Title

Pod indehiscence is a domestication and aridity resilience trait in common bean

Journal

New Phytologist, 225(1)

ISSN

0028-646X

Authors

Parker, Travis A
Mier y Teran, Jorge C Berny
Palkovic, Antonia
et al.

Publication Date

2020

DOI

10.1111/nph.16164

Peer reviewed

Pod indehiscence is a domestication and aridity resilience trait in common bean

Travis A. Parker, Jorge C. Berny Mier y Teran, Antonia Palkovic, Judy Jernstedt, and Paul Gepts

University of California, Department of Plant Sciences, 1 Shields Avenue, Davis, CA 95616-8780, USA.

Author for correspondence:
Paul Gepts
Tel: +1-530-752-7743
Email: plgepts@ucdavis.edu

ORCID
Travis Parker https://orcid.org/0000-0002-1233-7829
Jorge Berny Mier y Teran https://orcid.org/0000-0003-3709-9131
Judy Jernstedt https://orcid.org/0000-0003-0575-0938
Paul Gepts https://orcid.org/0000-0002-1056-4665

SOCIAL MEDIA

Name	Facebook	Twitter	Instagram
T. Parker	www.facebook.com/travis.parker.9803		
J. Berny	www.facebook.com/jcberny	@JorgeCBerny	
P. Gepts	www.facebook.com/paul.gepts	@plgepts	plgepts

TP:
Total word count: 6,162
Summary: 182
Word count:
- Introduction: 766
- Materials and Methods: 1,730
- Results: 2,129
- Discussion: 1203
- Acknowledgments: 113

Number of figures: 5 (all in color)
Number of tables: 1
Supporting information: 10 figures, 4 tables

Summary

- Plant domestication has strongly modified crop morphology and development. Nevertheless, many crops continue to display atavistic characteristics that were advantageous to their wild ancestors but are deleterious under cultivation, such as pod dehiscence (PD). Here, we provide the first comprehensive assessment of the inheritance of PD in common bean (*Phaseolus vulgaris*), a major domesticated grain legume.
- Using three methods to evaluate the PD phenotype, we identified multiple, unlinked genetic regions controlling PD in a biparental population and two diversity panels. Subsequently, we assessed patterns of orthology among these loci and those controlling the trait in other species.
- Our results show that different genes were selected in each domestication and ecogeographic race. A chromosome Pv03 dirigent-like gene, involved in lignin biosynthesis, showed a base-pair substitution that is associated with decreased PD. This haplotype may underlie the expansion of Mesoamerican domesticates into northern Mexico, where arid conditions promote PD.
- The rise in frequency of the decreased-PD haplotype may be a consequence of the markedly different fitness landscape imposed by domestication. Environmental dependency and genetic redundancy can explain the maintenance of atavistic traits under domestication.

Key words: adaptive domestication, aridity tolerance, dirigent, genome-wide association study, local adaptation, pod shattering, seed dissemination.

Introduction

Plant domestication was a transformative evolutionary process, which turned wild plants into crops adapted to the human-mediated environment starting some 10,000 years ago (Gepts 2004, 2014; Myer *et al.* 2012, Meyer and Purugganan 2013, Larson *et al.* 2014; Martínez-Ainsworth and Tenaillon 2016). Core domestication traits across a range of seed-propagated taxa include a) a reduction in seed dispersal, b) reduced seed dormancy, c) increased phenotypic diversity of harvested structures, including gigantism, d) changes in growth habit, and e) modified phenology, collectively called the domestication syndrome (Hammer 1984; Lenser & Theißen, 2013). Global food security is entirely dependent on crops that have undergone these changes. The domestication process has also served as a series of natural experiments in evolutionary biology and genetics, a role that has been recognized since the inception of these fields (Darwin 1859, Mendel 1866).

Effective seed dispersal is vital for spermatophytes. In the Fabaceae, the third largest family of flowering plants (Azani *et al.* 2017), seed dispersal is typically mediated by the explosive dehiscence ("shattering") of pods at maturity. While this mechanism is effective for the propagation of plants in the wild, it results in yield reduction and constrains the temporal window for harvest in the cultivated environment. This has led to selection for pod indehiscence during and after domestication across a range of legume taxa (Ogutcen *et al*. 2018, Rau *et al*. 2019). These cultivated forms generally display pod indehiscence, also known as PD-resistance.

Phaseolus beans are an exceptional experimental system to study domestication and the molecular evolution associated with this process. Humans domesticated members of this genus seven times (Gepts *et al.* 2008; Bitocchi *et al.* 2017), which are part of the 41 domestications in the Fabaceae (Harlan 1992). Common bean (*Phaseolus vulgaris* L.), a dietary staple for hundreds of millions of people worldwide (Singh 1999, Gepts *et al.* 2008), diverged into distinct Middle American and Andean gene pools approximately 87,000 years before present (Ariani *et al.* 2018), well before the first human migrations into the Americas some 16,000-23,000 years ago (Moreno-Mayar *et al.* 2018, Potter *et al.* 2018). It was domesticated independently in Middle America and the Andes, resulting in a replicated experiment in evolution. Each of the two domesticated gene pools of common bean is subdivided into several ecogeographic races. For example, the Middle American domesticated gene pool is comprised in part by race Durango (sometimes clustered with the genetically indistinguishable race Jalisco to form race

Durango/Jalisco), which is adapted to the arid, higher altitude regions of northern Mexico, and race Mesoamerica, adapted to the warmer, humid lowlands of southern Mexico and Central America (Singh *et al.* 1991). Atmospheric dryness has a strong PD-promoting effect in legumes, and mean annual precipitation is related to signatures of selection on PD-related candidate genes (Bandillo *et al*. 2017). Desiccation is also often used to induce pod fracture experimentally (Dong *et al.* 2014, Funatsuki *et al.* 2014).

Koinange *et al*. (1996) were the first to identify a pod fiber factor, namely a major gene on linkage group Pv02 (Freyre *et al.* 1998) in the recombinant inbred (RI) population derived from stringless cv. 'Midas' and wild accession G12873. This gene, called *Stringless (St),* maps near the common bean ortholog of *INDEHISCENT* (*PvIND*), but a low frequency of recombination is known to exist between the *PvIND* and the stringless trait, and no causal polymorphism is known to exist in the *PvIND* sequence (Gioia *et al.* 2013). *St* epistatically masks the effect of all other PD QTLs by dramatically decreasing fiber content but is only relevant in snap beans grown for pods as a vegetable. This locus does not explain any PD variation in the nutritionally important classes grown for grain. Recently, Rau *et al*. (2018) used QTL mapping to identify a single segregating locus on Pv05 in the same Midas x G12873 genetic background (Table 1). To date, a comprehensive evaluation of the genetic basis of PD in diverse germplasm has not yet been conducted and no molecular polymorphisms with a potential causal effect on PD have been described.

In the research reported here, we used high-precision phenotyping techniques, both in an RI population and diversity panels, to identify PD QTLs in common bean grown for nutritionally important dry seeds. We sequenced a locus underlying a major QTL to identify a possible causal polymorphism. We found that orthologous genes regulate PD among certain domesticated legumes. We were further able to identify associations between PD and the environmental backgrounds of common bean races. Alleles identified in this study will be valuable for developing common bean varieties suited to the increasingly arid climatic conditions of coming decades.

Materials and Methods

Germplasm

A recombinant inbred (RI) population (n = 238), developed from a cross between ICA Bunsi (domesticated, PD-susceptible, Middle American) and SXB 405 (domesticated, PD-resistant, Middle American), was used for QTL mapping (Assefa *et al.* 2013; Berny Mier y Teran *et al.* 2019). For association mapping, different panels were used. Two-hundred eight members of the Andean Diversity Panel (ADP, Cichy *et al.* 2015) and 278 members of the Middle American Diversity (MDP, Moghaddam *et al.* 2016) were grown and phenotyped. Sequencing was performed in a diverse panel of 90 varieties representing six species were acquired from the National Plant Germplasm System. Eighteen varieties commonly grown at UC Davis with known PD phenotypes were also genotyped. Stringless snap bean varieties were specifically excluded from the analysis to avoid the epistatic effect of the *Stringless (St*) locus on PD.

Microscopy

Pods of G12873 (wild, high dehiscence), ICA Bunsi (domesticated dry bean, dehiscence-susceptible) SXB 405 (domesticated dry bean, dehiscence-resistant), and Midas (domesticated snap bean, dehiscence-susceptible) were Vibratome-sectioned to identify anatomical differences that might be associated with PD. All sectioned pods were greenhouse-grown and harvested when pods were at full size with seeds filled, at the onset of pod color change. All sections were 100 micrometers thick and made in a transverse plane perpendicular to the fibers of interest. All sections were treated with Auramine O (aqueous, 0.01%) for at least 20 minutes to stain lignified tissue (Ursache *et al.* 2018). Fluorescence was visualized using an Olympus microscope.

RI population cultivation and PD phenotyping

The ICA Bunsi/SXB 405 (IxS) RI population of 238 RILs was field-grown during the spring and summer of 2014. The spring planting was an un-replicated trial conducted at Coachella, California. At maturity, plots were visually evaluated for the presence or absence of PD, and the data were used as a phenotype for QTL mapping. During the summer of 2014, the RI population was grown in a replicated field trial in Davis, California. At maturity, dried non-dehiscing pods from 191 RILs were harvested from each plot; these were evaluated for susceptibility to PD by two methods. First, all pods were desiccated at 65°C for seven days, and then returned to room temperature for a minimum of seven additional days. The

proportion of dehiscing pods after this process was recorded for each plot. Second, the amount of force required to induce pod fracture was measured using an Imada force measurement gauge (method modified from Dong *et al.,* 2014). Force measurements were taken on pods that had not dehisced during the desiccation treatment. A bit mounted to the gauge was used to press the ventral side of each pod at the most apical seed, and the peak force required to cause fracture at the apical end of the pod beak was recorded. Force required for PD was normalized to account for small but significant differences between note-takers, and the standardized score was used for QTL mapping. Pods that failed to produce seeds were excluded from all phenotyping analyses.

