Journal of Heredity 2013:104(2):273–286
doi:10.1093/jhered/ess102
Advance Access publication December 12, 2012

Identification and Characterization of a Homologue to the *Arabidopsis INDEHISCENT* Gene in Common Bean

Tania Gioia, Giuseppina Logozzo, James Kami, Pierluigi Spagnoletti Zeuli, and Paul Gepts

From the Scuola di Scienze Agrarie, Forestali, Alimentari ed Ambientali, Università degli Studi della Basilicata, Viale dell'Ateneo Lucano 10, 85100 Potenza, Italy (Gioia, Logozzo, Spagnoletti Zeuli); and the Department of Plant Sciences/MS1, University of California, 1 Shields Avenue, Davis, CA 95616–8780 (Kami and Gepts).

Address correspondence to Paul Gepts at the address above, or e-mail: plgepts@ucdavis.edu.

Abstract

Reduction in pod shattering represents a key component of the domestication syndrome in common bean (*Phaseolus vulgaris*) and makes this domesticate dependent upon the farmer for seed dispersal. Attempts to elucidate the genetic control of this process have led to the identification of a major gene (*St*) linked to the presence/absence of pod suture fibers affecting pod shattering. Although *St* has been placed on the common bean genetic map, the sequence and the specific functions of this gene remain unknown. The purpose of this study was to identify a candidate gene for *St*. In *Arabidopsis thaliana*, *INDEHISCENT* gene (*IND*) is the primary factory required for silique shattering. A sequence homologous to *IND* was successfully amplified in *P. vulgaris* and placed on the common bean map using two recombinant inbred populations (BAT93 × Jalo EEP558; Midas × G12873). Although *PvIND* maps near the *St* locus, the lack of complete cosegregation between *PvIND* and *St* and the lack of polymorphisms at the *PvIND* locus correlated with the dehiscent/indehiscent phenotype suggests that *PvIND* may not be directly involved in pod shattering and may not be the gene underlying the *St* locus. However, *PvIND* may be closely linked to an as yet unidentified regulatory element at the *St* locus. Alternatively, a more precise phenotyping method taking into account quantitative trait variation needs to be developed to more accurately map the *St* locus.

Key words: *basic helix-loop-helix (bHLH) protein, candidate gene, molecular markers, nucleotide diversity,* Phaseolus vulgaris, *pod shattering, seed dispersal*

Common bean (*Phaseolus vulgaris* L; 2n = 2x = 22) is one of the most important legumes for direct human consumption worldwide. It represents an essential source of calories, proteins, dietary fibers, minerals, and vitamins for millions of people in both developing and developed countries throughout the world (Broughton et al. 2003).

Based on morphology, seed storage protein variation, biochemical and molecular markers, domestication of *P. vulgaris* occurred independently in the Mesoamerican and Andean areas giving rise to two highly differentiated domesticated gene pools (Gepts and Debouck 1991; Gepts 1998) that are characterized by geographic and partial reproductive isolation (Gepts and Bliss 1985; Paredes and Gepts 1995). Those two gene pools are usually clearly distinguished in common bean collections, either by different kinds of molecular data (Gepts 1988; Koenig and Gepts 1989; Pallottini et al. 2004; McClean and Lee 2007; Kwak and Gepts 2009) or by morphological characters (Singh et al. 1991). During domestication, the common bean has evolved from having pods that shatter due to highly fibrous and parchmented pod walls to pods with less fiber that are less or not subject to shattering (Gepts and Debouck 1991; Gepts 1998). These marked pod organization differences between wild relatives and their crop descendants—together with other traits such as seed dormancy, growth habit, photoperiod sensitivity, and seed type—are collectively called the *domestication syndrome* (Hammer 1984).

Loss or reduction of natural seed dispersal was selected from a dehiscent wild ancestor during domestication. Pods of wild beans have fibers in the pods, both in their sutures (*string*) and the pod walls (Koinange et al. 1996). Complete loss of these fibers leads to indehiscence of the pods and lack of seed dispersal at maturity as observed in many modern snap bean (*stringless*) varieties. Partial or quantitative loss of fibers characterizes dry bean varieties. Domesticated beans have, thus, come to depend upon human intervention for their continued survival. In *P. vulgaris*, the physiology and biochemistry of the seed dispersal process are poorly

understood, although some investigations have been directed toward elucidation of the genetic control of the determination of seed dispersal trait. Koinange et al. (1996) carried out a detailed genetic analysis (including major genes and quantitative trait loci [QTLs]) in a recombinant inbred (RI) population derived from a cross between a domesticated Andean green bean (Midas) and a wild Mesoamerican accession (G12873). They found that the lack of pod suture fibers was controlled by a major gene (*St* locus) on common bean chromosome 2 (Pv02). Lack of pod wall fibers was also controlled by a single gene on chromosome Pv02 and was tightly linked or identical to the *St* gene.

One of the possible strategies to identify *St* is the candidate gene approach. Candidate genes are sequenced genes of known or presumed function segregating at a locus putatively responsible for the variation of the traits of interest and which could correspond to major loci or QTLs (Pflieger et al. 2001). In this context, genes isolated in model species represent putative candidate genes for agronomic species. Most of what we know about fruit dehiscence has been learnt by studying the model plant *Arabidopsis thaliana*. In *Arabidopsis*, several genes have been characterized and found to play important roles in silique opening regulation (Ferrándiz 2002; Liljegren et al. 2004) . Five transcription factors are implicated in the formation of the dehiscent zone in *Arabidopsis* siliques: *FRUITFULL* (*FUL*), *SHATTERPROOF1* (*SHP1*), *SHATTERPROOF2* (*SHP2*), *INDEHISCENT* (*IND*), and *ALCATRAZ* (*ALC*). *IND* was identified as the primary factor required for seed dispersal. *IND* is an atypical basic helix-loop-helix (bHLH) protein that is expressed in narrow strips of tissues closely overlapping with the valve margins in a silique. *IND*, *ALC*, *SHP*, and *FUL* interact to allow differentiation of the lignified valve layer, the spring-loaded mechanism of *Arabidopsis* fruit that promotes opening. Moreover, *IND* acts as the key regulator in a network including *SHP* and *ALC* that controls specification of the valve margin (Liljegren et al. 2004). Furthermore, preliminary observations had tentatively identified sequences homologous to the *Arabidopsis IND* sequence in the proximity of the *St* locus on chromosome Pv02 (Gepts P, unpublished data). Thus, we selected *IND* as a potential candidate gene for *St*. The purpose of this study was to provide further evidence for the role of *PvIND* as a candidate gene for *St* in common bean using genetic mapping and an association study. In this study, we have located *PvIND* on two common bean maps (Midas × G12873: Koinange et al. 1996; BAT93 × Jalo EEP558: Freyre et al. 1998). Furthermore, we have also examined the nucleotide polymorphism of the *IND* homologous gene sequence in a large sample of common bean (n = 157), which included both wild and domesticated accessions from the Andean and Mesoamerican gene pools, as well as green bean accessions, which have succulent pod walls and low pod fiber. Because the *St* gene may represent a typical gene under selection during common bean domestication, the identification of the candidate gene of *St* and analysis of its genetic diversity could improve our understanding of the effect of selection on the gene during common bean domestication.

Table 1 Degenerate primer sequences used to obtain sequence information

Primer	Sequence 5′–3′
INDconting1F	GYTAGGAGCHATGAAGGA
INDconting1R	ACTTRACATADCGAATGG
INDconting2F	GGAGCHATGAAGGARATGATG
INDconting2R	CAGGRACRAGTCTYTYTGRAGGA
INDconting3F	GGTAGGAGCHATGAAGGARATGATG
INDconting3R	CCRGGRACRAGTCTYTGRAGGA

Materials and Methods

Identification of a Common Bean Sequence Homologous to *Arabidopsis IND*

The amino acid sequence of the *Arabidopsis INDEHISCENT* gene (*AtIND;* At4g00120) was used to search the corresponding homologous of *Glycine*, *Lotus*, and *Medicago* by BLAST in GenBank. The sequences were aligned with ClustalX (Thompson et al. 1997). Using the online version of the Primer3 software (Rozen and Skaletsky 2000; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), degenerate primers were designed based on conserved motif among the sequences from those species (Table 1).

