the plant journal

SEB Society for Experimental Biology

The Plant Journal (2019) **97,** 693–714 doi: 10.1111/tpj.14155

Genomic dissection of pod shattering in common bean: mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species

Domenico Rau[1,*,†], Maria L. Murgia[1,†], Monica Rodriguez[1], Elena Bitocchi[2], Elisa Bellucci[2], Davide Fois[1], Diego Albani[1], Laura Nanni[2], Tania Gioia[3], Debora Santo[2], Luca Marcolungo[4], Massimo Delledonne[4], Giovanna Attene[1] and Roberto Papa[2,*]

[1]*Dipartimento di Agraria, Università degli Studi di Sassari, Via E. De Nicola, 07100 Sassari, Italy,*

[2]*Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, via Brecce Bianche, 60131, Ancona, Italy,*

[3]*Scuola di Scienze Agrarie, Forestali, Alimentari e Ambientali, Università degli Studi della Basilicata, viale dell'Ateneo Lucano 10, 85100 Potenza, Italy, and*

[4]*Dipartimento di Biotecnologie, Università degli Studi di Verona, Cà Vignal 1, Strada Le Grazie 15, 37134, Verona, Italy*

Received 9 May 2018; revised 14 October 2018; accepted 30 October 2018; published online 13 November 2018.
*For correspondence (e-mails rpapa@univm.it and dmrau@uniss.it).
†These authors should be considered joint first authors.

SUMMARY

The complete or partial loss of shattering ability occurred independently during the domestication of several crops. Therefore, the study of this trait can provide an understanding of the link between phenotypic and molecular convergent evolution. The genetic dissection of 'pod shattering' in *Phaseolus vulgaris* is achieved here using a population of introgression lines and next-generation sequencing techniques. The 'occurrence' of the indehiscent phenotype (indehiscent versus dehiscent) depends on a major locus on chromosome 5. Furthermore, at least two additional genes are associated with the 'level' of shattering (number of shattering pods per plant: low versus high) and the 'mode' of shattering (non-twisting versus twisting pods), with all of these loci contributing to the phenotype by epistatic interactions. Comparative mapping indicates that the major gene identified on common bean chromosome 5 corresponds to one of the four quantitative trait loci for pod shattering in *Vigna unguiculata*. None of the loci identified comprised genes that are homologs of the known shattering genes in *Glycine max*. Therefore, although convergent domestication can be determined by mutations at orthologous loci, this was only partially true for *P. vulgaris* and *V. unguiculata*, which are two phylogenetically closely related crop species, and this was not the case for the more distant *P. vulgaris* and *G. max*. Conversely, comparative mapping suggests that the convergent evolution of the indehiscent phenotype arose through mutations in different genes from the same underlying gene networks that are involved in secondary cell-wall biosynthesis and lignin deposition patterning at the pod level.

Keywords: *Phaseolus vulgaris*, domestication, pod shattering, convergent evolution, pool sequencing, genotype by sequencing.

INTRODUCTION

Convergent evolution defines the independent evolution of similar features in different evolutionary lineages (Losos, 2011). This is a relatively frequent phenomenon in nature (e.g., Morris, 2008). Although the relationship between convergence and adaptation is not always clear-cut, convergence in taxa that occupy similar selective environments is often the result of selection (Losos, 2011). Therefore, studying the genetic basis of convergent evolution can help to shed light on the genetic process of adaptive evolution.

In an agricultural context, the 'domestication syndrome' is perhaps the most evident case of convergent evolution: at different sites and at different times, diverse crops evolved similar phenotypic features from their wild

progenitors because of repeated and independent selection for adaptation to the agro-ecosystem (Doebley *et al.*, 2006; Purugganan and Fuller, 2009; Gaut, 2015; Bitocchi *et al.*, 2017). The loss or reduction of pod shattering is considered to have been a key event in the domestication of seed crops from their wild progenitors (Doebley *et al.*, 2006; Purugganan and Fuller, 2009). This trait evolved independently in different crops, and it represents one of the most evident cases of convergence of phenotypic adaptation to agro-ecosystems (Doebley *et al.*, 2006; Purugganan and Fuller, 2009). The genetic bases of this trait have been studied in several crops, to unravel the mechanisms of this convergent evolution at the molecular level (Lin *et al.*, 2012; Lenser and Theiβen, 2013; Olsen and Wendel, 2013a, b; Dong and Wang, 2015; Li and Olsen, 2016; Ballester and Ferrándiz, 2017).

In cultivated plants, the most comprehensive studies are those that have been conducted in cereals. Paterson *et al.* (1995) first noted that seed shattering was determined by quantitative trait loci (QTLs) that corresponded closely to maize (*Zea mays* L.), sorghum (*Sorghum bicolor* L.), and rice (*Oryza sativa* L.). Therefore, convergent molecular evolution was hypothesized for this trait. The discovery of a YABBY locus that confers shattering in maize, sorghum, and rice validated the hypothesis that genes for seed shattering were under parallel selection during the domestication of these three species (Lin *et al.*, 2012). However, Tang *et al.* (2013) observed that seed shattering in a wild sorghum is also conferred by a locus that is not related to domestication, which illustrates a case in which the genetic control of a trait in a wild relative fails to extrapolate even to closely related crops. Moreover, two BEL1-type homeobox genes, *qSH1* and *SH5*, induce the *SHAT1* and *Sh4* genes that are responsible for abscission-zone differentiation. However, *qSH1* and *SH5* act via at least two independent pathways to develop a non-shattering phenotype (Yoon *et al.*, 2014).

In legumes, studies on pod shattering have been conducted in alfalfa (*Medicago sativa*), common vetch (*Vicia sativa*), narrow-leafed lupin (*Lupinus angustifolius*), chickpea (*Cicer arietinum*), pea (*Pisum sativum*), lentil (*Lens culinaris*), soybean (*Glycine max*), cowpea (*Vigna unguiculata*), and common bean (*Phaseolus vulgaris*) (for reviews, see Dong and Wang, 2015; Li and Olsen, 2016; Ballester and Ferrándiz, 2017). Among all these species, the molecular bases that led to non-shattering pods have been characterized in detail only in soybean (Dong *et al.*, 2014). Together with cowpea and common bean, soybean belongs to the Phaseoleae tribe in which the lowest phylogenetic distance is between common bean and cowpea (Choi *et al.*, 2004). In soybean, a domestication shattering gene, SHATTERING1-5 (*SHAT1-5*), was mapped to chromosome 16 (chr 16) of soybean, and it was shown that, in domesticated plants, the indehiscent phenotype arises from excessive lignification of the 'fiber-cap cells' along the ventral suture of the pod valves (Dong *et al.*, 2014). Variability in the degree of shattering within the cultivated gene pools (i.e., landraces) of soybean has also been documented (Tsuchiya, 1987). This variability arises through another gene, Pod dehiscence 1 (*Pdh1*), the expression of which is correlated with lignin deposition in the inner sclerenchyma of the pod valves (Funatsuki *et al.*, 2014). *Pdh1* regulates dehiscence through increased twisting force in the pod wall at low humidity. The combined data of Dong *et al.* (2014) and Funatsuki *et al.* (2014) suggest that domestication and plant breeding acted on several shattering genes. A very large genome-wide association study (GWAS) also confirmed that shattering in soybean is mainly due to genes located on chr 16 (Zhou *et al.*, 2015).

For *Vigna*, the available knowledge of the genetic basis and mechanisms of pod shattering are more limited. Working with F_2 and backcross populations between *V. unguiculata* subsp. *sesquipedalis* and wild *V. unguiculata*, a major domestication QTL was mapped to linkage group 7, while two other additional QTLs were located on linkage groups 1 and 4; these QTLs also co-mapped with those for fiber content of the pod walls (Suanum *et al.*, 2016). Comparative genome analysis with *Vigna angularis* has indicated that the QTL on linkage group 7 contains a gene that encodes a MYB transcription factor, *MYB83*, which regulates fiber biosynthesis, while the QTL on linkage group 1 contains a gene that encodes cellulose synthase A7 (CESA7) (Suanum *et al.*, 2016). More recently, Lo *et al.* (2018) identified two novel domestication QTLs for pod shattering in *V. unguiculata*. These QTLs, named *CPshat3* and *CPshat5*, explained 37.7% and 30.3% of the phenotypic variation, respectively. Lo *et al.* (2018) indicated that two transcription factors that appear to be involved in secondary cell-wall biosynthesis – a NAC domain gene and the ortholog of *Arabidopsis thaliana MYB26* – underlie *CPshat3* and *CPshat5*, respectively.

In *P. vulgaris*, a monogenic basis for this trait was suggested by several early studies that were based on classical genetic approaches (Von Tshemark, 1901, 1902; Tjebbes and Kooiman, 1922; Wellensiek, 1922; Prakken, 1934), with less frequent evidence of its oligogenic control associated with a major gene that determines the presence or absence of the trait (Currence, 1930; Lamprecht, 1932). These studies also suggested histological differences between shattering and non-shattering phenotypes that were mainly due to the lignification patterns of the valve tissue. It was also suggested that the indehiscent phenotype emerged because of the loss of the 'strings' along the suture line of the pod valves (Prakken, 1934). More than 60 years later, a QTL for the presence of pod strings was identified on chr 2 of common bean (i.e., the *St* locus), and this was also proposed to control the differences in the shattering abilities between domesticated and wild

accessions (Koinange *et al.*, 1996). In the attempt to find the genes that underlie the *St* locus, the sequences homologous to the *A. thaliana SHATTERPROOF-1* (*SHP1*) and *INDEHISCENT* (*IND*) genes that are required for silique shattering were mapped on chr 6 (Nanni *et al.*, 2011) and chr 2 (Gioia *et al.*, 2012) of *P. vulgaris*, respectively. *P. vulgaris IND* (*PvIND*) was mapped near the *St* locus, but the lack of complete cosegregation between *PvIND* and *St* and the lack of polymorphisms at the *PvIND* locus correlated with the dehiscent/ indehiscent phenotype suggested that *PvIND* was not directly involved in pod shattering and was not the gene underlying the *St* locus (Gioia *et al.*, 2012).

Hagerty *et al.* (2016) used a dry bean × snap bean recombinant inbred population to map the snap bean pod and color traits, and a QTL for the string-to-pod-length ratio found on chr 2 explained 32% of the total genetic variation. Also, QTLs for pod height, width, and wall fiber and thickness were found clustered on chr 4, and these explained 26%, 18%, 21%, and 16% of the genetic variation of each of these respective traits. Another QTL for pod length was found on chr 9 that explained 5% of the genetic variation (Hagerty *et al.*, 2016).

Outside the Phaseoleae tribe, in *P. sativum* (tribe: Viceae), a locus name *Dpo* that is responsible for the loss of pod shattering was localized on linkage group III (Bordat *et al.*, 2011). Hradilová *et al.* (2017) suggested that in *P. sativum* the main candidate gene responsible for pod shattering and localized on linkage group III is a homolog of peptidoglycan-binding domain protein (PGDB) of *Medicago truncatula*. These proteins might have a general peptidoglycan-binding function, and this motif is found at the N- or C-terminus of a variety of enzymes involved in bacterial cell-wall degradation. Although not due to selection under domestication, in some species of the legume genus *Medicago* (tribe: Trifolieae), variations in pod morphology and shattering have been associated with variations in lignin deposition at the valve margin. This was attributed to a change in the protein sequence of *A. thaliana SHATTERPROOF* (*SHP*) orthologs (Fourquin *et al.*, 2013).

Recently, Murgia *et al.* (2017) carried out a comprehensive characterization of the pod shattering trait in common bean using a population of 257 introgression lines (ILs) that were developed with the specific purpose of studying pod shattering in *P. vulgaris*. They showed that this mainly behaves as a qualitative trait, with high shattering associated with a high fiber content of the pod valves and strong lignification of the ventral sheath and the inner sclerenchyma of the pod valves. In the present study, the same population of ILs that were phenotyped by Murgia *et al.* (2017) was exploited using next-generation sequencing (NGS) strategies of pool sequencing (pool-seq; Ferretti *et al.*, 2013) and genotype by sequencing (GBS; Poland *et al.*, 2012). To obtain the ILs, a wild-like, highly shattering recombinant inbred line (MG38) was backcrossed with a non-shattering Andean snap variety (MIDAS) as a recurrent parent. Different cycles of backcrossing and selfing were carried out, together with selection for the wild characteristics of the pods and seeds (i.e. high pod shattering, small pods and seeds, colored pods and seeds).

Here, we pursued two main aims. First, using QTL mapping, we dissected out the genetic architecture of the shattering trait in common bean to identify the genomic regions involved in the determination of 'occurrence' of pod shattering (i.e., presence versus absence), and in its 'level' (i.e. low versus high, for number of pods per plant) and 'mode' (i.e. presence versus absence of torsion, twisting, or spiral coiling of the pod valves after dehiscence). Secondly, we compared these data in *P. vulgaris* with those obtained in *V. unguiculata*, *G. max*, and *P. sativum*, to shed light on the genetic mechanisms that underlie this convergent phenotypic selection during the domestication of legume crops.

RESULTS

Pod shattering under field conditions

Details of the data for the phenotypic analysis of the pod shattering of these ILs were reported in Murgia *et al.* (2017). Briefly, phenotyping was conducted under field condition of Sardinia Island (Lat. 41°N, Long. 9°E, 81 m a.s.l.) during the spring–summer period (May–October) of 2014, in a typical Mediterranean climate. ILs were first classified into shattering or non-shattering ($SH_{Y/N}$). Therefore, for each IL, all the fertile pods were counted and the percentage of fully shattered pods per plant determined (SH%). Based on the presence of twisting, torsion/spiral coiling of the pod valves, the percentages of fertile shattering pods with twisting (TW%) or non-twisting valves (NTW%) were determined. For each IL, 'resistance to manual shattering' was evaluated on a scale from 1 (very low resistance) to 9 (very high resistance) (RES_{1-9}). Moreover, carbon content of pod valves (C%) was also determined for all the ILs, as this was highly contrasting between the two parental lines MG38 and MIDAS. ILs were further classified into two contrasting categories considering simultaneously $SH_{Y/N}$ and C% ($SH_{Y/N}$+C%).

For shattering occurrence ($SH_{Y/N}$), 29 ILs were non-shattering and 228 ILs were shattering. Among the 257 ILs, the shattering level was highly variable: SH% varied from 0 to 82%, and RES_{1-9} from very low resistance to very high resistance; mode of shattering (TW% and NTW%) varied from the absence to the strong prevalence of twisting pods; and carbon content (C%) of pod valves varied from 38.9 to 47.4% (Figures 3, 4 and 6a of Murgia *et al.*, 2017). When $SH_{Y/N}$ and C% were considered jointly ($SH_{Y/N}$+C%), two groups of ILs were defined: the first group (GR1) comprised 48 ILs with mean SH% of 1.4% and mean C% of 41.8%, while the second group (GR2) comprised 209 ILs

with mean SH% of 36% and mean C% of 44.6% (see also Figure 3 of the Results section).

When all of the ILs were considered, the heritability values were particularly high for $SH_{Y/N}$+C% (0.990), $SH_{Y/N}$ (0.866), RES_{1-9} (0.895), and SH% (0.792) (Table 1). When only the 210 ILs with higher SH% and C% were considered, the heritability for SH% and RES_{1-9} decreased (0.507, 0.683, respectively), although it remained moderate to high. The lower heritability values were seen for the mode of shattering particularly when SH% was used as covariate (0.180 in both cases; Table 1).

