Title

Loss of pod strings in common bean is associated with gene duplication, retrotransposon insertion and overexpression of PvIND

Journal

New Phytologist, 235(6)

ISSN

0028-646X

Authors

Parker, Travis A

Cetz, Jose

de Sousa, Lorenna Lopes

et al.

Publication Date

2022-09-01

DOI

10.1111/nph.18319

Peer reviewed

New Phytologist

Loss of pod strings in common bean is associated with gene duplication, retrotransposon insertion and overexpression of *PvIND*

Travis A. Parker[1], Jose Cetz[2], Lorenna Lopes de Sousa[1], Saarah Kuzay[1], Sassoum Lo[1], Talissa de Oliveira Floriani[1,3], Serah Njau[1,4], Esther Arunga[4], Jorge Duitama[5], Judy Jernstedt[1], James R. Myers[6], Victor Llaca[7], Alfredo Herrera-Estrella[2] and Paul Gepts[1]

[1]Department of Plant Sciences, University of California Davis, Davis, CA 95616-8780, USA; [2]National Laboratory of Genomics for Biodiversity, CINVESTAV, Irapuato, Guanajuato C.P. 36821, Mexico; [3]Department of Genetics, Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo, Piracicaba, SP 13418-900, Brazil; [4]Department of Water and Agricultural Resource Management, University of Embu, Embu 60100, Kenya; [5]Department of Systems and Computing Engineering, Universidad de los Andes, Bogotá, Colombia; [6]Department of Horticulture, Oregon State University, Corvallis, OR 97331, USA; [7]Corteva Agriscience, Johnston, IA 50131, USA

Author for correspondence:
Paul Gepts
Email: plgepts@ucdavis.edu

Received: *17 February 2022*
Accepted: *27 May 2022*

New Phytologist (2022) **235:** 2454–2465
doi: 10.1111/nph.18319

Key words: bHLH transcription factor, differential expression, domestication, gain-of-function mutation, gene duplication, *Phaseolus vulgaris*, pod dehiscence, *Ty1-copia* retrotransposon.

Summary

- Fruit development has been central in the evolution and domestication of flowering plants. In common bean (*Phaseolus vulgaris*), the principal global grain legume staple, two main production categories are distinguished by fibre deposition in pods: dry beans, with fibrous, stringy pods; and stringless snap/green beans, with reduced fibre deposition, which frequently revert to the ancestral stringy state. Here, we identify genetic and developmental patterns associated with pod fibre deposition.
- Transcriptional, anatomical, epigenetic and genetic regulation of pod strings were explored through RNA-seq, RT-qPCR, fluorescence microscopy, bisulfite sequencing and whole-genome sequencing.
- Overexpression of the *INDEHISCENT* ('*PvIND*') orthologue was observed in stringless types compared with isogenic stringy lines, associated with overspecification of weak dehiscence-zone cells throughout the pod vascular sheath. No differences in DNA methylation were correlated with this phenotype. Nonstringy varieties showed a tandemly direct duplicated *PvIND* and a *Ty1-copia* retrotransposon inserted between the two repeats. These sequence features are lost during pod reversion and are predictive of pod phenotype in diverse materials, supporting their role in *PvIND* overexpression and reversible string phenotype.
- Our results give insight into reversible gain-of-function mutations and possible genetic solutions to the reversion problem, of considerable economic value for green bean production.

Introduction

Novel forms of fruit-mediated seed dispersal have been important to the evolutionary success of flowering plants. These depend on unique developmental programmes, which have evolved across taxa. In the Fabaceae, the third most speciose plant family (LPWG, 2017), seed dispersal is mediated typically by explosive pod dehiscence, or shattering. For pod shattering to occur, multiple lignified pod tissues must develop properly, including vascular bundle sheath fibres, also called pod suture 'strings', and this is determined using a network of transcription factors and downstream cell wall modifying genes (Parker *et al.*, 2020b).

Similar to legumes, *Arabidopsis thaliana* produces dehiscent seed pods, called siliques, whose development is under the control of a network of transcription factors (Di Vittori *et al.*, 2019; Parker *et al.*, 2020b). Among these, *INDEHISCENT (IND)*, *SHATTERPROOF1/2* (*SHP1/2*), and *ALCATRAZ* (*ALC*) specify the valve margin region along which dehiscence occurs, and their expression is spatially restricted by genes such as *REPLUMLESS* (*RPL*) and *FRUITFUL* (*FUL*) (Gu *et al.*, 1998). In particular, *IND* is known to specify the area along which dehiscence occurs, such as the separation layer and lignified margin layer (Liljegren *et al.*, 2004; Girin *et al.*, 2010). These patterning genes ultimately promote the formation of secondary cell walls, which are strongly lignified cell wall materials added between the original (primary) lignified cell wall and the plasma membrane. The formation of these secondary cell walls is controlled by downstream NAC and MYB family transcription factors (Nakano *et al.*, 2015; Ohtani & Demura, 2019; Chen *et al.*, 2021; Gupta *et al.*, 2021), which are known to play a role in legume pod dehiscence (Rau *et al.*, 2019; Takahashi *et al.*, 2020; Watcharatpong *et al.*, 2020; Zhang & Singh, 2020). Other legume shattering-

controlling genes, such as the dirigent gene *PDH1* of soybean and *PvPdh1* of common bean, affect pod valve torsion without anatomical changes (Suzuki *et al.*, 2009; Parker *et al.*, 2020a, 2021) and are unlikely to regulate pod suture strings. Common bean genes homologous to Arabidopsis pod patterning genes (*IND*, *SHP1/2*, *ALC*, *RPL*, *FUL*) and their downstream NAC and MYB transcription factors are strong candidates for the control of pod string formation.

Members of the legume family have been independently domesticated at least 40 times (Hammer & Khoshbakht, 2014). Of these, common bean (*Phaseolus vulgaris*) is the largest source of plant protein and micro-nutrition for direct human consumption (Parker & Gepts, 2021). In wild beans, strongly lignified fibres exist at the pod sutures and inside the pod walls (Supporting Information Fig. S1). These fibres lead to explosive pod dehiscence (or 'shattering') at maturity and, therefore, ballistic seed dispersal. Following initial domestication-related selection for protein-rich dry beans (Piperno, 2012), pod wall fibre was significantly reduced but not eliminated. The suture fibres, also known as vascular bundle sheath fibres, are cells with strongly lignified primary and secondary cell walls. In the centre of this sheath of fibre cells is a narrow, weak dehiscence zone. The narrow dehiscence zone is the line along which the two pod walls separate and pod dehiscence and shattering occur. Dehiscence-zone cells have lignified primary cell walls but lack secondary cell walls entirely. Suture fibres were only partially reduced during the initial domestication of common bean, leaving suture fibres strong enough to facilitate threshing during harvest and to be removed as pod 'strings'. These were culinarily undesirable in vegetable green beans, and had to be removed by hand. The 19th century breeder Calvin Keeney identified a stringless mutation that led to loss of pod suture fibres in the cultivar 'Refugee Wax', eliminating the need to manually remove pod strings. Subsequent breeding efforts with this allele led to a novel commercial class, namely 'snap' beans (Wallace *et al.*, 2018). Nonstringy snap beans with this mutation have become the global market standard for types with edible pods for consumption as vegetables.

Nonstringy snap beans have lost lignified secondary cell wall thickening of the vascular bundle sheath and lack a discernible dehiscence zone (Prakken, 1934; Murgia *et al.*, 2017; Rau *et al.*, 2019; Parker *et al.*, 2020b). Snap bean varieties display frequent spontaneous reversion to high pod fibre content (Fig. 1), for pod strings and pod wall fibre. Reversions that affect either trait individually or both simultaneously occur in all known snap bean varieties (Smith *et al.*, 1997; Hagerty *et al.*, 2016). Here, *c.* 0.5–2.25% of plants in a population revert on average (Hagerty *et al.*, 2016). For commercial seed production, each plant in the field must be evaluated individually to remove ('rogue') revertants, which is a major expense (*c.* US$50 000 per cultivar, per year) for seed companies and a target for future breeding (Singh, 1989; Al-Bader, 2014). Improving the understanding of pod suture string inheritance would therefore be valuable commercially as well as scientifically. The high-frequency reversion to pod strings indicates that the trait could be controlled by transposable elements (Lisch, 2013; Hirsch & Springer, 2017), epigenetic factors such as DNA methylation (Miryeganeh & Saze, 2020), or by rapidly evolving sequence repeats (Gemayel *et al.*, 2012). This reversion also presents a highly isogenic system to study the basis of pod string development and the mechanism of this high-frequency reversion.

