factors and its regulation in secondary cell wall formation and lignin biosynthesis during xylem development. *International journal of molecular sciences* **22**, 3560 (2021).

591 51 Xu, X. *et al.* MdMYB52 regulates lignin biosynthesis upon the suberization process in apple.  
592 *Frontiers in plant science* **13**, 1039014 (2022).

593 52 Feng C, C. B., Hofer J, Shi Y, Jiang M, Song B, Cheng H, Lu L, Wang L, Howard A,  
594 Bendahmane A, Fouchal A, Moreau C, Sawada C, LeSignor C, Zhang C, Vikeli E, Tsanakas G,  
595 Zhao H, Cheema J, Barclay JE, Hou J, Sayers L, Wingen L, Vigouroux M, Vickers M, Ambrose  
596 M, Dalmais M, Higuera-Poveda P, Li P, Yuan Q, Spanner R, Horler R, Wouters R, Chundakkad  
597 S, Wu T, Zhao X, Li X, Sun Y, Huang Z, Wu Z, Deng XW, Steuernagel B, Domoney C, Ellis N,  
598 Chayut N, Cheng S. Genomic and genetic insights into Mendel's pea genes. *Nature* (2025).  
599 <https://doi.org/10.1038/s41586-025-08891-6>

600 53 Zhou, J., Lee, C., Zhong, R. & Ye, Z.-H. MYB58 and MYB63 are transcriptional activators of the  
601 lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. *The Plant Cell*  
602 **21**, 248-266 (2009).

603 54 Parker, T. A., de Sousa, L. L., de Oliveira Floriani, T., Palkovic, A. & Gepts, P. Toward the  
604 introgression of PvPdh1 for increased resistance to pod shattering in common bean. *Theor Appl*  
605 *Genet* **134**, 313-325 (2021). <https://doi.org/10.1007/s00122-020-03698-7>

606 55 Yang, C. *et al.* Arabidopsis MYB26/MALE STERILE35 regulates secondary thickening in the  
607 endothecium and is essential for anther dehiscence. *The Plant Cell* **19**, 534-548 (2007).

608 56 Takahashi, Y. *et al.* Same locus for non-shattering seed pod in two independently domesticated  
609 legumes, *Vigna angularis* and *Vigna unguiculata*. *Frontiers in Genetics*, 748 (2020).

610 57 Watcharatpong, P., Kaga, A., Chen, X. & Somta, P. Narrowing down a major QTL region  
611 conferring pod fiber contents in yardlong bean (*Vigna unguiculata*), a vegetable cowpea. *Genes*  
612 **11**, 363 (2020).

613 58 Murgia, M. L. *et al.* A comprehensive phenotypic investigation of the “pod-shattering syndrome”  
614 in common bean. *Frontiers in plant science* **8**, 251 (2017).

615 59 Gepts, P. in *Evolutionary biology* 51-94 (Springer, 1993).

616 60 Morrell, P. L. & Clegg, M. T. Genetic evidence for a second domestication of barley (*Hordeum*  
617 *vulgare*) east of the Fertile Crescent. *Proceedings of the national academy of sciences* **104**, 3289-  
618 3294 (2007).

619 61 Alo, F., Furman, B. J., Akhunov, E., Dvorak, J. & Gepts, P. Leveraging genomic resources of  
620 model species for the assessment of diversity and phylogeny in wild and domesticated lentil.  
621 *Journal of Heredity* **102**, 315-329 (2011).

622 62 Sanjur, O. I., Piperno, D. R., Andres, T. C. & Wessel-Beaver, L. Phylogenetic relationships  
623 among domesticated and wild species of Cucurbita (Cucurbitaceae) inferred from a mitochondrial  
624 gene: Implications for crop plant evolution and areas of origin. *Proceedings of the National*  
625 *Academy of Sciences* **99**, 535-540 (2002).

626 63 Matsuoka, Y. *et al.* A single domestication for maize shown by multilocus microsatellite  
627 genotyping. *Proceedings of the National Academy of Sciences* **99**, 6080-6084 (2002).

628 64 Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J. & Dickau, R. Starch grain and phytolith  
629 evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico.  
630 *Proceedings of the National Academy of Sciences* **106**, 5019-5024 (2009).

631 65 Ranere, A. J., Piperno, D. R., Holst, I., Dickau, R. & Iriarte, J. The cultural and chronological  
632 context of early Holocene maize and squash domestication in the Central Balsas River Valley,  
633 Mexico. *Proceedings of the National Academy of Sciences* **106**, 5014-5018 (2009).

634 66 Xu, G. *et al.* Population genomics of *Zea* species identifies selection signatures during maize  
635 domestication and adaptation. *BMC plant biology* **22**, 72 (2022).

636 67 Zizumbo-Villarreal, D. & Colunga-GarcíaMarín, P. Origin of agriculture and plant domestication  
637 in West Mesoamerica. *Genetic Resources and Crop Evolution* **57**, 813-825 (2010).

638 68 Zizumbo-Villarreal, D., Flores-Silva, A. & Colunga-GarcíaMarín, P. The food system during the  
639 formative period in West Mesoamerica. *Economic Botany* **68**, 67-84 (2014).



640 69 Zizumbo-Villarreal, D., Colunga-GarcíaMarín, P. & Flores-Silva, A. Pre-columbian food system  
641 in west Mesoamerica. *Ethnobotany of Mexico: Interactions of people and plants in Mesoamerica*,  
642 67-82 (2016).

643 70 Singh, S. P., Gepts, P. & Debouck, D. G. Races of common bean (*Phaseolus vulgaris*, Fabaceae).  
644 *Economic botany*, 379-396 (1991).

645 71 Hang, A., Silbernagel, M. & Miklas, P. Registration of Orca'black-and-white mottled Anasazi-  
646 type dry bean. *Crop science* **43**, 1882-1883 (2003).

647 72 Temple, S. UC Canario 707 Plant Variety Protection certificate. (2003).

648 73 Moghaddam, S. M. *et al.* Genome-wide association study identifies candidate loci underlying  
649 agronomic traits in a Middle American diversity panel of common bean. *The Plant Genome* **9**,  
650 plantgenome2016.2002.0012 (2016).

651 74 Cichy, K. A. *et al.* A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop*  
652 *Science* **55**, 2149-2160 (2015).