Mapping shattering loci

Pool sequencing: occurrence of shattering depends on a single major QTL on chromosome 5. The pool-sequencing (pool-seq) analysis compared two pools of ILs that, among all the 257 analyzed, were highly contrasted for shattering ability. The first pool, $Pool_{NSH}$, comprised 27 non-shattering ILs (like MIDAS; SH% = 0), and the second pool, $Pool_{SH}$, comprised 30 shattering ILs with shattering levels greater than or equal to those of MG38 (SH% $\geq$ 65% shattering pods). Within $Pool_{SH}$, the mode of shattering was variable among the plants, with the ratio between non-twisting (NTW%) and twisting (TW%) pods per plants varying from 1:2 to 3:1. Overall, between $Pool_{NSH}$ and $Pool_{SH}$, 51 280 single nucleotide polymorphisms (SNPs) were polymorphic. On average, there were 4662 SNPs per chromosome, with the minimum of 3146 SNPs for chr 6, and the maximum of 5960 SNPs for chr 8.

Pool-seq analysis revealed one genomic region on chr 5 that was highly divergent between these pools; i.e., that was strongly associated with the pod shattering trait (Figure 1). This was characterized by a ΔSNP index that reached the maximum value of 1.0, which indicated that this genomic segment contained SNPs that were 'diagnostic'; i.e. polymorphic between these pools, but monomorphic within both of these pools. This situation therefore allows perfect discrimination between shattering and non-shattering plants. We have named this region qPD5.1-Pv; i.e., QTLs affecting pod shattering on chr 5 of *P. vulgaris*.

Based on the ΔSNP index, four other genomic regions were found that were less divergent between these pools, and these can be ranked as follows: qPD1.1-Pv (ΔSNP index ~0.9), qPD3-Pv (ΔSNP index ~0.75), qPD5.2-Pv (ΔSNP index ~0.6), and qPD1.2-Pv (ΔSNP index ~0.55) (Figure S1).

The *qPD5.1-Pv* region spanned ~1.16 Mb (from positions 37990066 to 39153303), whereby two regions (named A and B; Figure 1) were characterized by a high ΔSNP index and were separated by a 'valley' of low ΔSNP index. The first region ('A') spanned ~0.70×10^6 bp, while the second ('B') spanned ~0.13×10^6 bp (Figure 1).

Genotype by sequencing. Narrowing the size of the QTL on chromosome 5—Overall, the genotype by sequencing (GBS) conducted with these 257 ILs resulted in 14196 polymorphic markers. The number and density of markers (markers/MB) per chromosome were in descending order of markers: 2670 and 50.0 (chr 3); 2221 and 58.1 (chr 9); 1741 and 33.8 (chr 1); 1610 and 25.5 (chr 8); 1557 and 49.8 (chr 6); 1540 and 31.0 (chr 2); 1076 and 22.4 (chr 4); 989 and 24.7 (chr 7); 380 and 8.6 (chr 10); 333 and 8.1 (chr 5); and 79 and 1.5 (chr 11). We observed linkage disequilibrium (LD) decay as a function of physical distance using GBS SNPs. Overall, we found low LD baseline, with $r^2_v < 0.10$, within 0.5 Mb. Chromosomes show different LD decay patterns, as shown by the LD halving distance that varied from ~0.1 Mb (chr 1) to ~0.5 Mb (chr 10). The r^2_v values became lower than 0.10 from below 0.2 Mb (chr 6) to below 1.2 Mb (chr 10). The low LD in the IL population suggests it can achieve high mapping definition.

This second mapping exercise clearly confirmed a QTL that underlies pod shattering in the distal part of chr 5 that co-mapped with qPD5.1-Pv, as was seen for pool-seq results (Figure 2). This mapping was first conducted considering shattering as a qualitative trait with two possible states, as yes or no ($SH_{Y/N}$). This resulted in a single highly significant QTL in position 38916019 ($P = 3.45 \times 10^{-8}$, $R^2 = 0.127$; Figure 2; Table 2). The QTL for C% overlapped with that for $SH_{Y/N}$, as 17 markers that started from position 38675127 and continued to position 39011678 showed the strongest and the same association strength ($P = 2.77 \times 10^{-8}$, $R^2 = 0.128$; Figure 2; Table 2). The mapping of the $SH_{Y/N}$+C% composite trait reinforced the evidence for the relevance of *qPD5.1-Pv* for pod shattering ($P = 2.27 \times 10^{-40}$; Figure 2; Table 2). It was possible to identify within *qPD5.1-Pv* two sub-regions (S1, S2) that

Table 1 Heritability estimates (h^2) for the various shattering traits considered in this study. For each trait, the heritability was obtained at step 0 in the MLMM model, when no markers were included in the model

Trait	Variable	h^2
Shattering occurrence	$SH_{Y/N}$	0.866
Carbon content of pod valves	C%	0.591
Combined $SH_{Y/N}$ and C%	$SH_{Y/N}$+C%	0.99
Resistance to manual shattering	RES_{1-9}	
257 ILs		0.895
210 ILs		0.683
Shattered pod per plant	SH%	
257 ILs		0.792
210 ILs		0.507
Twisting pods per plant	TW%	
Not considering SH% as covariate		0.355
Considering SH% as covariate		0.180
Non-twisting pods per plant	NTW%	
Not considering SH% as covariate		0.321
Considering SH% as covariate		0.180

ILs, introgression lines.

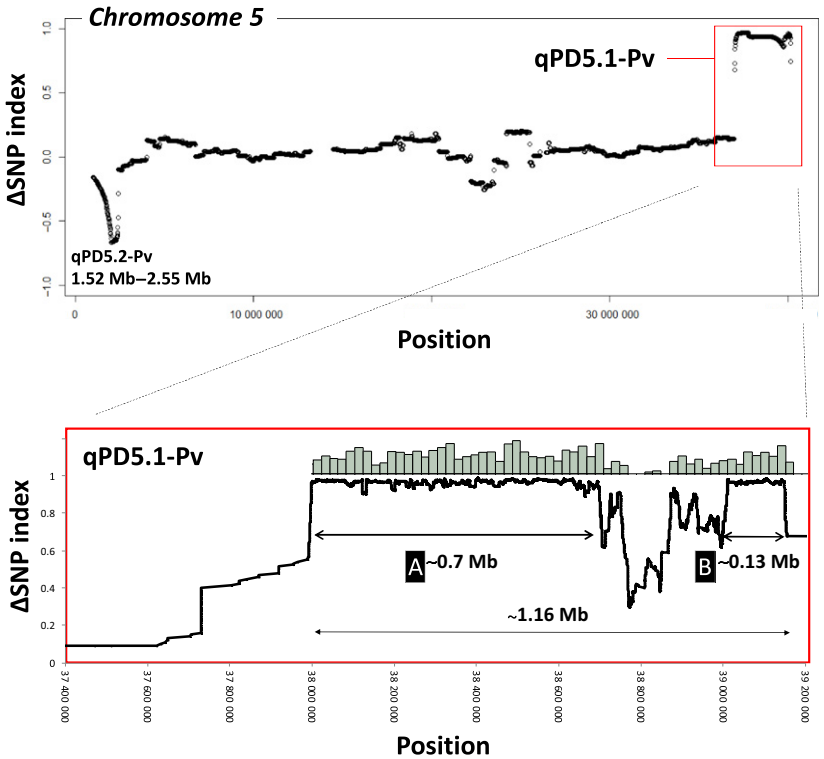


Figure 1. Genetic divergence (ΔSNPs index) along chr 5 between pools contrasting for shattering ability. ΔSNPs index: SNP index of the pool of the highly shattering lines minus the SNP index of the pool of the non-shattering lines. The SNP index is calculated as the fraction of reads per position that is attributable to MG38 (i.e., the highly shattering parental line). Top panel: To reduce noise, the plots represent the average values of sliding windows of 2 Mb, with a step of 10 kb. Bottom panel: The plots represent the average values of sliding windows of 10 000 bp, with a step of 1 kb. A, B, indicate the two regions within *qPD5.1-Pv*. The histogram illustrates the distribution of the diagnostic SNPs between the pools for adjacent genomic windows of 20 kb (bars: from a minimum of 0 SNPs to a maximum of 335 SNPs).

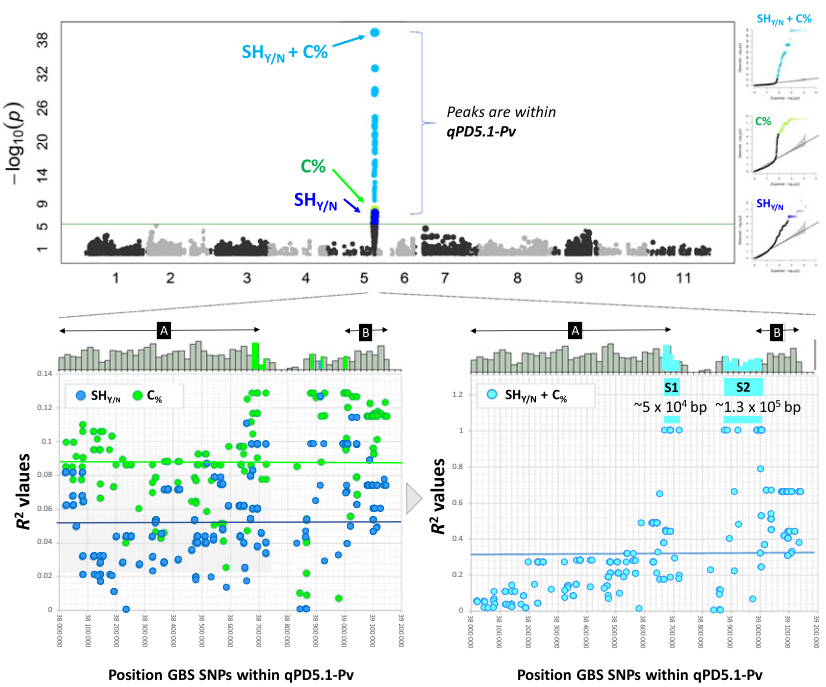


Figure 2. Top panels: Manhattan plot across all of the 11 chromosomes of common bean. The data from three mapping exercises are overlapped, and significant associations are colored: blue, shattering yes/no ($SH_{Y/N}$); green: carbon content of pod valves (C%); sky blue: $SH_{Y/N}$ and C%, considered simultaneously ($SH_{Y/N}$+C%). No significant associations are in gray or black. Right: Quantile–quantile plots for MLM (including correction for kinship) for each of the three traits. Bottom panels: Zoom into the *qPD5.1-Pv* region identified by pool-seq to show associations found by GBS. Left: Results for $SH_{Y/N}$ and C% traits. Right: Results for $SH_{Y/N}$+C%; highlighted: two genomic windows (S1, S2) overall of ~1.8×10^5 bp that contain SNPs in perfect association with the $SH_{Y/N}$+C% trait. The histograms at the top of each panel report the distributions of the diagnostic markers between $Pool_{SH}$ and $Pool_{NSH}$.

spanned a total of ~1.8×10^5 bp. These comprised 17 SNPs (the same that were found for C%) that resulted in perfect association ($R^2 = 1.00$) with the $SH_{Y/N}$+C% trait (Table 2), which was obtained by classifying the ILs into two groups (G1, G2) considering simultaneously $SH_{Y/N}$ and C%. Consequently, the allelic variation at these SNPs separated two strongly divergent groups (G1, G2) of the ILs in terms of shattering ability, as either occurrence, $SH_{Y/N}$ or level, SH%, and carbon content, as C% (Figure 3).

Table 2 Association mapping using SNPs from genotype by sequencing and 257 introgression lines (ILs). Data are presented for several different measures of pod shattering and for different sets of ILs, to analyze the trait as 'occurrence', 'level', and 'mode'. For quantitative variables, the data are presented for two different statistical models. For twisting (TW%) and non-twisting (NTW%) pod valves, the data in parentheses are for the analysis repeated considering SH% as a cofactor in the association mapping

Shattering component	Variable	Sample size	Model	Chromosome	Position	*P*	R^2
Occurrence (qualitative trait)	$SH_{Y/N}$	257	Tassel (MLM+kinship)	5	38916019	3.45E-08	0.127
	C%		Tassel (MLM+kinship)	5	[a]38675127–39011678	2.77E-08	0.128
	$SH_{Y/N}$+C%		Tassel (MLM+kinship)	5	[b]38675127–39011678	2.27E-40	1.000
Occurrence, level, mode (quantitative trait)	RES_{1-9}	257	Tassel (MLM+kinship)	5	38916019	2.99E-06	0.089
			MLMM (kinship)	5	38916019	1.55E-07	0.550
	SH%	257	Tassel (MLM+kinship)	5	[c]38916011–39146588	2.35E-07	0.110
			MLMM (kinship)	5	38708645	1.19E-11	0.461
Level, mode	RES_{1-9}	210	Tassel (MLM+kinship)	6	22061911	2.95E-07	0.152
			MLMM (kinship)	6	24311992	5.13E-07	0.150
				6	23989634	3.55E-06	0.051
	SH%	210	Tassel (MLM+kinship)	6	23828803	1.32E-04	0.089
			MLMM (kinship)	5	1022962	3.21E-17	0.221
				4	44198457	1.55E-09	0.121
				9	29702346	5.10E-07	0.062
Mode	TW%	210	Tassel (MLM+kinship)	3 (2)	23649132 (48261529)	4.42E-5 (9.20E-5)	0.100 (0.042)
			MLMM (kinship)	4	44613224 (n.d)	6.27E-12 (n.d.)	0.220
				6	20794299 (n.d.)	2.02E-06 (n.d)	0.181
	NTW%	210	Tassel (MLM+kinship)	8 (2)	51548749 (48261529)	7.55E-5 (9.20E-5)	0.095 (0.042)
			MLMM (kinship)	n.d (n.d)	n.d (n.d)	n.d (n.d)	–

[a]17 other positions between these two extremes had the same statistical significance.
[b]17 other positions between these two extremes were completely associated with the trait.
[c]17 other positions between these two extremes had the same statistical significance.

Shattering was also mapped as a quantitative trait considering SH% and resistance to manual shattering (RES_{1-9}), using the data for all of the 257 ILs (Figure 4a; Table 2). These data mainly confirmed the relevance of the identified SNPs (38916019: RES_{1-9} both for Tassel and multi-locus mixed model [MLMM]; Table 2) or that they were placed within the already identified regions (38708645: SH% with MLMM; Table 2).

Single-locus analysis of SH% with Tassel defined a position (38916011; $R^2 = 0.111$; $P = 2.35 \times 10^{-7}$) that was adjacent to a SNP (38916019) that was also associated with $SH_{Y/N}$ and RES_{1-9} (Table 2). However, this analysis also moved away from S1 and S2 of *qPD5.1-Pv* (Figure 4b). Indeed, this analysis found 17 additional markers, from position 39080874 to position 39146588 (encompassing 65714 bp; S3) that were all associated with the SH% trait with the same strength ($P = 2.35 \times 10^{-7}$) (Table 2; Figure 4b).

However, a very strong haplotype structure was observed at the *qPD5.1-Pv* locus (Figure S2). The 242 GBS markers that covered the *qPD5.1-Pv* region from position 38022400 to position 39146588 were in strong LD, with an overall mean $R^2 = 0.91$, with minimum $R^2 = 0.20$, and maximum $R^2 = 1.00$ (Figure 5). As expected, markers in complete LD were more often also in close proximity although, in some cases, complete LD extended over a greater range (Figure 5). This was particularly noted for positions between 3.86×10^7 bp and 3.87×10^7 bp, where the highest number of marker pairs in complete LD ($R^2 = 1.00$) was also seen (Figure 5).