Emerson (1904) determined that the presence of pod strings in common bean were recessive unlike other wild-type, ancestral traits, and only partly fit Mendelian segregation ratios. Drijfhout (1978) proposed that the dominant *String* (*St*) allele was required for any reduction in pod string, with a dominant hypostatic allele *Temperature Sensitive* (*Ts*) able to recover partial pod string in the presence of *St* at elevated temperatures. Koinange *et al.* (1996) mapped *St* to chromosome Pv02. In their analysis, pod wall fibre and pod suture fibre were genetically co-located, although other authors have found them unlinked (Emerson,

Fig. 1 Pod phenotypes of plant materials of common bean (*Phaseolus vulgaris*). (a–f) Six stringless cultivars of snap bean. No suture string is present in pods at seed fill stage (stage R8; Fernández *et al.*, 1983) or dry mature pods. (g–l) Stringy revertants of each of the varieties shown above. When these pods are broken, a strong string can easily be removed from the sutures at both maturity stages. (a, g), cv 'Pismo'; (b, h), cv 'Prevail'; (c, i), cv 'Hystyle'; (d, j), cv 'Galveston'; (e, k), cv 'BBL156'; (f, l), cv 'Huntington'; (m), stringless wax bean cv 'Midas', which has been frequently included in studies of pod fibre traits (Koinange *et al.*, 1996; Gioia *et al.*, 2013; Murgia *et al.*, 2017; Rau *et al.*, 2019; Di Vittori *et al.*, 2021; Parker *et al.*, 2021).

1904; Prakken, 1934; Hagerty *et al.*, 2016). Gioia *et al.* (2013) described a *P. vulgaris* orthologue of *IND*, known as *PvIND*, which mapped 7.8 cM from *St*. However, the occasional 'recombination' between *St* and *PvIND*, as well as a lack of explanatory genetic variation at the locus and 1 kb of promoter, suggested that this specific gene sequence might not be responsible for control of pod strings. Hagerty *et al.* (2016) subsequently identified flanking markers for *St* spanning *c.* 500 kbp, from 43 984 700 to 44 472 300 (*P. vulgaris* reference genome G19833 v.2.1, https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1). This region is therefore of major interest for the control of pod suture string formation in common bean.

Here, we investigate the genetic and transcriptional control of pod suture string development in common bean. To this end, we compare transcriptional patterns of diverse genotypes, including stringless/revertant pairs, identify anatomical effects of differentially expressed candidate genes, screen the methylation state of select regions of interest and explore sequence variation in candidate regions across *P. vulgaris*.

Materials and Methods

Plant materials

RNA-seq was conducted on four genotypes (G12873, ICA Bunsi, SXB 405, and Midas), which span the full range of pod fibre and shattering properties found in common bean (Table S1). For RT-qPCR, anatomical studies and bisulfite sequencing, eight pairs of stringy revertant pods and nonrevertant stringless controls were collected at Syngenta facilities in 2017 (Table S2). These were subsequently planted in the glasshouse in Davis, California for pod sampling. All seeds bred true to type for pod and other traits, without further reversion or instability in subsequent generations. One nonrevertant snap type of the accession Hystyle was used for sequencing the full region between the flanking markers of the pod string locus of Pv02.

A set of 100 diverse *Phaseolus vulgaris* L. accessions were acquired from NPGS, University of California (UC) Davis, and Oregon State University, and were glasshouse grown in Davis, California. DNA was extracted from leaf material or glasshouse-grown seeds using a modified cetyltrimethylammonium bromide (CTAB) method (based on Allen *et al.*, 2006). Mature full-sized pods of each type were allowed to dry and were broken by hand to analyse pod string phenotype on a scale of 0 (no removable string) to 10 (string readily removable).

Pod string candidate and control genes

Candidate genes for RT-qPCR and qPCR included all genes between the pod string flanking markers which either (1) were closely homologous to known modulators of pod (silique) development in Arabidopsis or (2) belonged to the NAC or MYB transcription factor families, which are known to have a major effect on lignified secondary cell wall development. To identify Arabidopsis homologues, the amino acid sequences of the pod patterning genes *IND*, *SHP1/2*, *ALC*, *RPL* and *FUL* were downloaded from Phytozome 12 and compared by Blast to the common bean proteome. The top 10 most similar genes were screened to identify any homologues that might exist between the flanking markers for pod strings. The 56 gene models between *St* flanking markers (Phvul.002G269200 to Phvul.002G274700; Hagerty *et al.*, 2016) were accessed via the Legume Information System (Dash *et al.*, 2015) (https://legumeinfo.org/home). Gene family and Gene Ontology (GO) term data were downloaded through PMine (Goodstein *et al.*, 2012). Gene expression data for all 56 genes were retrieved through PhytoMine to verify expression of candidate gene classes were expressed in pods. In total, these systematic screens identified four candidate genes: two NAC family transcription factors ('*NAC 1*', Phvul.002G271700; and '*NAC 2*', Phvul.002G273100), one MYB family transcription factor ('*MYB*', Phvul.002G269900), and an atypical bHLH transcription factor closely related to *IND* (*PvIND*, Phvul.002G271000). These were each expressed in pod tissue and were therefore considered candidate genes for RNA-seq and RT-qPCR. For RT-qPCR, *Act11* (Phvul.008G011000) and *Ukn1* (Phvul.011G023200) were used as stably expressed reference gene controls based on Borges *et al.* (2012) and O'Rourke *et al.* (2014). Amino acid sequences of *IND* homologues in *P. vulgaris* and Arabidopsis were compared using a fast minimum evolution tree based on the Grishin protein matrix on the NCBI website (blast.ncbi.nlm.nih.gov).

RNA-seq

Gene annotation and GO data of the *Phaseolus vulgaris* v.2.1 genome were downloaded from Phytozome (http://phytozome.jgi.doe.gov/). TopGO v.2.26.0 was used to determine GO term enrichments. Transcriptomes were characterised using three pod replicates with pairwise comparisons between each of three different stages (Fernández *et al.*, 1983): pod formation (R7), pod fill (R8) and pod maturation (R9). Formation of lignified secondary cell walls begins in the R8 stage, soon after pods had reached full length, and peaks in the R9 stage. RNA sequencing libraries were prepared following the Illumina® TruSeq® Stranded Total RNA Sample Preparation kit instructions. In total, 36 TruSeq libraries were sequenced using the Illumina NextSeq500 platform in the 1 × 75 single-end mode, obtaining an average of 11.4 million raw reads per library. Raw RNA-seq data were processed using FastQC v.0.11.2. Sequences with QC below 20 were trimmed using Trimmomatic (Bolger *et al.*, 2014) and adapters and overrepresented sequences were eliminated, obtaining *c.* 10.8 million high-quality reads per library. The resulting reads of the good quality libraries were mapped with Kallisto to *P. vulgaris* v.2.1 from Phytozome. DEGs were determined using edgeR (Robinson *et al.*, 2010) (v.3.16.5) in R Core Team (2018; v.3.3.2) using a two-fold change threshold and a false-discovery rate (FDR) < 0.05.

To identify transcripts that were differentially expressed between stringy and nonstringy accessions, the expression patterns of ICA Bunsi, SXB 405 and G12873 were compared against those of Midas to establish GO functional category enrichments. Expression of candidate loci was also compared by analysis of variance (ANOVA) of the linear model using genotype and maturity stage as fixed variables.

RT-qPCR

Pods were harvested for RT-qPCR at 5 and 21 d after flowering (DAF). All samples from each developmental stage were harvested at the same time and date. Whole pods were harvested at 5 DAF; at 21 DAF, pod cross-sections 1-cm thick were sampled at the first seed, with seed material immediately removed. Samples were flash frozen in liquid N_2 then kept frozen at −80°C. For RNA extraction, pods were ground in liquid N_2 using a mortar and pestle, and extracted using the RNeasy Plant Mini kit (Qiagen). Concentration and quality were checked by NanoDrop and bleach gel. Here, 500 µg of RNA were used per sample for cDNA synthesis. Reverse transcription was conducted with the SuperScript IV VILO Master Mix using the ezDNase kit (Invitrogen). Transcript identifiers were used to generate primers for 70–150-bp amplicons using NCBI Primer BLAST (Table S4; please refer to later paragraphs), with specificity checking enabled to avoid amplifying nontarget transcripts. Intron-spanning primers were used for multiexon genes ('*NAC 1*', '*NAC 2*', *Act11*, *Ukn1*). This was not possible for the single-exon genes '*MYB*' and *PvIND*, so a control not treated with reverse transcriptase was included for these. qPCR primer efficiency was checked on pooled cDNA from each pod harvest date. All primers performed with an efficiency > 1.00. C_T values of reference genes were subtracted from those of candidate genes to generate ΔC_T data. These values are logarithmically related to RNA quantity, so they were converted to $2^{-\Delta C_T}$ values. The mean, standard deviation, and standard errors of $2^{-\Delta C_T}$ data for each phenotypic class was calculated for each gene comparison. Expression differences between the stringy and nonstringy groups were then compared by *t*-test.