Genotyping

Genomic DNA was extracted from parents and RILs of the IxS population using a modified CTAB protocol. DNA quality was confirmed using a NanoDrop spectrophotometer. The IxS population was genotyped using the Illumina Infinium II BARCBean6K_3 BeadChip (Song *et al.* 2015); 382 segregating SNPs were identified in the population. Primers spanning the transcribed sequence of Phvul.003G252100, also known as *Phaseolus vulgaris Pod Dehiscence 1* (*PvPdh1*), a candidate gene underlying a major QTL identified in this study, were developed using the NCBI Primer-BLAST tool. Several differences in the genomic sequence exist between the Middle American and Andean gene pools, so a mixture of two forward primers was introduced into each PCR with a common reverse: PvPDH1 ALL Middle American Forward: CATCTCCCCCATTTTCCCCC; PvPDH1 ALL Andean Forward: CATCTCTCCCATTTTCTCCT; PvPDH1 ALL common Reverse: AACACGTGGAAGAGGAGGATT. PCR conditions for this amplification included an initial denaturation at 95°C for 180s, 38 cycles of 95°C for 30s, 51°C for 30s, and 68°C for 60s, and a final elongation step of 68°C for 300s. Another set of primers was developed to specifically improve the amplification and sequencing of Andean common beans, with the sequences: PvPDH1 Andes Forward: TTTTTCTTGTGAGCAAAATTGAGTT; PvPDH1 Andes Reverse: GCAGAGGAAAAACACGTGGA. This primer set was amplified with an initial denaturation at 95°C for 300s, 34 cycles of 95°C for 30s, 46°C for 30s, and 72°C for 70s, and a final elongation step of 72°C for 300s. PCR products were cleaned using a GeneJET PCR Purification Kit and sequenced at the UC DNA Sequencing Facility by Sanger sequencing.

QTL mapping

Composite interval mapping was conducted using the R package R/qtl (Broman *et al.* 2003). Field dehiscence score, proportion dehiscing in a desiccator, and force measurements were separately used to identify PD QTLs marked by SNPs. The maximum LOD score of 1000 randomized permutations of the

data was used as a significance threshold. Single QTL scans were performed using the scanone function. Multiple QTL mapping was conducted using the scantwo function in R/qtl and by running the analysis with RILs subset by genotype at the most significant marker near *PvPdh1* on Pv03. QTL mapping results were based on maximum likelihood via the EM algorithm (Lander and Botstein 1989).

Validation of QTL mapping results using association mapping

Two hundred and eight accessions of the ADP were grown in Davis, CA during summer 2016. PD in the field, proportion dehiscing in a desiccator, and force required for fracture were recorded. Principal component analysis was conducted on SNP data for the population, and the results were used as covariates to account for population structure. Two hundred seventy-eight members of the MDP were phenotyped for PD by desiccation in 2017. Association mapping was conducted using GLMs in TASSEL via SNiPlay (Bradbury *et al.*, 2007; Dereeper *et al.*, 2011). A minor allele frequency of 0.1 was used as a threshold for SNPs, and these SNPs were evaluated for significance based on a Bonferroni-corrected alpha of 0.05. QTL regions of significance were determined as the area between the first and last significant SNP on a chromosome arm. Individual significant SNPs without significant neighbors in the same population or others were not given further consideration, as these are likely All results were visualized using the qqman R package (Turner, 2018), including the Bonferroni-corrected significance thresholds at alpha=0.05 and 0.01 were shown, along with the positions of major candidate genes.

Expression and synteny mapping

Gene expression information from a variety of tissues and developmental stages was extracted from published data (O'Rourke *et al.* 2014) and visualized independently using R base graphics (R Core Team, 2013). Candidate genes related to PD were identified in significant QTL intervals based on definition line terms for gene families related to PD, which were downloaded with the PhytoMine interface of Phytozome 12 (Goodstein *et al.* 2012). Subsequent comparisons were made using the Basic Local Alignment Search Tool (BLAST) function with known amino acid sequences from related species. Synteny comparisons between common bean and soybean (*Glycine* max) were made using the Legume Information System 2.0 (Rice *et al.* 2015); these were verified using available literature (McClean *et al.* 2010, Schmutz *et al.* 2014). The CoGe SynMap (Lyons *et al.* 2008) and LegumeIP 2.0 (Li *et al.*, 2016) synteny tools were used to compare syntenic regions between *Arabidopsis* (Col-0, TAIR10), common bean (G19833, Pvulgaris_V1.0_218; Schmutz *et al.* 2014), and soybean (Williams 82, Release 1.1;

Schmutz *et al.*, 2010). A neighbor-joining tree was produced to determine the pattern of homology between a common bean candidate gene (*PvPdh1*), a related soybean gene (*GmPDH1*), and other members of the dirigent gene family in these two species. The amino acid sequence of these proteins was BLASTed against the *G. max* and *P. vulgaris* proteomes identify closely related genes. These were then compared using a multiple BLASTP to develop a distance tree based on a Grishin protein distance matrix (Grishin 1995). A fast-minimum evolution tree (Desper & Gascuel 2004) was generated based on a maximum sequence difference of 0.85.

Amino acid conservation analyses

The complete amino acid sequence of *PvPdh1* from accession G19833 was compared via BLASTP against the NCBI proteome database, using a BLOSUM62 matrix for comparison and existence and extension costs of 11 and 1, respectively (Altschul *et al.* 2005). The COnstraint-Based multiple ALignment Tool (COBALT; Papadopoulos & Agarwala, 2007, https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi) was used to align the most similar proteins known among several plant taxa and identify conserved residues based on the BLASTP results. The Protein Variation Effect Analyzer (PROVEAN; Choi & Chan, 2015) v1.1.3 software tool was used to estimate the effect of mutations of interest using default settings, including a cutoff threshold of -2.5 for identifying deleterious alleles.

Validation of the role of *PvPdh1* in a wider population

Genomic DNA was extracted using a modified CTAB method; amplification and Sanger sequencing of PvPdh1 were conducted as described previously. An indel was identified between positions 646 and 647 of the PvPdh1 transcript reference sequence. Varieties of known Andean ancestry, including the reference accession G19833, lack two base pairs found in varieties of Middle American ancestry. This indel occurs in the gene's 3' UTR and therefore does not affect the protein product's reading frame. The indel was used to distinguish Andean from Middle American varieties; only Middle American varieties included the mutant PvPdh1 allele. After sequencing, Middle American varieties were separated based on amino acid at position 162 of PVPDH1. The degree of dehiscence between these groups was evaluated by Student's t-test. Pod shatter phenotype data from the Germplasm Resource Information Network (GRIN: https://npgsweb.ars-grin.gov/gringlobal/descriptordetail.aspx?id=83053) was compared with our sequencing data for varieties acquired from NPGS.

245

246 Landrace ecogeography

247 Precipitation across the native range of Middle American beans was mapped in QGIS 2.18.19
248 using data from worldclim2 (Fick & Hijmans, 2017). National boundaries and coastlines were
249 added using shapefiles available through Natural Earth (Kelso & Patterson, 2010). USGS
250 topographical global raster data grids were also used to improve the visualization of coastlines
251 (https://topotools.cr.usgs.gov/gmted_viewer/gmted2010_global_grids.php). Landraces genotyped
252 by Kwak and Gepts (2009) were filtered by their ecogeographic race, those with values of 0.5 in
253 STRUCTURE groups K6 (race Mesoamerica) and K9 (race Durango/Jalisco) were used for
254 subsequent analysis. Delimited text layers were added in QGIS for varieties with latitude and
255 longitude data that belonged to one of the ecogeographic races of interest. The average annual
256 precipitation and elevation of the region where each landrace was collected using the "add raster
257 values to points" function in QGIS, and the values between ecogeographic races were compared
258 by student's t-test.

Results

Anatomical analysis of developing pods

Clear differences in pod anatomy were found between domesticated snap bean, domesticated dry bean, and wild common bean (Fig. 1). Wild beans produce a lignified wall fiber layer (LFL) in the pods that is thicker than the vascular bundle sheaths (VS, or suture string) layer, while the LFL is greatly reduced in domesticated varieties. Stringless snap beans have a weakly lignified VS at the suture, with a reduction in the number of lignified cells and the extent of secondary cell wall deposition in each cell, as reported previously (Prakken, 1934; Rau *et al.*, 2018). In stringless beans, the LFL is typically absent. In contrast to the clear anatomical differences between these three groups, no variation between PD-resistant and PD-susceptible domesticated dry bean pods was observed (Fig. 1B, 1C), which parallels the pattern caused by the soybean gene *POD DEHISCENCE 1* (*PDH1,* Suzuki *et al.*, 2009; Tiwari & Bhatia, 1995). This observation suggests that the genetic change responsible for reduction of PD among dry beans may have been related to a modification of fiber composition or structure (e.g., lignin) rather than the total quantity of lignin or cell fate in the relevant pod structures.