Level and mode of shattering depend on several additional genes that have minor effects. The variables SH% and RES_{1-9} depend on the occurrence, level, and mode of shattering. Therefore, as a further step, mapping was repeated considering that the ILs belonged to GR2 identified by the $SH_{Y/N}$+C% trait, which showed occurrence of shattering, SH%, from 4.4 to 82.6%, and C% of pod valves starting from 41.5% (Figure 3). Interestingly, *qPD5.1-Pv* was not detected anymore when either the SH% or RES_{1-9} trait was considered and single-locus or multi-locus analyses were performed (Table 2; Figure S3). This situation indicated that *qPD5.1-Pv* per se was not relevant to explain the level and mode of shattering; i.e. it is mainly correlated with the occurrence of shattering.

The data for RES_{1-9} indicated the relevance of chr 6. The best association by Tassel ($R^2 = 0.152$; $P = 2.95 \times 10^{-07}$) was position chr6_22061911, while MLMM indicated a two-locus model that involved positions chr6_23989634 and chr6_24311992 (Table 2). The data for SH% partially supported this finding. Indeed, with Tassel, the best association was again with a SNP on chr 6 in position 23828803; this, however, did not reach significance after Bonferroni

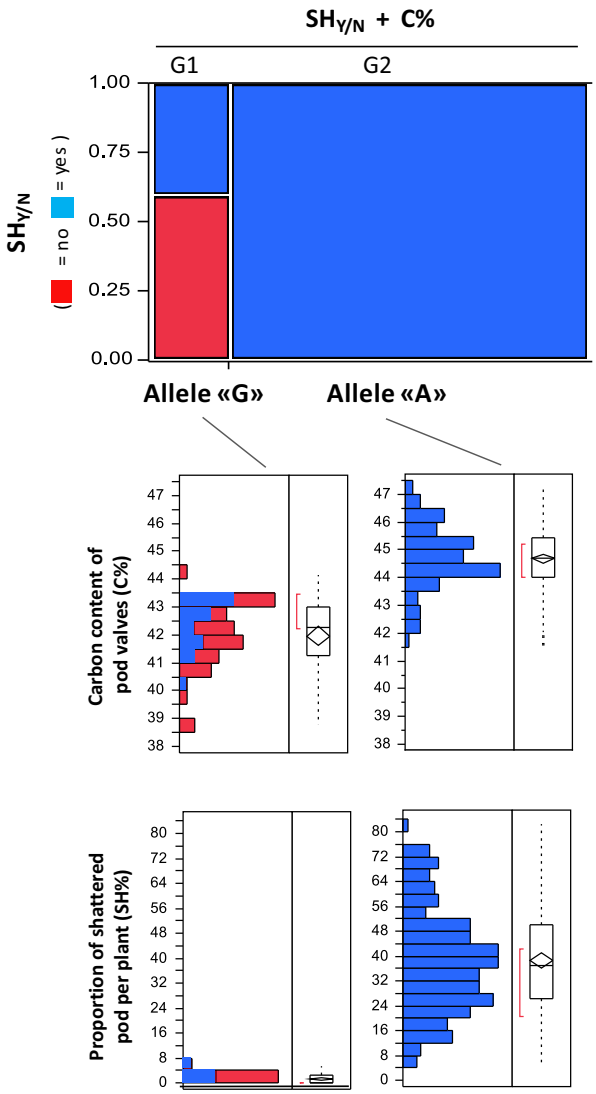


Figure 3. Effects of the alleles at locus chr 05_38675127 in perfect association (R^2 = 1) with SH$_{Y/N}$+C%. Top: Mosaic plot showing that allele segregation is associated with separation of all of the non-shattering ILs from the vast majority of shattering ILs. Centre: The ILs with very low shattering ability (>0%, <4.4%) were attributed to the class of indehiscent types. Bottom: The two allelic classes show almost disjointed distributions of carbon content of pod valves (C%) and shattering and non-shattering ILs of the same allelic class 'G' showed similar C%.

correction ($P = 1.32 \times 10^{-4}$). By contrast, MLMM suggested a three-gene model to explain SH%; a SNP on chr 5 initially entered the model with a high level of probability (R^2 = 0.22; $P = 3.21 \times 10^{-17}$), then a locus on chr 4 (R^2 = 0.12), and then another on chr 9 (R^2 = 0.06), so that cumulatively the model explained 40% of the total variance for SH% (Table 2).

For the mode of shattering (i.e., TW% versus NTW%), none of the analyses reached statistical significance after Bonferroni correction, except for MLMM with TW%. In this case, a two-locus model was defined that involved SNPs on chr 4 and chr 6 (Table 2). These positions are ~0.08–0.35 and ~0.41 Mb relative to those previously found on the same chromosomes for SH% and RES$_{1-9}$, respectively. With Tassel, the best associations were on chr 3 (TW%) and chr 8 (NTW%) (Table 2). When the mapping of TW% and NTW% was repeated considering SH% as a cofactor, all of the *P*-values decreased, MLMM did not show any significant association, and Tassel indicated the same position on chr 2 for both TW% and NTW%. This indicates that the mode of shattering is mainly determined by the same genes that underlie the level of shattering, except for other minor genes, such as those detected on chr 3, chr 8, and chr 2.

To further test these associations, all of the SNPs in Table 2 were used to model SH% using partition analysis, for the data for all of the 257 ILs. These results are presented in Figure 6. Overall, this analysis confirmed that shattering ability can be (conservatively) modeled with a major QTL on chr 5 (R^2 = 0.481) and at least two genes that control the level of shattering on chr 5 (R^2 = 0.187) and chr 4 (R^2 = 0.055). Cumulatively, these explained 72.4% of the phenotypic variance of the trait. This model identifies four groups of ILs with means of 1.26, 31.7, 42.6, and 61.7% SH%, and these were well separated in the Tukey–Kramer honest significant difference (HSD) test ($P < 0.05$). With three additional splits (Figure S4), the model QTLs that were entered showed very small effects on chr 9 (29702346, R^2 = 0.009), chr 4 (44613224, R^2 = 0.013), and chr 6 (20794299; R^2 = 0.006). For the models with four or five loci, the group means were still well separated in the Tukey–Kramer HSD test ($P < 0.05$), and for the model with six loci, the group means were not clearly separated any more (Figure S4). All of this indicates that a three-gene model appears to be parsimonious, and that two additional loci might also have roles (i.e., on chr 9 and chr 4).

For all of the three possible pairs of loci of the three-locus model, there were significant interactions between alleles at different loci; i.e., epistatic interactions (Table S1; Figure 7). The interaction between chr5_38916011 and chr5_1022962 was not possible to estimate with reliability, as the two loci were in slight LD (R^2 = 0.05; $P < 10^{-3}$), and almost all of the individuals that inherited the 'indehiscent' allele 'G' at chr5_38916011 also had the allele associated with low shattering ('C') at chr5_1022962 (Figure 7). In the two other cases, chr5_38916011-chr4_44198457 and chr5_1022962-chr4_44198457, the two loci were in LD, all four gametic types were well represented, and it was possible to test the interaction. The strongest interaction was seen for the chr5_1022962-chr4_44198457 pair (Table S1; Figure 7).

Genes underlying the QTLs identified

Pattern of polymorphism and gene content within qPD5.1-Pv. *qPD5.1-Pv* comprises 138 genes and 9785

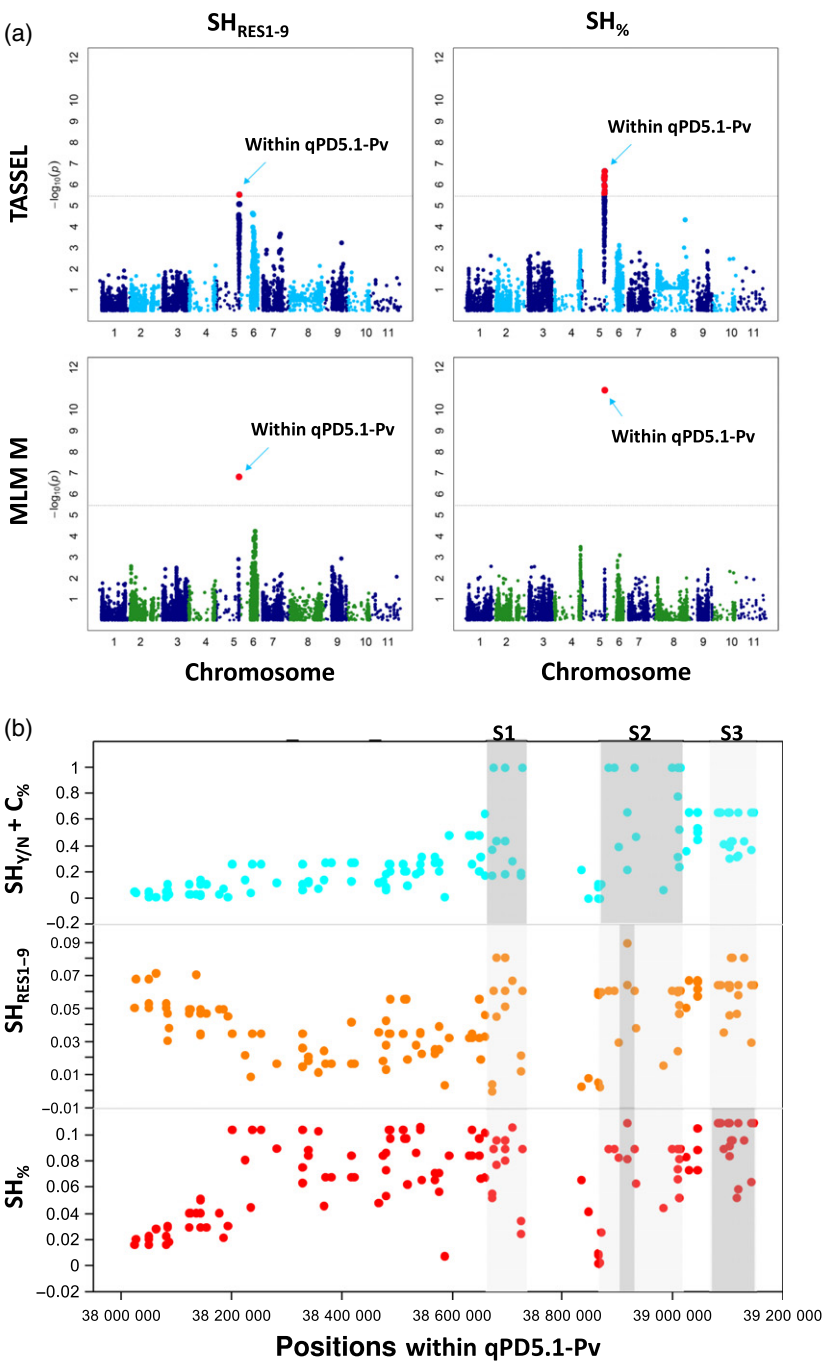


Figure 4. (a) Mapping of the resistance to manual shattering (RES_{1-9}) and percentage of shattered pods per plant (SH%).
(b) Comparison of the single-locus association analyses conducted with Tassel (with MLM model and kinship) for the traits $SH_{Y/N}$+C%, RES_{1-9}, and SH%. For each trait, the dark gray boxes indicate the peaks of R^2 for the trait, the light gray boxes indicate the corresponding region in the other traits.

diagnostic SNPs between pools. The majority of the SNPs (7325, 74.6%) were intergenic. Among the genic polymorphisms, non-coding SNPs prevailed over coding SNPs (Table S2). All of the SNPs were categorized based on their sequence ontology (http://www.sequenceontology.org), with the subjective classification of the severity of the variant consequence carried out (https://www.ensembl.org/Help/Glossary?id=535). The four categories were: high, as a variant assumed to have high (disruptive) impact in the protein (potentially protein truncation or loss of function, or triggering nonsense-mediated decay); moderate, as a non-disruptive variant (potentially changing protein effectiveness); low, as the least disruptive (unlikely to change protein behavior); and modifier, as non-coding variants or variants affecting non-coding genes (when predictions are difficult, or there is no evidence of impact) (https://www.ensembl.org/Help/Glossary?id=535). The distribution of the SNPs classified as modifiers and of low impact was relatively even (Figure S5). In contrast, it appeared that the regions that were significantly associated with C%, $SH_{Y/}$

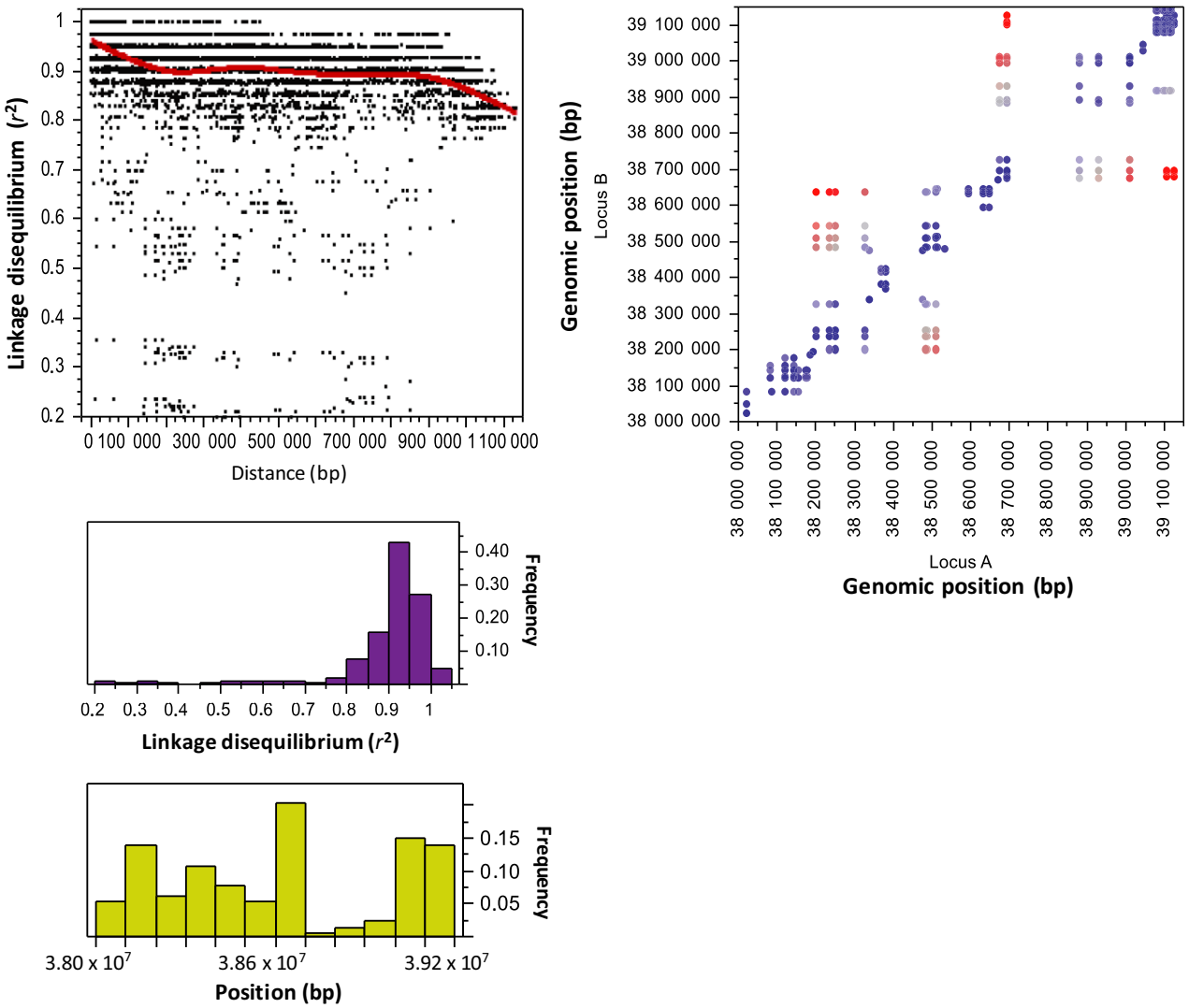


Figure 5. Linkage disequilibrium (LD) within *qPD5.1-Pv*.
(a) Level of LD between pairs of SNPs as a function of physical distance.
(b) SNPs that are in complete LD (R^2 = 1), as a function of distance.
(c) Distribution of LD values (R^2) and of the number of SNPs in complete LD (R^2), as a function of physical distance.