Microscopy

Full-sized green pods (stage R8) were harvested from the stringless/revertant pairs used for RT-qPCR, and 100 µm transverse sections were made using a Vibratome. These were treated with Auramine O (0.01%) and Calcofluor (0.007%) for 20 min (Lo *et al.*, 2021). Auramine O stains hydrophobic compounds, including lignin, in green, whereas Calcofluor stains cellulose in blue (Fig. 2). Fluorescence was visualised using an Olympus BH2-RFL microscope (Waltham, MA, USA) with the ultraviolet filter set (UG-1 and DM-400 + L-420).

Conserved *PvIND* promoter motifs

Upstream regulatory sequences of *PvIND* orthologues were compared with identified conserved elements with a potential role in transcriptional regulation. The *PvIND* amino acid sequence was retrieved from PHYTOZOME 13 (phytozome-next.jgi.doe.gov). and PBLAST was used to find highly similar proteins in Arabidopsis and all legume species with available proteomes. DNA sequences upstream of these genes were downloaded and aligned with 2500 bp of the comparable *PvIND* region using NCBI BLASTN. In total, sequences upstream of 21 gene models representing 16 species were compared. The *PvIND* promoter was screened for enhancer activity using ENHANCERPRED (Jia & He, 2016). *PvIND* homologues in common bean and Arabidopsis were also compared using the NCBI BLAST tree viewer to compare relationships between these related proteins.

Bisulfite sequencing

Bisulfite sequencing was conducted to analyse DNA methylation patterns potentially related to reversible phenotypic change.

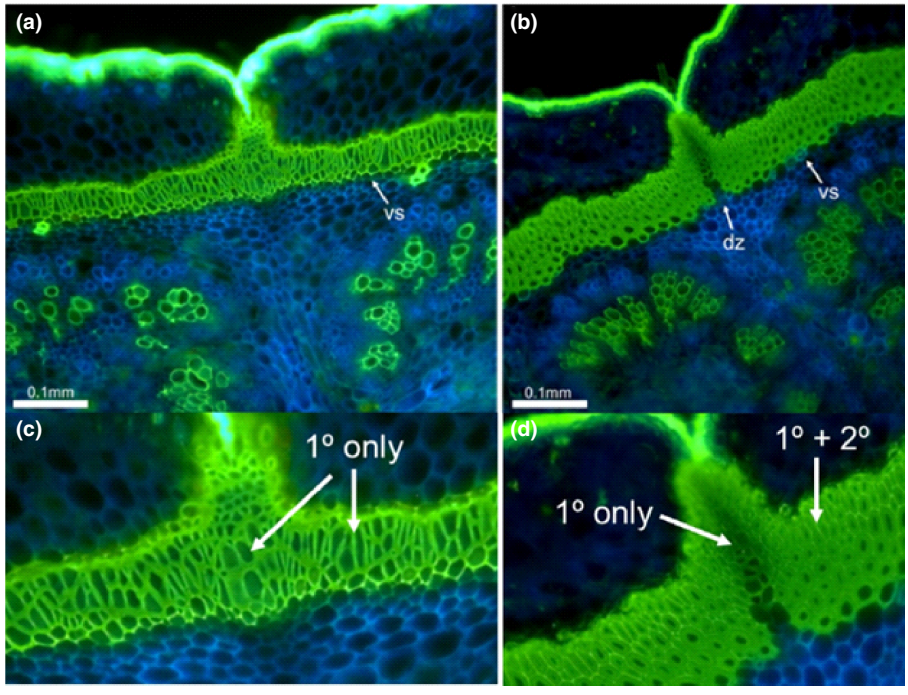

Fig. 2 Plants of common bean (*Phaseolus vulgaris*) with elevated *PvIND* expression levels produce an over-abundance of weak 'dehiscence-zone' (dz) cells throughout the 'vascular bundle sheath' (vs) fibre layer, leading to loss of pod strings. Anatomical comparison of (a) nonstringy lines with elevated *PvIND* expression and (b) stringy revertant pods with lower *PvIND* expression. In (a) nonstringy accessions, vs cells are weak, with little to no secondary cell wall thickening and just 1–3 lignified cell layers, whereas (b) stringy revertants have strong secondary thickening and 3–6 fibre cell layers, except in the weak central dz along which dehiscence occurs in susceptible varieties. In nonstringy types with *PvIND* overexpression, overspecification of these weak dz-like cells occurs throughout the vs, leading to the lack of pod suture string. Samples (a) 'BBL156' (Nampa-20, nonrevertant) and (b) 'Hystyle' (Nampa-3, revertant) are shown. Photographs (c, d) are more highly magnified images of photographs (a, b), with primary (1°) and secondary (2°) cell walls labelled.

These areas included three conserved motifs upstream of *PvIND*, the area surrounding the *PvIND* start site, a region within the *PvIND* gene body, and part of the 3′UTR (Table S5, please refer to later paragraphs). Genomic DNA was extracted using a modified CTAB protocol. Bisulfite treatment was conducted with the EZ DNA Methylation-Lightning Kit. Primers were designed using Zymo Bisulfite Primer Seeker 12S (Table S3), polymerase chain reaction (PCR) products were checked on a gel, cleaned using a QIAquick PCR Purification Kit, and genotyped using Sanger sequencing at the UC Davis DNA sequencing facility. FASTQ reads were converted to FASTA and aligned using NCBI BLAST to identify sequence variation. Methylation status of all cytosines was compared between nonstringy and revertant types.

Sequencing 500 kb surrounding *PvIND*

To systematically search for variation that might cause differential expression of *PvIND*, the entire region between the *St* flanking markers (Hagerty *et al.*, 2016) was sequenced and scaffolded in the stringless cultivar 'Hystyle' (sample Nampa-5) by Corteva Agrisciences (Johnston, IA, USA). DLS BioNano mapping was combined with HiFi PacBio sequencing to create a single hybrid scaffold spanning the region. The scaffold was then aligned with the seven *Phaseolus* reference genomes of the genus *Phaseolus* on PHYTOZOME to identify unique sequence features. Structural variation was evaluated using a panel of 100 stringy and nonstringy beans. PCR primers were developed using NCBI Primer BLAST (Table S6, please refer to later paragraphs) to span (1) the retrotransposon site and (2) the tandem duplication splice site, which also included the retrotransposon insertion. PCR was conducted using ExTaq (TaKaRa Bio, Kusatsu, Japan) and amplicons were visualised on a 1.4% agarose gel.

Reanalysis of Illumina whole-genome sequencing (WGS) data

Publicly available Illumina WGS reads for nonstringy 'Midas', the Middle American stringy genotype 'VAX3', and the Andean stringy accession G4627 were downloaded and reanalysed to assess if the duplication/insertion events identified for 'Hystyle' were shared by other nonstringy genotypes. The reference genome (G19833 v.2.1; PHYTOZOME 13: https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1) was augmented including the retrotransposon sequence. Reads were mapped against this augmented reference genome using BOWTIE2 (Langmead & Salzberg, 2012). Alignments were sorted by reference coordinates using PICARD (https://broadinstitute.github.io/picard/) and visualised with the Integrative Genomics Viewer (https://software.broadinstitute.org/software/igv/) to identify reads of the accession Midas that spanned the retrotransposon boundary near *PvIND*.