653 75 Korani, W. *et al.* De novo QTL-seq identifies loci linked to blanchability in peanut (*Arachis*  
654 *hypogaea*) and refines previously identified QTL with low coverage sequence. *Agronomy* **11**,  
655 2201 (2021).

656 76 Xie, W. *et al.* Parent-independent genotyping for constructing an ultrahigh-density linkage map  
657 based on population sequencing. *Proceedings of the National Academy of Sciences* **107**, 10578-  
658 10583 (2010).

659 77 R Core Team. (2025).

660 78 Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses.  
661 *bioinformatics* **19**, 889-890 (2003).

662 79 Taylor, J. & Butler, D. R package ASMap: efficient genetic linkage map construction and  
663 diagnosis. *arXiv preprint arXiv:1705.06916* (2017).

664 80 Haley, C. S., Knott, S. A. & Elsen, J. M. Mapping quantitative trait loci in crosses between  
665 outbred lines using least squares. *Genetics* **136**, 1195-1207 (1994).

666 81 Wickham, H. ggplot2. *Wiley interdisciplinary reviews: computational statistics* **3**, 180-185  
667 (2011).

668 82 Andrews, S. FastQC: a quality control tool for high throughput sequence data. (*No Title*) (2010).

669 83 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence  
670 data. *Bioinformatics* **30**, 2114-2120 (2014).

671 84 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform.  
672 *bioinformatics* **25**, 1754-1760 (2009).

673 85 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

674 86 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24-26 (2011).

675 87 Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic*  
676 *acids research* **40**, D1178-D1186 (2012).

677 88 Lo, S. *et al.* Genetic, anatomical, and environmental patterns related to pod shattering resistance  
678 in domesticated cowpea [*Vigna unguiculata* (L.) Walp]. *Journal of Experimental Botany* **72**,  
679 6219-6229 (2021).

680 89 Jackson, D., Veit, B. & Hake, S. Expression of maize KNOTTED1 related homeobox genes in  
681 the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot.  
682 *Development* **120**, 405-413 (1994).

683 90 Xu, X. *et al.* Single-cell RNA sequencing of developing maize ears facilitates functional analysis  
684 and trait candidate gene discovery. *Developmental Cell* **56**, 557-568. e556 (2021).

685 91 Fernandez, F., Gepts, P. L. & López, M. Vol. Serie 04EB-09.03 (Programa de las Naciones  
686 Unidas (PNUD), 1983).

687 92 Goering, H. K. *Forage fiber analyses (apparatus, reagents, procedures, and some applications)*.  
688 (US Agricultural Research Service, 1970).

689 93 Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment,  
690 interactive sequence choice and visualization. *Briefings in bioinformatics* **20**, 1160-1166 (2019).

691 94 Minh, B. Q. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in  
692 the genomic era. *Molecular biology and evolution* **37**, 1530-1534 (2020).  
693 95 Chernomor, O., Von Haeseler, A. & Minh, B. Q. Terrace aware data structure for phylogenomic  
694 inference from supermatrices. *Systematic biology* **65**, 997-1008 (2016).  
695 96 Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree  
696 display and annotation tool. *Nucleic acids research* **52**, W78-W82 (2024).  
697 97 Huson, D. H. & Bryant, D. The SplitsTree App: interactive analysis and visualization using  
698 phylogenetic trees and networks. *Nature methods* **21**, 1773-1774 (2024).  
699 98 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer  
700 datasets. *Gigascience* **4**, s13742-13015-10047-13748 (2015).  
701 99 Pfeifer, B., Wittelsb rger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient  
702 Swiss army knife for population genomic analyses in R. *Molecular biology and evolution* **31**,  
703 1929-1936 (2014).  
704 100 O'Rourke, J. A. *et al.* An RNA-Seq based gene expression atlas of the common bean. *BMC*  
705 *genomics* **15**, 1-17 (2014).  
706 101 Yu, D., Chen, C. & Chen, Z. Evidence for an important role of WRKY DNA binding proteins in  
707 the regulation of NPR1 gene expression. *The Plant Cell* **13**, 1527-1540 (2001).  
708 102 Abe, H. *et al.* Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional  
709 activators in abscisic acid signaling. *The Plant Cell* **15**, 63-78 (2003).  
710 103 Huang, T.-H., Hsu, W.-H., Mao, W.-T. & Yang, C.-H. The Oncidium Ethylene Synthesis Gene  
711 Oncidium 1-Aminocyclopropane-1 Carboxylic Acid Synthase 12 and Ethylene Receptor Gene  
712 Oncidium ETR1 Affect GA-DELLA and Jasmonic Acid Signaling in Regulating Flowering  
713 Time, Anther Dehiscence, and Flower Senescence in Arabidopsis. *Frontiers in plant science* **13**,  
714 785441 (2022).  
715 104 Wang, Y. *et al.* MYB transcription factors and their roles in the male reproductive development  
716 of flowering plants. *Plant Science* **335**, 111811 (2023).  
717

## 718 Acknowledgements

719 This project was made possible by germplasm from the United States Department of Agriculture  
720 National Plant Germplasm System (Pullman, Washington, U.S.A.) and the gene bank of the  
721 International Center for Tropical Agriculture (CIAT, Cali, Colombia). We thank Daniel Debouck  
722 for important germplasm advice. Judy Jernstedt provided microscopy equipment. Maizie Flores,  
723 Anham Rafique, and Troy Williams assisted in population development and phenotyping. This  
724 work was funded by Kirkhouse Trust A23-1518-001 to T.P. and USDA-AFRI award 2023-  
725 67013-40001 to T.P. X.X. acknowledges support from the University of California, Davis (UC  
726 Davis) new faculty start-up funds and California Agricultural Experiment Station/National  
727 Institute of Food and Agriculture (CA-D-PLB-2850-H).  
728

## 729 Author contributions

730 T.P., P.M., and P.G. conceived the research; T.P., B.C., and P.M. designed the  
731 experiments; T.P., P.M., and A.F. acquired funding; B.C., T.P., J.R., A.F., S.E., X.Y.,

732 X.X., and P.M. performed the analyses. B.C. and T.P. drafted the first version of the  
733 paper. All authors contributed to discussions and revisions of the paper.