$_N$+C%, and RES_{1_9} were richer in SNPs with 'moderate' effects. This region also comprised four out of the six variants that had been classified as 'high' effects (Table S3), and that were also characterized by the ratio of the numbers of non-synonymous/ synonymous mutations >1, while this ratio for the entire QTL was <1 (444/581, 0.764) (Figure S5).

Comparisons among legume crops. In Table 3, we report the results of comparative mapping for the shattering genes of *V. unguiculata*, *G. max*, *P. sativum*, and *M. truncatula* against *P. vulgaris*. It is useful to recall here that the three species *P. vulgaris*, *V. unguiculata*, and *G. max* belong to the Phaseoleae tribe, and the phylogenetic distance between *P. vulgaris* and *V. unguiculata* is lower than between *P. vulgaris* and *G. max* (Choi *et al.*, 2004). *Pisum sativum* and *M. truncatula* were phylogenetically more distant, as they belong to the tribes of Viceae and Trifolieae, respectively (Choi *et al.*, 2004).

We found that the major shattering QTL of *P. vulgaris*, *qPD-5.1-Pv*, was related to one QTL for pod shattering in *V. unguiculata*. Indeed, *qPD-5.1-Pv* is related to the QTL on chr 5 of *V. unguiculata*, *CPshat5,* which was described by Lo *et al.* (2018). The *CPshat5* region spanned 7.74 cM, which corresponded to ~1.60 Mb and which is ~1.5-fold the size of qPD5.1_Pv (1.16 Mb). Among the 204 genes within *CPshat5*, *Vigun05g273500* (annotated as Myb domain protein 26) was considered the best candidate (Lo *et al.*, 2018). Indeed, *AtMYB26* regulates secondary cell-wall formation in anther endothecium, which underlies anther

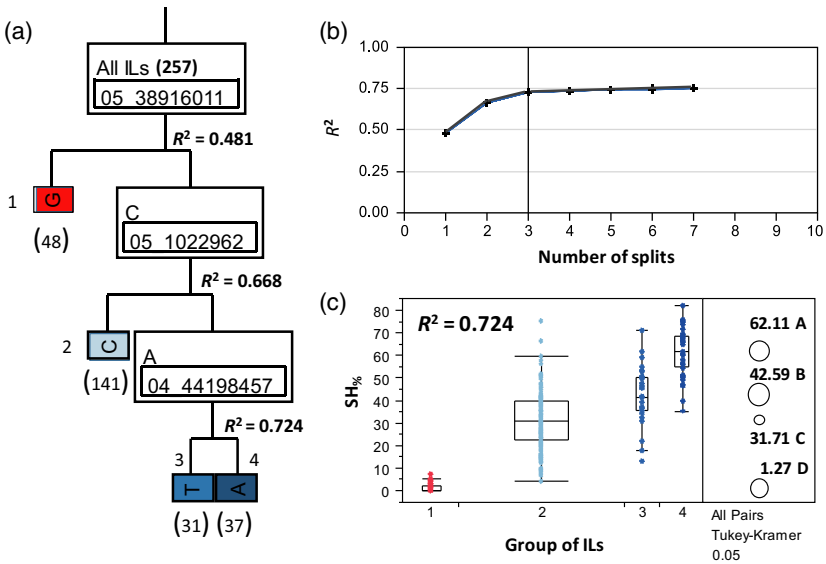
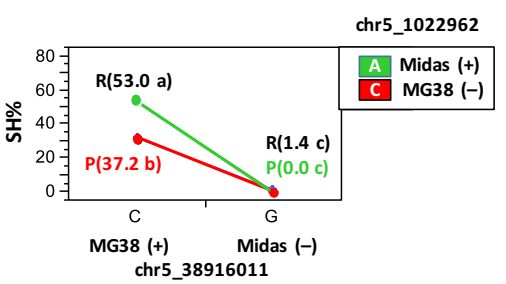
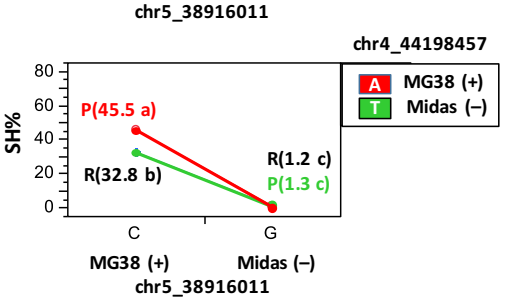


Figure 6. A three-gene model to explain pod shattering in common bean.
(a, b) Partition history for three splits. Different colors represent different sets of ILs grouped based on the genetic information.
(c) Means with different letters are significantly different ($P < 0.05$; Tukey–Kramer HSD tests).

chr5_38916011 \ chr5_1022962	A	C	Total
C	67	141	208
G	3	45	48
Total	70	186	256



chr5_38916011 \ chr4_44198457	A	T	Total
C	94	114	208
G	24	24	48
Total	118	138	256



chr5_1022962 \ chr4_44198457	A	T	Total
A	36	34	70
C	82	104	186
Total	118	138	256

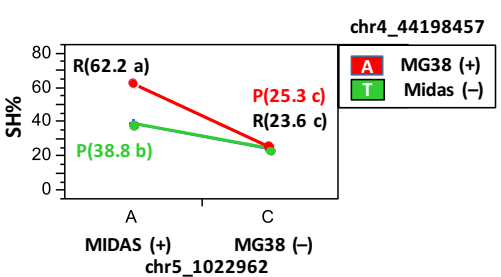


Figure 7. Left: Contingency tables. Right: Interaction plots for the three possible pairs of loci for the three-locus model presented in Figure 6, to explain the shattering ability as a percentage of the shattered pods per plant (SH%). For each allele, the parental donor (MG38, MIDAS) and its effect, as plus (+) or minus (–), are specified. Within each interaction plot, the parental (P) or recombinant (R) gametes are specified. The average shattering levels (as SH%) are reported in parentheses; within each plot, average phenotypic values with different letters are separated ($P < 0.05$; Tukey–Kramer HSD multiple comparison tests).

indehiscence (Wilson *et al.*, 2011). Moreover, it has been shown that *AtMYB26* regulates the NAC domain transcription factors *NST1* and *NST2* that act as master regulators in cell-wall biosynthesis (Yang and Wang, 2016).

The best match of *Vigun05g273500* against the *P. vulgaris* genome was with *Phvul.005G157600* (E value = 1.22e-146) which is on chr 5 and in position 38337097–38339199, i.e. well within the *qPD5.1-Pv* interval that extends from position 37990066 to position 39153303. However, albeit within *Phvul.005G157600* (*Phaseolus MYB26*) there were diagnostic SNPs between pools of contrasting shattering ability, we found that the GBS SNPs within *Phvul.005G157600* were not among the best associations with shattering variables ($SY_{Y/N}$: $R^2 = 0.052$, $P = 2 \times 10^{-4}$;

C%: $R^2 = 0.079$, $P = 1.13 \times 10^{-5}$; $SH_{Y/N}$+C%: $R^2 = 0.135$, $P = 1.36 \times 10^{-8}$).

Moreover, *qPD5.1-Pv* does not comprise the candidate genes underling three other major QTLs for pod shattering on: (1) chr 1, chr 3, and chr 4 of *V. unguiculata* (Table 3); (2) homologs of the soybean shattering genes *SHAT1-5* and *PDH1* (Table 3); and (3) homologs to the *SHATTERPROOF* gene of *M. truncatula* or to *P. sativum* MTR2 g079050 (Table 3). Furthermore, *qPD5.1-Pv* did not contain genes that were orthologous to other known shattering genes of non-leguminous crop species, such as rice, wheat, barley, and tomato (Table S4).

Therefore, only one correspondence among QTLs of different species was found, this was within the Phaseoleae tribe and at the lowest phylogenetic distance considered (*P. vulgaris* – *V. unguiculata*).

What genes does qPD5.1-Pv contain?. Figure 8 shows the genes comprised in the sub-regions S1, S2, and S3 of qPD5.1-Pv, and the associations with the shattering variables. The distribution of the SNPs that were diagnostic between these pools is provided in Figure S6. Overall, there were 38 SNPs that were best associated with at least one shattering variable. Six of these SNPs were intergenic, and 32 were genic. The latter are distributed across 14 genes, of which 13 were annotated.

LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASES (LRR-RLKs) are highly represented here, and across sub-regions S1, S2, and S3 there are seven LRR-RLKs, Phvul.005G162-200/ 210/ 220, Phvul.005G163900, Phvul.005G164-000/ 700, Phvul.005G165900, which together comprised 14 associations across the five shattering traits (Figure 8). Among these seven LRR-RLKs, Phvul.005G162000 within sub-region S1 is homologous to the immune receptor FLAGELLIN-SENSING 2 (FLS2) of Arabidopsis (Figure 8). This gene was also tagged by pool-seq, which defined several mutations within this gene. These included 20 missense mutations and a start loss (which was classified as of potentially high impact). Regarding the start loss, $Pool_{SH}$ was homozygote as the reference, while $Pool_{NSH}$ was homozygote for the alternative alleles. As indicated by the positive value of the ΔSNP index, ILs of $Pool_{SH}$ inherited the genomic segment from MG38 (i.e., the high-shattering parental line), while ILs of $Pool_{NSH}$ inherited the genomic segment from MIDAS (i.e., the parental line with complete absence of shattering). Therefore, the start loss was of MIDAS (the non-shattering parental line) and of the non-shattering ILs. The remaining six LRR-RLKs did not show other homologies with genes of potential interest for pod shattering. However, among the others, one LRR-RLK in sub-region S2, *Phvul.005G164000*, carried four different SNPs that were among the best associated with the five shattering traits (Figure 8). Moreover, *Phvul.005G164000* is surrounded by other sites and genes that were associated with the shattering traits and renders this gene and its neighborhood interesting for future studies.

Phvul.005G161900, which is annotated as a BASIC HELIX–LOOP–HELIX (b-HLH) TRANSCRIPTION FACTOR, is adjacent to the FLS2 homolog. GBS did not find any SNPs within the b-*HLH* gene, but pool-seq found three polymorphisms that were diagnostic between the shattering and non-shattering pools. Two polymorphisms were in the 5′UTR of the gene, a region that in general might be important for the regulation of translation of a transcript. Among these, one was a premature start codon gain and another was a missense mutation. In Arabidopsis, a b-*HLH* gene, *INDEHISCENT* (*AtIND*), is essential for silique shattering. However, the best match of *AtIND* against the genome of *P. vugaris* was on chr 2 (E = 1E-47; Table 3), while *Phvul.005G161900* was the sixth best match (E = 5E-24) after four matches on chr 2, chr 9, and chr 10 (E values between 3E-47 and 4E-24), where QTLs for pod shattering in common bean were not found (Table 3). Interestingly three other b-*HLH* genes, among which the homologs of *A. thaliana* ALCATRAZ were found by pool-seq, were within qPD1.1-Pv (Table S5a).

An ortholog of *AtDOF4.7* (*Phvul.005G161200*), a gene that in Arabidopsis is probably involved in initiating abscission, was found at 46 319 bp from the b-HLH transcription factor. By pool-seq analysis, this gene contained diagnostic markers between pools, however GBS did not find any SNPs within this gene, and the closest SNP (4633 bp upstream) was not very strongly associated with shattering traits ($SH_{Y/N}$: $R^2 = 0.100$, $P = 9 \times 10^{-4}$; C%: $R^2 = 0.086$, $P = 4.3 \times 10^{-6}$; $SH_{Y/N}$+C%: $R^2 = 0.267$, $P = 6.75 \times 10^{-15}$).

An ATP-BINDING CASSETTE TRANSPORTER (PDR) that carried three strong associations and was the ortholog of *AtPDR1* was found within sub-region S3 (Figure 8). Based on pool-seq, this was the gene with the highest number of diagnostic SNPs between pools among all of the genes comprised within sub-regions S1, S2, and S3. Variants were mainly upstream, in introns, and downstream, with some mutations also observed in the 5′UTR and coding region sequence (CDS). Moreover, at ~30 kb from the above-cited ATP-BINDING CASSETTE TRANSPORTER (PDR), there was a MEKHLA domain gene (*Phvul.005G166900*) that is an ortholog of the *AtHb15* gene, a class III HD-ZIP TRANSCRIPTION FACTOR. There were no SNPs from GBS that covered this gene but, based on pool-seq, this contained diagnostic markers between the shattering and non-shattering pools.

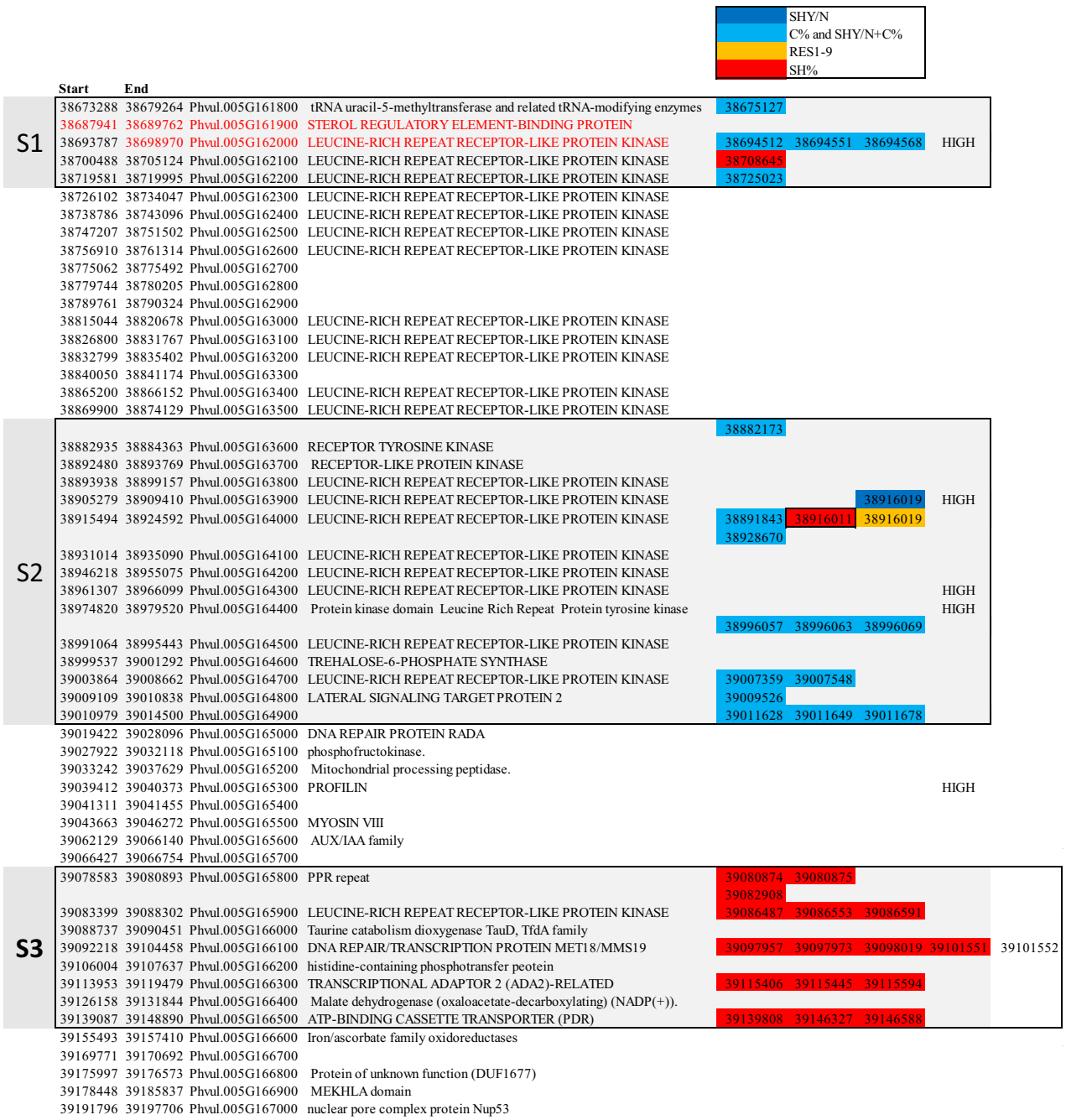
Among the other five genes carrying shattering-associated SNPs, there was a DNAJ HOMOLOG (*Phvul.005G166300*), which has a MYB-like domain (Figure 8).