Results

Transcriptional characterisation

RNA-seq successfully characterised pod transcriptomes of four varieties across three pod development stages (Fig. S2; Tables S1, S2). Differentially expressed genes (DEGs) between stringy and nonstringy accessions were considered to identify GO enrichments for pod string formation (Fig. S3). In total, 237 genes were obtained from the R8 stage (Pod fill) (Fernández *et al.*, 1983) vs R7 stage (Initial pod development) comparison, 409 DEGs from the R9 (Pod maturation) vs R8 comparison (Pod fill), and 230 DEGs for the R9 vs R7 comparison. Patterns of gene expression moving into the R9 stage, when lignification is most pronounced, were particularly unique in the stringless accession Midas, which had the greatest number of unique genes with differential expression patterns at the R8 vs R9 comparison (2418 genes) and the R7 vs R9 comparison (3381 genes, Fig. S3). By contrast, at the R7 vs R8 maturity stage, before major lignification, Midas had the second most genes with altered expression (1768 genes), behind SXB 405 (2421 genes; Fig. S3). Functional enrichment analysis of GO terms of stringy accessions (Fig. S4) showed the activity of phosphatidylinositol phosphate kinases, *N*-acetyltransferases, phosphotransferases and the union of transition metal ions as the most representative between stages R8 and R7. GO terms related to transferase, phosphotransferase and carbohydrate activity were mainly registered in the comparison R9 vs R8 in both gene sets. The enriched categories for R9 vs R7 included tetrapyrrole binding, oxidation–reduction activity, actin binding, binding to rRNA and nucleotidyl-transferase activity. Among all stringy accessions, membrane coat categories are included as enriched between R8 and R7, while the R9 vs R8 comparison did not present changes in these categories. Photosynthetic components, thylakoids, and membrane proteins were enriched in the comparison of stages R9 and R7 between stringy and stringless accessions.

Among the four Pv02 candidate genes, *PvIND* (Phvul.002G271000.1; Figs S5, S6) was significantly more strongly expressed in the stringless accession than in the three stringy accessions ($P = 6.8 \times 10^{-6}$, ANOVA of linear model). The results were independently significant at each maturity stage (R7: $P = 1.0 \times 10^{-4}$; R8: $P = 4.3 \times 10^{-3}$; R9: $P = 1.2 \times 10^{-5}$; *t*-test). *PvIND* expression was nonoverlapping between stringy and stringless varieties within each maturity stage. Of the other candidate genes, the NAC transcription factor Phvul.002G273100.1 was differentially expressed at the latest maturity stage (R9), with a 3.6-fold higher expression in Midas than in the stringy varieties on average (*P*-value = 0.01, *t*-test). Phvul.002G273100.1 expression at the other maturity stages individually and cumulatively across all stages was insignificantly different ($P > 0.05$, *t*-test). Similarly, the two other Pv02 candidate genes were insignificantly differentially expressed between phenotypic categories at any maturity stage.

PvIND expression is predictive of string formation

RT-qPCR of nonstringy/stringy revertant pairs (Tables S3, S4) determined that *PvIND* expression was highly significantly correlated with pod strings across both sampling time points (Fig. 3). At 5 DAF, *PvIND* was *c.* five-fold more expressed in stringless accessions than in stringy revertants, whether the control gene was *Act11* (4.9-fold difference, $P = 8.5 \times 10^{-5}$) or *Ukn1* (5.3-

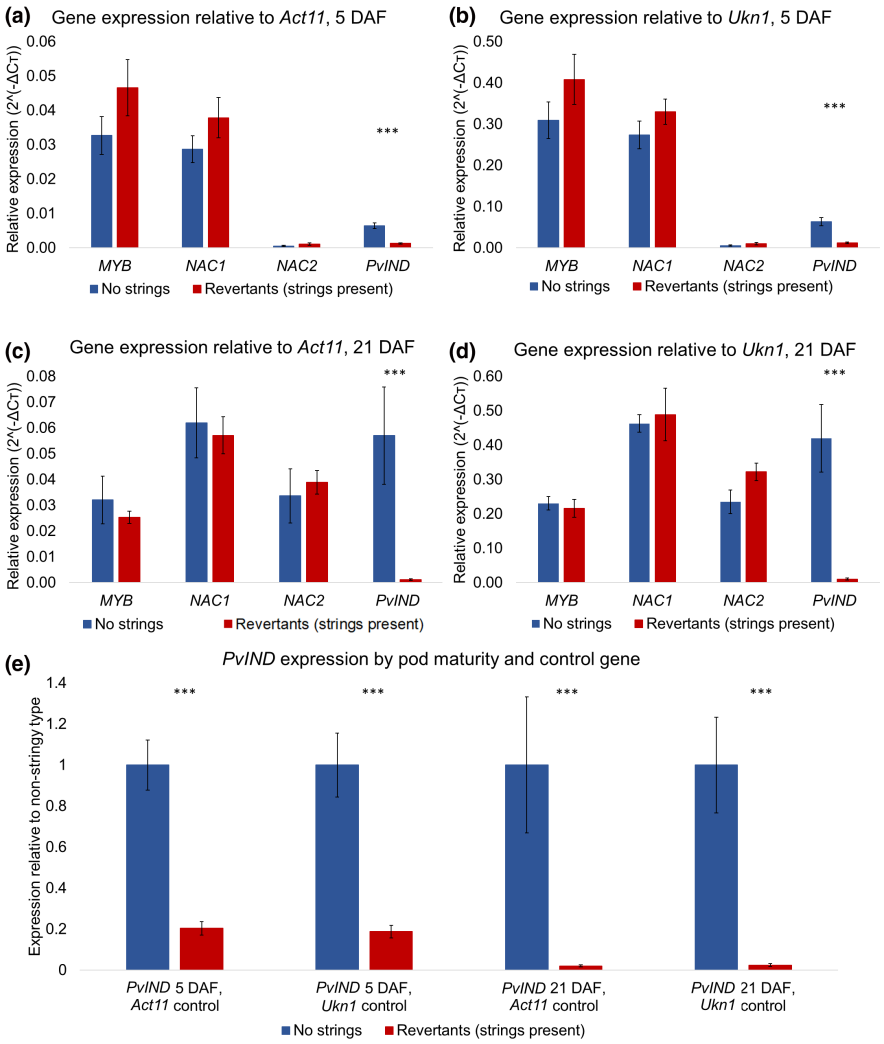

Fig. 3 Significant differences in *PvIND* expression exist between cultivars of common bean (*Phaseolus vulgaris*) with and without pod strings. At 5 d after flowering (DAF), a highly significant difference exists in *PvIND* expression whether (a) *Act11* or (b) *Ukn1* is used as a reference. No significant expression difference exists for other candidate genes. (c, d) At 21 DAF, the difference in *PvIND* expression has increased relative to each reference gene. No significant difference exists in the expression of other genes. (e) *PvIND* expression from previous panels re-scaled, relative to expression in stringy revertants. The difference in *PvIND* expression is greater later in pod maturity and shows similar patterns regardless of the reference gene. Three asterisks indicate $P < 0.001$, no asterisks indicate $P > 0.05$. Error bars represent standard error of the mean of six replicates of each phenotypic class. *PvIND*: Phvul.002G271000; *NAC 1*: Phvul.002G271700; *NAC 2*: Phvul.002G273100; *MYB*: Phvul.002G269900; *Act11*: Phvul.008G011000; *Ukn1*, Phvul.011G023200. Significance of differences was determined by *t*-tests.

fold difference, $P = 4.5 \times 10^{-4}$). By 21 DAF, *PvIND* expression was 48-fold higher in stringless types when *Act11* was used as a reference ($P = 2.5 \times 10^{-6}$) and 41-fold greater when *Ukn1* was used ($P = 1.3 \times 10^{-6}$). By contrast, the three candidate genes of the NAC and MYB families were not significantly differentially expressed between phenotypic classes at either time point, regardless of which reference gene was used as a control. The minimum *P*-value of any of these comparisons was 0.06 ('*NAC 2*' vs *Ukn1*, 21 DAF).

PvIND expression associated with changes in cell identity

Major differences in secondary cell wall development distinguished stringless and stringy revertant types (Figs 2, S7). Revertant vascular sheaths were nearly identical to those of wild or dry beans (Fig. 2b,d; Prakken, 1934; Parker *et al.*, 2021). In revertant pods, the vascular bundle sheath is primarily composed of 3–6 layers of thickly lignified fibre cells. At the centre of the sheath is a narrow strip of weakly lignified cells approximately two cells wide (Fig. 2b,d), known as the dehiscence zone. By contrast, stringless varieties produce weak dehiscence-zone-like cell layers with only primary cell walls throughout the vascular sheath (Fig. 2a,c). These are lignified, but lack secondary cell wall thickening and have a small fraction of the cross-sectional cell wall surface area of fibre cells. The total area of the vascular sheath is also reduced, with only 1–3 cell layers in nonstringy varieties (Fig. 2a, c). Only one type, the Prevail revertant, showed strong pod wall fibre along with pod strings. The stringless form of Prevail also shows subtle wall fibre deposition (Fig. S7).