Variation in the ICA Bunsi/SXB 405 (IxS) population

Segregation for PD was first determined in a RI population derived from PD-susceptible cv. 'ICA Bunsi' and PD-resistant breeding line SXB 405 (Assefa *et al.*, 2013). Both parental genotypes belong to the Middle American domesticated gene pool. Three phenotyping approaches were used to evaluate PD (Supplementary Fig. S1) and each had a unique segregation pattern (Supplementary Fig. S2). These phenotypes were strongly correlated (Supplementary Fig. S3). RI lines that dehisced in the field had higher rates of PD after desiccation at 65°C (two-tailed t-test, $p=3.1*10^{-8}$) and required lower levels of force to induce fracture at the sutures (two-tailed t-test, $p=1.2*10^{-9}$). Similarly, the proportion dehiscing in the desiccator and force required to cause PD were negatively correlated ($r^2 = 0.71$ simple linear model, $p < 2*10^{-16}$).

QTL mapping by composite interval mapping identified a major, PD-related QTL peak located in the same position on linkage group Pv03 using each of the three phenotyping methods (Fig. 2). The QTL mapped between SNP markers ss715639553 and ss715639323 (Table 1).

Force measurement produced the most significant results (LOD score 53.3), followed by desiccation (LOD score 42.7), and field notes (LOD score 8.9). Each phenotyping method produced results that were statistically significant based on 1000 randomized permutations of the data. The allele at the most significant SNP explained 17% of the variation in PD based on field notes, 59% of the variation based on desiccation, and 64% of the variation in fracture force in the population. Analyses to find additional PD QTLs failed to identify other regions of interest in the IxS population.

Validation through association mapping

Next, we examined whether the Pv03 QTL affected PD in a broader cross-section of the dry bean gene pool. A genome-wide association study (GWAS), conducted using the desiccation method in the Middle American Diversity Panel (MDP), indicated that the most significant SNP (S1_149243152) was located in the QTL interval on Pv03 (Fig. 3A, MAF threshold = 0.1). This SNP was less than 5.7 kb from a candidate gene, *PvPdh1* (see next section). This association analysis also revealed loci significantly associated with PD on chromosomes Pv06 and Pv08 (Fig. 3A).

GWAS was similarly conducted in the Andean diversity panel (ADP) to determine which loci control PD in this independently domesticated population. Chromosomes Pv03, Pv05, Pv08, and Pv09 all included major regions significantly associated with PD (Fig. 3B). The QTL on chromosome Pv08 was in an overlapping physical position with the QTL from the MDP (Fig. 3B, Table 1). The QTLs on chromosome Pv03 in the ADP and MDP appear to be only partially overlapping, and different candidate genes can be invoked (see next sections).

In both the Andean and Middle American gene pools, PD varied greatly among market classes (Supplementary Table S1). GWAS using only members of race Mesoamerica (MDP with PC1 > 50) showed that the Pv08 QTL was most closely associated with PD in this race (Supplementary Fig. S4). SNP S1_329543689, near the center of this interval of interest, was used for subsequent analyses. The region near *PvPdh1* did not include significant SNPs in this race, further indicating that races Durango and Mesoamerica rely on different genes for PD resistance.

To visualize the correlation between PD and population substructure in the MDP, PD was plotted against the first principal component of the genetic data. Each point was color-coded by

its allele at the GWAS SNP peaks on Pv03 (S1_149243152, 5.7kb from *PvPdh1*) and Pv08 (SNP S1_329543689) (Fig. 4A, B). Members of the MDP with the Pv03 PD resistance allele exhibited mean PD in the desiccator of 0.0067, with a maximum value of 0.14. Members of the MDP with the Pv08 PD resistance allele showed a mean PD of 0.021 and a maximum value of 0.08. In genotypes with no known resistance allele, the mean level of PD was 0.206 and ranged up to 0.63 (Fig. 4B). The mutations on Pv03 and Pv08 likely reflected independent selection for reduced PD in their respective environments (highland vs. lowland). No synergistic gene action was observed between these two loci (Fig. 4B).

Identification of a candidate gene for the Pv03 QTL

The most significant SNP from the MDP GWAS (Fig. 3A) was located in an intergenic region well within the QTL mapping interval. One of the genes directly flanking this intergenic region was of immediate interest due to its unique expression pattern. The gene, Phvul.003G252100, is transcribed solely in developing pods (Supplementary Fig. S5; data from O'Rourke *et al.*, 2014), indicating that its function is unique to this structure. This gene encodes a dirigent-like protein, a family believed to regulate PD in soybean (Funatsuki *et al.,* 2014). Due to the close phylogenetic relationship and extensive microsynteny between *P. vulgaris* and *G. max* (McClean *et al.*, 2010; Schmutz *et al.*, 2014), further analyses were conducted to determine the degree of synteny and orthology between common bean and soybean QTLs related to PD. The LegumeIP 2.0 synteny tool (Li *et al.*, 2016) indicated that strong synteny exists between the soybean region surrounding *GmPdh1* in soybean and the common bean QTL on Pv03 (Supplementary Table S2), in agreement with previous synteny analyses (McClean *et al.*, 2010; Schmutz *et al.*, 2014). An amino acid BLAST of GmPDH1 (cv. Toyosume) against the *P. vulgaris* G19833 proteome (Schmutz *et al.*, 2014) indicated that the most similar common bean protein is encoded by the Phvul.003G252100 gene model, which was immediately adjacent to our most significant GWAS SNP. A neighbor-joining tree of common bean and soybean dirigent proteins indicates that GmPDH1 and the protein product of Phvul.003G252100 cluster together (Supplementary Fig. S6). Together, these results suggest that Phvul.003G252100 is orthologous to *GmPDH1*. Phvul.003G252100 is hereafter referred to as *PvPdh1*.

350 Sequencing of *PvPDH1*

351 Sequencing of *PvPdh1* in ICA Bunsi and SXB 405 revealed a non-synonymous single-base-pair
352 substitution at position 485 of the gene's coding sequence (Supplementary Fig. S7A). This
353 substitution leads to a threonine/asparagine polymorphism (T162N) in the protein product
354 (Supplementary Fig. S7B). The 11 RILs with recombination between the most significant
355 markers from QTL mapping showed complete co-segregation between the threonine/asparagine
356 polymorphism and the PD phenotype (Supplementary Table S3). To investigate the functional
357 importance of T162N, we evaluated the extent of its sequence conservation, surveyed literature
358 related to this position in closely related dirigent proteins, and used PROVEAN to predict the
359 effect of this substitution at the position. Sequencing of *PvPdh1* in several species of wild and
360 domesticated *Phaseolus* from NPGS and UC Davis showed that the asparagine at this position
361 was unique to the Middle American domesticated gene pool (Supplementary Table S4). No
362 polymorphism in the Andean gene pool was consistently associated with PD. In the Middle
363 American gene pool, PD was significantly higher among genotypes with a threonine at position
364 162 than an asparagine (t-test: $p=9.97 \times 10^{-5}$, n=47, Supplementary Fig. S8). This threonine was
365 strictly conserved in Andean domesticated common bean, Middle American and Andean wild
366 common bean, and the closely related *P. dumosus* and *P. lunatus* (Supplementary Table S4).
367 In addition, the threonine residue is present in 99 of the 100 most similar proteins in the
368 NCBI database (Supplementary Fig. S9A), indicating its functional importance. The protein that
369 lacks a threonine at this position is found in *Trifolium subterraneum*, a legume that produces
370 pods that mature underground. PD is not relevant for seed dispersal in this species and the gene
371 may be undergoing pseudogenization. This threonine is also conserved in the 19 most similar
372 proteins of *Selaginella moellendorffii* (Supplementary Fig. S9B), a member of the first diverging
373 group of lignin-containing plants, indicating that the residue has been conserved since before the
374 lycophyte-euphyllophyte divergence 400 million years ago (Soltis *et al.*, 2002; Zimmer *et al.*,
375 2007). No comparable protein could be found in the proteome of *Physcomitrella patens*, a non-
376 lignified moss. Studies of closely related dirigent proteins indicate that this threonine is a
377 component of one of the protein's active sites, and that its substitution eliminates protein
378 function. An analysis with PROVEAN (Choi & Chan, 2015) predicted that the T162N mutation
379 would have a deleterious effect (score: -4.587, cutoff = -2.5).
380

Candidate genes for other QTLs identified by association mapping

Association mapping revealed several other dehiscence-related QTLs across the gene pools and races of common bean (Table 1). Our ADP association mapping identified significant Pv03 SNPs in an interval that is syntenic with a region controlling dehiscence in cowpea (Lo *et al.*, 2018). *NAC* family and C2H2-type zinc finger transcription factors are found in this region (Table 1) and members of these families affect PD in soybean (Dong *et al.*, 2014) and rapeseed (Tao *et al.*, 2017), respectively. Orthologs of these genes may also affect dehiscence in cowpea (Lo *et al.*, 2018). The QTL identified in the ADP is large enough to include *PvPdh1*, although the QTLs discovered in Middle American beans and cowpeas are non-overlapping (Table 1).

Another major QTL for PD in Andean beans maps to Pv05, as described recently (Rau *et al.*, 2018), and several genes in this region are candidates for future study. Rau *et al*. (2018) noted that an ortholog of *MYB26* exists in the qPD5.1-Pv region of interest on Pv05, which may be responsible for variation in PD. Significant Pv05 SNPs from our association mapping completely envelope the qPD5.1-Pv interval, supporting this result. Our most significant Pv05 SNPs in the ADP are found just 22kb from *MYB46. MYB46* is involved in the same pathway as *MYB26* and the soybean PD resistance gene *SHAT1-5* (Dong *et al.*, 2014; McCarthy *et al.*, 2009). *MYB46* also works redundantly with *MYB83*, a gene that may play a role in cowpea pod development (Suanum *et al.*, 2016; Lo *et al.*, 2018), making *MYB46* another potential subject of future study.