What genes do the QTLs for shattering level and mode contain?. The results of the survey of gene functions for QTLs for the level and mode of shattering are summarized in Table 4. For RES_{1-9}, the best associated SNP was within

Table 3 Blast search for the shattering genes of leguminous species against *Phaseolus vulgaris*

Source	Reference	Sequence name	Gene	*Phaseolus* chromosome	Locus tag	Location	E value
Vigna unguiculata	Suanum *et al.* (2016)	Vigan 07 g046100	*MYB46/83*	8	Phvul.008G211900	52353945..52364172	0
				6	Phvul.006G009100	4748326..4763798	0
	Suanum *et al.* (2016)	Vigan 01 g359600	**CESA7**	9	Phvul.009G205100	30314266..30319989	0
				3	Phvul.003G154600	35952680..35958487	0
				5	**Phvul.005G022100**	**1967500..1975402**	**0**
				11	Phvul.011G211500	49542339..49550516	0
				2	Phvul.002G268200	43340178..43347095	0
				9	Phvul.009G094200	14499200..14505378	0
				9	Phvul.009G242700	35604655..35611113	0
				2	Phvul.002G188600	34389354..34394716	0
				7	Phvul.007G081700	8016723..8023135	0
				4	Phvul.004G093300	25319515..25328549	0
				9	Phvul.009G090100	13993463..13999583	0
				9	**Phvul.009G205200**	**30325929..30331572**	**0**
				5	**Phvul.005G010400**	**904542..912610**	**0**
				2	Phvul.002G240200	40683746..40689705	0
				7	Phvul.007G190300	42660453..42665383	0
				2	Phvul.002G136300	26843699..26848700	0
				11	Phvul.011G020100	1578058..1583372	0
				2	Phvul.002G040200	3841996..3847312	0
Vigna unguiculata	Lo *et al.* (2018)	Vigun 03 g306000.1	NAC	3	Phvul.003G217600	43378154..43384221	0
				2	Phvul.002G061000	6758318..6763329	1.72e-178
				11	Phvul.011G005700	416772..423422	4.07e-162
				2	Phvul.002G110900	23745541..23750187	2.80e-107
				3	Phvul.003G260500	48840801..48844236	1.70e-106
	Lo *et al.* (2018)	Vigun03 g 302600.1	C2H2-type zinc finger	3	Phvul.003G221000	43812834..43813679	1.29e-180
				2	Phvul.002G058900	6165939..6167517	5.15e-102
				8	Phvul.008G143900	26243012..26244166	4.89e-48
				10	Phvul.010G071300	19145148..19147366	2.76e-41
				9	Phvul.009G070800	11852677..11853726	1.31e-13
	Lo *et al.* (2018)	**Vigun05 g 273500.1**	***MYB26***	**5**	**Phvul.005G157600**	**38337097..38339199**	**1.22e-146**
				10	Phvul.010G137500	40979193..40980996	1.04e-120
				5	Phvul.005G047400	5333946..5335592	4.33e-112
				11	Phvul.011G059800	5205325..5207669	7.17e-44
				11	Phvul.011G212000	49615520..49618217	1.18e-39
Glycine max	Dong *et al.* (2014)	Glyma.16G 019400	*SHAT1–5*	10	Phvul.010G118700	38725057..38727208	0
				5	Phvul.005G044600	4613761..4616889	4.94e-93
				3	Phvul.003G217600	43378154..43384221	6.02e-92
				11	Phvul.011G003900	288840..291061	1.28e-87
				11	Phvul.011G005700	421441..426171	8.74e-87
	Funatsuki *et al.* (2014)	Glyma.16G 141300	*PDH1*	3	Phvul.003G252200	47996213..47997847	0
				2	Phvul.002G027500	2913459..2915012	0
				8	Phvul.008G143400	26081402..26084421	1E-59
				10	Phvul.010G072900	21305619..21309475	3E-59
				10	Phvul.010G072700	20939461..20942539	7E-58
Medicago truncatula	Ferrándiz & Fourquin (2014)	*MTR8 g 007270*	STP (WRKY)	8	Phvul.008G048100	4258051.. 4259563	1.97e-66
				9	Phvul.009G138600	20401295..20404607	2.29e-47
				2	Phvul.002G266400	43213877..43215924	1.38e-45
				10	Phvul.010G046500	7105698..7106642	1.51e-43
				10	Phvul.010G104700	36413066..36414301	2.33e-36
Pisum sativum +++ sativum sativum sativum	Hradilová *et al.* (2017)	*MTR2 g 079050*	PGDB	2	Phvul.002G076700	11373048..11374448	8.04e-19
				2	Phvul.002G076400	11298435..11299694	1.46e-18
				3	Phvul.003G055800	7115837..7117359	3.38e-14
				2	Phvul.002G076600	11322075..11323466	9.80e-12
				5	Phvul.005G073200	12648450..12650220	2.42e-11

Bold indicates the matches between candidate genes and the QTLs mapped in this experiment.



	Start	End							
S1	38673288	38679264	Phvul.005G161800	tRNA uracil-5-methyltransferase and related tRNA-modifying enzymes	38675127				
	38687941	38689762	Phvul.005G161900	STEROL REGULATORY ELEMENT-BINDING PROTEIN					
	38693787	38698970	Phvul.005G162000	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE	38694512	38694551	38694568	HIGH	
	38700488	38705124	Phvul.005G162100	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE	38708645				
	38719581	38719995	Phvul.005G162200	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE	38725023				
	38726102	38734047	Phvul.005G162300	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38738786	38743096	Phvul.005G162400	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38747207	38751502	Phvul.005G162500	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38756910	38761314	Phvul.005G162600	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38775062	38775492	Phvul.005G162700						
	38779744	38780205	Phvul.005G162800						
	38789761	38790324	Phvul.005G162900						
	38815044	38820678	Phvul.005G163000	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38826800	38831767	Phvul.005G163100	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38832799	38835402	Phvul.005G163200	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38840050	38841174	Phvul.005G163300						
	38865200	38866152	Phvul.005G163400	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38869900	38874129	Phvul.005G163500	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
S2					38882173				
	38882935	38884363	Phvul.005G163600	RECEPTOR TYROSINE KINASE					
	38892480	38893769	Phvul.005G163700	RECEPTOR-LIKE PROTEIN KINASE					
	38893938	38899157	Phvul.005G163800	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38905279	38909410	Phvul.005G163900	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE			38916019	HIGH	
	38915494	38924592	Phvul.005G164000	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE	38891843	38916011	38916019		
					38928670				
	38931014	38935090	Phvul.005G164100	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38946218	38955075	Phvul.005G164200	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38961307	38966099	Phvul.005G164300	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE				HIGH	
	38974820	38979520	Phvul.005G164400	Protein kinase domain Leucine Rich Repeat Protein tyrosine kinase				HIGH	
					38996057	38996063	38996069		
	38991064	38995443	Phvul.005G164500	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE					
	38999537	39001292	Phvul.005G164600	TREHALOSE-6-PHOSPHATE SYNTHASE					
	39003864	39008662	Phvul.005G164700	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE	39007359	39007548			
	39009109	39010838	Phvul.005G164800	LATERAL SIGNALING TARGET PROTEIN 2	39009526				
	39010979	39014500	Phvul.005G164900		39011628	39011649	39011678		
	39019422	39028096	Phvul.005G165000	DNA REPAIR PROTEIN RADA					
	39027922	39032118	Phvul.005G165100	phosphofructokinase.					
	39033242	39037629	Phvul.005G165200	Mitochondrial processing peptidase.					
	39039412	39040373	Phvul.005G165300	PROFILIN				HIGH	
	39041311	39041455	Phvul.005G165400						
	39043663	39046272	Phvul.005G165500	MYOSIN VIII					
	39062129	39066140	Phvul.005G165600	AUX/IAA family					
	39066427	39066754	Phvul.005G165700						
S3	39078583	39080893	Phvul.005G165800	PPR repeat	39080874	39080875			
					39082908				
	39083399	39088302	Phvul.005G165900	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE	39086487	39086553	39086591		
	39088737	39090451	Phvul.005G166000	Taurine catabolism dioxygenase TauD, TfdA family					
	39092218	39104458	Phvul.005G166100	DNA REPAIR/TRANSCRIPTION PROTEIN MET18/MMS19	39097957	39097973	39098019	39101551	39101552
	39106004	39107637	Phvul.005G166200	histidine-containing phosphotransfer peotein					
	39113953	39119479	Phvul.005G166300	TRANSCRIPTIONAL ADAPTOR 2 (ADA2)-RELATED	39115406	39115445	39115594		
	39126158	39131844	Phvul.005G166400	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+)).					
	39139087	39148890	Phvul.005G166500	ATP-BINDING CASSETTE TRANSPORTER (PDR)	39139808	39146327	39146588		
	39155493	39157410	Phvul.005G166600	Iron/ascorbate family oxidoreductases					
	39169771	39170692	Phvul.005G166700						
	39175997	39176573	Phvul.005G166800	Protein of unknown function (DUF1677)					
	39178448	39185837	Phvul.005G166900	MEKHLA domain					
	39191796	39197706	Phvul.005G167000	nuclear pore complex protein Nup53					

Figure 8. Genes within the three sub-regions S1, S2, and S3 of *qPD5.1-Pv*, and positions of the GBS SNPs associated with the shattering traits. HIGH, genes with polymorphism classified as of putative high impact based on pool-seq analysis.

a gene for a HOMEOBOX-LEUCINE ZIPPER PROTEIN ATHB-14-RELATED, which is a homolog of *AtHB14*, Arabidopsis *PHABULOSA* (*PHA*).

For SH%, the best associated SNP is within a *PECTINESTERASE/PECTINESTERASE INHIBITOR* gene on chr 5. Moreover, at about 0.1 Mb from this gene and not covered by GBS SNP, a homolog of *A. thaliana* cellulose synthase A7 (*AtCESA7*) was found. The *PECTINESTERASE/PECTINESTERASE INHIBITOR* gene between pools was also found with the SNPs diagnostic within qPD3.1-Pv (Table S5b). There was also a homolog of *AtCESA7* found with the SNPs diagnostic between pools that was within qPD5.2-Pv (Table S5c).

Regarding the mode of pod shattering, the best association with TW%, was found in an intergenic position (Table 4), while for NTW% the best association was with a

Table 4 Summary of the genes with interesting functions matching with the QTLs for the level and mode of pod shattering. In the last column x/y indicate the distance from the peak and the R^2

Component	Trait	Method	Chr	Peak position	Genes	Annotations	R^2
Level, mode	RES_{1-9}	Tassel	6	22061911	–	Intergenic	0.152
		MLMM	6	24311992	Phvul.006G128600	HOMEOBOX-LEUCINE ZIPPER PROTEIN, ortholog toAtHB14 (PHABULOSA)	0.150
			6	23989634	Phvul.006G125100	PROTEIN AAGR-1	0.051
	SH%	Tassel	6	23828803	Phvul.006G122900	(Acetate–CoA ligase / Acyl-activating enzyme // Butyrate–CoA ligase /	0.089
		MLMM	5	1022962	Phvul.005G011900	PECTINESTERASE/PECTINESTERASE INHIBITOR 25-RELATED	0.221
				−105194	Phvul.005G010400	PX-BOX TRANSCRIPTION FACTOR-RELATED homolog to AtCESA7	n.c.
			4	44198457	Phvul.004G160300	K06891 - ATP-dependent Clp protease adaptor protein ClpS (clpS)	0.121
			9	29702346	Phvul.009G200500	PROTEIN F53H1.3, ISOFORM A	0.062
Mode	TW%	Tassel	3	23649132	–	Intergenic	0.100
			(2)	(48261529)	(Phvul.002G32370)	LEUCINE-RICH REPEAT-CONTAINING PROTEIN	0.042
		MLMM	4	44613224	–	Intergenic	0.220
			6	20794299		Intergenic	0.181
	NTW%	Tassel	8	51548749	Phvul.008G204400	ANKYRIN REPEAT FAMILY PROTEIN	0.095
			(2)	(48261529)	(Phvul.002G32370)	LEUCINE-RICH REPEAT-CONTAINING PROTEIN	0.042

gene coding an ANKIRIN REPEAT FAMILY PROTEIN. When SH% was used as a cofactor, both TW% and NTW% were best associated to a polymorphism within a gene coding for LEUCINE-RICH REPEAT-CONTAINING PROTEIN (Table 4).

DISCUSSION

Identification of the genetic basis of shattering is relevant for both evolutionary studies and plant breeding (Swain *et al.*, 2011; Lin *et al.*, 2012; Dong and Wang, 2015; Li and Olsen, 2016). Here, we have dissected out the 'architecture' of the genetic control of pod shattering in common bean through mapping the genes that condition the shattering/non-shattering phenotype, and the level and mode of shattering in common bean. The most characteristic element of this architecture is a major locus on chr 5 that determines whether the pod valves can separate. Two other loci on chr 5 and chr 4, and two additional loci with smaller effects on chr 6 and chr 9, determine the level and the mode of shattering, and explain the phenotypic variation observed in shattering lines. Moreover, we have shown that the pod shattering phenotype depends not only on the single effects of the genes, but also on their epistatic interactions. We also discuss our findings for *P. vulgaris* in comparison with *V. unguiculata* (Suanum *et al.*, 2016; Lo *et al.*, 2018), *G. max* (Dong *et al.*, 2014; Funatsuki *et al.*, 2014), *M. truncatula* (Fourquin *et al.*, 2013; Ferrándiz and Fourquin, 2014) and *P. sativum* (Hradilová *et al.*, 2017), to shed light on the genetic mechanisms of convergent evolution under parallel selection and domestication.