Conservation in *PvIND* homologue promoters

Three main regions of high similarity were identified among promoters of *PvIND* homologues. These spanned from *c.* 1514–1643 bp, 906–1048 bp and 293–430 bp before the *PvIND* transcription start site, and were enumerated as motifs 1, 2 and 3, respectively (Fig. S8). While motifs 1 and 3 were conserved among all legumes analysed, motif 2 was conserved only among the Phaseoleae. At the middle of these were core sequences with very high conservation. In motif 1, the sequence CCCTAGGAT TTCAGTGC was identified without substitution or gaps for 17 of 21 gene models, while in the other four there were no more than

two single nucleotide polymorphisms (SNPs). Motif 3 included the sequence (ATGCTTTTTGCAGTSASW(C)$_{0\text{-}1}$CCCCTTTCAG TAAAAAC) across all species with above-ground pods. The end of this conserved sequence and 50 bp immediately following it were predicted to have enhancer activity using EnhancerPred, while this was not the case anywhere else in the 2.5-kb region upstream of *PvIND*. The comparison of *IND* homologues in common bean and Arabidopsis indicated that *PvIND* most closely clusters with *AtIND* rather than other similar proteins (Fig. S6).

DNA methylation of *PvIND* promoter is not predictive of pod strings

Bisulfite sequencing (Table S5) returned reads at three conserved motifs upstream of *PvIND*, 246 bp flanking the transcriptional start site, 270 bp in the gene body, and 48 bp of 3′UTR. Relative to the transcription start site, these ranged from −1652 to −1474 bp, −1136 to −906 bp, −489 to −215 bp, −186 to 102 bp, 406 to 676 bp and 1187 to 1234 bp. No surveyed positions were predictive of pod phenotype (Fig. S8b). Methylated cytosines existed in motif 1, with two methylated residues; as well as the gene body, with two identical methylated cytosines in both stringy and nonstringy variants of Hystyle. Across surveyed cytosines, methylation patterns tended to be highly consistent among accessions, and at no position did methylation status predict pod string phenotype.

PvIND duplication and retrotransposon insertion

HiFi/BioNano mapping identified several unique features of the nonstringy accession 'Hystyle' in the region surrounding *PvIND*, including a tandem duplication of *PvIND* and an insertion of a retrotransposon between the two copies (Fig. 4). PacBio HiFi sequencing averaged a depth of 45×, with median read lengths of 17.3 kb. These sequences were combined with the Bionano mapping results to create a single scaffold containing the full 500 kb region of Hystyle between the *PvIND* flanking markers identified by Hagerty *et al.* (2016). All seven available *Phaseolus* reference genomes are of accessions with stringy pods, with a single copy of *PvIND* and no retrotransposon in the surrounding region. By contrast, Hystyle contains a 12-kb tandem duplication of *PvIND* and its upstream promoter. The duplicated copies (*PvINDa* and *PvINDb*) are identical in putative transcribed sequence. Furthermore, the second tandem repeat included a *Ty1-copia* family retrotransposon insertion *c.* 10 kb upstream of *PvINDb* and 400 bp downstream of *PvINDa*. The retrotransposon is 2760 bp in length and includes insertion site repeats and terminal inverted repeats typical of the transposable element family. Finally, CT repeats of 278 bp and 170 bp immediately follow *PvINDa* and *PvINDb*, respectively, whereas these repeats are no longer than 140 bp in any other *Phaseolus* reference genome. Reanalysis of previously sequenced Illumina data for the nonstringy accession 'Midas' (Lobaton *et al.*, 2018) also supports the retrotransposon insertion for this accession (Fig. S9). The *PvIND* coding DNA sequence of the stringless accession Hystyle (of Andean origin) was identical between both tandem repeats, and was also identical to the reference genome of the Andean accession G19833. By contrast, there were a small number of polymorphisms between the Hystyle *PvIND* sequence and the sequence of Middle American accessions 5–593 (843/849 bp identical), Labor Ovalle (842/849 bp identical) and UI 111 (841/849 bp identical).

Sequence variation associated with pod strings in revertants and other accessions

Among 100 diverse accessions, tandem duplication of *PvIND* was always associated with retrotransposon insertion and vice versa (Fig. 5). Tandem duplication and retrotransposon insertion were extremely predictive of pod string phenotype, without

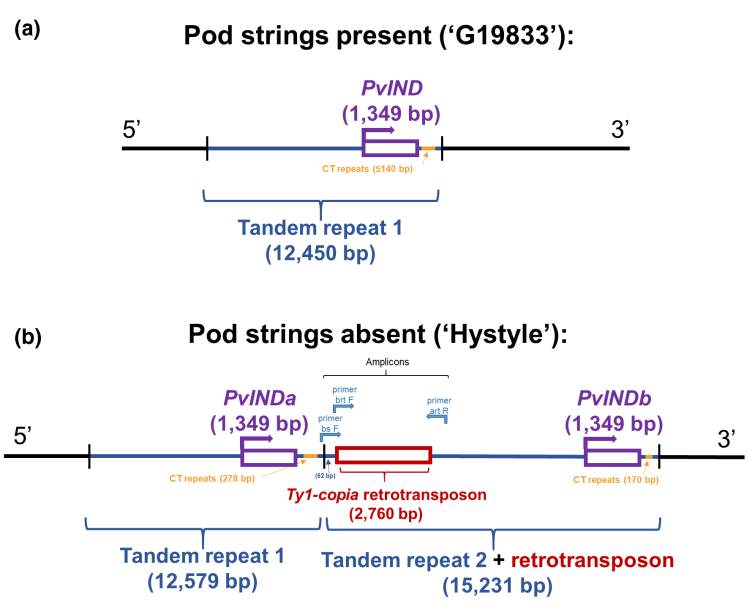

Fig. 4 *PvIND* sequence structure differences between stringy and nonstringy accessions of common bean (*Phaseolus vulgaris*). (a) Stringy accessions such as G19833 have a single copy of *PvIND* without retrotransposon. By contrast, (b) nonstringy accessions such as Hystyle show a tandem duplication of *PvIND*, with a retrotransposon inserted at the beginning of the duplicated region. These sequence features may cause the *PvIND* overexpression associated with loss of pod strings. Primer binding sites and amplicons used for evaluating tandem duplication and retrotransposon presence are shown.

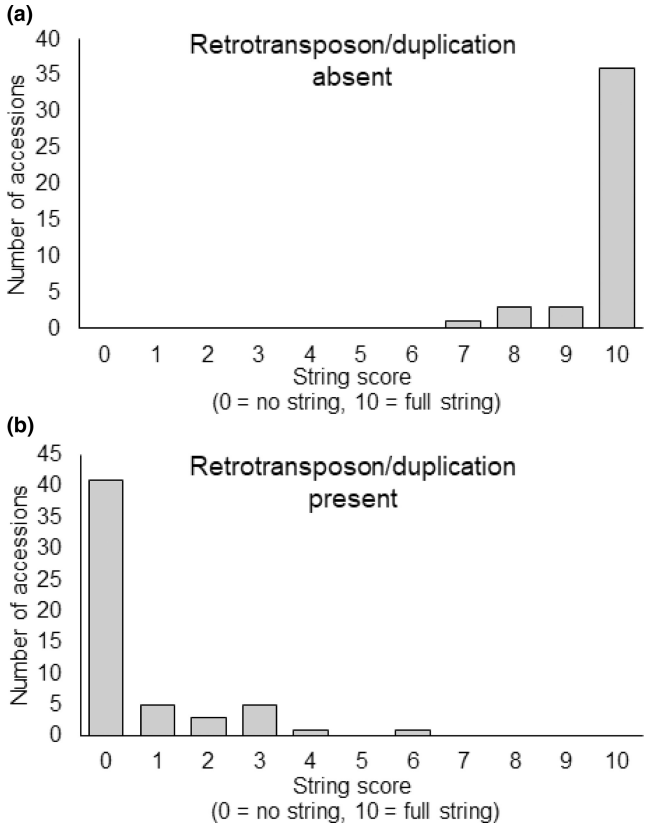

Fig. 5 Among 100 diverse accessions of common bean (*Phaseolus vulgaris*), pod string score is strongly related to *Ty1-copia* retrotransposon presence near *PvIND* and duplication status of the gene. (a) Accessions lacking these sequence features tend to produce strong suture strings, whereas (b) accessions with the retrotransposon insertion and gene duplication produce little to no pod strings.

phenotypic overlap between genotypic categories (Figs 5, S10; Tables S6, S7). This determined that previously reported 'recombination' between *PvIND* and pod strings (Gioia *et al.*, 2013) may have been due to errors in phenotyping based on immature pod (instead of mature pod) observations and the assumption of co-inheritance of pod suture and wall fibre (Koinange *et al.*, 1996), as our results clearly matched the sequence in the *PvIND* region and the string phenotype in analysed RILs (Table S7). Intriguingly, stringy revertants did not include the tandem duplication and retrotransposon insertion, unlike their isogenic non-stringy lines, indicating that the reversion process involves loss of the second *PvIND* tandem repeat, including the retrotransposon insertion.