Several genes of interest exist near the middle of the ADP's Pv08 GWAS peak. These include a MYB family transcription factor with similarity to *A. thaliana MYB17,* three *WRKY* family transcription factors, which are related to genes involved in sorghum dehiscence (Tang *et al.*, 2013) and a polygalacturonase, a group known to influence PD in *A. thaliana* (Ogawa *et al.*, 2009) (Table 1).

The Pv09 GWAS peak found in the ADP included a gene predicted to be *cellulose synthase A7* (*CESA7*, Table 1). *CESA7* may play a role in fiber development in cowpea (Suanum *et al.*, 2016). Similarly, two polygalacturonases are found in this interval, and members of this family are known to affect seed dispersal in *A. thaliana* (Ogawa *et al.*, 2009). These genes may regulate dehiscence by altering the breakdown of cell wall material in developing pods.

411 Associations between ecogeographic race, environment of origin, and PD
412 In landraces genotyped by Kwak and Gepts 2009, individuals belonging primarily to race
413 Durango (genetically indistinguishable from race Jalisco) came from regions with significantly
414 lower rainfall (709mm/yr vs. 1215mm/yr, Student's t-test $p=2.3*10^{-5}$) and higher elevations
415 (1312m vs. 1879m, student's t-test p=0.002) than landraces primarily belonging to race
416 Mesoamerica (Fig. S10). These results are in agreement with previous analyses (Singh *et al.*
417 1991).
418 The PD-resistant allele of *PvPdh1* on Pv03 is found exclusively in genotypes with
419 ancestry from ecogeographic race Durango (Fig. 4A, Table 1), which evolved in the northern,
420 semiarid highlands of Mexico. The conditions in these areas cause pods to become dry and
421 brittle, which exacerbates PD. The non-functional *PvPdh1* allele (caused by the replacement of a
422 threonine in position 162 by an asparagine) rose to very high frequency in this ecogeographic
423 race. In contrast, race Mesoamerica is adapted to humid lowlands, where environmental
424 conditions mask PD and reduce selection pressure against it. In this race, the loss-of-function
425 *PvPdh1* allele remains at low frequency and PD is widespread (Figs. 4A, 5).
426
427

Discussion

Associations with environmental conditions

Pod dehiscence (PD) in common bean is correlated with environmental parameters (Fig. S10). Common bean was domesticated twice, once in the Andes and once in the western region of Middle America (Gepts *et al.* 1986, Kwak *et al.* 2009, Bitocchi *et al.* 2013). From the Middle American center of origin, race Durango developed as cultivated common bean spread north into the semi-arid highlands of northern Mexico and the southwestern United States. In contrast, race Mesoamerica formed as the crop spread south into the lowland tropics of southern Mexico and Central America (Fig. 5; Singh *et al.* 1991, Kwak *et al.* 2009). These variable environmental conditions may have led to strong differences in selection pressure among the races, including differences in selection against PD. The arid conditions of northern Mexico are highly conducive to PD, which could lead to major yield losses. In the tropical lowlands, environmental humidity masks susceptibility to PD, reducing selection pressure against it. The wild-type *PvPdh1* allele may also be responsible for the ease of threshing that has been noted in race Mesoamerica. In humid environments, the wild type *PvPdh1* allele may facilitate separation of seeds from pod material, while PD in the field remains low. In northern Mexico, the semi-arid climate facilitates threshing but increases PD in the field. Under these conditions, the PD-resistance allele may be advantageous. Therefore, variation in *PvPdh1* allele frequency may be the result of selection for local adaptation based on this tradeoff (Fig. 5). Nevertheless, the existence of varieties that displayed low levels of PD despite having no known PD-resistance allele indicates that there could be incomplete PD expressivity or additional PD-resistance loci that remain to be identified. Future work could identify detailed spatial patterns of *PvPdh1* allele frequency across a broad panel of Mexican landraces of known geographic origins. Alleles that prevent PD will be valuable in coming decades, which are predicted to be increasingly arid (Sherwood & Fu, 2014).

The markedly different fitness landscape of domestication

The strict conservation of the threonine at position 162 in *PvPdh1* highlights its functional importance in wild populations and species over hundreds of millions of years. Yet, in a remarkable example of parallelism, independent loss-of-function mutations in this gene at some time in the last 10,000 years since domestication are found in certain domesticated populations in

soybean and common bean, both species being subjected to selection for reduced dehiscence. This highlights the strong differences in selection pressure between the wild and cultivated environments, which in turn modify the fitness landscapes of the wild and cultivated environments. Whereas the wild environment favors PD, the cultivated environment favors pod indehiscence: a single locus with a single amino acid substitution is sufficient to bridge these two fitness peaks. The threonine to asparagine substitution further provides an additional example of strongly convergent phenotypic and molecular evolution (Lenser & Theißen, 2013). Similar examples of parallel evolution in common bean include the determinacy trait (*fin* or *PvTFL1y*; Repinski *et al.*, 2012; Kwak *et al.*, 2012), absence of pigmentation (*P*; McClean *et al.*, 2018), and photoperiod adaptation (Weller *et al.*, 2019). In contrast, the major QTL on Pv05 discovered in a biparental population by Rau *et al.* (2018) and confirmed here in a diverse panel of Andean beans is not closely orthologous to PD-related loci yet described in other species. Future investigations may find that this locus has also been subject to parallel molecular evolution among taxa.

Our results serve as a note of caution when assessing the 'cost of domestication' on the basis of supposedly deleterious mutations identified by sequence variation alone. This cost refers to the load of harmful mutations that accumulates as a consequence of linkage, selection, and genetic drift during and after domestication. Several studies have documented this cost, for example, in horse (Schubert *et al.* 2014), sunflower, globe artichoke, and cardoon (Renaut & Rieseberg, 2015), and rice (Liu *et al.*, 2017). Conversely, our results indicate that non-synonymous mutations may also be responsible for advantageous changes that have occurred during crop domestication and dispersal beyond the species' native range. Thus, these bioinformatic studies should be complemented by studies measuring fitness under specific environments reflecting both the ancestral, wild and the derived, domesticated environments.

Further research is needed to identify the biochemical and biophysical aspects responsible for differences in PD in domesticated dry beans. Notably, our results could shed light on the fundamental process of lignin synthesis and fate under different environmental conditions. Dirigent-like genes, including *PvPdh1,* encode non-enzymatic proteins that guide the dimerization of lignin and lignan monomers (Davin *et al.*, 1997). The role of these proteins in lignin synthesis has been debated, with suggestions that polymerization is guided (Davin & Lewis, 2005; Hosmani *et al.*, 2013) or unguided (Ralph *et al.*, 1999, 2008). Varieties of common

bean with mutations in *Pdh1* could be used to elucidate the role of this protein family in lignin synthesis generally.

Redundancies in genetic control and maintenance of atavistic traits

Crosses between races have tremendous potential for crop improvement (for example, between races Durango and Mesoamerica: Singh *et al.*, 1993), but can also result in problematic gene complementation in the progeny of crosses between parental lines with different PD resistance genes. Because several genes influence PD redundantly, progenies descended from crosses between these parents could show complementation allowing the expression of PD. In the absence of selection against PD, in a humid environment, for example, PD could reappear in breeding programs in spite of the deleterious effects of PD in a domesticated environment. Complementation and environmental dependency of PD are the cause for the maintenance of atavistic traits in a domesticated gene pool in the absence of sympatric wild populations, and are responsible for the high levels of dehiscence seen in some cultivars of common bean.

In conclusion, our results depict crop domestication as a complex phenomenon, going beyond a single process that took place in a single, geographically and temporally circumscribed area. Domestication embraced the genetic complexity of higher plants wherein the same phenotype can be based on contrasting molecular foundations and interactions, in addition to spatially and temporally variable environments. This stands in contrast to many earlier studies, which have been based on the assumption that domestication occurred in a very specific geographic and temporal range within any given species (e.g. Matsuoka *et al.* 2002, Kwak *et al.* 2009, Huang *et al.* 2012, Bitocchi *et al.* 2013). It also highlights the importance of studying the genetic basis of domestication traits in genetically diverse populations. Our results depict domestication as including adaptations to a series of radically different environments, in which long-standing selection regimes in the wild can be reversed and replaced by new selective paradigms and alternate monomorphisms under domestication. Our results further highlight the fact that even core domestication traits, such as seed retention, can be found in a variable state in well-domesticated species. Crop domestication was a complex process of adaptation to a range of new environments, with multiple genetic paths to increased fitness in each environment, and without a single fixed solution for overcoming any given obstacle. This genetic complexity brings the

investigation of plant domestication beyond the realm of an academic exercise, and has serious implications for plant breeding and future food security.

Acknowledgments

S. Beebe (CIAT, Cali, Colombia) provided seeds of the IxS population. Seeds of the ADP and MDP were provided by R. Lee and P. McClean (North Dakota State University). Undergraduates Mayara Rocha, Poliana Silva Rezende, Guilherme Coelho Portilho, Natalie Hamada, Emily Yang, Ariel Herrera, Jose Pimentel, Matthew Bustamante, Emily White, Julia Gonzales, and Paige Augello contributed to DNA extractions, pod phenotyping, and other lab protocols. Paola Hurtado, Andrea Ariani, and other members of the Gepts group provided ideas for data analysis. Funding for T.A.P. was provided through a Clif Bar Family Foundation Seed Matters fellowship and Lundberg Family Farms research support. We would like to thank two anonymous reviewers for excellent, in-depth reviews.