Convergent evolution

Although the domestication genes involved in the shattering of *P. vulgaris* (the present study) and in other leguminous species are generally different, they share the feature of being directly or indirectly involved in a gene network that is related to the regulation of cell-wall deposition and/or lignin patterning (Dong *et al.*, 2014; Suanum *et al.*, 2016; Ballester and Ferrándiz, 2017; Lo *et al.*, 2018) (Figure 9). In this context, the architecture of the genetic control of pod shattering in *P. vulgaris* is particularly 'original'. Convergent domestication often proceeds via mutations at orthologous loci (Lenser and Theiβen, 2013; Ballester and Ferrándiz, 2017). However, in this comparison of *P. vulgaris* with other leguminous crops, we have shown that this was only partially true for *P. vulgaris* and *V. unguiculata*, the two closest crop species in the Phaseoleae tribe, and it was not the case for the more distant bean and soybean. However, within the Viceae tribe, *P. sativum* and *L. culinaris* have a gene that controls pod shattering that maps to a syntenic region, which suggests that the same genes might have been modified during domestication of these two cool-season legumes (Weeden *et al.*, 2002; Weeden, 2007). Overall, this indicates that within leguminous crop species, similar genetic solutions are more likely at close phylogenetic distances (i.e. within the same tribe) but that, overall, leguminous species have often evolved different genetic solutions to the same selective pressure imposed by the agro-ecosystem; i.e. the need to reduce yield losses due to shattering.

The approximate evolutionary distance between *Phaseolus* and *Vigna* is 8 MY, and that between *Phaseolus* and soybean is 19.2 MY (Lavin *et al.*, 2005). In cereals, the *Sh1* genes for seed shattering were under parallel selection during the domestication of sorghum, rice and maize, a correspondence that transcends 65 million years of reproductive isolation (Paterson *et al.*, 1995; Li *et al.*, 2002). This also suggests lower constraints on adaptation within leguminous crop species compared with cereals, and it might be explained as a consequence of the 'contingencies of history' (Gould, 2002): the ancestral populations of different crops might have experienced different constraints prior to passage to the new selective regime of the agroecosystem, and the order in which particular mutations occurred might also have been different in different populations (Losos, 2011). However, Li and Gill (2006) suggested that there are multiple genetic pathways for seed shattering also in grasses, and more recently this was demonstrated in rice (Yoon *et al.*, 2014) and barley (Pourkheirandish *et al.*, 2015).

Pod shattering loss in common bean

A single shattering QTL, qPD5.1-Pv, determines the indehiscent phenotype in common bean. The first relevant outcome of the present study is that pod shattering/non-shattering depends on a single locus that is located on the distal part of the long arm of chr 5, and that interacts epistatically with at least two other loci to modulate the phenotypic expression of the trait. This is consistent for both pool-seq and GBS analyses. This QTL is not involved in the determination of either the 'level' or the 'mode' of pod shattering.

Single locus control of pod shattering was reported for *L. culinaris* (Ladizinsky, 1998), while two loci were found in

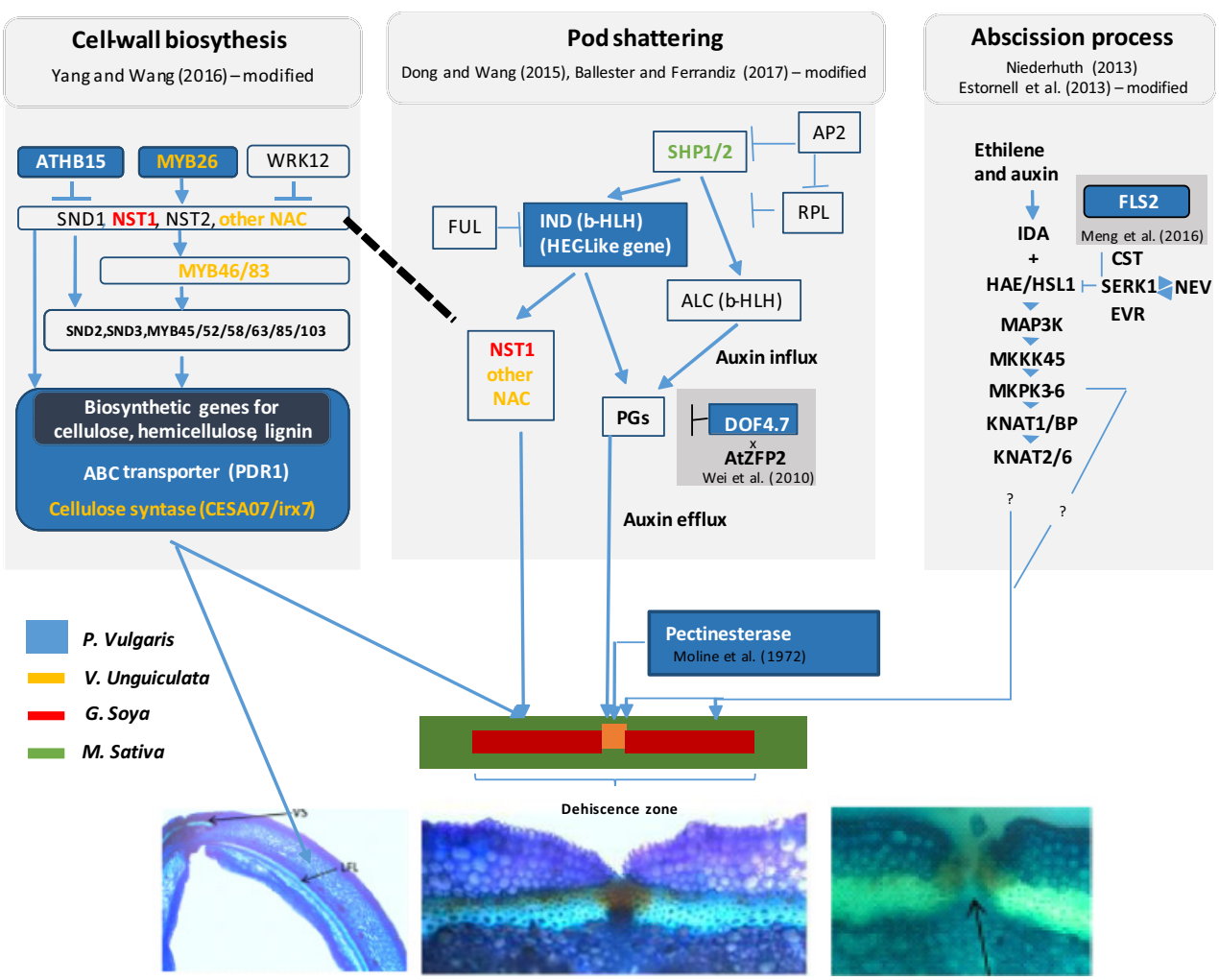


Figure 9. Genes for pod shattering in bean (blue background), *Vigna unguiculata* (yellow font), *G. soya* (red font), and *Medicago truncatula* (green font). Genes are shown in the context of the networks for cell wall, pod shattering, and the abscission process. The image at the bottom of the figure shows the ventral sheath of the pod valves of the MG38 line (highly shattering). Lignified cells are sky blue.

Vigna radiata (Isemura *et al.*, 2012) and *V. unguiculata* subsp. *sesquipedalis* (Kongjaimun *et al.*, 2012). The existence of a 'switching' mechanism as the basis of pod shattering in common bean was predicted by Murgia *et al.* (2017) and was also proposed in pioneering studies by Tshemark (1901, 1902), Emerson (1904), Wellensiek (1922), and Tjebbes and Kooiman (1922). Prakken (1934) used classical genetics experiments to also provide evidence that this trait is under the control of a single locus. Other studies have suggested oligogenic control (Lamprecht, 1932), with the more complex models including epistatic and gene-environmental interactions (Currence, 1930).

Koinange *et al.* (1996) indicated a major locus (*St*) that controlled pod string shattering on chr 2 of the common bean. This finding is not in agreement with our data of a single major locus (qPD5.1-Pv) on chr 5. The *St* locus was mapped in an F2 population that was derived from a cross between the domesticated Andean variety MIDAS and the wild Mesoamerican line G12873 (Koinange *et al.*, 1996). Our population of ILs was derived from the cross MIDAS × MG38, where MG38 is a recombinant inbred line selected from a cross of MIDAS × G12873, and is characterized by wild-like pod/seed traits. Therefore, MG38 inherited the shattering ability from G12873, and in the present study, we expected *a priori* to fit the same genomic region mapped by Koinange *et al.* (1996). The lack of agreement between our study and that of Koinange *et al.* (1996) might be explained by considering the lower marker density of the Koinange study, and the phenotyping method they adopted. Indeed, starting from the assumption that shattering ability is conditioned by the presence of fibers in the pods, in both their sutures ('strings') and walls, Koinange *et al.* (1996) reported 'the presence of fibers in pod sutures and pod walls [...] by breaking the pod beak or pod wall, respectively, and examining the break surface for the presence of fibers.' It is, however, possible that the trait considered by Koinange *et al.* (1996) per se was not necessarily strongly correlated with shattering ability. Working within the domesticated gene pool, a QTL that controls 32% of the total genetic variation for string-to-pod length ratio was found on chr 2 of common bean (Hagerty *et al.*, 2016). It can also be speculated that qPDV-5.1 alleles are 'complex alleles'; i.e. clusters of tightly linked variants. In other words, the large effect of the qPDV-5.1 locus might be due to multiple associated polymorphisms on different genes, rather than to larger individual mutations. Interestingly, in barley, shattering is conferred by mutations in two adjacent, dominant and complementary genes (*Btr1* and *Btr2*) that were also subjected to spatially and temporally independent selection (Pourkheirandish *et al.*, 2015).

qPD5.1-Pv co-maps with a single major QTL that underlies the carbon content of pod valves. Co-mapping between a QTL for shattering and the carbon content has been observed in cross and backcross populations of wild and domesticated *V. unguiculata* (Suanum *et al.*, 2016). Moreover, it was shown that C% is strongly associated with the lignin content of pod valves and that, in turn, high shattering is associated with the high lignin content of the cell wall of pod valves (Murgia *et al.*, 2017). Therefore, our data also indicate that in common bean the genes that underlie pod shattering are likely to be involved in the lignification patterns of the pod valve tissues, and possibly connected to the regulation of the secondary metabolism of the cell wall. This is reinforced by histological analyses in which shattering and non-shattering genotypes showed differential lignification patterns of the pod valve tissues (Prakken, 1934; Murgia *et al.*, 2017). Interestingly, the identified shattering genes of soybean, *SHAT1*-5 (Dong *et al.*, 2014) and *PDH1* (Funatsuki *et al.*, 2014), were both involved in cell-wall lignification, similar to that expected for *V. unguiculata* (Suanum *et al.*, 2016; Lo *et al.*, 2018) and *P. vulgaris* (the present study). Interestingly, in some species of the genus *Medicago*, increased shattering ability has also been associated with increased lignin deposition at the valve margin. This was attributed to a change in the protein sequence of a SHATTERPROOF ortholog (Ferrándiz and Fourquin, 2014).

At least two additional minor QTLs determine the level and maybe the mode of shattering. While the determination of the occurrence of shattering ($SH_{Y/N}$) is conditioned by a single QTL, the control of its level (low versus high shattering; SH%) and mode (twisting versus non-twisting; TW% versus NTW%) might instead be more complex. Two QTLs on chr 5 and chr 4 that also interact epistatically appear to be more relevant, although other QTLs with minor effects have been detected. However, as shown by the heritability values, environmental effects influence the level and mode of shattering much more than its occurrence. This scenario appears to confirm the suggestion of Lamprecht (1932), who hypothesized that a major factor influences the shattering trait, while three other genes act synergistically to tune the expression of the trait, along with more complex models that include epistatic and environmental effects and gene × environment interactions (Currence, 1930; Drijfhout, 1970). However, as indicated for soybean (Dong and Wang, 2015), it is apparent that selection under domestication might have targeted multiple loci for the trait that is also common in bean.

Candidate genes for pod shattering in common bean. Shattering occurrence—Several genes within *qPD5.1-Pv* are homologous to genes implicated in the process of cell-wall biosynthesis, lignin deposition, and organ dehiscence processes. For instance, we found LRR-RLK that was homologous to the immune receptor FLAGELLIN-SENSING 2 (FLS2) of Arabidopsis (Meng *et al.*, 2016). Moreover, *RLK*s can have roles in the induction of abscission, signaling,

membrane trafficking, and post-abscission processes (Cho *et al.*, 2008; Stenvik *et al.*, 2008; Tucker and Yang, 2012; Niederhuth *et al.*, 2013).

We show here that *qPD5.1-Pv* contains a *b-HLH* transcription factor. Three *b-HLH* transcription factors are known to be involved in shattering of Arabidopsis*: INDEHISCENT* (*AtIND*), *ALCATRAZ* (*ALC*) and *SPATULA* (*SPT*) (for review, see Dong and Wang, 2015). Among these three, the *b-HLH* within *qPD5.1-Pv* is best related to *AtIND*, which directs the differentiation of the silique dehiscence zone (Dong and Wang, 2015). As reviewed by Dong and Wang (2015) and Ballester and Ferrándiz (2017), *AtIND* orthologs are specific to Brassicaceae and their role in shattering has been acquired through a recent neofunctionalization that occurred in the Brassicaceae *HECATE3* (*HEC3*) gene clade. However, in Arabidopsis, other HEC genes are involved in genetic routes with similar components to the dehiscence zone network (Ballester and Ferrándiz, 2017). Therefore, the polymorphisms in other HEC genes may be associated with shattering ability (Dong and Wang, 2015), which make them 'still good candidates [. . .] in non-Brassicaceae' (Ballester and Ferrándiz, 2017). Other interesting candidates include a homolog of *AtPDR1* that was shown to transport *p*-coumaryl alcohol, a monolignol lignin precursor (Bienert *et al.*, 2014), and a DNAJ homolog, which contains a MYB-like domain. MYB factors are involved in shattering in both model and crop species, and are known to interact with b-HLH transcription factors (Feller *et al.*, 2011; Dong and Wang, 2015).

Finally, the homologs of the transcription factors *AtDOF4.7* and *AtHb15* merit some attention, although they are not covered by GBS SNPs, they contained diagnostic SNPs between pools. The first of these genes is probably involved in initiating abscission (Wei *et al.*, 2010; Wang *et al.*, 2016), while the latter will have a part to play in the regulation of secondary cell-wall biosynthesis (Yang and Wang, 2016).

Shattering level and mode—Loci implicated in the determination of shattering level and mode contain genes that have annotated functions that reinforce their involvement in cell-wall biosynthesis. For instance, we found that *PECTINESTERASE/PECTINESTERASE INHIBITOR* is involved in the pathway of pectin degradation and in the events leading to chemical and structural alterations of an existing cell wall. Pectins are a family of complex polysaccharides in the cell wall that are important components of the adhesive materials. Pectin methylesterases are associated with the dehiscence zone also in common bean (Moline *et al.*, 1972), and are likely to contribute to the degradation of the middle lamella at valve separation. This will make pectin accessible to other hydrolytic enzymes, including polygalacturonases, for further breakdown. However, their precise requirement in the dehiscence processes has not been characterized yet (Jaradat *et al.*, 2014; Ballester and Ferrándiz, 2017). Other interesting genes include the ortholog of *AtHB14 (PHABULOSA)* that acts upstream in the network of regulation of cell-wall depositions (Yang and Wang, 2016). Close to this gene, we observed a homolog of *AtCESA7*. Interestingly, *AtCESA7* was also found underlying a QTL for pod shattering in *V. unguiculata* (Suanum *et al.*, 2016).