Discussion

Pod strings are controlled by *PvIND* and surrounding sequence

Our results demonstrate that the absence of pod strings in common bean (Fig. 1) is uniformly associated with a tandem duplication of *PvIND* and a retrotransposon insertion between the two gene copies (Figs 4, 5, S9). Accessions with these sequence features consistently express *PvIND* transcripts at abundances at least 40-fold higher than isogenic revertant lines (Fig. 3), which lack these features. *PvIND* expression level between stringless and revertant lines is related to overspecification of weak pod dehiscence-zone tissue throughout the vascular bundle sheath, in cell types that would alternatively produce strong secondary cell wall deposition (Figs 2, S7). Phenotypic reversion can be explained by the loss of the second *PvIND* tandem repeat, which is absent in revertant lines. The *PvIND* gene duplication leading to increased expression is by no means unique in *P. vulgaris*. For example, genes coding for seed proteins, which require intense expression during the pod fill phase (R8 stage) have been duplicated in tandem to form a complex locus, as illustrated by phaseolin seed protein on chromosome Pv07 (Slightom *et al.*, 1985; Llaca & Gepts, 1996) and the APA seed protein locus on chromosome Pv06 (Kami *et al.*, 2006).

Our results in common bean show that a major gain in *PvIND* expression has occurred in the development of stringless forms, which corresponds with the dominant nature of the apomorphic allele. Previous studies had mapped variation for pod strings to chromosome Pv02 (Koinange *et al.*, 1996), in the vicinity of *PvIND* (Gioia *et al.*, 2013), and specifically to a 500-kb region surrounding *PvIND* (Hagerty *et al.*, 2016). Our RNA-seq results, which included comparisons of genes in this region and throughout the transcriptome, show that Midas has by far the most unique pattern of gene expression in pods among the four analysed accessions. This is consistent with the idea that loss of pod strings led to more extensive pod remodelling than the initial reduction in of pod shattering, a core element of legume domestication (Parker *et al.*, 2020b). Our RNA-seq results parallel those of Di Vittori *et al.* (2021), who also identified *PvIND* among genes differentially expressed between Midas and G12873. The distant relationship between these accessions, which descend from distinct gene pools (Andean vs Middle American, respectively; Parker & Gepts, 2021), necessitated an analysis of gene expression in a controlled genetic background. Our RT-qPCR results conducted in six isogenic backgrounds demonstrate unequivocally that differential *PvIND* expression is not only correlated with string deposition but is also qualitatively predictive of string formation among all samples and at both sampled time points (Fig. 3). This is compelling evidence that *PvIND* expression regulates pod string formation in common bean.

We next analysed the anatomical effects of *St* reversion in the isogenic stringless/revertant pairs. In all cases, stringless accessions demonstrated an expansion of the weak dehiscence-zone cells throughout the vascular bundle sheath, the fibrous structure that is removed as a pod string (Fig. 2). This result is consistent with the known role of *IND* in Arabidopsis. In Arabidopsis, *IND* regulates the development of the valve margin layer where breaking occurs, including weak cells of the separation layer (Liljegren *et al.*, 2004; Girin *et al.*, 2010). These cells are homologous to the weak dehiscence layer in common bean. Loss of pod strings could therefore be the result of ectopic *PvIND* increased expression throughout the bundle sheath, leading to overspecification of weak dehiscence-zone cells (which lack secondary cell wall

biosynthesis) throughout the bundle sheath. Methods such as RNA *in situ* hybridisation and laser capture microdissection RT-qPCR could be useful to test this possibility. Additionally, five of six revertants showed no appreciable increase in pod wall fibre deposition, indicating that the effect of *St* is typically specific to the string region in most genetic backgrounds. The identification of wall fibre deposition in the 'Prevail' revertant indicates that *PvIND* may prevent pod wall fibre development in some environments and/or genetic backgrounds. Several regions upstream of *PvIND* have been conserved base-by-base over 10s of millions of years of legume evolution. This indicates that they play a role critical to the conservation of pod shattering and, therefore, the survival of many wild legumes. Bisulfite sequencing of these and other regions in and around *PvIND* found no patterns of DNA methylation predictive of pod string formation, ruling out their role in the reversion and regulation of string formation.

The *PvIND* tandem duplication and retrotransposon insertion are both logical drivers of *PvIND* increased expression in non-stringy types. Gene duplication frequently leads to enhanced expression due to increased gene copy number (Gemayel *et al.*, 2012), while transposable elements typically contain transcription-enhancing motifs with long-range effects (Lisch, 2013; Hirsch & Springer, 2017). Both gene duplications and transposable element insertions frequently lead also to ectopic gene expression, which could explain the gain of dehiscence-zone cell identity throughout vascular bundle sheath layers in stringless varieties. The existence of the retrotransposon in only the second tandem repeat indicates that its insertion occurred after gene duplication. Intermediate forms with *PvIND* duplication but without the retrotransposon would be useful to separate the role of each sequence feature, although these types have not yet been identified.

Alternatively, retrotransposon-mediated duplication of coding DNA sequences is known to occur via readthrough transcription (Xiao *et al.*, 2008; Jiang *et al.*, 2009), but this process involves semirandom reinsertion back into the genome, making tandem duplication of a region by this mechanism unlikely. Furthermore, duplication by readthrough transcription involves duplication of the retrotransposon itself, while the duplicated region at *PvIND* includes only a single retrotransposon. Together, these factors make readthrough transcription an unlikely mechanism for producing the sequence features associated with stringlessness. The identical coding DNA sequence of *PvIND* in Hystyle and the Andean landrace G19833 (both originating in the Andean gene pool), in contrast with Middle American types, indicates that the stringless mutation originated in an Andean genetic background. From there, it was subsequently used broadly in breeding programmes with diverse Andean and Middle American germplasm (Myers & Davis, 2002; Lobaton *et al.*, 2018; Wallace *et al.*, 2018).

The dominance of the stringless *St* mutation fits well with *PvIND*'s increased gene expression, duplication and retrotransposon insertion. By contrast, during domestication and breeding of diploid autogamous species, most new characteristic states are mediated by recessive alleles based on loss-of-function mutations. For example, of 13 surveyed mutations leading to reduced seed shattering among Near Eastern domesticates, all but one are recessive (Ladizinsky, 1985). Similarly, among all alleles reducing pod shattering in the 40 domesticated legume species, *PvIND* is unique in its genetic dominance (Parker *et al.*, 2020b). In common bean, mutations for other key domestication traits such as determinacy, photoperiod insensitivity, loss of seed dormancy, white-seededness, and reduction in pod shattering were all mediated by loss-of-function mutations (Parker & Gepts, 2021). In pea (*Pisum sativum* L.), the only other legume species with a loss of pod strings, the stringless allele is recessive (McGee & Baggett, 1992). This contrast with common bean indicates that a different genetic mechanism may be involved between the two legume species.

PvIND tandem duplication as the source of instability, and potential solutions

The uniform loss of the *PvIND*-associated tandem duplication and retrotransposon across all eight stringy revertant lines is an interesting find. Although excision of DNA transposons occurs regularly, this process is not known to occur among retrotransposons, which replicate via reverse transcription. Instead, loss of one tandem repeat may be responsible for reversion. Genotypic and the associated phenotypic change patterns similar to those of *PvIND* and pod strings have been identified in other systems, such as the *Bar* gene of *Drosophila* (Sturtevant & Morgan, 1923; Sturtevant, 1925; Wolfner & Miller, 2016). Tandem duplication of *Bar* leads to dominant overexpression of the gene and a 'bar'-type eye shape. This gene frequently reverts to the wild-type due to excision of a single copy, at a ratio of *c.* 1 per 1000 individuals (Wolfner & Miller, 2016). These patterns closely parallel those of the *PvIND* locus and pod string regulation. This pattern of gene duplication, gain of function and reversion due to single copy excision has been found in numerous other systems, such as maize (Veit *et al.*, 1990, Du *et al.*, 2021), *Ipomoea* (Park *et al.*, 2004), mice (Gondo *et al.*, 1993), other genes in *Drosophila* (Ishimaru *et al.*, 1995), as well as in other species. These reversions can be the result of unequal crossing over, or, by contrast, can be mediated by looping out of one repeat based on specific microhomologies (Karess, 1982; Ishimaru *et al.*, 1995; Park *et al.*, 2004). Pod string reversion may be the first recorded example of this pattern in an agriculturally important context.