Primers were tested in a preliminary sample of four individuals: BAT93, G12873, Jalo EEP558, and Midas (parents of the mapping populations used in this study). PCR conditions were adjusted empirically to obtain mainly fragments of the expected size. The PCR products were analyzed in a 2% agarose gel in 1× TBE (45 mM Tris-base, 45 mM boric acid, 1 mM EDTA, pH 8.0), and bands of the expected size were excised and gel purified with the QIAquick Gel Extraction kit (QIAGEN, Valencia, CA). Sequence analyses of the PCR-amplified fragments revealed that a fragment with high homology with a conserved region of *AtIND* had been amplified in common bean (henceforth called *PvIND).*

Sequence of this fragment was used to design *PvIND*-specific primers targeting the coding region (5' TCAGACGAATATCTCAGGATGAA 3' and 5' TATCCTAA CAAACCTTTCCCCTCAC 3'). Additional primers were then used to extend the genomic region flanking the coding region by identifying *PvIND* sequence-containing clones in a bacterial artificial chromosome (BAC) library of the common bean inbred line BAT93 developed by Kami et al. (2006). The library contains 110 592 clones with an average insert size of 125 kb and an estimated coverage of 20.8 haploid genome equivalents. PCR-based BAC-library screening and BAC DNA pool preparation were done as described in Yim et al. (2007). BAC plasmid DNA was isolated using Qiagen Plasmid Midi kit (QIAGEN). DNA was resuspended in 10 mM Tris–HCl (pH 8.5). Insert size and approximate DNA concentration of the identified BACs were estimated from *Not*I digests of BAC DNA (Marek and Shoemaker 1997). Individual BAC identity was confirmed by PCR using BAC DNA as template. BAC-end sequences were generated using the BAC DNA as template with the universal primers T7 (5' TAATACGACTCACTATAGGG 3') and M13 reverse (5' CAGGAAACAGCTATGACC 3') following the methods of Ammiraju et al. (2006).

Plant Materials

A collection of 157 *P. vulgaris* genotypes, including 77 wild and 80 domesticated accessions, and 4 accessions of closely related *Phaseolus* species were used to evaluate nucleotide variation at the candidate gene locus (Table 2). The accessions were chosen to be representative of the major steps of the evolutionary history of *P. vulgaris*: they included wild forms from Ecuador and northern Peru characterized by the ancestral type I phaseolin (Debouck et al. 1993; Kami et al. 1995), Mesoamerican and Andean wild and domesticated forms, and modern dry and green bean cultivars grown in the United States. The *P. vulgaris*, *P. acutifolius*, and *P. lunatus* accessions were obtained from the *Phaseolus* World Collection at CIAT (Cali, Colombia), the Plant Introduction Station of the USDA at Pullman, WA, and from collaborators of the Bean Coordinated Agricultural Project (BeanCAP). The *P. coccineus* accessions were obtained from Johnsons Seed (Kentford, Suffolk, UK). Passport information (e.g., race type, phaseolin type, seed color and pattern, seed weight) was available for the entire sample of genotypes. Because of the exchange between the Andean and Mesoamerican gene pools after domestication and the subsequent dissemination across different regions or continents, the geographic origin of individual accessions is not a reliable indicator of the origin of domestication (Andean vs. Mesoamerican; Gepts et al. 1986; Gepts 1998). Thus, the gene pool designation was based not only on passport data but also on previous studies of the molecular diversity of common bean (Beebe et al. 2000, 2001; McClean et al. 2004; Blair et al. 2006; Díaz and Blair 2006; McClean and Lee 2007). According to this information, out of the 80 domesticated accessions used in this study, 44 were identified as Mesoamerican and 36 as Andean (Supplementary Table S1). Furthermore, 105 common bean accessions out of 157 were scored phenotypically for 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 (Koinange et al. 1996).

For mapping the candidate gene onto the common bean molecular linkage map, two RI populations were used: BAT93 × Jalo EEP 558 (BJ, $n = 80$) and Midas × G12873 (MG, $n = 58$). The BJ RI is the core linkage mapping population of the common bean (Freyre et al. 1998), whereas the MG RI population was developed by Koinange et al. (1996) to investigate the inheritance of domestication syndrome traits in common bean. The MG population shows several segregating traits related to domestication, including seed dispersal (*Sd*), which does not segregate in the BJ population. Nevertheless, because the map of MG is a low-density map, the homologous sequence was mapped in both the BJ and MG populations. Seeds of the mapping population are maintained at the University of California.

Table 2 Number of accessions of *Phaseolus* used in this study

Phaseolus species	No. of accessions
P. vulgaris	157
Wild	77
Domesticated	80
P. acutifolius	1
P. lunatus	1
P. coccineus	2
Total	161

DNA Isolation, Primer Design, PCR Amplification, and Sequence Analysis

Tissue samples were taken from young, healthy leaves of greenhouse-grown plants and stored overnight at –80 °C. The frozen leaf tissue samples were lyophilized in a VirTis Sentry 2.0 for approximately 48 h and ground to a fine powder. Genomic DNA was then extracted from the leaf tissue powder using the DNeasy Plant kit (QIAGEN) and following the protocols provided by the manufacturer. DNA concentration was determined using a DYNA Quant 200 fluorometer (Hoefer Pharmacia Biotech, San Francisco, CA). The final concentration of genomic DNA was adjusted to 20 ng μL^{-1} in Tris–EDTA buffer (pH 8.0) and then stored at –20 °C until use.

Based on the nucleotide sequence of the *PvIND* homologous fragments, specific primers were designed for amplification of the candidate gene from parental lines of the mapping population (Table 3). Genomic fragments were amplified using PCR with *Taq* DNA polymerase (New England Biolabs, Ipswich, MA). PCR reaction mixtures contained approximately 50 ng of total DNA, 0.2 mM of

Table 3 Primer sequences used for genotyping

Primer	Primer sequence 5′–3′	Product size (bp)
pvIND1F-R	F: TCAGACGAATATCTCAGGATGAA R: TGGTGTGATCATATGCTTGAAA	1060
INDcont1F-3R	F: AGTAAAAACCCGTTCCTAATACTTC R: GAAGAAGGTTGGGGTTGTGA	700
INDcont2F-1R	F: TTCCTTTTCACCCAATTAATCTTT R: TATCCTAACAAACCTTTCCCCTCAC	1000
pvIND2FA-2RA	F: TGTGACTTTGAGGAGGAGGCAATC R: CAAGCGATTAGTTGCTTCAGA	850
pvIND4FA-4R	F: TTGACCATTTAATATGCTGTTTTTCT R: CCGGATCTCACTTTTGAAACA	966
pvIND5FA-5RA	F: ATGCCATCTAGGCATTGGAT R: TTCACCGAAAATTTGCCATA	804

dNTP, 0.2 μM of forward and reverse primers, standard *Taq* buffer with 1.5 mM $MgCl_2$, and 1 unit of *Taq* polymerase in a total volume of 30 μL reaction. The PCR cycle consisted of 5 min at 94 °C and 35 cycles of 40 s at 94 °C, 1 min at 56 °C, and lastly, 2 min at 72 °C followed by a 5 min extension at 72 °C. PCR products were separated by electrophoresis in a 1.5% agarose gels in 1× TBE and then visualized by ethidium bromide staining. Bands of the expected size were sequenced, directly from the PCR product after purification. The PCR product was purified using the QIAquick PCR Purification kit (QIAGEN). Fragments were sequenced in both directions by the DNA core facility in the Division of Biological Sciences at the University of California in Davis using an ABI PRISM 3700 sequencer (Applied Biosystems) and the BigDye-terminator v3.1 chemistry. To ensure accuracy of low-frequency polymorphic nucleotides, individual sequence chromatograms were assembled and edited using the PREGAP4 and GAP4 software from the Staden package (Staden 1996). Following corrections, the sequences were aligned using the ClustalX algorithm (Thompson et al. 1997) as implemented in BioEdit version 7.0 (Hall 1999). If necessary, additional manual refinements were made. A GenBank BLASTx search was conducted on the nucleotide sequences to confirm homology with *AtIND*. Gene structure was predicted using the software FGENESH (http://linux1.softberry.com/berry.phtml), whereas the search for ORFs was performed using ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/). To identify homology with peptide sequences, a tBLASTn analysis was carried out against the TIGR Plant Transcript Assembly (http://blast.jcvi.org/euk-blast/plantta_blast.cgi) using the *Arabidopsis* database.

Linkage Mapping

Single nucleotide polymorphisms (SNPs) creating a differential restriction enzyme recognition sites between parental lines of the two mapping population BJ and MG were used to develop cleaved amplified polymorphic sequence (CAPS) markers. Genomic fragments of the candidate gene amplified in the BJ population (primers listed in Table 3) were further digested with the *Msp*I restriction endonuclease, whereas the fragments of the MG population were digested with *Bsm*I. Digestions followed the manufacturer's instructions. CAPS markers were separated on 2% agarose gels, stained, and analyzed for the presence or absence of specific alleles.