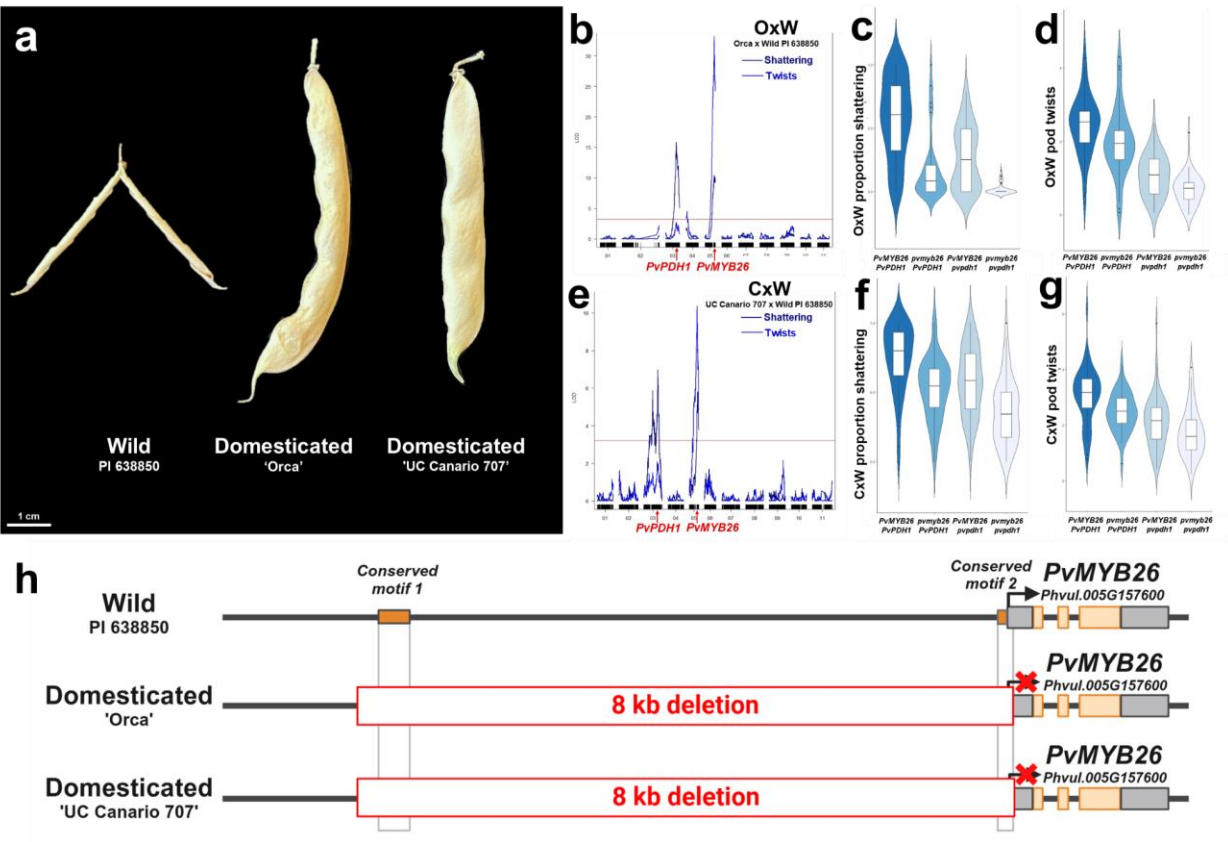
## 734 Data deposition

735 Sequences aligning within 300 kb of *PvMYB26* and used for population genetics  
736 screens will be submitted to the NCBI SRA. Mapping data will be submitted to Data  
737 Dryad. Other published datasets used in this study are listed in the related legends,  
738 Online Methods, and/or supplement.