Finally, the positions of the QTLs involved in the level and mode of shattering did not match the important processing traits, such as pod wall fiber, height, width, wall thickness, and length (see Figure 1 of Hagerty *et al.*, 2016).

Domestication and post-domestication shattering genes

The occurrence of several genes involved in the genetic control of the level and mode of shattering in bean suggests that the process of domestication was most likely to be associated to a long sequence of changes from which our fully domesticated crop was derived. This supports (Allaby *et al.*, 2008), the protracted domestication hypothesis as was also suggested from the results of metabolomics changes associated to the domestication of tetrapoloid wheat (Beleggia *et al.*, 2016). Moreover, an interesting question is which gene(s) were first domesticated and which one(s) were selected during the post-domestication process during crop diversification and expansion (Meyer and Purugganan, 2013; Abbo *et al.*, 2014). To fully answer this question, an in-depth analysis of the molecular and phenotypic diversity of the domesticated common bean needs to be done to identify different alleles and their effects at the various genes involved. However, with the exception of snap bean types, most of the common bean varieties have a certain degree of shattering (needed to facilitate seed threshing), which suggests that the 'switching' allele associated with the indehiscent phenotype mapped in our study on *Pv_Ch5* might have appeared relatively late during the post-domestication process, and probably only in the Andean gene pool, as suggested by Gepts (1998). Moreover, even if indehiscent Mesoamerican genotypes are also observed, they are probably derived from hybridisation with an Andean snap bean cultivar (Gepts, 1998). Moreover, Bitocchi *et al.* (2013) observed that a bottleneck occurred before domestication in the Andes that strongly impoverished wild germplasm, leading to a more minor effect of the subsequent domestication bottleneck (i.e. sequential bottleneck). Therefore it is conceivable that, in the Andean beans, the indehiscent phenotype evolved when farmers were selecting on narrow genetic base.

Perspectives

At least two main research questions remain open. First, it will be necessary to close the net around a few candidates,

and ultimately to clone the genes responsible for pod shattering in common bean. This will lead to an understanding of the molecular bases of shattering in common bean, through the definition of the direct main effects of the genes, which will also allow an understanding of the molecular bases of the interactions among the shattering genes.

Second, based on the data presented in the present study, Leguminosae species appear to have evolved different molecular mechanisms that underlie their shattering abilities. This finding begs the question whether different shattering mechanisms have differential associated costs (e.g., physiological). In this regard, Murgia *et al.* (2017) suggested that shattering in common bean comes with a 'cost', as it is associated (albeit weakly) with low pod size, low seed weight per pod, high pod weight, and low seed to pod-valves ratio. Therefore, it might be interesting to compare different legume species in dedicated studies.

EXPERIMENTAL PROCEDURES

Plant materials

Among the 287 ILs, for 257 ILs it was possible to determine both the shattering phenotype and the molecular features. This was a representative fraction of a larger set of about 1200 ILs that were mainly developed by Papa and colleagues (Università Politecnica delle Marche, Ancona, Italy) in collaboration with the Attene group (Università degli Studi di Sassari, Sassari, Italy).

The population was developed starting from a cross between the MG38 line and the Andean MIDAS variety. MG38 is a recombinant inbred line that was obtained from a cross between the G12873 wild Mesoamerican (shattering) genotype and the Andean (non-shattering) MIDAS variety. To obtain the ILs, MG38 was backcrossed with MIDAS as the recurrent parent. Several cycles of backcrossing and selfing were carried out, together with selection for the shattering trait. Among the 257 lines analyzed in the present study, 62 belong to the BC_3/F_4:F_5 families, and 195 to the BC_3/F_6:F_7 families. Therefore, a high homozygosity level is expected within each family.

Phenotyping

The phenotyping under field conditions was carried out in 2014, between May and October (sowing date 19 May 2014). The experiment was conducted at the 'Mauro Deidda' experimental farm (Lat. 41°N, Long. 9°E, 81 m a.s.l.) of the Università degli Studi di Sassari, Sardinia, Italy.

The detailed data for this phenotyping were reported by Murgia *et al.* (2017). Briefly, a layout of eight rows was adopted, with 35–38 holes per row, a distance between rows of 1.5 m, and a distance between holes within the rows of 0.8 m. The positions of the ILs in the field were completely randomized. Each IL was represented by a single plant. A plastic sheet was positioned along each row to facilitate weed control.

Each plant was initially defined in terms of fertile and sterile pods. Fertile pods were further classified into three different typologies: non-shattering; 'fissured' (i.e., with valves that were not perfectly closed along the ventral suture); and fully shattering. The shattering pods were defined as those with non-twisting and twisting valves based on the presence of torsion (twisting/ spiral coiling) of the pod valves after shattering (Lamprecht, 1932). For each plant, the number of pods in each of these categories was counted and expressed as percentages of the total number of fertile pods produced by the plant. The percentage variables of the shattered pods per plant (SH%), non-twisting pods (NTW%), and twisting pods (TW%) were considered for mapping for the present study. Furthermore, for each line separately, non-shattering pods were manipulated by hand to evaluate the 'resistance to manual shattering' on a scale from 1 (very low resistance, when valves abruptly shatter under very light pressure on the distal part of the pod) to 9 (very strong resistance, when valves do not separate, and it was necessary 'to break' them) ($RES_{1–9}$).

The chemical compositions of the pod valves were also investigated. This element composition analysis included the carbon, hydrogen, and nitrogen contents, which were determined using an element analyzer (LECO CHN 628; Leco Corporation, St. Joseph, MI, USA) (Murgia *et al.*, 2017).

Molecular analysis

DNA extraction. Genomic DNA was extracted from the young leaves of each plant, using ~100 mg of tissue. The plant tissue was ground (TissueLyserII; Qiagen), and its DNA was extracted using DNeasy 50 mini plant kits (Qiagen). The quantity and purity of the DNA were determined using a spectrophotometer (Genequant II; Pharmacia Biotech Ltd). The DNA stocks were stored at −20°C until they were processed.

Pool-sequencing analysis. For the pool-seq analysis, two DNA pools were created through the selection of individual samples with contrasting pod shattering phenotypes for the shattering trait. The pool of non-shattering (NSH) individuals ($Pool_{NSH}$) was created by mixing equal amounts of DNA from 27 completely indehiscent plants (i.e., like MIDAS; SH% = 0.0). The second, shattering pool ($Pool_{SH}$) was created by mixing DNA from 30 highly shattering (SH) plants (i.e., SH% between 65% and 82%; as higher than MG38, where SH% is 65%). Within $Pool_{SH}$, the mode of shattering was variable among the plants, with the ratio between non-twisting (NTW%) and twisting (TW%) pods per plant varying from 1:2 to 3:1, respectively. The DNA quality and concentrations were measured by electrophoresis on 0.8% agarose gels, and the final DNA concentration was adjusted to 100 ng μl^{-1}.

The genomic DNA digestion and amplification, the fragment selection, extraction, and amplification, and the sequencing were performed by the NGS Service of the Centre for Functional Genomics of the University of Verona (Verona, Italy). Four libraries were prepared, as two for the pools and two for the parental lines. The two pools (i.e., $Pool_{NSH}$, $Pool_{SH}$) were processed at 65.2× coverage, as also for the MG38 parental line; for MIDAS, the coverage was 2.1×.

The divergence between the two pools was estimated according to the SNP index of the pool of the highly shattering lines minus the SNP index of the pool of non-shattering lines, as the ΔSNP index. The SNP index was calculated as the fraction of reads per position that was attributable to the MG38 parental line. Therefore, a positive ΔSNP index indicates that the lines of $Pool_{SH}$ inherited the genomic segment from MG38 (i.e., high-shattering parental line), and similarly, that the lines of $Pool_{NSH}$ inherited the genomic segment from MIDAS (i.e., the parental with a complete absence of shattering). Conversely, a negative ΔSNP index indicated that the lines of $Pool_{SH}$ inherited the genomic segment from MIDAS, and similarly again, that the lines of $Pool_{NSH}$ inherited the genomic segment from MG38.

All the SNPs were categorized based on their sequence ontology (http://www.sequenceontology.org). The severity of the variant consequence was classified into four categories: high, moderate, low, and modifier (https://www.ensembl.org/Help/Glossary?id=535). The significance of these categories is given in the Results section.

Genotype-by-sequencing analysis. Before library preparation, the quantities and purities of the extracted genomic DNA samples were re-assessed. This was carried out using spectrofluorimetry (Spark 10M multimode microplate reader) with the benzimidazole derivative H33258 (Hoechst). DNA degradation was determined using gel electrophoresis, with ethidium bromide for DNA visualization. The libraries for the NGS were prepared according to the original GBS protocol of Elshire *et al.* (2011), with major adaptations only for the multiplexing and by applying size-selection filtering for fragments of 300–700 bp in length. Multiplexing was achieved as a nested adapter design using 24 barcoded adapters and 13 indices (Illumina; one per 24-sample pool). The libraries were pair-end sequenced as 150 bp in length on a sequencing system (HiSeq 2000; Illumina) in the INRA facility in Toulouse (France). Library preparation was carried out in the SupAgro Facilities of INRA, Montpellier (France). The raw reads were processed with the GATK pipeline by the NGS Service of the Centre for Functional Genomics of the University of Verona (Verona, Italy). All the analyses were performed using version 1.0 of the bean genome. We did not apply a filter for the minimum number of reads, but we adopted the standard 'hard filtering' procedure of GATK 'ReadPosRankSum,' with the following parameters: QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 4.0, HaplotypeScore > 13.0, MQRankSum < −12.5 (https://software.broadinstitute.org/gatk/documentation/article.php?id=3225).

The original SNP count was 170 868. As expected, the average level of heterozygosity was overall low and under control being 5% (markers-wise) and 7% sample-wise. Indeed, to obtain the ILs, MG38 was backcrossed with MIDAS as a recurrent parent, and several cycles of backcrossing and selfing were carried out together with selection for the wild characteristics of the pods and seeds. Therefore, a relatively high level of homozygosity was expected. However, heterozygotes were treated as missing data. Moreover, 23 127 loci were removed as multiallelic. After removing markers and samples with excessive missing, a final dataset of 257 samples and 14 195 markers with a 30% of missing data rate was obtained. This was the subjected to an imputation procedure using Beagle v4.1 software (Browning and Browning, 2007). The final imputed dataset used to conduct the marker-trait association study had 0.89% of heterozygous loci (introduced by Beagle) and 0% of missing data.

Marker-trait association analysis. We determined the intra-chromosomal LD using r^2 corrected by the relatedness of the individuals (r_v^2). This method is implemented in the 'LDcorSV' R package (Mangin *et al.*, 2012). We plotted the LD r^2_v data against the genetic distance, and fitted the LD decay line as in Marroni *et al.* (2011), adapting an R script to our data (https://fabiomarroni.wordpress.com/). The regression function is based on Hill and Weir (1988), and the parameter C was calculated using SneP, a program designed to estimate effective population sizes from genome-wide SNP data or directly from LD levels (Barbato *et al.*, 2015).

To dissect out the genetic architecture of pod shattering, the step-wise approach suggested by Murgia *et al.* (2017) was adopted. For the first step, shattering was considered as a two-state qualitative trait, and the 'occurrence' of shattering was mapped as presence (Yes, SH% > 0) or absence (No; SH% = 0) ($SH_{Y/N}$). For this analysis, all 257 ILs were considered, and the 29 non-shattering versus 228 dehiscent ILs were compared (i.e. shattering ILs were grouped regardless of their 'degree' of shattering). Murgia *et al.* (2017) observed clear-cut differences between shattering and non-shattering ILs in terms of the carbon content of the pod valves (C%) and suggested that the complementation of field characterization and chemical element composition analysis can lead to more precise and alternative or complementary phenotyping. Therefore, this C% trait was mapped here, and the data were compared with those for $SH_{Y/N}$. Furthermore, the K means clustering method at $K = 2$, and considering the two variables $SH_{Y/N}$ and C% identified two groups of ILs with contrasting characteristics. These two groups comprised 48 ILs (G1) and 209 ILs (G2) and, as expected, they were contrasting for both variables (for more details, see Results section). These two groups were considered as variants of a shattering trait that is here referred to as $SH_{Y/N}$+C%, which was also mapped.

For the second step, shattering was considered as a quantitative trait, and the variables of percentage shattering pods per plant ('level'; SH%) and resistance to manual shattering (RES_{1-9}) were mapped considering all of the 257 ILs. These analyses were therefore carried out to identify genes that are involved in the 'occurrence' ($SH_{Y/N}$) and/or 'level' (SH%) of shattering. To disentangle the occurrence and level, the mapping of SH% and RES_{1-9} traits was repeated considering only 210 of the shattering lines. This subset of 210 lines was also used to map the 'twisting' (TW%) and 'non-twisting' (NTW%) pod traits, with or without considering SH% as a cofactor, to allow the level and mode of shattering to be untangled.

TASSEL 5.2.9 (Bradbury *et al.*, 2007) was used to detect possible marker-trait associations. A mixed linear model that accounted for kinship (K) was used to analyze each phenotypic trait, and the final data were also checked based on the QQ-plot data. The Bonferroni-corrected threshold at $P = 0.05$ was used to identify associated loci. MLMM (Segura *et al.*, 2012) was also used, and implemented in R. Compared with traditional single-locus approaches, this MLMM method increases the detection power and reduces the false discovery rate. It can therefore provide better evaluation of the trait architecture. Indeed, MLMM uses a step-wise mixed-model regression with forward inclusion and backward elimination, with re-estimation of the phenotypic variance components of the model at each step (Segura *et al.*, 2012). The model selection criterion adopted was the multiple-Bonferroni criterion (mBonf) defined as the largest model in which all of the cofactors have a *P*-value below a Bonferroni-corrected threshold (threshold used, 0.05; for details, see Segura *et al.*, 2012). Manhattan plots were created using a modified R script based on the related GAPIT (Lipka *et al.*, 2012) and MLMM scripts.

ACKNOWLEDGEMENTS

This study was supported by the Italian Government (PRIN2008 ref. 20083PFSXA_001 and NextBEAN FIRB project RBFR13IDFM_001). DR and MLM express their gratitude for assistance in the field trial from the technicians of the 'Mauro Deidda' experimental farm of the Dipartimento di Agraria of the Università degli Studi di Sassari (Sardinia, Italy).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Designed the experiments: RP and DR. Managed the project: DR, GA, and RP. Wrote the manuscript: DR and MLM.

All co-authors contributed to the drafting and critical revision of the manuscript. Contributed plant materials: EBi, EBe, LN, GA, and RP. Performed phenotypic characterization: MLM, DF, DA, TG, and DR. Sequencing: LM, DS and MD. Mapping: MR and DR. Analyzed and interpreted data: DR, MLM, RP, and GA. Edited the manuscript: DR, MM, MR, EBi, DA, GA, and RP. All authors have approved the final version of the manuscript.

This study includes part of the PhD project carried out by MLM at the Doctoral School of Science and Biotechnology of Agricultural and Forestry Science and Food Production, curriculum Crop Productivity of the University of Sassari (Supervisor: DR; Tutor: GA). MLM gratefully acknowledges the Sardinia Regional Government for partial funding of her PhD scholarship (P.O.R. Sardegna F.S.E. 2007–2013—Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività I.3.1. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013—Axis IV Human Resources, Objective I.3, Line of Activity I.3.1.).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Genetic divergence (ΔSNP index) along chromosome 1 and chromosome 3 between the pools with contrasting shattering ability.