Author contributions

TAP prepared the manuscript and conducted laboratory phenotyping, QTL mapping, GWAS, microscopy, and sequencing. JCBMT genotyped the IxS population, gathered field phenotypes, co-conducted QTL mapping, and provided guidance for other procedures. AP assisted with field and greenhouse trials. JJ led the sectioning and microscopy studies. PG conceived the initial project and provided guidance. All authors edited the manuscript.

Data availability

Segregation data of pod shattering data (oven test proportion, force, and shattering in the field) as well as SNP markers in the ICA Bunsi x SXB405 population have been deposited in the UC Davis Dash public database: https://doi.org/10.25338/B8TW2N (Parker *et al.* 2019). Genotype data for the Middle American Diversity Panel (Moghaddam *et al.* 2016) can be accessed at http://arsftfbean.uprm.edu/beancap/research/. Genotype data for the Andean Diversity Panel can be accessed at http://arsftfbean.uprm.edu/bean/?p=472 (Cichy *et al.* 2015). Coding DNA sequences of *PvPdh1* have been deposited in the NCBI database: accessions MN094634-MN094748.

References

Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, Yu YK. 2005. Protein database searches using compositionally adjusted substitution matrices. *The FEBS Journal* **272:** 5101-5109.

Ariani A, Berny Mier y Teran J, Gepts P. 2018. Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Molecular Biology and Evolution* **35:** 119-131.

Assefa T, Beebe SE, Rao IM, Cuasquer JB, Duque MC, Rivera M, Battisti A, Lucchin M. 2013. Pod harvest index as a selection criterion to improve drought resistance in white pea bean. *Field Crops Research* **148:** 24-33.

Azani N, Babineau M, Bailey CD, Banks H, Barbosa AR, Pinto RB, Boatwright JS, Borges LM, Brown GK, Bruneau A, *et al.* 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny The Legume Phylogeny Working Group (LPWG). *Taxon* **66:** 44-77.

Ballester P, Ferrándiz C. 2017. Shattering fruits: variations on a dehiscent theme. *Current Opinion in Plant Biology* **35:** 68-75.

Bandillo NB, Anderson JE, Kantar MB, Stupar RM, Specht JE, Graef GL, Lorenz AJ. 2017. Dissecting the genetic basis of local adaptation in soybean. *Scientific Reports* **7:** 17195.

Berny Mier y Teran JC, Konzen ER, Palkovic A, Tsai SM, Rao IM, Beebe S, Gepts P. 2019. Effect of drought stress on the genetic architecture of photosynthate allocation and remobilization in pods of common bean (*Phaseolus vulgaris* L.), a key species for food security. *BMC Plant Biology* **19:** 171.

Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Zeuli PS, Gioia T, Logozzo G, Attene G. 2013. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist* **197:** 300-313.

Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, Santo D, Nanni L, Attene G, Papa R. 2017. Beans (*Phaseolus* ssp.) as a model for understanding crop evolution. *Frontiers in Plant Science* **8:** 722.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23:** 2633-2635.

Broman KW, Wu H, Sen Ś, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19:** 889-890.

Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31:** 2745-2747.

Cichy KA, Porch TG, Beaver JS, Cregan P, Fourie D, Glahn RP, Grusak MA, Kamfwa K, Katuuramu DN, McClean P, et al. 2015. A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop Science* **55:** 2149-2160.

Darwin C. 1859. On the origin of species. J. Murray, London

Davin LB, Lewis NG. 2005. Lignin primary structures and dirigent sites. *Current Opinion in Biotechnology* **16:** 407-415.

Davin LB, Wang H-B, Crowell AL, Bedgar DL, Martin DM, Sarkanen S, Lewis NG. 1997. Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science* **275:** 362-367.

Dereeper A, Nicolas S, Le Cunff L, Bacilieri R, Doligez A, Peros J-P, Ruiz M, This P. 2011. SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. *BMC Bioinformatics* **12**: 134.

Desper R, Gascuel O. 2004. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Molecular Biology and Evolution* **21**: 587-598.

Dong Y, Wang Y-Z. 2015. Seed shattering: from models to crops. *Frontiers in Plant Science* **6**: 476.

Dong Y, Yang X, Liu J, Wang B-H, Liu B-L, Wang Y-Z. 2014. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nature Communications* **5**: 3352.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**: 4302-4315.

Freyre R, Skroch P, Geffroy V, Adam-Blondon A-F, Shirmohamadali A, Johnson W, Llaca V, Nodari R, Pereira P, Tsai S-M, et al. 1998. Towards an integrated linkage map of common bean. 4. Development of a core map and alignment of RFLP maps. *Theor. Appl. Genet.* **97**: 847-856.

Funatsuki H, Suzuki M, Hirose A, Inaba H, Yamada T, Hajika M, Komatsu K, Katayama T, Sayama T, Ishimoto M, et al. 2014. Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proceedings of the National Academy of Sciences* **111**: 17797-17802.

Gepts P. 2004. Crop domestication as a long-term selection experiment. *Plant Breeding Reviews* **24 (Part 2)**: 1-44.

Gepts P. 2014. The contribution of genetic and genomic approaches to plant domestication studies. *Current Opinion in Plant Biology* **18**: 51-59.

Gepts P, Aragão FJL, Barros Ed, Blair MW, Brondani R, Broughton W, Galasso I, Hernández G, Kami J, Lariguet P, et al. 2008. Genomics of *Phaseolus* beans, a major source of dietary protein and micronutrients in the Tropics. In: Moore PH, Ming R eds. *Genomics of Tropical Crop Plants*. Berlin: Springer, 113-143.

Gepts P, Berny Mier y Teran JC, Konzen ER, Palkovic A, Tsai SM, Rao IM, Beebe S 2019. Dataset - Effect of drought stress on the genetic architecture of photosynthate allocation and remobilization in pods of common bean *(Phaseolus vulgaris* L.), a key species for food security. https://doi.org/10.25338/B8TW2N. Davis, CA, USA: University of California Dash.

Gioia T, Logozzo G, Kami J, Spagnoletti Zeuli P, Gepts P. 2013. Identification and characterization of a homologue to the Arabidopsis *INDEHISCENT* gene in common bean. *Journal of Heredity* **104**(2): 273-286.

Grishin NV. 1995. Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *Journal of Molecular Evolution* 41: 675-679.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**(D1): D1178-D1186.

Hammer K. 1984. Das Domestikationssyndrom. *Die Kulturpflanze* **32**: 11-34.

Harlan JR. 1992. *Crops and man, 2nd ed.* Madison, WI: American Society of Agronomy.

Hosmani PS, Kamiya T, Danku J, Naseer S, Geldner N, Guerinot ML, Salt DE. 2013. Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. *Proceedings of the National Academy of Sciences* **110**: 14498-14503.

Huang X, Kurata N, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497.

Kelso NV, Patterson T. 2010. Introducing natural earth data - naturalearthdata.com. *Geographia Technica* **Special issue 2010**: 82-89.

Koinange EMK, Singh SP, Gepts P. 1996. Genetic control of the domestication syndrome in common-bean. *Crop Science* **36**: 1037-1045.

Kwak M, Gepts P. 2009. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theoretical and Applied Genetics* ***118***: 979-992.

Kwak M, Kami JA, Gepts P. 2009. The putative Mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma–Santiago Basin of Mexico. *Crop Science* **49**: 554-563.

Kwak M, Toro O, Debouck D, Gepts P. 2012. Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris* L.). *Annals of Botany* **110**: 1573-1580.

Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185-199.

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Climer Vigueira C, Denham T, Dobney K, et al. 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* **111**: 6139-6146.

Lenser T, Theißen G. 2013. Molecular mechanisms involved in convergent crop domestication. *Trends in Plant Science* **18**: 704-714.

Li J, Dai X, Zhuang Z, Zhao PX. 2016. LegumeIP 2.0—a platform for the study of gene function and genome evolution in legumes. *Nucleic Acids Research* **44**(D1): D1189-D1194.

Liu Q, Zhou Y, Morrell PL, Gaut BS. 2017. Deleterious variants in Asian rice and the potential cost of domestication. *Molecular Biology and Evolution* **34**: 908-924.

Lo S, Muñoz-Amatriaín M, Boukar O, Herniter I, Cisse N, Guo Y-N, Roberts PA, Xu S, Fatokun C, Close TJ. 2018. Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp.). *Scientific Reports* **8**: 6261.

Lyons E, Pedersen B, Kane J. 2008. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the Rosids. *Tropical Plant Biology* **1**: 181–190.

Martínez-Ainsworth NE, Tenaillon MI. 2016. Superheroes and masterminds of plant domestication. *Comptes Rendus Biologies* **339**: 268-273.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences* **99**: 6080-6084.

McCarthy RL, Zhong R, Ye Z-H. 2009. MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant and Cell Physiology* **50**: 1950-1964.

McClean P, Mamidi S, McConnell M, Chikara S, Lee R. 2010. Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics* **11**: 184.

McClean PE, Bett KE, Stonehouse R, Lee R, Pflieger S, Moghaddam SM, Geffroy V, Miklas P, Mamidi S. 2018. White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the *P* (pigment) gene. *New Phytologist* **219**: 1112-1123.