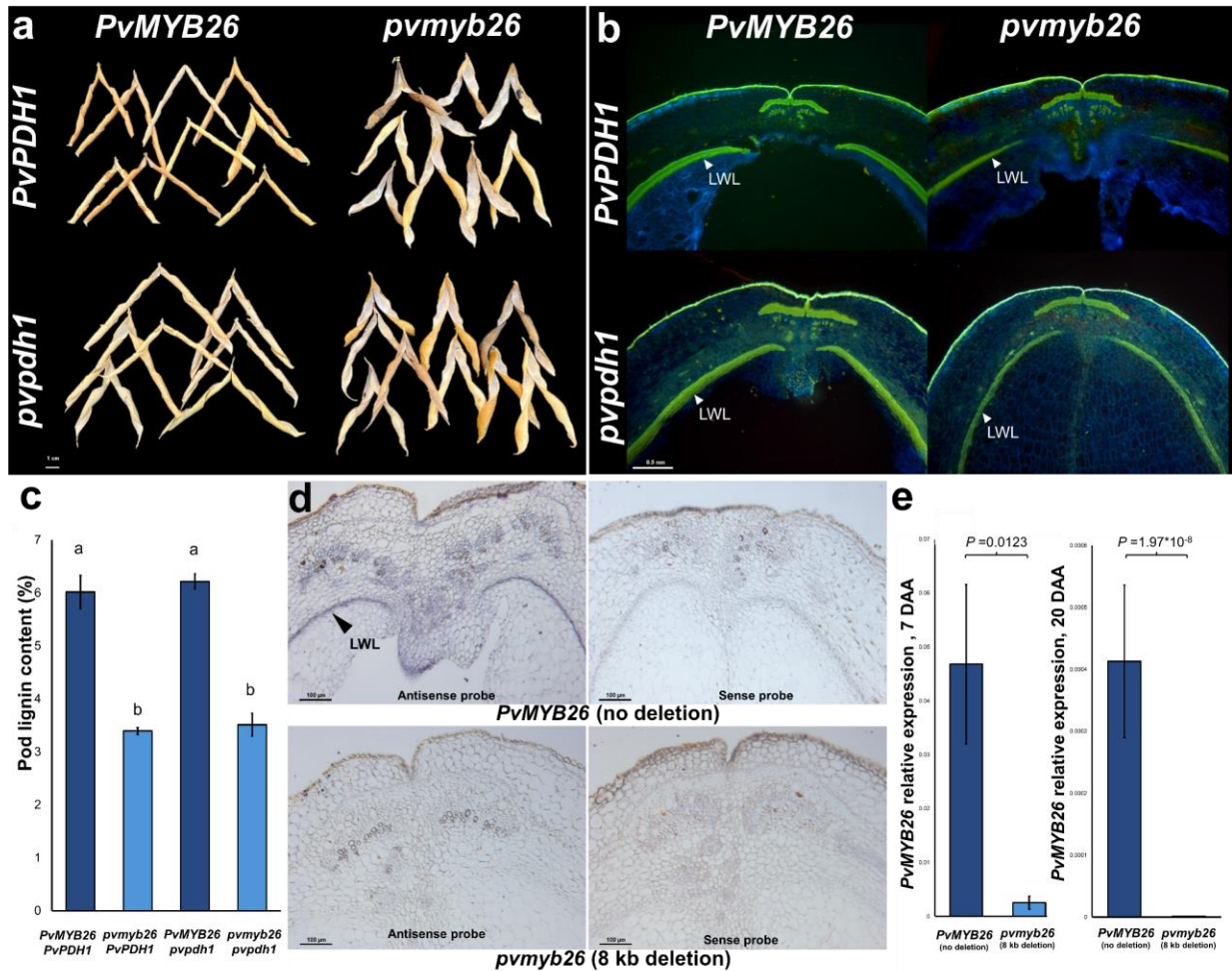
739

740

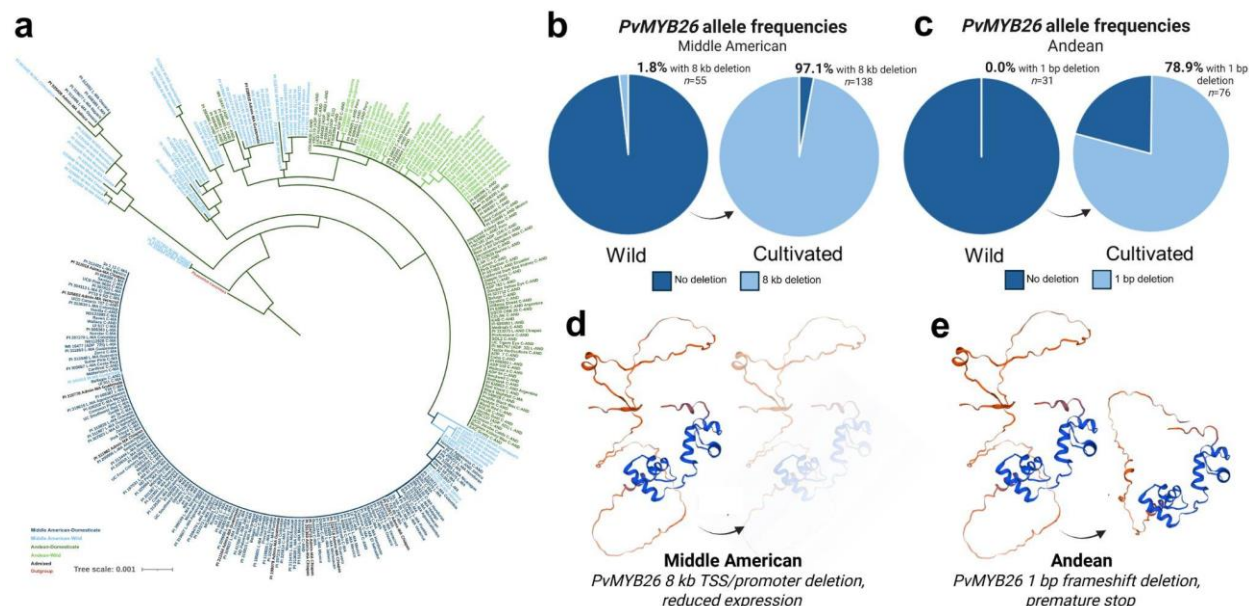
741 Figures



**Fig. 1** | A *PvMYB26* deletion is found at a locus for reduced pod shattering and pod twists in domesticated *P. vulgaris*—a, Pod phenotypes of parental lines. The wild PI 638850 has high shattering and pod twists, while the domesticates Orca and UC Canario 707 have reduced shattering and pod twists. b-d, QTL mapping and allele effect plots in the Orca x PI 638850 (OxW) population identified significant QTL for pod shattering and pod twists, mapping to the candidate genes *PvPDH1* and *PvMYB26*. e-g, The UC Canario 707 x PI 638850 (CxW) population also had two significant QTL for pod shattering and two for pod twists, mapping to the same two candidate genes. h, Whole genome sequencing of the parents identified a large deletion eliminating the transcription start site, 61 bp of 5'UTR, and over 8 kb of promoter, including two motifs conserved throughout the legume family.