Figure S2. Haplotype structure at the *qPD5.1-Pv* locus (see the relative xlsx file).

Figure S3. Mapping level (RES_{1-9}, SH%) and mode (TW%, NoTW%) of bean pod shattering, as analyzed using both single locus (Tassel) and multi-locus (MLMM) methods.

Figure S4. Genetic models for pod shattering in common bean. The results of partition analysis for 4, 5 and 6 splits.

Figure S5. Distribution of SNP diagnostics between $Pool_{SH}$ and $Pool_{NSH}$ along *qPD5.1-Pv*. The distributions are presented as overall SNP types and for each category. For the meaning of each category, see Experimental procedures. The distribution of the non-synonymous/ synonymous ratio is also provided. The areas shaded in gray are those where significant associations with shattering variables were found (see also Figures 2 and 4b).

Figure S6. Diagnostic SNPs between pools. The genes in sub-regions S1, S2, and S3 of *qPD5.1-Pv* are shown.

Table S1. Two-way ANOVA to test for interactions between alleles at different loci; i.e., the existence of epistatic interactions.

Table S2. Mutation patterns within qPD5.1 QTL based on the pool-seq analysis. Variants were categorized based on their position and assumed impact.

Table S3. Genes with variants classified to be of potentially high impact.

Table S4. List of candidate genes and locations of their best matches on the common bean genome (see relative .xls file).

Table S5. Genes within *qPD1.1-Pv* (A), *qPD3.1-Pv* (B), and *qPD5.2-Pv* (C) that were 'diagnostic' between $Pool_{SH}$ and $Pool_{NSH}$ and that might be involved in shattering, based on their annotated function.

REFERENCES

Abbo, S., van-Oss, R.P., Gopher, A., Saranga, Y., Ofner, I. and Peleg, Z. (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* **19**, 351–360.

Allaby, R.G., Fuller, D.Q. and Brown, T.A. (2008) The genetic expectations of a protracted model for the origins of domesticated crops. *Proc. Natl Acad. Sci. USA*, **105**, 13982–13986.

Ballester, P. and Ferrándiz, C. (2017) Shattering fruits: variations on a dehiscent theme. *Curr. Opin. Plant Biol.* **35**, 68–75.

Barbato, M., Orozco-terWengel, P., Tapio, M. and Bruford, M.W. (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* **6**, 109.

Beleggia, R., Rau, D., Laidò, G. *et al.* (2016) Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol. Biol. Evol.* **33**, 1740–1753.

Bienert, M.D., Baijot, A. and Boutry, M. (2014) ABCG transporters and their role in the biotic stress response. *In Plant ABC Transporters, Signaling and Communication in Plants 22* (Geisler M., ed.). Cham: Springer International Publishing, pp. 137–162.

Bitocchi, E., Bellucci, E., Giardini, A., et al. (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, M.L., Gioia, T., Santo, D., Nanni, L., Attene, G. and Papa, R. (2017) Beans (Phaseolus ssp.) as a model for understanding crop evolution. *Front. Plant Sci.* **8**, 722.

Bordat, A., Savois, V., Nicolas, M. *et al.* (2011) Translational genomics in legumes allowed placing in silico 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. *G3*, **1**, 93–103.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Browning, S.R. and Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097.

Cho, S.K., Larue, C.T., Chevalier, D., Wang, H., Jinn, T.-L., Zhang, S. and Walker, J.C. (2008) Regulation of floral organ abscission in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **105**, 15629–15634.

Choi, H.K., Mun, J.H., Kim, D.J. *et al.* (2004) Estimating genome conservation between crop and model legume species. *Proc. Natl Acad. Sci. USA*, **101**, 15289–15294.

Currence, T.M. (1930) Inheritance studies in *Phaseolus vulgaris*. *Tech. Bull. Minn. Agric. Expt. Stn.* **68**, 1–28.

Doebley, J.F., Gaut, B.S. and Smith, B.D. (2006) The molecular genetics of crop domestication. *Cell*, **127**, 1309–1321.

Dong, Y. and Wang, Y.-Z. (2015) Seed shattering: from models to crops. *Front. Plant Sci.* **6**, 476.

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

Drijfhout, E. (1970) Influence of temperature on string formation of beans (*Phaseolus vulgaris*). *Euphytica*, **19**, 145–151.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Emerson, R.A. (1904) Heredity in bean hybrids. *Ann. Rep. Nebr. Agric. Exp. Stn.* **17**, 33–78.

Feller, A., Machemer, K., Braun, E.L. and Grotewold, E. (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **66**, 94–116.

Ferrándiz, C. and Fourquin, C. (2014) Role of the FUL–SHP network in the evolution of fruit morphology and function. *J. Exp. Bot.* **65**, 4505–4513.

Ferretti, L., Ramos-Onsins, S.E. and Pérez-Enciso, M. (2013) Population genomics from pool sequencing. *Mol. Ecol.* **22**, 5561–5576.

Fourquin, C., del Cerro, C., Victoria, F.C., Vialette-Guiraud, A., de Oliveira, A.C. and Ferrándiz, C. (2013) A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in Medicago genus. *Plant Physiol.* **162**, 907–917.

Funatsuki, H., Suzuki, M., Hirose, A. *et al.* (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc. Natl Acad. Sci. USA*, **111**, 17797–17802.

Gaut, B.S. (2015) Evolution is an experiment: assessing parallelism in crop domestication and experimental evolution: (Nei Lecture, SMBE 2014, Puerto Rico). *Mol. Biol. Evol.* **32**, 1661–1671.

Gepts, P. (1998) Origin and evolution of common bean: past events and recent trends. *Hort. Sci.* **33**, 1124–113.

Gioia, T., Logozzo, G., Kami, J., Spagnoletti Zeuli, P. and Gepts, P. (2012) Identification and characterization of a homologue to the Arabidopsis INDEHISCENT gene in common bean. *J. Hered.* **104**, 273–286.

Gould, S.J. (2002) *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Hagerty, C.H., Cuesta-Marcos, A., Cregan, P., Song, Q., McClean, P. and Myers, J.R. (2016) Mapping snap bean pod and color traits, in a dry bean × snap bean recombinant inbred population. *J. Am. Soc. Hortic. Sci.* **141**, 131–138.

Hill, W.G. and Weir, B.S. (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Pop. Biol.*, **33**, 54–78.

Hradilová, I., Trněný, O., Válková, M. *et al.* (2017) A combined comparative transcriptomic, metabolomic, and anatomical analyses of two key domestication traits: pod dehiscence and seed dormancy in pea (*Pisum* sp.). *Front. Plant Sci.* **8**, 542.

Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., Jo, U., Vaughan, D.A. and Tomooka, N. (2012) Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE*, **7**, e41304.

Jaradat, M.R., Ruegger, M., Bowling, A., Butler, H. and Cutler, A.J. (2014) A comprehensive transcriptome analysis of silique development and dehiscence in Arabidopsis and Brassica integrating genotypic, interspecies and developmental comparisons. *GM Crops Food*, **5**, 302–320.

Koinange, E.M., Singh, S.P. and Gepts, P. (1996) Genetic control of the domestication syndrome in common bean. *Crop Sci.* **36**, 1037–1045.

Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D.A. and Srinives, P. (2012) The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. unguiculata cv.-gr. sesquipedalis. *Ann. Bot.* **109**, 1185–1200.

Ladizinsky, G. (1998) *Plant Evolution under Domestication*. Dordrecht: Kluwer Academic Publishers.

Lamprecht, H. (1932) Beiträge zur genetik von *Phaseolus vulgaris*. *Hereditas*, **16**, 169–211.

Lavin, M., Herendeen, P.S. and Wojciechowski, M.F. (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst. Biol.* **54**, 575–594.

Lenser, T. and Theißen, G. (2013) Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* **18**, 704–714.

Li, W. and Gill, B.S. (2006) Multiple genetic pathways for seed shattering in the grasses. *Funct. Integr. Genomics*, **6**, 300–309.

Li, L.-F. and Olsen, K.M. (2016) To have and to hold: selection for seed and fruit retention during crop domestication. *Curr. Top. Dev. Biol.* **119**, 63–109.

Li, J., Wen, J., Lease, K.A., Doke, J.T., Tax, F.E. and Walker, J.C. (2002) BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell*, **110**, 213–222.

Lin, Z., Li, X., Shannon, L.M. *et al.* (2012) Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* **44**, 720.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z. (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, **28**, 2397–2399.

Lo, S., Munoz-Amatriain, M., Boukar, O. *et al.* (2018) Identification of genetic factors controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp). *Scientific Reports*, **8**, 6261.

Losos, J.B. (2011) Convergence, adaptation, and constraint. *Evolution*, **65**, 1827–1840.

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P. and Cierco-Ayrolles, C. (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, **108** (3), 285.

Marroni, F., Pinosio, S., Zaina, G., Fogolari, F., Felice, N., Cattonaro, F. and Morgante, M. (2011) Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. *Tree Genet. Genomes*, **7**(5), 1011–1023.

Meng, X., Zhou, J., Tang, J., Li, B., de Oliveira, M.V., Chai, J., He, P. and Shan, L. (2016) Ligand-induced receptor-like kinase complex regulates floral organ abscission in Arabidopsis. *Cell Rep.* **14**, 1330–1338.

Meyer, R.S. and Purugganan, M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**, 840.

Moline, H.E., LaMotte, C.E., Gochnauer, C. and McNamer, A. (1972) Further comparative studies of pectin esterase in relation to leaf and flower abscission. *Plant Physiol.* **50**, 655–659.

Morris, S.C. (2008) *The deep structure of biology: is convergence sufficiently ubiquitous to give a directional signal*. Templeton: Foundation Press.

Murgia, M.L., Attene, G., Rodriguez, M. *et al.* (2017) A comprehensive phenotypic investigation of the "pod-shattering syndrome" in common bean. *Front. Plant Sci.* **8**, 251.

Nanni, L., Bitocchi, E., Bellucci, E., Rossi, M., Rau, D., Attene, G., Gepts, P. and Papa, R. (2011) Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **123**, 1341–1357.

Niederhuth, C.E., Cho, S.K., Seitz, K. and Walker, J.C. (2013) Letting go is never easy: abscission and receptor-like protein kinases. *J. Integr. Plant Biol.* **55**, 1251–1263.

Olsen, K.M. and Wendel, J.F. (2013a) A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* **64**, 47–70.

Olsen, K.M. and Wendel, J.F. (2013b) Crop plants as models for understanding plant adaptation and diversification. *Front. Plant Sci.* **4**, 290.

Paterson, A.H., Lin, Y.-R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R., Liu, S.-C., Stansel, J.W. and Irvine, J.E. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science*, **269**, 1714–1718.

Poland, J.A., Brown, P.J., Sorrells, M.E. and Jannink, J.-L. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Pourkheirandish, M., Hensel, G., Kilian, B. *et al.* (2015) Evolution of the grain dispersal system in barley. *Cell*, **162**, 527–539.

Prakken, R. (1934) Inheritance of colours and pod characters in *Phaseolus vulgaris* L. *Genetica*, **16**, 177–296.

Purugganan, M.D. and Fuller, D.Q. (2009) The nature of selection during plant domestication. *Nature*, **457**, 843.

Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, Ü., Long, Q. and Nordborg, M. (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.*, **44**, 825.

Stenvik, G.-E., Tandstad, N.M., Guo, Y., Shi, C.-L., Kristiansen, W., Holmgren, A., Clark, S.E., Aalen, R.B. and Butenko, M.A. (2008) The EPIP peptide of INFLORESCENCE DEFICIENT IN ABSCISSION is sufficient to induce abscission in Arabidopsis through the receptor-like kinases HAESA and HAESA-LIKE2. *Plant Cell*, **20**, 1805–1817.

Suanum, W., Somta, P., Kongjaimun, A., Yimram, T., Kaga, A., Tomooka, N., Takahashi, Y. and Srinives, P. (2016) Co-localization of QTLs for pod fiber content and pod shattering in F2 and backcross populations between yardlong bean and wild cowpea. *Mol. Breed.* **36**, 80.

Swain, S., Kay, P. and Ogawa, M. (2011) Preventing unwanted breakups: using polygalacturonases to regulate cell separation. *Plant Signal. Behav.* **6**, 93–97.

Tang, H., Cuevas, H.E., Das, S. *et al.* (2013) Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. *Proc. Natl Acad. Sci. USA*, **110**, 15824–15829.

Tjebbes, K. and Kooiman, H.N. (1922) Erfelijkheidsonderzoekingen bij boonen. *Genetica*, **4**, 447–456.

Von Tshemark, E. (1901) Weitere Beiträge über Verschiedenwertigkeit der Merkmale bei Kreuzung yon Erbsen und Bohnen. ZeitsChrf. d. landwirtsch. Versuchsw. in *Oesterreich*, **IV**, 641–735.

Von Tshemark, E. (1902) Über die Gesetzmäissige Gestaltungsweise der Mischlinge. ZeitsChrf. d. landwirtsch. Versuchsw. in *Oesterreich*, **V**, 781–861.

Tsuchiya, T. (1987) Physiological and genetic analysis of pod shattering in soybean. *JARQ*, **21**, 166–175.

Tucker, M.L. and Yang, R. (2012) IDA-like gene expression in soybean and tomato leaf abscission and requirement for a diffusible stelar abscission signal. *AoB Plants*, **2012**, pls035. https://doi.org/10.1093/aobpla/pls035

Wang, G.-Q., Wei, P.-C., Tan, F., Yu, M., Zhang, X.-Y., Chen, Q.-J. and Wang, X.-C. (2016) The transcription factor AtDOF4. 7 is involved in ethylene- and IDA-mediated organ abscission in Arabidopsis. *Front. Plant Sci.* **7**, 863.

Weeden, N.F. (2007) Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the 'domestication syndrome' for legumes? *Ann. Bot.* **100**, 1017–1025.

Weeden, N.F., Brauner, S. and Przyborowski, J.A. (2002) Genetic analysis of pod dehiscence in pea (*Pisum sativum* L.). *Cell. Mol. Biol. Lett.* **7**, 657–664.

Wei, P.C., Tan, F., Gao, X.Q., Zhang, X.Q., Wang, G.Q., Xu, H., Li, L.-J., Chen, J. and Wang, X.-C. (2010) Overexpression of AtDOF4. 7, an Arabidopsis DOF family transcription factor, induces floral organ abscission deficiency in Arabidopsis. *Plant Physiol.* **153**, 1031–1045.

Wellensiek, S.J. (1922) De erfelikheid van het al of niet bezit van "draad" bij rassen van *Phaseolus vulgaris*. *Genetica*, **4**, 443–446.

Wilson, Z.A., Song, J., Taylor, B. and Yang, C. (2011) The final split: the regulation of anther dehiscence. *J. Exp. Bot.* **62**, 1633–1649.

Yang, J.H. and Wang, H. (2016) Molecular mechanisms for vascular development and secondary cell wall formation. *Front. Plant Sci.* **7**, 356.

Yoon, J., Cho, L.H., Kim, S.L., Choi, H., Koh, H.J. and An, G. (2014) The BEL1-type homeobox gene SH5 induces seed shattering by enhancing abscission-zone development and inhibiting lignin biosynthesis. *Plant J.* **79**, 717–728.

Zhou, Z., Jiang, Y., Wang, Z. *et al.* (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**, 408.