Mendel G. 1866. Versuche über Pflanzenhybriden. Verhandlungen des Naturforschenden Vereins in Brunn, 4, 3-47.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet* **14**: 840-852.

Meyer RS, DuVal AE, Jensen HR. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist* **196**: 29-48.

Moghaddam SM, Mamidi S, Osorno JM, Lee R, Brick M, Kelly J, Miklas P, Urrea C, Song Q, Cregan P, et al. 2016. Genome-wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean. *The Plant Genome* **9**(3).

Moreno-Mayar JV, Vinner L, de Barros Damgaard P, de la Fuente C, Chan J, Spence JP, Allentoft ME, Vimala T, Racimo F, Pinotti T, et al. 2018. Early human dispersals within the Americas. *Science* **362**: eaav2621.

O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, McClean PE, Li J, Dai X, Zhao PX, et al. 2014. An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* **15**: 866.

Ogawa M, Kay P, Wilson S, Swain SM. 2009. ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are polygalacturonases required for cell separation during reproductive development in *Arabidopsis*. *The Plant Cell* **21**: 216-233.

Ogutcen E, Pandey A, Khan M, Marques E, Penmetsa R, Kahraman A, von Wettberg E. 2018. Pod shattering: a homologous series of variation underlying domestication and an avenue for crop improvement. *Agronomy* ***8***: 137.

Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**: 1073-1079.

Parker T, Berny Mier y Teran JC, Palkovic A, Jernstedt J, Gepts P 2019. Dataset - Pod indehiscence is a domestication and aridity resilience trait in common bean. Davis, CA: UC Davis Dash. doi: https://doi.org/10.25338/B8ZG68

Potter BA, Baichtal JF, Beaudoin AB, Fehren-Schmitz L, Haynes CV, Holliday VT, Holmes CE, Ives JW, Kelly RL, Llamas B, Malhi RS. 2018. Current evidence allows multiple models for the peopling of the Americas. *Science Advances* ***4***: eaat5473.

Prakken R. 1934. Inheritance of colours and pod characters in *Phaseolus vulgaris* L. *Genetica* **16**: 177-294.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Ralph J, Brunow G, Harris PJ, Dixon RA, Schatz PF, Boerjan W 2008. Lignification: are lignins biosynthesized via simple combinatorial chemistry or via proteinaceous control and template replication. In: Daayf F, Lattanzio V eds. *Recent Advances in Polyphenol Research*. Chichester, UK: Blackwell, 36-66.

Ralph J, Peng J, Lu F, Hatfield RD, Helm RF. 1999. Are lignins optically active? *Journal of Agricultural and Food Chemistry* **47**: 2991-2996.

Rau D, Murgia ML, Rodriguez M, Bitocchi E, Bellucci E, Fois D, Albani D, Nanni L, Gioia T, Santo D, et al. 2018. Genomic dissection of pod shattering in common bean: mutations at nonorthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *The Plant Journal* **97:** 693-714.

Renaut S, Rieseberg LH. 2015. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Molecular Biology and Evolution* **32**: 2273-2283.

Repinski SL, Kwak M, Gepts P. 2012. The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis TFL1*. *Theoretical and Applied Genetics* **124**(8): 1539-1547.

Rice AG, Umale PE, Dash S, Farmer AD, Cleary AM, Wilkey AP, Campbell JD, Karingula V, Huang W, Cannon SB, et al. 2015. Legume Information System (LegumeInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Research* **44**(D1): D1181-D1188.

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178-183.

Schmutz J, McClean P, Mamidi S, Wu G, Cannon S, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, et al. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics* **46**: 707-713.

Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, Albrechtsen A, Dupanloup I, Foucal A, Petersen B, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences* **11**: E5661-E5669.

Sherwood S, Fu Q. 2014. A drier future? *Science* **343**: 737-739.

Singh S. 1999. *Bean breeding for the 21st century*. Dordrecht, the Netherlands: Kluwer.

Singh SP, Gepts P, Debouck DG. 1991. Races of common bean (*Phaseolus vulgaris* L., Fabaceae). *Economic Botany* **45**: 379-396.

Singh SP, Molina A, Urrea CA, Gutiérrez JA. 1993. Use of interracial hybridization in breeding the race Durango common bean. *Canadian Journal Plant Science* **73**: 785-793.

Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough TG. 2002. Rate heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils. **99**: 4430-4435.

Song Q, Jia G, Hyten DL, Jenkins J, Hwang E-Y, Schroeder SG, Osorno JM, Schmutz J, Jackson SA, McClean PE, et al. 2015. SNP assay development for linkage map

construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3: Genes|Genomes|Genetics* **5**(11): 2285-2290.

Suanum W, Somta P, Kongjaimun A, Yimram T, Kaga A, Tomooka N, Takahashi Y, Srinives P. 2016. Co-localization of QTLs for pod fiber content and pod shattering in F_2 and backcross populations between yardlong bean and wild cowpea. *Molecular Breeding* **36**: 1-11.

Suzuki M, Fujino K, Funatsuki H. 2009. A major soybean QTL, *qPDH1*, controls pod dehiscence without marked morphological change. *Plant Production Science* **12**: 217-223.

Tang H, Cuevas HE, Das S, Sezen UU, Zhou C, Guo H, Goff VH, Ge Z, Clemente TE, Paterson AH. 2013. Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. *Proceedings of the National Academy of Sciences* **110**: 15824-15829.

Tao Z, Huang Y, Zhang L, Wang X, Liu G, Wang H. 2017. BnLATE, a Cys2/His2-type zinc-finger protein, enhances silique shattering resistance by negatively regulating lignin accumulation in the silique walls of *Brassica napus*. *PLOS ONE* **12**: e0168046.

Tiwari SP, Bhatia VS. 1995. Characters of pod anatomy associated with resistance to pod-shattering in soybean. *Annals of Botany* **76**: 483-485.

Turner S. 2018. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *Journal of Open Source Software* **3**: 731.

Ursache R, Andersen TG, Marhavý P, Geldner N. 2018. A protocol for combining fluorescent proteins with histological stains for diverse cell wall components. *The Plant Journal* **93**: 399-412.

Weller JL, Vander Schoor JK, Perez-Wright EC, Hecht V, González AM, Capel C, Yuste-Lisbona FJ, Lozano R, Santalla M. 2019. Parallel origins of photoperiod adaptation following dual domestications of common bean. *Journal of Experimental Botany* **70**: 1209-1219.

Zimmer A, Lang D, Richardt S, Frank W, Reski R, Rensing SA. 2007. Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Molecular Genetics and Genomics* **278**: 393-402.

Competing interests

The authors declare no competing interests.

Table 1. Summary of common bean pod fiber or dehiscence quantitative trait loci (QTLs), their genome locations, potential candidate genes, and homologies with other species.

Chromosome or Linkage Group	Gene pool	Ecogeographic race, if available (Singh et al. 1991)	QTL location (bp, V1.0, Schmutz et al. 2014)	Potential candidate genes (when identified)	Source in *Phaseolus vulgaris*	Homologies in other species (when known)
Pv02	Andean	Nueva Granada	43,425,893-43,900,872	*PvIND*	(Koinange *et al.*, 1996; Gioia *et al.*, 2013; Hagerty *et al.*, 2016)	*Arabidopsis* (Liljegren *et al.*, 2004)
Pv03	Middle American	Durango	47,527,006-48,475,205	*PvPdh1*: dirigent family	This research	Soybean (Funatsuki *et al.*, 2014)
Pv03	Andean		39,768,300-48,451,789	*NAC* family, C2H2 zinc finger	This research	Cowpea (Lo *et al.*, 2018)
Pv04	Middle American		42,310,662		Hagerty *et al.*, 2016	
Pv05	Andean	Nueva Granada	35,000,893-39,497,309	*MYB26*, *MYB46*	Rau *et al.*, 2018; this research	Cowpea (Suanum *et al.*, 2016 ; Lo *et al.*, 2018) ; *Arabidopsis* (McCarthy *et al.*, 2009)
Pv08	Andean & Middle American	Mesoamerica	330,345-9,215,942	*MYB* family, *WRKY* family, polygalacturonase	This research	Sorghum (Tang *et al.*, 2013); *Arabidopsis* (Ogawa *et al.*, 2009)
Pv09	Andean		29,587,741-37,450,759	*CESA7*, polygalacturonases	This research	Cowpea (Suanum *et al.*, 2016)

806

Legends to figures in the main text

Fig. 1 Variation in PD-related structures in common bean. (A) Cross-section of the ventral suture of G12873, a wild Middle American bean. Wild beans show very high pod dehiscence (PD) and extensively lignified vascular sheath (VS) and fiber layer (LFL) deposition in pod walls. (B) In pod dehiscence-susceptible domesticated dry beans (cv. ICA Bunsi shown), LFL deposition is reduced relative to wild types, indicating that these cells may be related to Middle American common bean domestication. (C) Pod dehiscence (PD)-resistant dry beans (cv. SXB 405 shown) are anatomically similar to PD-susceptible domesticated types (see B). (D) Stringless varieties (cv. Midas shown) display a reduction in VS lignification, including a reduction in secondary cell wall thickening. The LFL is absent in these varieties. Stained with 0.01% Auramine O. Scale bars represent 100μm.

Fig. 2 Pod dehiscence (PD) QTL mapping based on three phenotyping methods. (A) Genome-wide and (B) Pv03-specific mapping results. All methods produced statistically significant results in the same region of chromosome Pv03. The significance threshold, determined by 1000 randomized permutations of the data, is shown as a black bar at LOD=5.80. The common bean ortholog of *Pdh1*, which regulates PD in soybean, is located between the most significant markers from quantitative trait locus (QTL) mapping (Supplementary Table S3).