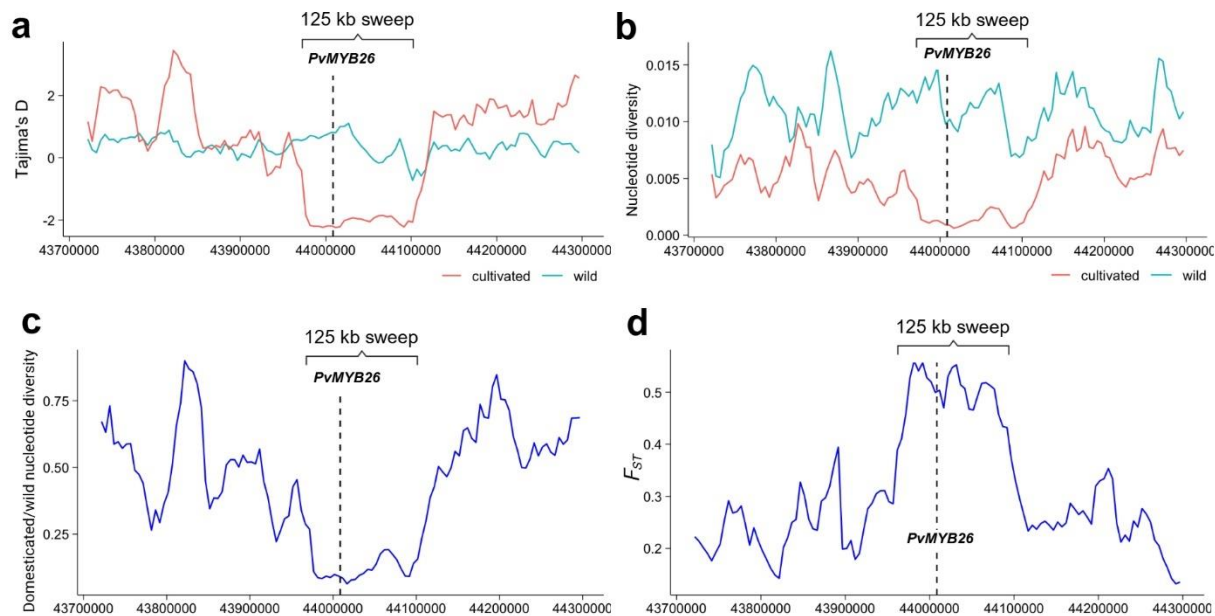


**Fig. 2 | *PvMYB26* mutation effects.** **a**, Visually distinct pod coiling habits in among randomly sampled pods of each allelic category, **b**, Lignified wall layer (“LWL”, hydrophobic lignin stained green) fiber deposition is visibly reduced among *pvmyb26* mutants compared to mutants, whereas *PvPDH1* has no anatomical effect, and **c**, *pvmyb26* mutants produce 44% less lignin than wild types. **d**, RNA in situ hybridization shows that *PvMYB26* expression in the pod walls is confined to the LWL, with weak additional expression in the vascular bundles. No discernible in situ hybridization signal is visible in any tissue when the 8 kb promoter deletion is present. **e**, *pvmyb26* mutants have 18.3-fold reduced *PvMYB26* expression at 7 days after anthesis (DAA) (heteroscedastic two-tailed *t*-test;  $P=0.012$ ) and 306-fold reduced expression by 20 DAA (heteroscedastic two-tailed *t*-test;  $P=1.9 \times 10^{-8}$ ). Data and images from OxW RILs.



**Fig. 3 |** Sequence diversity at *PvMYB26* among domesticated and wild populations. **a**, A dendrogram of sequence diversity based on 4 kb of sequence at *PvMYB26*, colored by accession gene pool and domestication status. **b**, The 8 kb deletion is found in only 1.8% of sampled Middle American wild types (PI 201016, with admixture and large seeds). In contrast, the mutation is found in 97.1% of surveyed cultivated lines with a Middle American haplotype. **c**, A distinct 1 bp mutation arose among Andean domesticates, which was not found in any wild type but was present in 78.9% of domesticated lines with an Andean haplotype. **d**, Whereas the Middle American 8 kb TSS/promoter deletion drastically reduces gene expression, **e**, the Andean one-bp mutation disrupts the final 79 amino acid chain and truncates its length by 45 residues. The extreme delineation between wild and cultivated types at *PvMYB26* indicates that mutations at the gene likely have differential fitness for plants in wild vs. cultivated ecosystems, and are not readily exchanged between wild and domesticated populations.

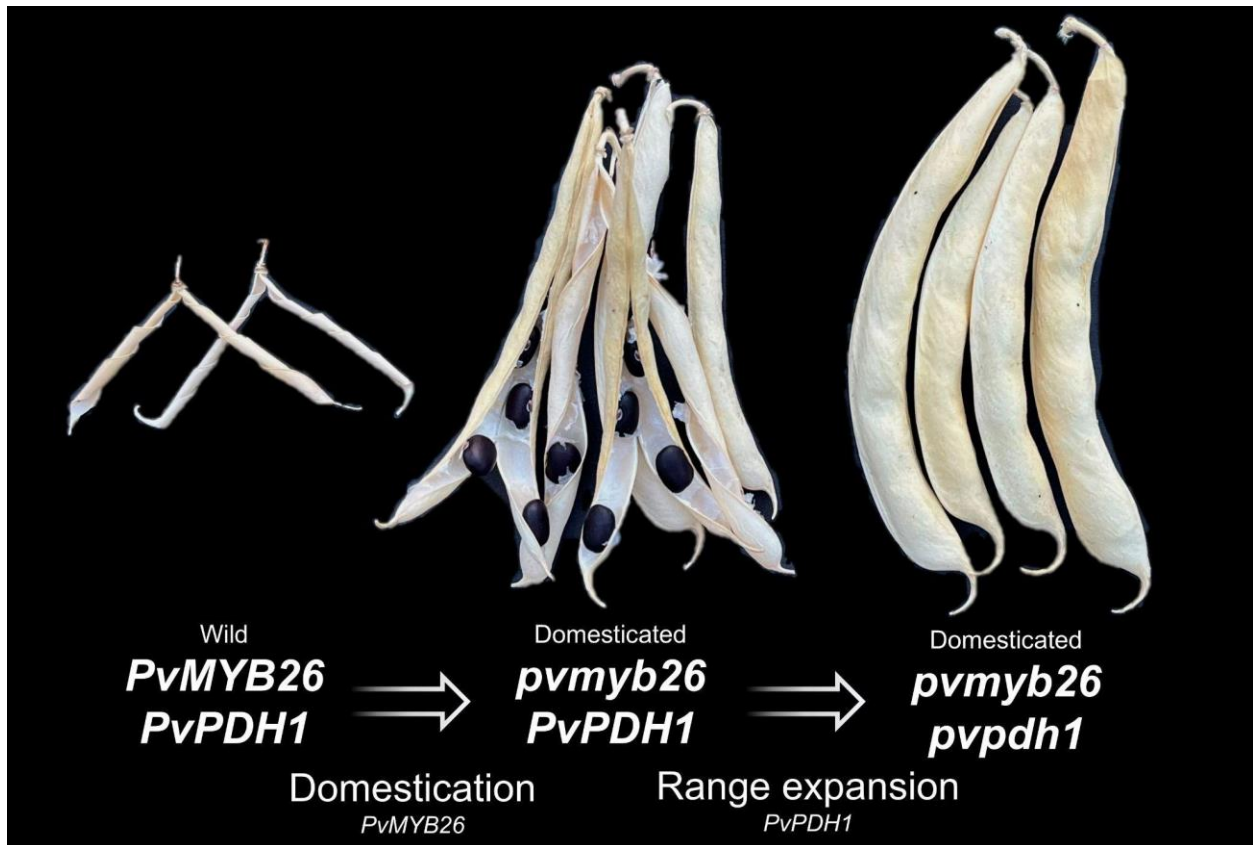




**Fig. 4** | A 125 kb selection sweep occurred at *PvMYB26* during the domestication of Middle American common beans. **a**, Tajima's D in wild and domesticated populations shows that values drop around *PvMYB26* among domesticates, while wild type values are stable, **b**, and **c**, Wild lines have stable nucleotide diversity in the region, while nucleotide diversity drops around *PvMYB26* among domesticates, **d**,  $F_{ST}$  between wild and domesticated populations is highest in the selective sweep region, showing high distinctiveness between populations. Coordinates relative to genome 5-593 v1.1.