Segregation data for the CAPS markers were added to the corresponding data set of the markers already mapped on the two RI populations (Koinange et al. 1996; Freyre et al. 1998). Linkage mapping was performed using the software package MapDisto 1.7 (Lorieux 2012). Assignment of the candidate gene to the common bean chromosomes was done with the FIND GROUPS command with a logarithm of the odd's ration (LOD) ≥ 3.0 The commands ORDER SEQUENCE, RIPPLE, and CHECK INVERSIONS were then used to assign the position of the candidate gene in the previously established marker order. Recombination units were converted into map distances using the Kosambi function (Kosambi 1944).

Nucleotide Diversity and Neutrality Test

Nucleotide diversity analysis was performed for the 161 genotypes (Table 2) using three pairs of PCR primers: pvIND1F-R, INDcont1F-3R, and INDcont2F-1R (Table 3). The DnaSP version 5.10 (Librado and Rozas 2009) software package was used to calculate population genetics parameters. Polymorphic sites (*S*), the number of singleton variable sites, the number of haplotypes, and haplotype diversity (*Hd*; Nei 1987) were estimated. Two parameters of nucleotide diversity were computed: π, the expected heterozygosity per nucleotide site (Nei 1987) derived from the average number of sequence differences in the sample, and θ_w, Watterson's theta estimator, an estimate of the neutral mutation parameter $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per nucleotide (Watterson 1975). Although DnaSP does not deal with indels, this situation was unlikely to have affected our analysis as only a single indel, corresponding to the loss of a codon (without impacting on the reading frame), was found in the coding region for 42 wild and 36 domesticated Andean accessions. Selection on the *PvIND* homologous gene was tested using multiple methods. Excess and deficiency of mutations were tested at segregating nucleotide sites with Tajima's *D* (Tajima 1989) and Fu and Li's *D** and *F** (Fu and Li 1993). Tajima's *D* measures the difference between π and θ_w and tends to be negative when there is an excess of low-frequency variants in the sample and positive when there is an overrepresentation of intermediate frequency variants. Fu and Li's *D** and *F** test the discrepancy between the number of polymorphic sites in external branches (polymorphism unique to an extant sequence) and number of polymorphic sites in internal phylogenetic branches (polymorphism shared by extant sequences). These tests were calculated at all polymorphic sites. Fu's F_S was applied to test the neutrality of mutations (Fu 1997). The hitchhiking effect was tested using the Fay and Wu *H* test (Fay and Wu 2000) and *P. acutifolius* as an outgroup. The Fay and Wu *H* test predicts positive selection on the basis of a site-frequency spectrum comparing the proportion of alleles of intermediate versus high frequency and is indicated by a negative H test (Fay and Wu 2000). Significance levels of Fay and Wu *H* test were calculated using 1000 coalescent simulations, with no recombination and a nominal threshold of $P = 0.05$ (Hudson 2000). The minimum number of recombination events was estimated using the four-gamete test (Hudson and Kaplan 1985).

Furthermore, selection across the coding region of the *PvIND* protein was studied by estimation of rates of nonsynonymous (dN) and synonymous (dS) substitutions at each codon site. A dN/dS ratio < 1 indicates negative selection and a conserved gene, whereas a dN/dS ratio > 1 indicates positive selection and a gene susceptible to rapid evolutionary change. To assess the role of selection among *PvIND*, we performed two tests of selection. We first used the PARtitioning approach for Robust Inference of Selection (PARRIS) method to determine if positive selection was detected in the entire alignment. For this analysis, we used the same data set used for the phylogenetic and nucleotide diversity analysis (157 sequences of *P. vulgaris*). The PARRIS method test was run on the HyPhy software package hosted

at the Datamonkey server. Second, we tested for residue-specific positive selection using three complementary maximum-likelihood methods (Pond and Frost 2005), all available at the Datamonkey web interface: single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), and random effects likelihood (REL). All three methods incorporate flexible models of nucleotide substitution bias and variation in both nonsynonymous and synonymous substitution rates across sites, facilitating the comparison between the methods (Kosakovsky Pond and Frost 2005).

To test for differentiation between the wild and domesticated populations from the different gene pools, the F_{ST} statistic (Hudson et al. 1992) permutation test with 1000 replicates (Hudson 2000) was computed. F_{ST} reflects differences in allele frequencies among samples and increases as allele frequency differences between population samples become more pronounced (Wright 1950). All calculations were performed using DnaSP (Librado and Rozas 2009).

Analysis of Genealogy and Population Structure

The genetic divergences among 157 *P. vulgaris*, two *P. coccineus*, one *P. acutifolius,* and one *P. lunatus* accessions were calculated by MEGA software version 4 (Tamura et al. 2007) according to the Kimura two-parameter distances model (Kimura 1980). Based on the pairwise nucleotide sequence divergences, a neighbor-joining (NJ) tree was constructed. A bootstrap analysis of the inferred NJ tree was performed with 1000 resampling. All nucleotide positions containing gaps and missing bases were eliminated from the data set (the complete deletion option). NETWORK (version 4.6.0.0; available at http://www.fluxus-engineering.com/sharenet.htm) was used for reconstructing a median-joining (MJ) network (Bandelt et al. 1999).

To further identify the genetic population structure of our sample (157 *P. vulgaris* accessions), we conducted a Bayesian clustering analysis method described in Corander et al. (2003). The software implementation of this method, Bayesian Analysis of Population Structure (BAPs) version 5.3, identifies hidden population structure by clustering individuals into K genetically distinguishable groups on the basis of nucleotide frequencies (Corander et al 2003; Corander and Marttinen 2006; Corander and Tang 2007).

We carried out a genetic mixture analysis to determine the most probable number of populations (K) for all of the 157 accessions. We estimated individual clustering using K values ranging from 1 to 10 as the assumed maximum number of populations present in the sample. The admixture analysis was then applied to estimate individual admixture proportions with regards to the most likely number of K clusters identified. We used 200 reference individuals/population and 100 iterations to estimate the admixture coefficients of the reference individuals. Each run was repeated three times to judge the consistency of the simulation results.

Association Analysis

To identify significant associations with pod fiber presence of absence, a trait association analysis was conducted on 105 accessions using the *PvIND* sequence with 2043 sites. A total of 59 Simple Sequence Repeats markers (Gioia T, Gepts P, unpublished data) and phenotypic data for seed dispersal traits from the 105 accessions were used for association analyses. Population structure was performed using the program STRUCTURE version 2.2 (Pritchard et al. 2000). Bayesian clustering analyses with the admixture models were used. We set K (the number of subpopulations) from 1 to 10 and performed 10 runs for each K value. For each run, a burn in of 5000 iterations was followed by an additional 50 000 iterations. To detect the optimal clustering number (K), the ad hoc statistic ΔK, which is based on the rate of change in the log probability of data between successive K values, was calculated using the STRUCTURE-sum program (Evanno et al. 2005). ΔK identified the optimal K number as 3 based on the lowest ΔK value and no further significant decrease beyond K = 3. The Q matrix, which describes the percentage of subpopulation parentage for each accession, was further incorporated into the association mapping models where the effect of population structure was considered. A general linear model (GLM) was applied for analysis in TASSEL software version 2.1 (Yu et al. 2006; Bradbury et al. 2007). The GLM analyses were performed using the population structure Q matrix based on 59 SSR data, phenotypic data from 105 accessions, and 22 SNP. Positive tests were reported using a significance threshold of $P < 4.5 \times 10^{-4}$, based on a stringent Bonferroni correction of 1% divided by 22 SNPs tested.

Results

Sequence of a Common Bean Gene Homologous to *Arabidopsis IND*

Using the degenerated primers described in the Materials and Methods, we isolated three genomic fragments from common bean. Fragment identities were determined with the BLAST algorithm by searches in nonredundant databases. One of the fragments showed sequence homology with a conserved region of the bHLH domain of *AtIND*.

Additional primers were then used to extend the coding region by identifying clones that contain *PvIND* sequences in a BAC library of the common bean inbred line BAT93 (Kami et al. 2006). Two positive BAC clones were identified (19D10 and 22G24). Sequencing both ends of the 19D10 clone yielded a total of 5280 bp. The *PvIND* gene (GenBank accession number KC192374) is predicted to encode a 282-aa product. The predicted polypeptide was found to be highly similar to *AtIND* (e value 1.9×10^{-35}). Like *AtIND*, the *PvIND* homologue is composed of a single small exon. The *AtIND* and *PvIND* gene products share extensive protein sequence similarity in the bHLH domain, as well as in a 30-aa N-terminal extension of this region. Similarity in this 30-aa region leads to *PvIND* being grouped as a member of a subfamily of bHLH proteins (Heim et al. 2003). Protein alignment of the *PvIND* with *AtIND* and more distantly related bHLH factors reveals that *PvIND*, like *AtIND*, lacks a conserved glutamate at position 9 of the basic domain. This glutamate has been shown to be important for DNA

binding (Ellenberger et al. 1994; Ma et al. 1994). Thus, if the *PvIND* and *AtIND* regulate gene transcription through DNA binding, they are likely to do so through the use of other residues (Gremski et al. 2007). These results convinced us that we had successfully sequenced a common bean sequence highly homologous to the *Arabidopsis IND* gene.