Fig. 3. GWAS of PD in independently domesticated common bean populations. (A) In the Middle American Diversity Panel (MDP), the most significant single-nucleotide polymorphism (SNP) is located 5.7kbp from the *PvPdh1* putative causal polymorphism. Pv06 and Pv08 also included loci of interest. (B) In the Andean Diversity Panel (ADP), chromosomes Pv03, Pv05, Pv08, and Pv09 include major regions of interest. SNPs located near *PvMYB26* (Rau *et al.* 2018) on Pv05 were highly significant. Horizontal red and blue lines indicate the Bonferroni-corrected significance threshold for an alpha of 0.01 and 0.05, respectively. Based on the proportion of pods dehiscing in a desiccator, with correction for population structure by principal component analysis.

Fig. 4 The relationship between pod dehiscence (PD), ecogeographic race, and resistance alleles. (A) The first principal component of genetic data for the Middle American Diversity Panel (MDP) separates race Durango (at left) from race Mesoamerica (at right). Members of race Durango have low susceptibility to PD relative to members of race Mesoamerica. Accessions are color-coded by genotype at the GWAS peaks on Pv03 and Pv08. (B) A violin plot showing of the extent of PD by allele in the MDP. Alleles are color-coded in the same way as in A. Accessions with these PD resistance loci have significantly lower levels of PD than accessions with neither allele. Letters “a” and “b” distinguish significantly different groups (Tukey’s Honestly Significant Difference).

Fig. 5 *PvPdh*ariation is correlated with range expansion and local adaptation in common bean. Pod dehiscence (PD) is nearly absent in Race Durango, a group adapted to the hot, dry environments of northern Mexico (see also Fig. S10), where environmental aridity exacerbates PD. The loss of function *PvPdh1* allele is nearly at fixation in this population. In contrast, race Mesoamerica is adapted to humid lowlands (Fig. S10), where more humid conditions mask PD susceptibility. PD may have been selected against less strongly in this population and the wild type *PvPdh1* predominates. For detailed information on the geographic distribution of these races, see Singh *et al*. (1991) and Kwak and Gepts (2009).

Legends to Supporting Information

Fig. S1. Pod dehiscence phenotyping methods.

Fig. S2. Phenotyping distributions in the ICA Bunsi/SXB 405 RI population.

Fig. S3. Correlations between phenotyping methods in the IxS RI population.

Fig. S4. GWAS of pod dehiscence (PD) in Race Mesoamerica (MDP, PC1>50) using GLM in SNiPlay/TASSEL.

Fig. S5. Expression of Phvul.003G252100.1 (*PvPdh1*) is unique to pods in *P. vulgaris* cv. Negro Jamapa.

Fig. S6. A rooted neighbor joining tree based on sequence of GmPDH1, PHAVU_003G252100g, and the most similar dirigent proteins of *G. max* and *P. vulgaris* in the NCBI database.

Fig. S7. A polymorphism exists in *PvPdh1* between the parents of the RI population.

Fig. S8. Dehiscence in Middle American GRIN NPGS accessions.

871 Fig. S9. The threonine at position 162 is a highly conserved component of the active site for dirigent-like
872 genes.

873 Fig. S10. The ecogeographic distribution of Race Durango and Race Mesoamerica landraces genotyped
874 by Kwak and Gepts 2009.

875 Table S1. PD after desiccation, by market class, gene pool, and ecogeographic race (Singh et al. 1991).

876 Table S2. Synteny near *Pdh1* in *G. max* and *P. vulgaris* – sharing of gene models. *PvPDH1*
877 (Phvul.003G252100.1) is in bold.

878 Table S3. Co-segregation between dehiscence phenotype and position 162 in *PvPdh1*. The 11 RILs with
879 recombination between the markers flanking the Pv03 QTL for pod dehiscence showed perfect
880 correspondence between phenotype and genotype at this position.

881 Table S4. Sequencing of *PvPdh1* in several species of wild and domesticated *Phaseolus*.

882

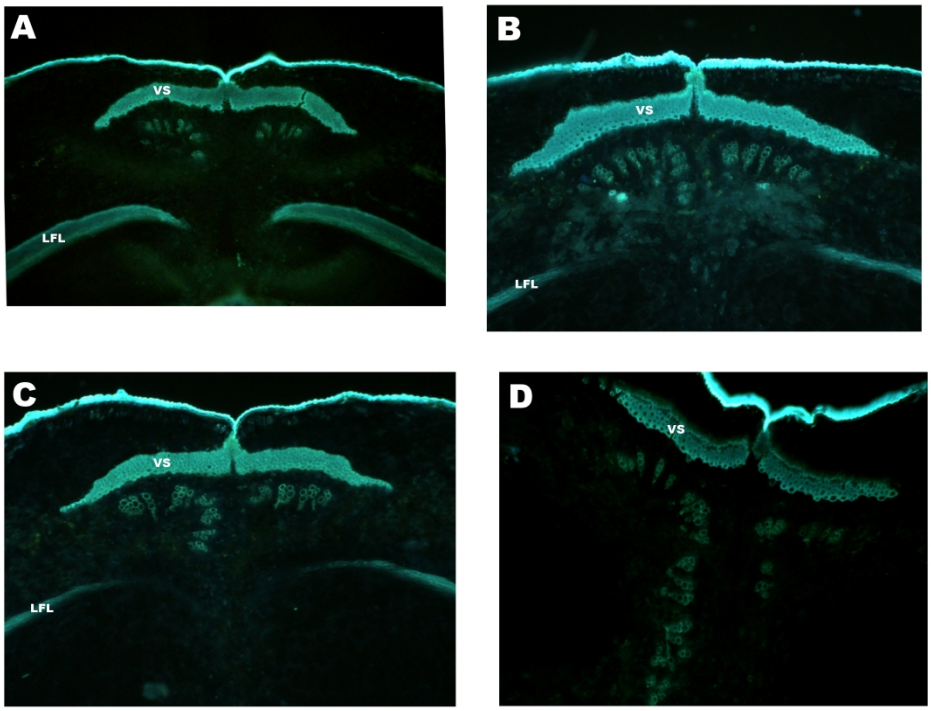

Fig. 1 Variation in PD-related structures in common bean. (A) Cross-section of the ventral suture of G12873, a wild Middle American bean. Wild beans show very high pod dehiscence (PD) and extensively lignified vascular sheath (VS) and fiber layer (LFL) deposition in pod walls. (B) In pod dehiscence-susceptible domesticated dry beans (cv. ICA Bunsi shown), LFL deposition is reduced relative to wild types, indicating that these cells may be related to Middle American common bean domestication. (C) Pod dehiscence (PD)-resistant dry beans (cv. SXB 405 shown) are anatomically similar to PD-susceptible domesticated types (see B). (D) Stringless varieties (cv. Midas shown) display a reduction in VS lignification, including a reduction in secondary cell wall thickening. The LFL is absent in these varieties. Stained with 0.01% Auramine O. Scale bars represent 100μm.

426x325mm (300 x 300 DPI)