**Fig. 6** | Overview of pod changes during the domestication and range expansion of the Middle American common bean. During domestication, a mutation in *PvMYB26* on Pv05 led to a reduction in the quantity of wall fiber between wild and domesticated Middle American beans with intermediate shattering. Varieties of the ecogeographic race Durango are adapted to arid conditions where pod shattering is most strongly expressed and have accumulated a mutation in *PvPDH1* on chromosome Pv03<sup>33</sup>.

808    **Supplementary Tables**

**Supplementary Table 1.** Significant QTL for pod shattering and pod twists in two wild x domesticated recombinant inbred populations.

Popu- lation	Trait (two-year average)	Chr	Most significant SNP	LOD	Variatio n Explaine d	Upstream flanking SNP	Downstrea m flanking SNP	Candidate gene
CxW	Pod twists	5	38,800,513	10.57	17.2%	38,579,173	38,825,585	<i>PvMYB26</i>
CxW	Pod shattering	3	38,463,766	5.88	3.5%	38,110,482	38,638,664	Unknown
CxW	Pod shattering	3	48,566,812	6.98	5.6%	48,514,102	49,117,061	<i>PvPDH1</i>
CxW	Pod shattering	5	39,470,090	5.81	10.1%	39,424,572	39,529,682	<i>PvMYB26</i>
OxW	Pod twists	4	1,640,562	4.52	4.6%	1,626,404	1,865,823	Unknown
OxW	Pod twists	5	38,633,111	33.18	39.8%	38,567,439	38,781,607	<i>PvMYB26</i>
OxW	Pod shattering	3	49,199,252	15.85	20.9%	48,927,734	49,301,961	<i>PvPDH1</i>
OxW	Pod shattering	5	38,567,439	10.43	13.6%	38,398,277	40,141,908	<i>PvMYB26</i>

810

**Supplementary Table 2.** *PvMYB26* allele frequencies in Wild, Middle American domesticated, and Andean domesticated common beans.

	8 kb deletion present	8 kb deletion absent	Total genotyped	Percent with deletion
Wild	2	88	90	2.2%
Middle American Diversity Panel	109	1	110	99.1%
Andean Diversity Panel	14	86	100	14%

811

812

813

**Supplementary Table 3.** Middle American selective sweep statistics in a 600 kb region centered on *PvMYB26*. A 125 kb hard sweep was identified from 43,976,690 to 44,101,690 on Pv05 (5\_593 genome v1.1). Data reflects a 600 kb region centered on the *PvMYB26*.

Test statistic	43,976,690- 44,101,690 mean $\pm$ SEM	43,709,324-43,976,690; 44,101,690-44,309,325 mean $\pm$ SEM
Nucleotide diversity (wild)	0.0110 $\pm$ 0.000441	0.0109 $\pm$ 0.000245
Nucleotide diversity (domesticated)	0.00129 $\pm$ 0.000106	0.00570 $\pm$ 0.000194
Nucleotide diversity– (domesticated/wild)	0.116 $\pm$ 0.00730	0.534 $\pm$ 0.0166
Tajima’s D (wild)	0.405 $\pm$ 0.0906	0.358 $\pm$ 0.0285
Tajima’s D (domesticated)	-2.05 $\pm$ 0.0267	1.16 $\pm$ 0.0996
<i>F<sub>ST</sub></i> (wild vs domesticated)	0.495 $\pm$ 0.0108	0.251 $\pm$ 0.00648

814

**Supplementary Table 4.** Andean selective sweep statistics selective sweep statistics in a 600 kb region centered on *PvMYB26*. No hard sweep was identifiable throughout the region. The coordinates 43,976,690 to 44,101,690 on Pv05 (5\_593 genome v1.1) are the region of the hard sweep in Middle American lines for comparison.

Test statistic	43,976,690-44,101,690 mean $\pm$ SEM	43,709,324-43,976,690; 44,101,690-44,309,325 mean $\pm$ SEM
Nucleotide diversity (wild)	0.00302 $\pm$ 0.000268	0.00260 $\pm$ 0.000142
Nucleotide diversity (domesticated)	0.00240 $\pm$ 0.000111	0.00266 $\pm$ 0.000120
Nucleotide diversity (domesticated/wild)	0.942 $\pm$ 0.0823	1.10 $\pm$ 0.0401
Tajima's D (wild)	-0.819 $\pm$ 0.0898	-0.859 $\pm$ 0.0596
Tajima's D (domesticated)	-1.14 $\pm$ 0.0542	-0.690 $\pm$ 0.0785
$F_{ST}$ (wild vs domesticated)	0.0900 $\pm$ 0.0177	0.118 $\pm$ 0.0125

815

**Supplementary Table 5.** Co-dominant primers showing 8 kb deletion status in *PvMYB26*.

Name	Sequence	Primer length	Primer $T_m$
MYB26-F1	TCCCAGCCCTTAACTCTACC	20 bp	66
MYB26-F2	AAGCTTCCTTTCATGGGGGA	20 bp	66
MYB26-R	AGAGTGATGCCCCATATGTTCA	22 bp	66

816

817

818

**Supplementary Table 6.** Primers for candidate gene (*PvMYB26*) and control gene (*Act11*)<sup>100</sup> used for RT-qPCR.

Target	Gene	Forward primer	Reverse primer
<i>Phvul.002G269900</i>	<i>ACT11</i>	TTTGCCTTCCATCACCTCG	CTCCAGCAAATCCAGCCTTA
<i>Phvul.005G157600</i>	<i>PvMYB26</i>	GAGTTCAGTCCCAAAGCTTGC	GAGCCCACCTGTTGCC