All of the revertants identified and genotyped in this study lacked the second *PvIND* tandem repeat (*PvINDb*), including the retrotransposon, while the first tandem repeat (*PvINDa*) remained intact (Figs 4, S9). This indicates that types lacking the first tandem repeat may be phenotypically similar to those with both repeats and were therefore not identified in our revertant screens. If so, the second tandem repeat alone may be sufficient to cause the absence of pod strings, but could be resistant to reversion due to the lack of tandem duplication. The identification of these types could be of major commercial value in developing snap bean varieties resistant to phenotypic reversion. Ultimately, the sequence features identified in this study will be important to improve the stability of snap beans for commercial production and shed light on the basis of reversible genetic variation.

Acknowledgements

We thank two anonymous reviewers for their excellent comments and suggestions. This work was supported by funding from Clif Bar/Seed Matters, Lundberg Family Farms, Kirkhouse Trust, UC Davis Henry Jastro Graduate Research Scholarships, the Universidad de los Andes FAPA initiative led by the Vice-presidency of Research and Knowledge Creation, CAPES Postdoctoral Program/Process no. 88881.170593/2018-01, UC MEXUS grant CN-17-46 and USDA-NIFA Regional Hatch project W3150. Dan Wahlquist and Syngenta generously allowed us to collect stringless/revertant types at their facilities.

Competing interests

None declared.

Author contributions

Conceptualisation and interpretation: PG, AH-E, TP, EA, JRM; Molecular biology and microscopy: TP, JC, LLdS, SK, SL, SN, TOF, JJ; Bioinformatics: VL, JD, TP, JC, AH-E; Funding acquisition: PG, AH-E, TP; Writing – original draft: TP, JC, PG. All authors revised and approved the final manuscript.

ORCID

Esther Arunga https://orcid.org/0000-0001-5265-9075
Jose Cetz https://orcid.org/0000-0002-1887-3737
Jorge Duitama https://orcid.org/0000-0002-9105-6266
Talissa de Oliveira Floriani https://orcid.org/0000-0003-1380-1037
Paul Gepts https://orcid.org/0000-0002-1056-4665
Alfredo Herrera-Estrella https://orcid.org/0000-0002-4589-6870
Judy Jernstedt https://orcid.org/0000-0003-0575-0938
Saarah Kuzay https://orcid.org/0000-0003-2789-7906
Victor Llaca https://orcid.org/0000-0003-4822-2924
Sassoum Lo https://orcid.org/0000-0002-0385-5691
James R. Myers https://orcid.org/0000-0003-0976-144X
Serah Njau https://orcid.org/0000-0001-8797-4622
Travis A. Parker https://orcid.org/0000-0002-1233-7829
Lorenna Lopes Sousade https://orcid.org/0000-0002-1154-5394

Data availability

RNA-seq data have been deposited in the NCBI database under Gene Expression Omnibus (GEO) accession number GSE179985. The 500-kb Hystyle sequence between the *PvIND* flanking markers has been submitted to NCBI under accession number ON164667. FASTA analysis of bisulfite-converted DNA has been deposited with DataDryad: https://datadryad.org/stash/share/H2e0nbREJ-9MfpXAorMTlb6tXKhCb8tDqUeanmuMnxY.

References

Al-Bader N. 2014. *Rogue pod traits in* Phaseolus vulgaris. MS Master's thesis. Corvallis, OR, USA: Oregon State University.

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols* **1**: 2320–2325.

Bolger AM, Lohse M, Usadel B. 2014. TRIMMOMATIC: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Borges A, Tsai SM, Caldas DGG. 2012. Validation of reference genes for RT-qPCR normalization in common bean during biotic and abiotic stresses. *Plant Cell Reports* **31**: 827–838.

Chen L, Wu F, Zhang J. 2021. NAC and MYB families and lignin biosynthesis-related members identification and expression analysis in *Melilotus albus*. *Plants* **10**: 303.

Dash S, Campbell JD, Cannon EKS, Cleary AM, Huang W, Kalberer SR, Karingula V, Rice AG, Singh J, Umale PE *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**: D1181–D1188.

Di Vittori V, Bitocchi E, Rodriguez M, Alseekh S, Bellucci E, Nanni L, Gioia T, Marzario S, Logozzo G, Rossato M *et al.* 2021. Pod indehiscence in common bean is associated with the fine regulation of *PvMYB26*. *Journal of Experimental Botany* **72**: 1617–1633.

Di Vittori V, Gioia T, Rodriguez M, Bellucci E, Bitocchi E, Nanni L, Attene G, Rau D, Papa R. 2019. Convergent evolution of the seed shattering trait. *Genes* **10**: 68.

Drijfhout E. 1978. Influence of temperature-dependent string formation in common bean (*Phaseolus vulgaris* L.). *Netherlands Journal of Agricultural Science* **26**: 99–105.

Du Y, Lunde C, Li Y, Jackson D, Hake S, Zhang Z. 2021. Gene duplication at the Fascicled ear1 locus controls the fate of inflorescence meristem cells in maize. *Proceedings of the National Academy of Sciences* **118**: e2019218118.

Emerson RA. 1904. Heredity in bean hybrids (*Phaseolus vulgaris*). In: *Agricultural Experiment Station of Nebraska, Seventeenth Annual Report.* Lincoln, NE, USA: College of Agriculture, University of Nebraska, 33–68.

Fernández F, Gepts P, López M. 1983. *Etapas de desarrollo de la planta de frijol común: guía de estudio [Stages of development of the common bean plant: study guide] Serie 04EB-09.03.* Cali, Colombia: CIAT.

Gemayel R, Cho J, Boeynaems S, Verstrepen KJ. 2012. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes* **3**: 461–480.

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**: 273–286.

Girin T, Stephenson P, Goldsack CMP, Kempin SA, Perez A, Pires N, Sparrow PA, Wood TA, Yanofsky MF, Østergaard L. 2010. Brassicaceae *INDEHISCENT* genes specify valve margin cell fate and repress replum formation. *The Plant Journal* **63**: 329–338.

Gondo Y, Gardner JM, Nakatsu Y, Durham-Pierre D, Deveau SA, Kuper C, Brilliant MH. 1993. High-frequency genetic reversion mediated by a DNA duplication: the mouse pink-eyed unstable mutation. *Proceedings of the National Academy of Sciences, USA* **90**: 297–301.

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**: D1178–D1186.

Gu Q, Ferrandiz C, Yanofsky MF, Martienssen R. 1998. The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development* **125**: 1509–1517.

Gupta K, Gupta S, Faigenboim-Doron A, Patil AS, Levy Y, Carrus SC, Hovav R. 2021. Deep transcriptomic study reveals the role of cell wall biosynthesis and organization networks in the developing shell of peanut pod. *BMC Plant Biology* **21**: 509.

Hagerty CH, Cuesta-Marcos A, Cregan P, Song Q, McClean P, Myers JR. 2016. Mapping snap bean pod and color traits, in a dry bean × snap bean recombinant inbred population. *Journal of the American Society for Horticultural Science* **141**: 131–138.

Hammer K, Khoshbakht K. 2014. A domestication assessment of the big five plant families. *Genetic Resources and Crop Evolution* **62**: 665–689.

Hirsch CD, Springer NM. 2017. Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**: 157–165.

Ishimaru S, Green MM, Saigo K. 1995. An intragenic tandem duplication of genomic DNA is responsible for the f3N mutation of Drosophila melanogaster. *Proceedings of the National Academy of Sciences, USA* **92**: 2999–3003.

Jia C, He W. 2016. EnhancerPred: a predictor for discovering enhancers based on the combination and selection of multiple features. *Scientific Reports* **6**: 38741.

Jiang N, Gao D, Xiao H, Van Der Knaap E. 2009. Genome organization of the tomato *sun* locus and characterization of the unusual retrotransposon *Rider*. *The Plant Journal* **60**: 181–193.

Kami J, Poncet V, Geffroy V, Gepts P. 2006. Development of four phylogenetically-arrayed BAC libraries and sequence of the APA locus in *Phaseolus vulgaris*. *Theoretical and Applied Genetics* **112**: 987–998.