Molecular Mapping and Detection of Polymorphisms within the Candidate Gene

Sequence analysis of the genomic fragments amplified from parental lines of the BJ and MG RI populations revealed that most of the polymorphisms involved base substitutions, resulting in the gain or loss of a restriction site rather than length variation. Thus, it was possible to identify SNPs within enzyme recognition sites and enabling their conversion into two CAPS markers. These two CAPS markers were then used to map the *PvIND* sequence homologues of *AtIND* into the common bean molecular linkage map. Two RI populations were used to integrate the *PvIND*, namely BJ and MG (Koinange et al. 1996; Freyre et al. 1998). Segregation of these markers in the RI populations was determined by a test of goodness-of-fit to a 1:1 ratio (χ^2) at a significance level of $P = 0.05$. Of the 64 BJ lines tested, 27 showed a BAT93 allele and 37 showed a Jalo EEP558 allele ($\chi^2 = 1.56$; no significant difference from a 1:1 segregation ratio). Whereas, of the 51 MG lines tested, 18 showed a *Midas* allele and 33 showed a G12873 allele ($\chi^2 = 4.41$; significantly different from a 1:1 ratio at $P = 0.05$). In both the BJ and the MG maps, *PvIND* was located on chromosome Pv02 (Figure 1). In the MG population, the *PvIND* locus was located close to the *St* locus for seed dispersal (mapped by Koinange et al. 1996), but did not cosegregate with it ($r = 0.078$; LOD score = 6.49).

Genealogical Analysis and Population Structure

Using the software MEGA 4.0 and a bootstrap test with 1000 replicates based on the SNPs of the *PvIND* gene, we were able to estimate the phylogenetic relationships of the 157 common bean accessions for this gene with two *P. coccineus*, one *P. lunatus,* and one *P. acutifolius* accessions used as outgroup (Figure 2). Members of the *P. coccineus*, *P. lunatus,* and *P. acutifolius* species were included into a separate cluster, with bootstrap values as high as 99–100%, whereas the *P. vulgaris* accessions were grouped into five weakly supported subgroups, with bootstrap values ≤70% in most cases. This structure of genetic diversity agrees with previous information on the evolution of *P. vulgaris* (Kami et al. 1995; Freyre et al. 1996) Within *P. vulgaris,* clusters from the Mesoamerican and Andean gene pools grouped together and formed subclusters well separated from each other, with several exceptions. In the Mesoamerican cluster, three subclusters were identified (Figure 2). The first Mesoamerican subcluster (C1) (97% bootstrap value) is composed of 45 domesticated genotypes and 4 wild genotypes from Mexico and Colombia. Furthermore, C1 is divided into three main parts. The first part of this group comprises medium-seeded Pinto and Great Northern accession of race Durango; the second group includes small-seeded Navy and Black beans, whereas the third group is a mix between the two races.

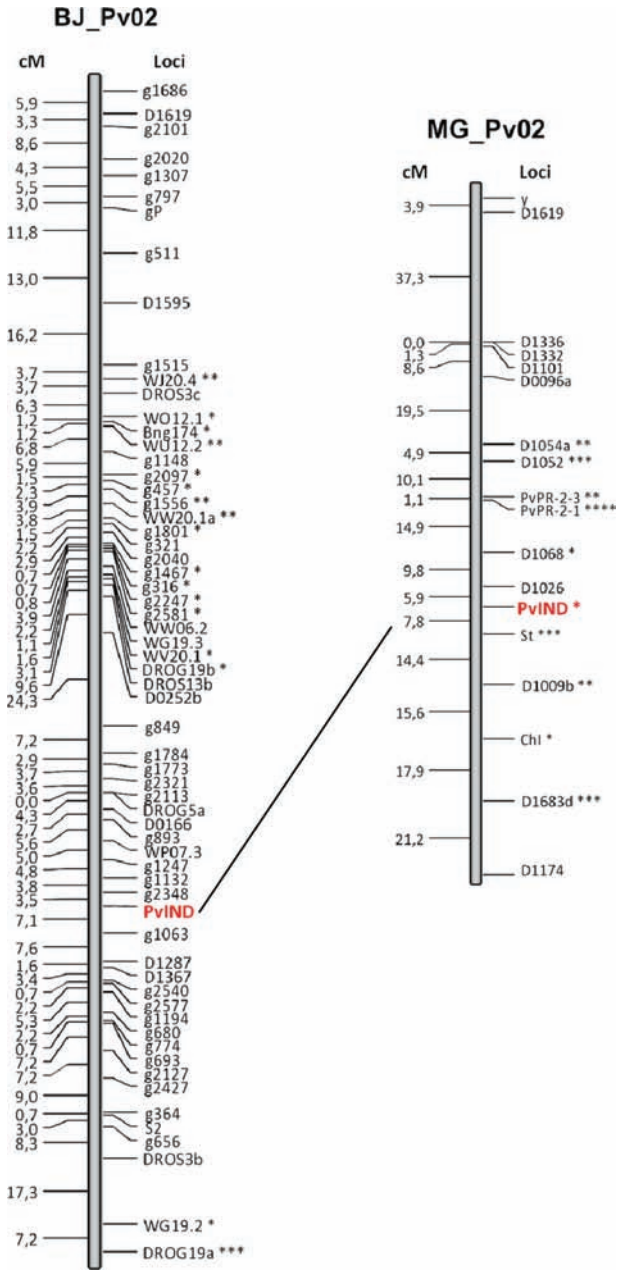

Figure 1. Pv02 chromosome: position of *PvIND* homologues to the *Arabidopsis IND* gene. BJ: BAT93 × Jalo EEP 558; MG: Midas × G12873. Genetic distances in Kosambi units.

The second Mesoamerican subcluster (C2) (66% bootstrap value) consists of four wild genotypes, all from Mexico. The third Mesoamerican subcluster (C3) (94% bootstrap value) was composed of four wild genotypes, three from Mexico, and one from northern Peru (the latter characterized by the ancestral I phaseolin type). In the Andean cluster, two main subclusters were identified. The first Andean subcluster (C4) (<60% bootstrap value) was composed of 23 wild accessions: 14 accessions from south and central Peru and 9 accessions from Mexico, Colombia, and Guatemala, identified as Mesoamerican accessions. In the second Andean subcluster