819

820

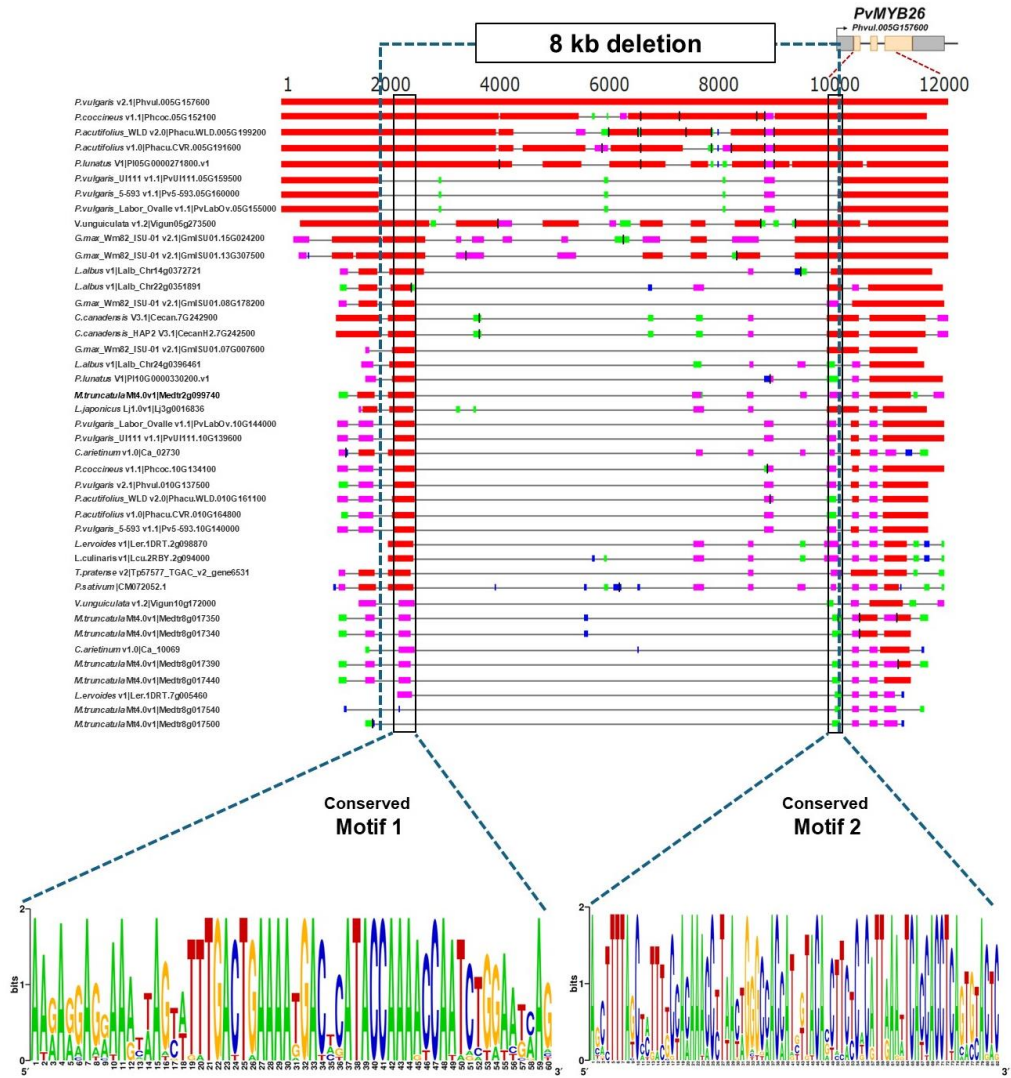
821

822

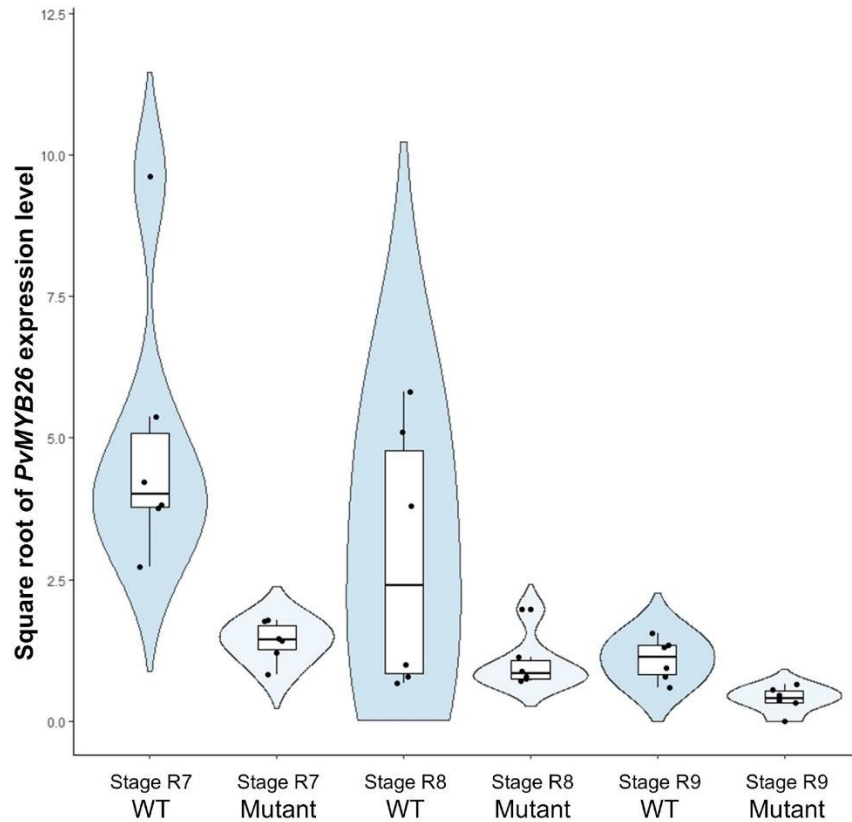
# Supplementary Figures



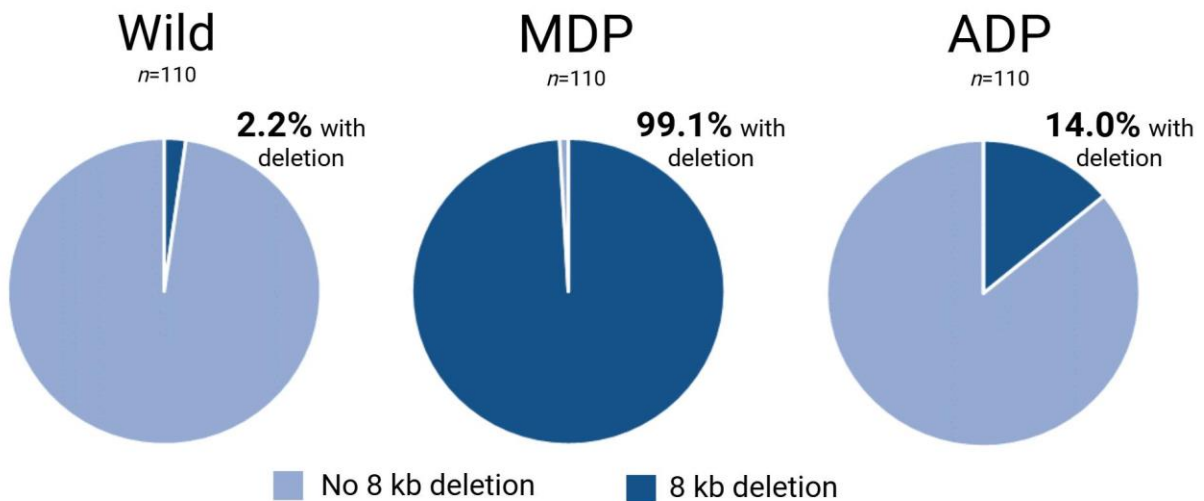
**Supplementary Fig. 1** | An IGV screenshot shows *PvMYB26* (*Phvul.005G157600*, transcribed region in blue at top left, exons thickened) and whole-genome-sequencing alignments with selected lines. The top track (PI 638850) is the wild parent of RIL populations and lacks the 8 kb deletion. The following six tracks show types with Middle American domesticated haplotypes, including the 8 kb deletion that eliminates the *PvMYB26* transcription start site and promoter. This includes the UC Canario 707 and Orca domesticated population parents. The last six tracks are types from the independently domesticated Andean gene pool, which lack the 8 kb deletion. Gray indicates the read depth in each accession at each base pair. A total absence of reads aligning to the reference indicates a deletion of the region in the sequenced experimental line. All sequences were aligned to the G19833 v2.1 genome<sup>19</sup> ([https://phytozome-next.jgi.doe.gov/info/Pvulgaris\\_v2\\_1](https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1)).



**Supplementary Fig. 2** | Two sequence motifs upstream of *PvMYB26* are highly conserved across the legume family, but are deleted in the domesticated RIL population parents and other domesticates with a Middle American haplotype at the gene. Motif 1 includes recognition sites of WRKY binding proteins (TTGAC)<sup>101</sup> and MYC (bHLH) and MYB transcription factors (CANNTG and WAACCA)<sup>102</sup>. In Arabidopsis, the WRKY/MYC genes and auxin response factors (ARFs; ARF6/8) act upstream of the jasmonic acid (JA) biosynthesis and *MYB26* gene<sup>103,104</sup>. Disruption