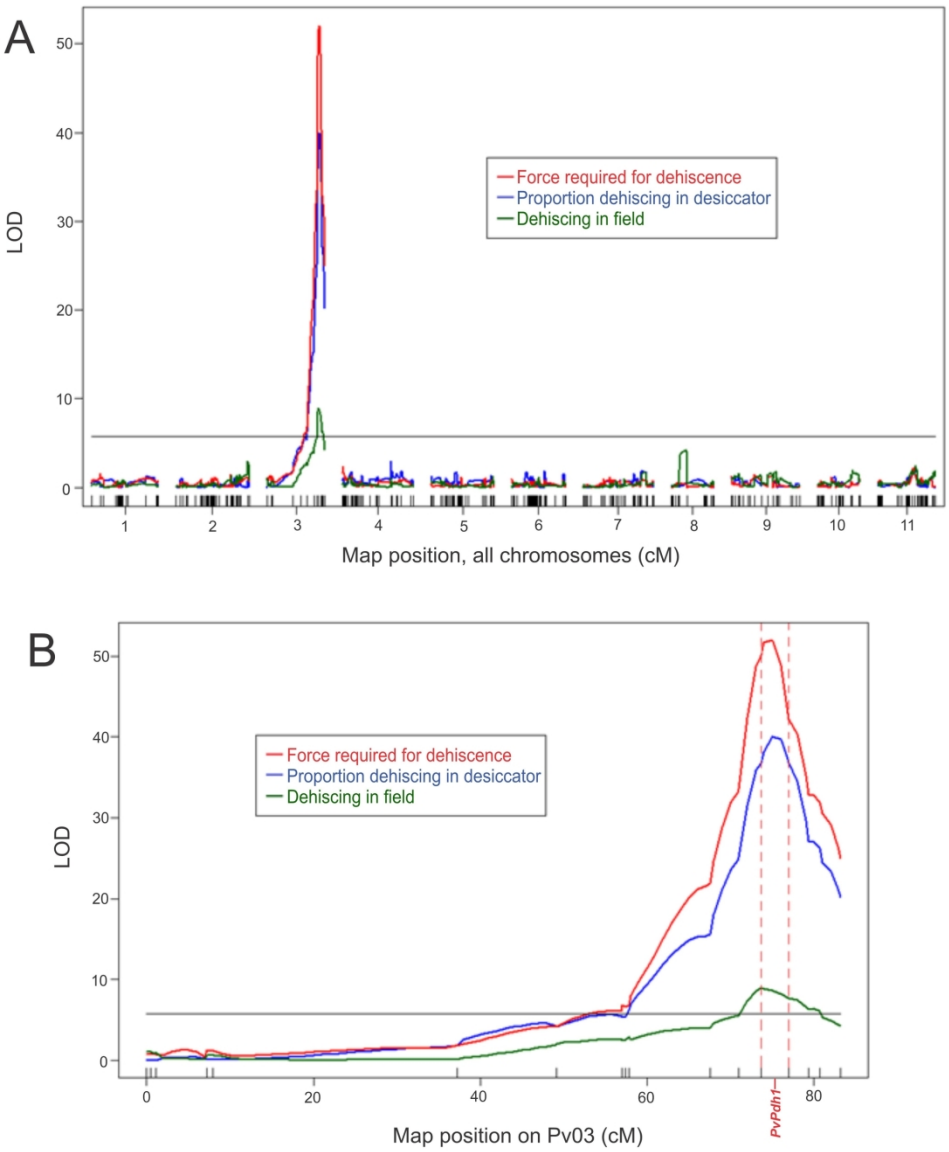

Fig. 2 Pod dehiscence (PD) QTL mapping based on three phenotyping methods. (A) Genome-wide and (B) Pv03-specific mapping results. All methods produced statistically significant results in the same region of chromosome Pv03. The significance threshold, determined by 1000 randomized permutations of the data, is shown as a black bar at LOD=5.80. The common bean ortholog of Pdh1, which regulates PD in soybean, is located between the most significant markers from quantitative trait locus (QTL) mapping (Supplementary Table S3).

184x218mm (300 x 300 DPI)

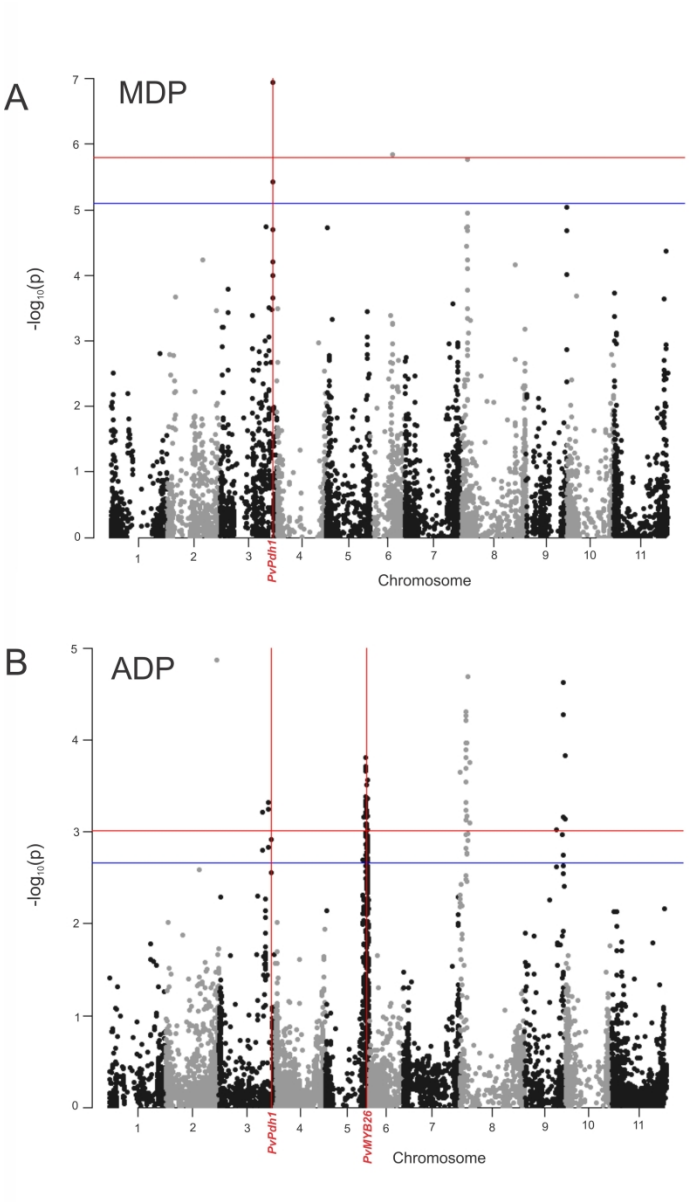

Fig. 3. GWAS of PD in independently domesticated common bean populations. (A) In the Middle American Diversity Panel (MDP), the most significant single-nucleotide polymorphism (SNP) is located 5.7kbp from the PvPdh1 putative causal polymorphism. Pv06 and Pv08 also included loci of interest. (B) In the Andean Diversity Panel (ADP), chromosomes Pv03, Pv05, Pv08, and Pv09 include major regions of interest. SNPs located near PvMYB26 (Rau et al. 2018) on Pv05 were highly significant. Horizontal red and blue lines indicate the Bonferroni-corrected significance threshold for an alpha of 0.01 and 0.05, respectively. Based on the proportion of pods dehiscing in a desiccator, with correction for population structure by principal component analysis.

195x325mm (300 x 300 DPI)

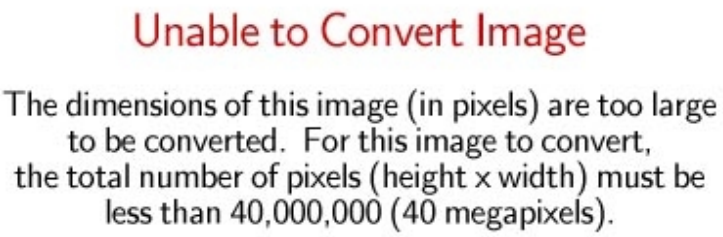

Fig. 4 The relationship between pod dehiscence (PD), ecogeographic race, and resistance alleles. (A) The first principal component of genetic data for the Middle American Diversity Panel (MDP) separates race Durango (at left) from race Mesoamerica (at right). Members of race Durango have low susceptibility to PD relative to members of race Mesoamerica. Accessions are color-coded by genotype at the GWAS peaks on Pv03 and Pv08. (B) A violin plot showing of the extent of PD by allele in the MDP. Alleles are color-coded in the same way as in A. Accessions with these PD resistance loci have significantly lower levels of PD than accessions with neither allele. Letters "a" and "b" distinguish significantly different groups (Tukey's Honestly Significant Difference).

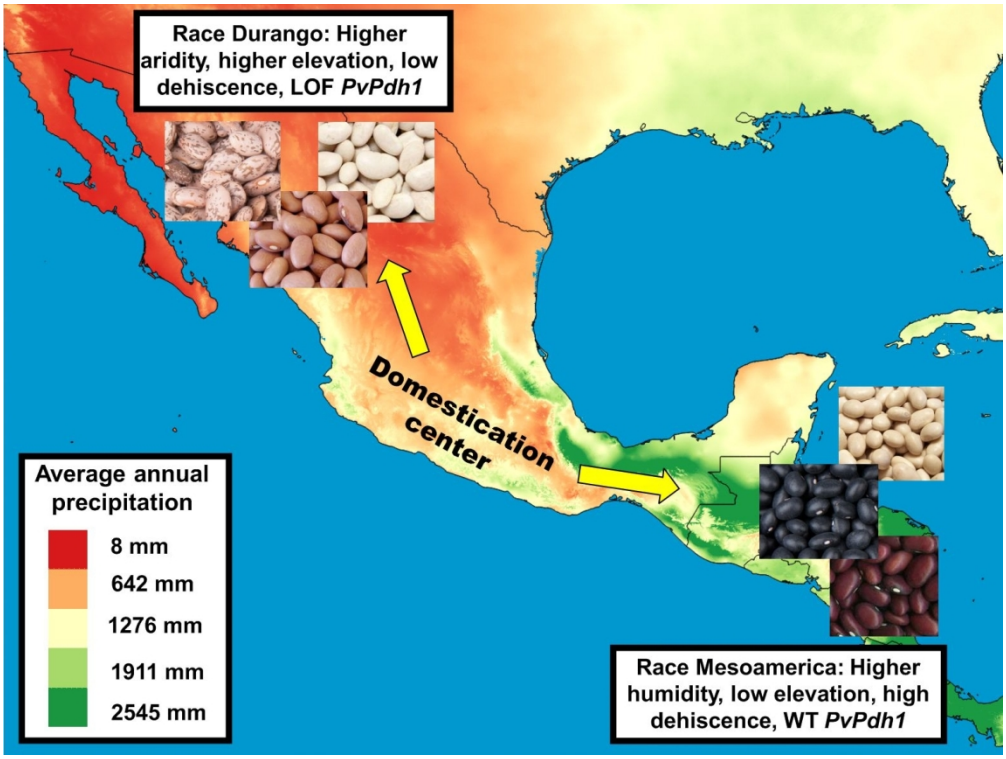

Fig. 5 PvPdhariation is correlated with range expansion and local adaptation in common bean. Pod dehiscence (PD) is nearly absent in Race Durango, a group adapted to the hot, dry environments of northern Mexico (see also Fig. S10), where environmental aridity exacerbates PD. The loss of function PvPdh1 allele is nearly at fixation in this population. In contrast, race Mesoamerica is adapted to humid lowlands (Fig. S10), where more humid conditions mask PD susceptibility. PD may have been selected against less strongly in this population and the wild type PvPdh1 predominates. For detailed information on the geographic distribution of these races, see Singh et al. (1991) and Kwak and Gepts (2009).

132x99mm (300 x 300 DPI)