Karess R. 1982. A small tandem duplication is responsible for the unstable white-ivory mutation in drosophila. *Cell* **30**: 63–69.

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

Ladizinsky G. 1985. Founder effect in crop-plant evolution. *Economic Botany* **39**: 191–198.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Liljegren SJ, Roeder AHK, Kempin SA, Gremski K, Østergaard L, Guimil S, Reyes DK, Yanofsky MF. 2004. Control of fruit patterning in Arabidopsis by INDEHISCENT. *Cell* **116**: 843–853.

Lisch D. 2013. How important are transposons for plant evolution? *Nature Reviews Genetics* **14**: 49–61.

Llaca V, Gepts P. 1996. Pulsed field gel electrophoresis analysis of the phaseolin locus region in *Phaseolus vulgaris*. *Genome* **39**: 722–729.

Lo S, Parker T, Muñoz-Amatriaín M, Berny-Mier Y, Teran JC, Jernstedt J, Close TJ, Gepts P. 2021. Genetic, anatomical, and environmental patterns related to pod shattering resistance in domesticated cowpea [*Vigna unguiculata* (L.) Walp]. *Journal of Experimental Botany* **72**: 6219–6229.

Lobaton J, Miller T, Gil J, Ariza D, de la Hoz J, Soler A, Beebe S, Duitama J, Gepts P, Raatz B. 2018. Re-sequencing of common bean identifies regions of inter-gene pool introgression and provides comprehensive resources for molecular breeding. *The Plant Genome* **11**: 170068.

LPWG. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny – The Legume Phylogeny Working Group (LPWG). *Taxon* **66**: 44–77.

McGee RJ, Baggett JR. 1992. Inheritance of stringless pod in *Pisum sativum* L. *Journal of the American Society for Horticultural Science* **117**: 628–632.

Miryeganeh M, Saze H. 2020. Epigenetic inheritance and plant evolution. *Population Ecology* **62**: 17–27.

Murgia M, Attene G, Rodriguez M, Bitocchi E, Bellucci E, Fois D, Nanni L, Gioia T, Albani D, Papa R *et al.* 2017. A comprehensive phenotypic investigation of the 'pod-shattering syndrome' in common bean. *Frontiers in Plant Science* **8**: 251.

Myers JR, Davis JW. 2002. Molecular phylogenetics of snap beans. *Annual Report of the Bean Improvement Cooperative* **45**: 16–17.

Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M. 2015. NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Frontiers in Plant Science* **6**: 288.

Ohtani M, Demura T. 2019. The quest for transcriptional hubs of lignin biosynthesis: beyond the NAC-MYB-gene regulatory network model. *Current Opinion in Biotechnology* **56**: 82–87.

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.

Park K-I, Choi J-D, Hoshino A, Morita Y, Iida S. 2004. An intragenic tandem duplication in a transcriptional regulatory gene for anthocyanin biosynthesis confers pale-colored flowers and seeds with fine spots in *Ipomoea tricolor*. *The Plant Journal* **38**: 840–849.

Parker TA, Berny Mier Y, Teran JC, Palkovic A, Jernstedt J, Gepts P. 2020a. Pod indehiscence is a domestication and aridity resilience trait in common bean. *New Phytologist* **225**: 558–570.

Parker TA, de Sousa LL, de Oliveira FT, Palkovic A, Gepts P. 2020b. Toward the introgression of *PvPdh1* for increased resistance to pod shattering in common bean. *Theoretical and Applied Genetics* **134**: 313–325.

Parker TA, Gepts P. 2021. Population genomics of *Phaseolus* spp.: a domestication hotspot. In: Rajora OP, ed. *Population genomics: crop plants*. Cham, Switzerland: Springer Nature, 1–83.

Parker TA, Lo S, Gepts P. 2021. Pod shattering in grain legumes: emerging genetic and environment-related patterns. *Plant Cell* **33**: 179–199.

Piperno DR. 2012. New archaeobotanical information on early cultivation and plant domestication involving microplant (phytolith and starch grain) remains. In: Gepts P, Famula T, Bettinger R, Brush S, Damania A, McGuire P, Qualset C, eds. *Biodiversity in agriculture – domestication, evolution, and sustainability*. Cambridge, UK: Cambridge University Press, 136–159.

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

R Core Team. 2018. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL https://www.R-project.org [accessed 8 February 2022].

Rau D, Murgia ML, Rodriguez M, Bitocchi E, Bellucci E, Fois D, Albani D, Nanni L, Gioia T, Santo D *et al.* 2019. 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.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Singh SP. 1989. Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). *Economic Botany* **43**: 39–57.

Slightom JL, Drong RF, Klassy RC, Hoffman LM. 1985. Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. *Nucleic Acids Research* **13**: 6483–6498.

Smith DS, Cash JN, Nip WK, Hui YHE, eds. 1997. *Processing vegetables: science and technology*. Lancaster, PA, USA: CRC Press/Technomic Publications.

Sturtevant AH. 1925. The effects of unequal crossing over at the *bar* locus in *Drosophila*. *Genetics* **10**: 117–147.

Sturtevant AH, Morgan TH. 1923. Reverse mutation of the *Bar* gene correlated with crossing over. *Science* **57**: 746–747.

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.

Takahashi Y, Kongjaimun A, Muto C, Kobayashi Y, Kumagai M, Sakai H, Satou K, Teruya K, Shiroma A, Shimoji M *et al.* 2020. Same locus for non-shattering seed pod in two independently domesticated legumes, *Vigna angularis* and *Vigna unguiculata*. *Frontiers in Genetics* **11**: 748.

Veit B, Vollbrecht E, Mathern J, Hake S. 1990. A tandem duplication causes the Kn1-O allele of Knotted, a dominant morphological mutant of maize. *Genetics* **125**: 623–631.

Wallace L, Arkwazee H, Vining K, Myers J. 2018. Genetic diversity within snap beans and their relation to dry beans. *Genes* **9**: 587.

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

Wolfner MF, Miller DE. 2016. Alfred Sturtevant walks into a *Bar*: gene dosage, gene position, and unequal crossing over in *Drosophila*. *Genetics* **204**: 833–835.

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**: 1527–1530.

Zhang J, Singh AK. 2020. Genetic control and geo-climate adaptation of pod dehiscence provide novel insights into soybean domestication. *G3: Genes|Genomes|Genetics* **10**: 545–554.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Differences in the presence/absence of suture strings and wall fibre between vegetable snap beans and dry beans.

Fig. S2 Transcriptome profile of pod development in a wild dehiscent *Phaseolus vulgaris* (G12873), a domesticated dehiscent dry bean cultivar (ICA Bunsi), a domesticated indehiscent dry bean breeding line (SXB 405), and a domesticated stringless snap bean cultivar (Midas).

Fig. S3 Venn diagram of differentially expressed genes observed for G12873, ICA Bunsi, SXB 405 and Midas when comparing the transcriptome at the indicated growth stages.

Fig. S4 Gene ontologies in terms of Biological process, Molecular function or Cellular component for differentially expressed genes during pod development.

Fig. S5 RNA-seq of four candidate genes between markers flanking the *St* locus on chromosome Pv02.

Fig. S6 Distance tree of amino acid sequence data of Phvul.002G271000.1 (PvIND) and closest relatives in *Phaseolus vulgaris* and *Arabidopsis thaliana*.

Fig. S7 Pod anatomy of stringless and revertant-stringy pairs.

Fig. S8 Bisulfite sequencing of conserved *PvIND* promoter elements.

Fig. S9 Reads previously sequenced from the genotypes G4627, Midas and VAX 3, aligned to a *Ty1-copia* retrotransposon found near *PvIND* and to the sequence upstream of *PvIND*, which lacks the retrotransposon.

Fig. S10 Amplification of retrotransposon-related sequence near PvIND in revertant pairs and diverse dry and green bean accessions of common bean (*Phaseolus vulgaris* L.).

Table S1 Plant material information for the four *Phaseolus vulgaris* accessions used for RNA-seq.

Table S2 Number of DEGs in four *Phaseolus vulgaris* accessions between stages R7 (Pod formation), R8 (Pod fill) and R9 (Pod maturation)

Table S3 Pairs of stringy and nonstringy *Phaseolus vulgaris* lines with the same genetic background.

Table S4 Primers used for RT-qPCR.

Table S5 Summary of primers and amplicons used for bisulfite sequencing.

Table S6 Primers used for characterising duplication and retrotransposon insertion.

Table S7 Genotype and pod phenotype of tested *Phaseolus vulgaris* accessions.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.