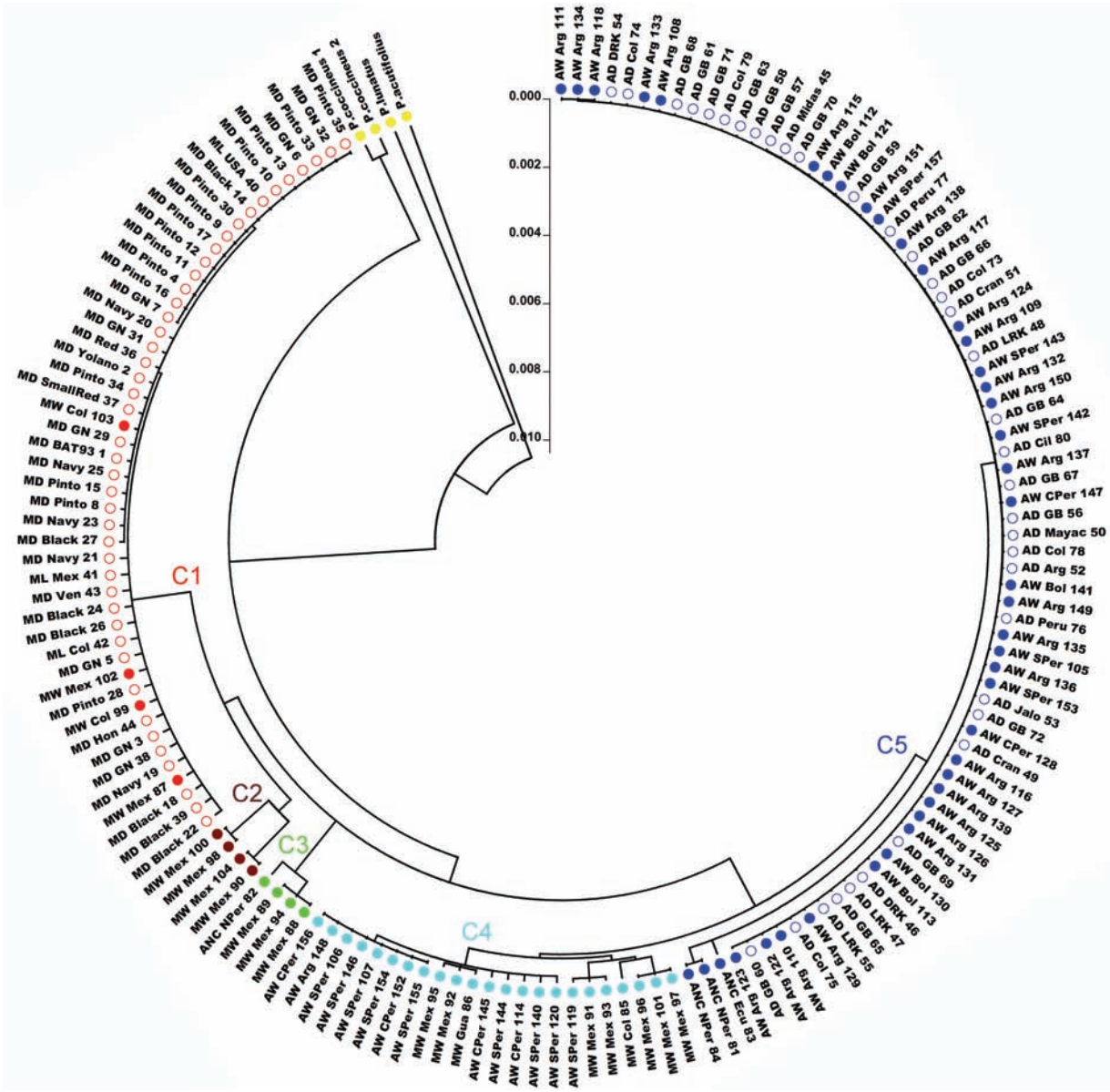

Figure 2. Unrooted NJ tree with 1000 bootstrapped Kimura two-parameter distances based on *PvIND* sequences data of the 157 genotypes of *P. vulgaris*, 2 accessions of *P. coccineus,* and 1 accession of *P. lunatus* and *P. acutifolius*. The main subclusters of *P. vulgaris* genotypes are indicated: in red, C1; brown, C2; green, C3; light blue, C4; blue, C5. Open dots represent domesticated accessions, and full dots represent wild accessions. For each accession, the first letter of the label indicates the gene pool (M = Mesoamerican; A = Andean; ANC = Ancestral), and the second letter indicates the form (W = wild; L = landraces; D = domesticated).

(C5) (69% bootstrap value), a total of 78 Andean accessions were identified. C5 is further divided into two groups. The first of these groups comprises three wild accessions from northern Peru and Ecuador, characterized by the I phaseolin, whereas the second group includes 36 domesticated and 39 wild accessions from the Andean gene pool.

The population structure based on haplotype data from the *PvIND* locus was also investigated. We used the software BAPs to assign each domesticated accession and wild individual to a cluster of origin without prior information regarding the geographic origin of individual samples. This analysis suggested that our sample is most likely made up of five genetically distinguishable subgroups (K = 5) as shown in Figure 3. Comparing the BAPs assignment results to the NJ groups, we found that the BAPs analysis parallels the results of the NJ genealogical analysis with no further information on the population structure.

Nucleotide Diversity and Neutrality Test of *PvIND*

To better understand the level of nucleotide diversity at the *PvIND* gene, we isolated and sequenced a region of 2043 bp

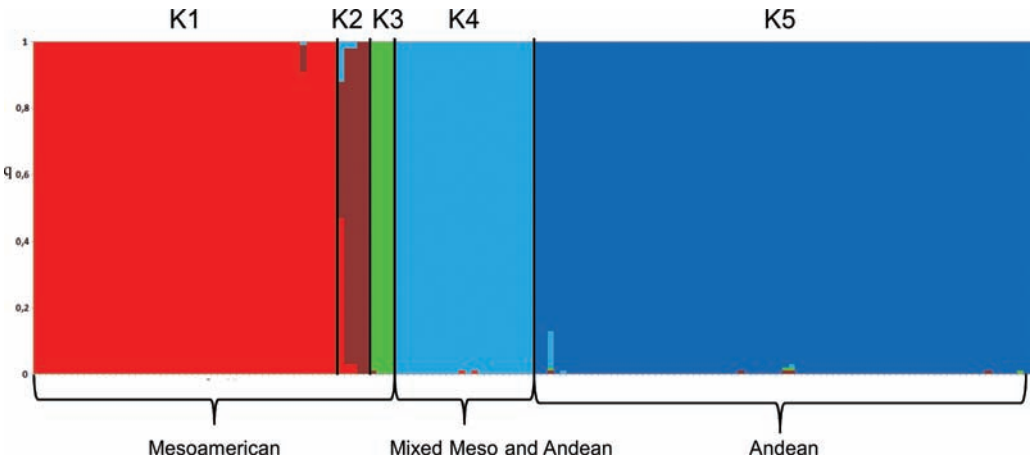

Figure 3. Population structure of *P. vulgaris* samples as estimated with the BAPs software. Each accession is represented by a vertical histogram partitioned into K = 5 colored segments that represent the estimated membership of each individual. Accessions were ordered by gene pool (Mesoamerican and Andean). Clusters are colored as in the NJ tree.

using three PCR primer pairs: pvIND1F-R, INDcont1F-3R, and INDcont2F-1R (Table 3). The length of the aligned sequence included 849 bp of the coding region, 1052 bp of 5' upstream region, and 142 bp of 3' downstream region. In total, we obtained *PvIND* sequences from 157 *P. vulgaris* accessions, 80 domesticated and 77 wild. Two accessions from *P. coccineus* and one each from *P. lunatus* and *P. acutifolius* were also sequenced to be used as an outgroup for subsequent analyses. Our PCR amplifications were successful 100% of the time, although low-quality sequences were sometimes produced because of a specific gene region (long T mononucleotide run in the DNA template).

We used the software DnaSP version 5.10 to compute several measures of DNA sequence variation within and between populations and performed some neutrality tests. The nucleotide diversity was analyzed considering the population subdivisions in gene pools (Mesoamerican and Andean groups) and type (wild and domesticated). These subdivisions are based on the geographic distributions of the samples and previous studies of the molecular diversity of common bean. Based on nucleotide variation, a total of 37 SNPs were found in the sequences of 157 genotypes of *P. vulgaris*. The average SNP frequency was one per 55 bp (Table 4). There was an average of one SNP per 45 bp within the noncoding sequences, whereas one SNP per 76 bp was found within the coding sequences. SNP numbers were higher in wild accessions (35) compared with domesticated accessions (12) and in Mesoamerican accessions (30) compared with Andean accessions (23). Of the 37 SNPs, 8 were nonsynonymous, whereas 3 were synonymous; the other 26 SNPs occurred in noncoding DNA. All 37 nucleotide substitutions sites were biallelic having only two alternative nucleotides. Singletons and doubletons accounted for 27% of variant sites in wild accessions.

Twenty-five haplotypes, defined by 37 polymorphic sites, were observed among the 157 genotypes (Table 5 and Figure 4). Wild accessions had a greater number of haplotypes (24) compared with domesticated accessions (4). Collectively, the Mesoamerican gene pool included 15 haplotypes: 12 in wild accessions, 1 in domesticated accessions, and 2 shared among wild and domesticated accessions; 2 wild accessions from Colombia and 1 wild accession from Mexico showed 2 haplotypes in common with the domesticated accessions. In the Andean gene pool, 14 haplotypes were found: 12 specific to the wild accessions, 1 for the domesticated accessions, and 1 shared between wild and domesticated forms. The haplotype tree in Figure 4 confirmed the strong separation between Andean and Mesoamerican domesticated types. It also showed that there is a reduction in diversity during domestication because there are more haplotypes in the wild types compared with the domesticated group. The teal-colored group had two interesting features: 1) it contained wild accessions of both Mesoamerica and the Andes and 2) the two closest *Phaseolus* species—*P. coccineus* and *P. acutifolius*—were attached to it.