of this pathway results in suppression of *MYB26* expression and absence of secondary thickening, resulting in the production of non-opening anthers<sup>103,104</sup>. Motif 2 includes MYB recognition sites (TAAC TG and YAACKG)<sup>102</sup>. Motif 2 includes an 82 bp sequence conserved in identical form across all *Phaseolus* species with available data (*P. acutifolius*, *P. coccineus*, *P. lunatus*, and wild-type *P. vulgaris*), as well as *Vigna angularis* and *Vigna unguiculata*, but 51 bp of which is deleted in Middle American domesticates.

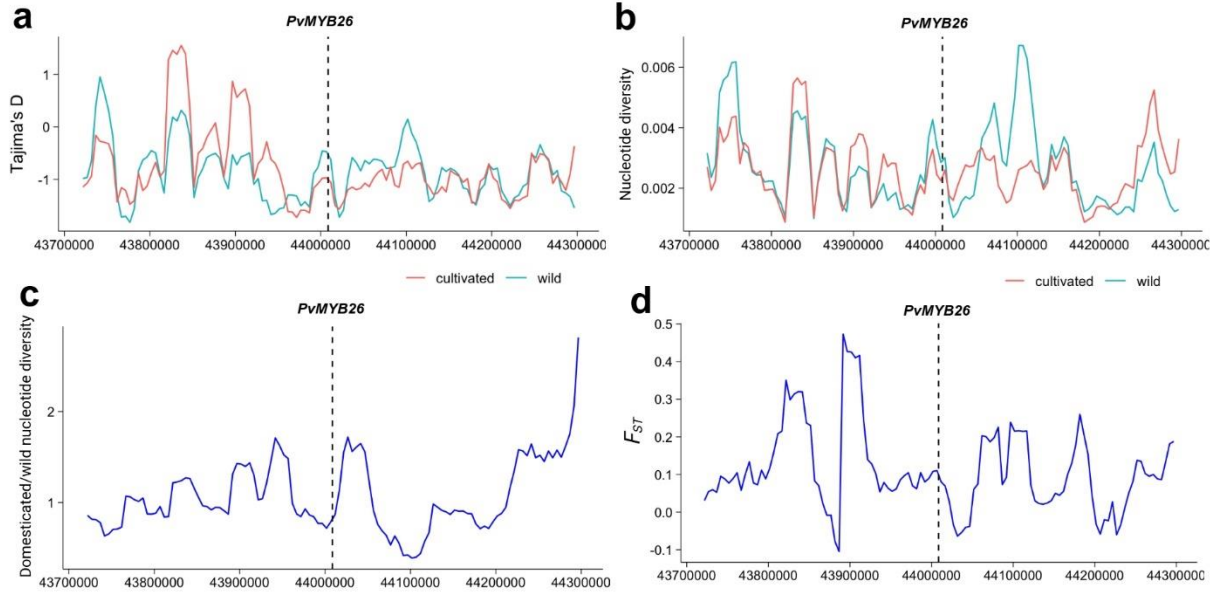


**Supplementary Fig. 3** | RNA-seq of wild and mutant genotypes<sup>32</sup> across three pod maturity stages (Stage R7-9) showed significant differences in expression between *PvMYB26* mutants (accessions ICA Bunsu and SXB405) and non-mutants (accessions Midas and G12873) ( $P = 5.90 \times 10^{-5}$ ). Expression was lower in mutants at every sampling time point.





**Supplementary Fig. 4** | *PvMYB26* allele frequencies among wild common beans, members of the Middle American Diversity Panel (MDP), and the independently domesticated Andean Diversity Panel (ADP). The 8 kb *pvmvb26* deletion is nearly diagnostic for domestication status between wild and domesticated Middle American accessions of the species. While it originated in the Middle American gene pool, it has also been transferred to 14% of Andean domesticates.



**Supplementary Fig. 5** | Population genetics statistics around *PvMYB26* show no selective sweep in Andean materials, in contrast to the hard sweep identified in Middle American materials. Calculated statistics include. **a**, Tajima's D, **b-c**, nucleotide diversity, and **d**,  $F_{ST}$ . Coordinates based on genome 5-593 v1.1.