To describe sequence variation or nucleotide diversity, two common measures were calculated: π, the expected heterozygosity per nucleotide site, and θ_W, the number of polymorphic sites in a genotypic sample corrected for sample size. The estimated nucleotide diversity parameters are listed in Table 5. For the entire *PvIND* sequence, π and θ_W estimated over all accessions was 0.00295 and 0.00325, respectively. Total diversity varied among gene pool and forms. For both diversity estimators, variation was greater in the Mesoamerican than in the Andean gene pool. The highest diversity was seen in the Mesoamerican wild genotypes (π = 0.00377, θ_W = 0.00388). This was more than 2- to 3-fold greater than for the Andean wild gene pool (π = 0.00102, θ_W = 0.00249). The Mesoamerican domesticated genotypes showed a lower diversity (π = 0.00046, θ_W = 0.00023) compared with the Mesoamerican wild types, but higher than the Andean domesticated genotypes (π = 0.00014, θ_W = 0.00012). When all genotypes were considered as a population, nucleotide variation among domesticated accessions was slightly greater than among wild

Table 4 Summary of SNPs identified in the candidate gene in *P. vulgaris*

Test set	Total SNP sites	Average frequency of SNP (bp/SNP)	SNPs in coding region	SNPs in noncoding region	Silent SNPs[a]	Nonsynonymous SNPs
All (n = 157)	37	55	11	26	29	8
Wild (n = 77)	35	58	11	24	28	7
Domesticated (n = 80)	12	170	5	7	9	3
Mesoamerican (n = 67)	30	68	8	22	25	5
Wild (n = 23)	29	70	7	22	25	4
Domesticated (n = 44)	2	1022	2	0	0	2
Andean (n = 90)	23	89	7	16	19	4
Wild (n = 54)	23	89	7	16	19	4
Domesticated (n = 36)	1	2043	1	0	1	0

[a]Synonymous and noncoding.

Table 5 Estimates of nucleotide diversity for *PvIND* in the different gene pools of *P. vulgaris*

	Gene pool								
				Mesoamerican			Andean		
	All	Wild	Domesticated	All	Wild	Domesticated	All	Wild	Domesticated
Number of sequences	157	77	80	67	23	44	90	54	36
Number of total sites	2043	2043	2043	2043	2043	2043	2043	2043	2043
Polymorphic sites	37	37	12	30	29	2	23	23	1
Singletones sites	8	10	0	4	5	0	15	16	0
Parsimony informative sites	29	27	12	26	24	2	8	7	1
Haplotypes	25	24	4	15	15	3	14	13	2
Haplotype diversity (Hd)	0.806	0.860	0.705	0.831	0.953	0.680	0.604	0.742	0.286
Nucleotide variation									
Diversity (π/bp)	0.00295	0.00216	0.00286	0.00262	0.00377	0.00046	0.00072	0.00102	0.00014
Polymorphic (θ/bp)	0.00325	0.00372	0.00120	0.00310	0.00388	0.00023	0.00224	0.00249	0.00012
Neutrality test									
Tajima's D	−0.27049	−1.34098	3.82794*	−0.50116	−0.10290	1.89739	−2.03387**	−1.88595	0.27210
Fu and Li's D	−0.44286	−0.61972	1.47349	0.70244	0.65504	0.76072	−4.09225***	−3.69204***	0.57433
Fu and Li's F	−0.44601	−1.07265	2.72554***	0.30566	0.49470	1.26733	−3.96443***	−3.63023***	0.56436
Recombination									
Four−gamete types site pairs	24	24	0	15	15	0	2	2	0
Rm	2	2	0	2	2	0	1	1	0

*P < 0.05; **P < 0.01; ***P < 0.001

accessions (π = 0.00286 vs. π = 0.00216). Tenaillon et al. (2004) developed a metric based on the ratio of diversity estimates in two samples that represents the loss of variation ($L\pi = 1 - \pi_{dom}/\pi_{wild}$). Surprisingly, this diversity statistic was actually negative for the domesticated group (−0.32), which suggests that either we have not adequately sampled the domesticated variability at this locus or the breeding process is generating new variability not seen in the wild type. Considering the two gene pools separately, the domesticated Andean accessions showed a loss of diversity from the wild of 86%, whereas the Mesoamerican domesticated accessions showed a loss of 89%.

To test if the *PvIND* gene has been subjected to positive selection, we performed several tests of neutrality (Table 6). When comparing wild and domesticated forms, we found a significant positive Tajima's D value in domesticated compared with wild forms (Table 6), indicating a reduction of low-frequency polymorphism that can be the result of balancing selection, diversifying selection or population subdivision (Hartl and Clark 1997). Additionally, we observed a stronger reduction of the number of haplotypes in the domesticated compared with the wild group. This finding suggests that this gene was under selection. To confirm this hypothesis, we applied the Fu and Li D*, Fu

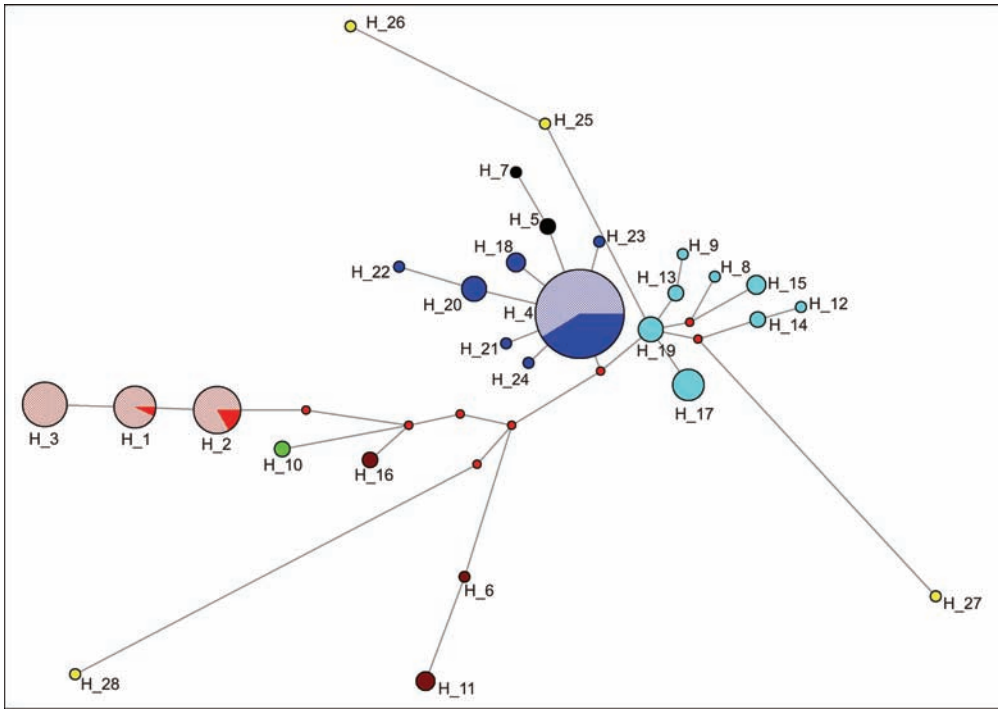

Figure 4. MJ network based on the *PvIND* sequence region from 161 individuals of *Phaseolus* spp. Numbers near circles indicate haplotypes. Circumference size is proportional to the haplotype frequency. Haplotype of each accession is indicated in Supplementary Table S1. Frequency of wild and domesticated accessions for each haplotype is shown as proportion of color in the node circle. Full color: wild accessions; hatched color: domesticated accessions. H_5 and H_7 (in black) correspond to accessions with ancestral phaseolin haplotype (Kami et al. 1995), whereas H_25, H_26, H_27, H_28 (in yellow) are *Phaseolus* spp. Red points represent median vectors. A median vector is a hypothesized (often ancestral) haplotype, which is not included in the current sample but is required to connect existing sequences within the network with maximum parsimony.

Table 6 Neutrality test for *PvIND*

	Tajima's *D*	Fu and Li's *D**	Fu and Li's *F**	Fu's F_S	Fay and Wu's *H*
All (n = 157)	−0.27049	−0.44286	−0.44601	−1.917	−3.70521*
Wild (n = 77)	−1.34098	−0.61972	−1.07265	−7.398**	−7.15038*
Domesticated (n = 80)	3.82794***	1.47349	2.72554**	13.361**	0.45506
Mesoamerican (n = 67)	−0.50116	0.70244	0.30566	−0.138	−5.09995*
Wild (n = 23)	−0.10290	0.65504	0.49470	−1.938	−2.31225
Domesticated (n = 44)	1.89739	0.76072	1.26733	1.781	−0.22622
Andean (n = 90)	−1.84289*	−3.92069**	−3.76599**	−6.032**	−5.88864*
Wild (n = 54)	−1.88595	−3.69204**	−3.63023**	−4.254**	−4.94200*
Domesticated (n = 36)	0.27210	0.57433	0.56436	0.735	0.22857

Fay and Wu's *H* was calculated using as an outgroup ortholog sequences from *P. acutifolius*.
Significance levels for neutrality test: *P < 0.05; **P < 0.01; ***P < 0.001.

and Li *F**, Fu F_S, and Fay and Wu *H* tests. In wild common bean, we observed significant negative Fu F_S and Fu and Li *D** values. This finding might be explained by an excess of singleton mutations (Ramos-Onsins and Rozas 2002), which is in agreement with the observed $\theta_W < \pi$. In domesticated common bean, we observed significant positive Fu F_S and Fu and Li *D** values. This is evidence for a deficiency of alleles, as would be expect from a recent population bottleneck or from overdominant selection. However, when we used the Fay and Wu *H* test, which compares the fitness of the observed derived allele frequency spectrum of the SNPs with that expected under neutrality (Table 6), it is significantly negative in the whole sample and in the wild populations but not in the domesticated population. A negative H value is indicative of genetic hitchhiking. Genetic hitchhiking occurs when a neutral mutation is tightly linked to a locus under positive selection. This linkage allows the mutation to rise to high frequency, which is indicated by a significant negative *H* value (Fay and Wu 2000). Tests of neutrality were calculated also in the two different gene pools, Mesoamerican and

Table 7 F_{ST} estimates between *P. vulgaris* groups determined by the gene pool (Andean and Mesoamerican) and the type (wild or domesticated)

	F_{ST}
Meso wild versus Andean wild	0.24***
Andean domesticated versus Meso domesticated	0.96***
Meso domesticated versus Meso wild	0.51***
Andean wild versus Andean dom	0.14***

F_{ST} estimates significantly different from zero.
***$P < 0.0001$ (1000 permutations; Hudson 2000).

Andean. Among the statistics, Tajima's *D*, Fu and Li's *D**, Fu and Li's *F**, and Fu's F_S tests were generally positive, but not statistically significant, for both Mesoamerican and Andean domesticated samples.

To identify past selection on individual codons, we also determined the rate of nonsynonymous (dN) and synonymous (dS) substitutions at each codon site. The global dN/dS ratios calculated with PARRIS was low (<0.759), indicating that most of the *PvIND* codons are under purifying selection. The three complementary maximum-likelihood methods (SLAC, FEL, and REL) detected no positively selected codon sites. FEL identified two of the amino acids as being under negative selection (codons 5 and 105), whereas REL identified four sites under negative selection (codons 5, 75, 105, and 106) ($P < 0.01$ in both analyses). All the sites identified as negatively selected were located outside the bHLH domain (Supplementary Figure 1). No negatively selected sites were found using the SLAC analysis. The estimation of the minimum number of recombination events tested showed that there was no evidence of recombination in the domesticated accessions. Recombination events were detected only for the wild genotypes in both gene pools.

We used the fixation index F_{ST} as a measure of genetic differences among the populations in our sample. F_{ST} is the most inclusive measure of population structure and is most useful for examining the overall genetic divergence among subpopulations. Application of the F_{ST} statistic to all *P. vulgaris* populations revealed significant geographic structure (Table 7). All the F_{ST} estimates were significantly different from 0 (permutation test; Hudson 2000). The highest value ($F_{ST} = 0.96$) was found between the Mesoamerican and Andean domesticated groups.

Association Analysis in *PvIND*

To determine whether polymorphisms at the *PvIND* locus are associated with observed phenotypic diversity for pod fiber presence among wild and domesticated accessions, we carried out an association analysis using GLM models in TASSEL software version 2.1 (Yu et al. 2006; Bradbury et al. 2007). One hundred five wild and domesticated accessions, a Q matrix from 59 SSRs, and 22 SNPs were used for the association analysis. Despite the close position of our *PvIND* candidate gene to the seed dispersal locus *St* on the common bean genetic map, no SNP controlling pod fiber presence was identified by GLM.

Discussion

As a first approach to identify an *St* candidate gene, we used degenerate primers to target conserved regions of a gene identified in *Arabidopsis* as playing an important role in the control of pod opening and seed dispersal (*IND*). Using this strategy, one putatively homologous sequence to *IND* of *Arabidopsis* was successfully amplified in *P. vulgaris*. The polymorphisms found in the *PvIND* sequence were useful to develop CAPS markers that we used for mapping the *PvIND* locus in two common bean mapping populations. Genetic mapping showed the *PvIND* gene to be located on chromosome Pv02. We found CAPS to be very useful and efficient for mapping the *PvIND* gene. CAPS are technically inexpensive and easy to use because after digestion, the PCR products are resolved in a normal agarose gel.

Because *PvIND* mapped in initial test on chromosome Pv02 is near the gene *St*, which is responsible for the presence or absence of pod string, we hypothesized that this gene homologous to *AtIND* could be the same gene as *St*. *St* represents a key component character of the domestication syndrome in common bean because it controls the presence or absence of pod fibers and, hence, pod dehiscence. This makes this species dependent upon the farmer for seed dispersal. The reduced or lack of pod shattering was selected from a dehiscent wild ancestor during the domestication of common bean. The domestication process typically leaves behind so-called signatures of selection in the form of reduced nucleotide diversity. In order to investigate these signatures of selection, we also analyzed nucleotide diversity data from the *PvIND* gene.

The DNA sequence diversity was sampled in the *PvIND* region by PCR amplification and subsequent sequencing of a 2043-bp segment amplified from genomic DNA from a collection of 157 genotypes of *P. vulgaris*, including both wild and domesticated accessions from the Andean and Mesoamerican gene pools. Two *P. coccineus*, one *P. lunatus*, and one *P. acutifolius* accessions were used as outgroup sequences. Across the entire ~2-kb region sequenced, which includes the immediate 5' and 3' noncoding sequence in addition to the coding sequence, 35 SNPs were evident within the wild sampling and 12 within the domesticated sampling. The frequency of one SNP every 55 bp found in this study was higher than previously reported range for common bean of one SNP every 88 bp (Gaitán-Solís et al. 2008).

Variation at the *PvIND* locus revealed a low level of diversity among the Andean genotypes ($\pi = 0.00072$) relative to the Mesoamerican genotypes ($\pi = 0.00262$). This result was consistent with previous studies in common bean obtained using different types of molecular markers, suggesting that the Andean gene pool might have a narrower genetic basis than the Mesoamerican gene pool (Beebe et al. 2001; McClean et al. 2004; Nanni et al. 2011). We found an average loss of nucleotide diversity in the Andean domesticated accessions of 86% and in the Mesoamerican domesticated accessions of 89%. These estimations are higher than those obtained in previous works. Indeed, using sequence data, the loss of diversity in the Andean domesticated accessions observed by Nanni et al. (2011) was 54%, whereas in the Mesoamerican, the loss in the domesticated accessions was 65–69%.

Genetic diversity measures based on pairwise differences (π) or Watterson's estimator (θ_W) are expected to yield approximately equal values at equilibrium under a standard neutral model (Watterson 1975; Tajima 1983). Departure from this neutral expectation are quantified by statistics such as Tajima's D, from which values different from zero may reflect the action of selection or demographic processes (Tajima 1989). Tajima's *D* based on all sites was calculated on domesticated and wild samples from both gene pools. Domesticated accessions showed a significant positive value of Tajima's *D* (3.82794, $P < 0.001$). This could be a signature of a recent bottleneck, population contraction with consequently a loss of low-frequency variants or demographic effects. Additionally, we observed a stronger reduction of the number of haplotypes of domesticated versus wild groups. This finding may suggest that this gene was under selection. To confirm this hypothesis, we applied several neutrality tests. Because only one gene was analyzed, the neutrality tests provide limited insights, as data from many loci and neutral genes are necessary to determine whether this is a general pattern in the genome, and more work is needed to define the reduction of diversity, if any, around the *PvIND* gene locus.

A NJ analysis was computed to investigate relationships among individual accessions using MEGA version 4 (Tamura et al. 2007). BAPs analysis was used to validate these results. The *PvIND* genetic structure is in general agreement with previous studies of the genetic architecture of common bean germplasm. Both the NJ tree and the BAPs population structure support the two-gene-pool concept (Gepts and Bliss 1986; Koenig and Gepts 1989). As with the previous molecular sequence studies, our data also show that the major divisions within the species are geographically based (McClean et al. 2004; Kwak and Gepts 2009; Nanni et al. 2011).

The MJ network analysis provided an interesting if unexpected image of the organization of genetic diversity at the *PvIND* locus. Of note was the existence of the teal-colored group (haplotypes H_8, H_9, H_12, H_13, H_14, H_15, H_17, and H_19; Figure 4), which corresponds to cluster C4 (Figure 2) and subpopulation K4 (Figure 3). It might be a more ancient group based on its widespread geographic distribution (in both Andean and Mesoamerican regions) and its closer connection to the *P. coccineus* and *P. lunatus* outgroups. Overall, the grouping in the MJ network corresponded well with the clustering in the NJ analysis and the BAPs analysis. The three analyses support the two-gene-pool concept (Gepts and Bliss 1986; Koenig and Gepts 1989). As with the previous molecular sequence studies, our data also show that the major divisions within the species are geographically based (McClean et al. 2004; Kwak and Gepts 2009; Nanni et al. 2011).

To determine whether polymorphism at the *PvIND* locus is associated with observed phenotypic diversity among wild and domesticated accessions, we carried out an association analysis. Association analysis provides high QTL resolution and, therefore, may be very useful for candidate gene validation. Despite the close linkage of our *PvIND* candidate gene with the pod fiber locus *St* on the common bean genetic map, no SNP was associated genetically with in the pod shattering phenotype of common bean. Our results were in agreement with those obtained by Nanni et al. (2011) from the analysis in common bean of the *Arabidopsis* SHATERPROOF (*SHP*) ortholog, which is a gene involved with *IND* in fruit shattering in *Arabidopsis*. They found no mutation in this gene associated with phenotypic variation. These results suggest that neither *IND* and *SHP* may play a role in common bean in pod shattering nor, alternatively, they play a more complex role in seed dispersal compared with the role of their homologues in *Arabidopsis*. Because common bean is a Fabaceae and *Arabidopsis* is a Brassicaceae it may be that different genetic controls are involved in the control of seed dispersal. The lack of complete cosegregation between *PvIND* and *St* suggests that *PvIND* is not directly involved in pod shattering and may not be the gene underlying the *St* locus. Rather, linked regulatory sequences or structural genes may control the expression of *PvIND*. Alternatively, a more precise phenotyping method needs to be developed to more accurately map the *St* locus.

In summary, our study was able to identify and characterize a homologous sequence to *IND* of *A. thaliana* that maps in the vicinity of a known major gene controlling pod dehiscence in common bean. However, we could not find polymorphisms in this sequence that correlate with the dehiscent/indehiscent phenotype. The colocalization of a candidate gene with a QTL or major gene does not provide a strong enough basis upon which to draw a conclusion with respect to the role of this candidate gene in phenotypic variation for pod fiber production (Pflieger et al. 2001). To answer whether seed dispersal in common bean involves molecular mechanisms similar or different to those described in *Arabidopsis* and this putative ortholog is in fact involved in the process, *PvIND* must be characterized further. Such analyses involve identifying selective sweeps in the *PvIND* region, measuring gene expression at the mRNA level and at the protein level as well as the determination of the anatomic and histological location of gene activity, coupled with more precise phenotypic observations on the dehiscence mechanism in individual lines and segregating populations.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

United States Department of Agriculture - National Institute of Food and Agriculture (USDA-NIFA) for the BeanCAP project, grant number #FAR-0015525-9; A Graduate Research Scholarship from Università degli Studi della Basilicata (Italy) to T.G.

Acknowledgments

We thank the BeanCAP participants for developing the sample of US dry and green bean lines. We also thank Drs. D. Debouck (CIAT, Cali, Colombia) and M. Welsh (USDA-ARS, Pullman, WA) for providing seed samples.

References

Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B, et al. 2006. The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. Genome Res. 16:140–147.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J. 2001. Diversity and origin of Andean landraces of common bean. Crop Sci. 41:854–862.

Beebe S, Skroch PW, Tohme J, Duque MC, Pedraza F, Nienhuis J. 2000. Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. Crop Sci. 40:264–273.

Blair MW, Giraldo MC, Buendía HF, Tovar E, Duque MC, Beebe SE. 2006. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet. 113:100–109.

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

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J. 2003. Beans (*Phaseolus* spp.) - model food legumes. Plant and Soil. 252:55–128.

Corander J, Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. Mol Ecol. 15:2833–2843.

Corander J, Tang J. 2007. Bayesian analysis of population structure based on linked molecular information. Math Biosci. 205:19–31.

Corander J, Waldmann P, Sillanpää MJ. 2003. Bayesian analysis of genetic differentiation between populations. Genetics. 163:367–374.

Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P. 1993. Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. Econ Bot. 47:408–423.

Díaz LM, Blair MW. 2006. Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. Theor Appl Genet. 114:143–154.

Ellenberger T, Fass D, Arnaud M, Harrison SC. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev. 8:970–980.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14:2611–2620.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics. 155:1405–1413.

Ferrándiz C. 2002. Regulation of fruit dehiscence in *Arabidopsis*. J Exp Bot. 53:2031–2038.

Freyre R, Ríos R, Guzmán L, Debouck D, Gepts P. 1996. Ecogeographic distribution of *Phaseolus* spp. (Fabaceae) in Bolivia. Econ Bot. 50:195–215.

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

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147:915–925.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133:693–709.

Gaitán-Solis E, Choi IY, Quigley C, Cregan P, Thome J. 2008. Single nucleotide polymorphism in common bean: their discovery and genotyping using a multiplex detection system. Plant Genome. 1:125–134.

Gepts P. 1988. A Middle American and an Andean common bean gene pool. In: Gepts P, editor. Genetic resources of *Phaseolus* beans. Dordrecht (The Netherlands): Kluwer. p. 375–407.

Gepts P. 1998. Origin and evolution of common bean: past events and recent trends. Hortic Sci. 33:1124–1130.

Gepts P, Bliss FA. 1985. F1 hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. J Hered. 76:447–450.

Gepts P, Bliss FA. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ Bot. 40:469–478.

Gepts P, Debouck DG. 1991. Origin, domestication, and evolution of the common bean, *Phaseolus vulgaris*. In: van Schoonhoven A, Voysest O, editors. Common beans: research for crop improvement. Oxon (UK): CAB International. p. 7–53.

Gepts P, Osborn TC, Rashka K, Bliss FA. 1986. Phaseolin protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ Bot. 40:451–468.

Gremski K, Ditta G, Yanofsky MF. 2007. The HECATE genes regulate female reproductive tract development in *Arabidopsis thaliana*. Development. 134:3593–3601.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 41:95–98.

Hammer K. 1984. The domestication syndrome (in German). Kulturpflanze. 32:11–34.

Hartl, DL, Clark AG. 1997. Principles of population genetics. 3rd ed. Sunderland (MA): Sinauer Associates, Inc.

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. 2003. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol. 20:735–747.

Hudson RR. 2000. A new statistic for detecting genetic differentiation. Genetics. 155:2011–2014.

Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting geographic subdivision. Mol Biol Evol. 9:138–151.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 111:147–164.

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

Kami J, Velásquez VB, Debouck DG, Gepts P. 1995. Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proc Natl Acad Sci USA. 92:1101–1104.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16:111–120.

Koenig R, Gepts P. 1989. Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor Appl Genet. 78:809–817.

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

Kosakovsky Pond SL, Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics. 21:2531–2533.

Kosambi DD. 1944. The estimation of map distances from recombination values. Ann Hum Genetics. 12:172–175.

Kwak M, Gepts P. 2009. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet. 118:979–992.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25:1451–1452.

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

Lorieux M. Forthcoming 2012. MapDisto: fast and efficient computation of genetic linkage map. Mol Breed.

Ma PC, Rould MA, Weintraub H, Pabo CO. 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell. 77:451–459.

Marek LF, Shoemaker RC. 1997. BAC contig development by fingerprint analysis in soybean. Genome. 40:420–427.

McClean PE, Kami J, Gepts P. 2004. Genomics and genetic diversity in common bean. In: Wilson RF, Stalker HT, Brummer EC, editors. Legume crop genomics. Urbana (IL): AOCS Press. p. 60–82.

McClean PE, Lee RK. 2007. Genetic architecture of chalcone isomerase non-coding regions in common bean (*Phaseolus vulgaris* L.). Genome. 50:203–214.

Nanni L, Bitocchi E, Bellucci E, Rossi M, Rau D, Attene G, Gepts P, 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.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Pallottini L, Garcia E, Kami J,Barcaccia G, Gepts P. 2004. The genetic anatomy of a patented yellow bean. Crop Sci. 44:968–977.

Paredes OM, Gepts P. 1995. Extensive introgression of Middle American germplasm into Chilean common bean cultivars. Genet Resour Crop Evol. 42:29–41.

Pflieger S, Lefebvre V, Causse M. 2001. The candidate gene approach in plant genetics: a review. Mol Breed. 7: 275–291.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. Mol Biol Evol. 19:2092–2100.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa (NJ): Humana Press. p. 365–386.

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

Staden R. 1996. The Staden sequence analysis package. Mol Biotechnol. 5:233–241.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437–460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol. 21:1214–1225.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276.

Wright S. 1950. The genetical structure of populations. Ann Hum Genetics. 15: 323–354.

Yim YS, Moak P, Sanchez-Villeda H, Musket TA, Close P, Klein PE, Mullet JE, McMullen MD, Fang Z, Schaeffer ML, et al. 2007. A BAC pooling strategy combined with PCR-based screenings in a large, highly repetitive genome enables integration of the maize genetic and physical maps. BMC Genomics. 8:47.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 38:203–208.

Received July 10, 2012; Revised October 29, 2012; Accepted November 11, 2012

Corresponding Editor: J Perry Gustafson