

Article

# Genetic Diversity, Population Structure, and Linkage Disequilibrium in a Spanish Common Bean Diversity Panel Revealed through Genotyping-by-Sequencing

Ana Campa \*<sup>ID</sup>, Ester Murube<sup>ID</sup> and Juan José Ferreira<sup>ID</sup>

Plant Genetics, Area of Horticultural and Forest Crops, SERIDA, 33300 Asturias, Spain; emurube@serida.org (E.M.); jjferreira@serida.org (J.J.F.)

\* Correspondence: acampa@serida.org

Received: 14 September 2018; Accepted: 19 October 2018; Published: 23 October 2018



**Abstract:** A common bean (*Phaseolus vulgaris*) diversity panel of 308 lines was established from local Spanish germplasm, as well as old and elite cultivars mainly used for snap consumption. Most of the landraces included derived from the Spanish common bean core collection, so this panel can be considered to be representative of the Spanish diversity for this species. The panel was characterized by 3099 single-nucleotide polymorphism markers obtained through genotyping-by-sequencing, which revealed a wide genetic diversity and a low level of redundant material within the panel. Structure, cluster, and principal component analyses revealed the presence of two main subpopulations corresponding to the two main gene pools identified in common bean, the Andean and Mesoamerican pools, although most lines (70%) were associated with the Andean gene pool. Lines showing recombination between the two gene pools were also observed, most of them showing useful for snap bean consumption, which suggests that both gene pools were probably used in the breeding of snap bean cultivars. The usefulness of this panel for genome-wide association studies was tested by conducting association mapping for determinacy. Significant marker–trait associations were found on chromosome Pv01, involving the gene *Phvul.001G189200*, which was identified as a candidate gene for determinacy in the common bean.

**Keywords:** *Phaseolus vulgaris*; common bean; diversity panel; GBS; association mapping

## Highlights

- A diversity panel (SDP) of 308 common bean lines was established for GWAS
- SDP gathered a wide genetic common bean diversity
- The main Spanish local diversity is represented in the SDP
- SDP is appropriate for the study of pod traits related to snap consumption

## 1. Introduction

The common bean (*Phaseolus vulgaris* L.) is one of the most important legumes for direct human consumption [1]. The common bean shows a broad phenotypic polymorphism, and both pods and seeds can be used for consumption depending on the genotype. Pods can be harvested before the seed development phase and consumed after cooking (green, French, or snap beans), while seeds can be harvested at physiological maturity (shell beans) or complete maturity (dry beans) and then consumed after re-hydrating and cooking.

The common bean is a diploid species ( $2n = 2x = 22$ ) native to America, where the wild forms are distributed from northern Mexico to northwestern Argentina [2]. The origin of the wild common bean remains a topic for debate. Recent works based on sequence data support a Mesoamerican

origin of wild forms [3,4], from where the common bean expanded to South America resulting in two major ecogeographically and genetically distinct wild gene pools, the Andean (AN) and the Mesoamerican (MA) pools. Domestication took place after the formation of these gene pools in independent events [3,5–12]. The self-pollinating nature of the species and their geographical and ecological separation over millennia have combined to achieve marked differences between the two gene pools at the morphological [13–16] and molecular levels [3,5,10,17,18]. The common bean probably arrived in Europe through the Iberian Peninsula around the 16th century, after the exploration of the Americas [19]. Although the two main gene pools, AN and MA, were introduced into Europe, a prevalence of the AN type was detected based on seed proteins and molecular markers [20–25].

In the Iberian Peninsula, the common bean has been traditionally grown and consumed since its arrival. A wide phenotypic diversity has been described among a set of 296 local accessions collected in Spain in the middle of the 20th century [26]; different growth habits, pod phenotypes, types of use (snap, dry or both), and seed phenotypes ranging in size between 115 to 488 seeds in 100 g, have been described, although those having white seeds and being used as dry beans were the most common. Most of this Spanish diversity is preserved at the National Gene bank in the Center for Plant Genetic Resources (Centro de Recursos Fitogenéticos, CRF, Madrid), in which a common bean collection of more than 3500 accessions collected in Spain since 1978, is maintained. A core collection of 211 accessions (the Spanish Core Collection; SCC) was established from the CRF collection based on two criteria, geographical origin and seed phenotype [24,27,28]. The SCC was evaluated for morpho-agronomic traits and for resistance to different local pathogens, showing a wide genetic diversity for all traits [29–31]. The SCC was also evaluated with a small number of molecular markers, including variation in the major seed protein phaseolin [24], and results distinguished between the two main gene pools of the common bean, AN and MA. The SCC was later updated; it currently consists of 202 accessions. Massive genotyping of the updated SCC could supply data about the origin and relationships among these local materials, and about the diversity of the species and should contribute to its rationalization.

Advances in next-generation sequencing (NGS) technologies have provided fast and cost-efficient methods for DNA sequencing, to the point where genotyping-by-sequencing (GBS) technology is now feasible for species with high diversity and large genome [32]. Genotyping-by-sequencing generates a large number of single nucleotide polymorphism (SNP) datasets that can be physically positioned when a reference genome of the species is available (massive genotyping). In the common bean, the first chromosome-scale genome reference has been available since 2014 [33] and GBS is widely used for high-density linkage map construction, diversity analysis, and genome-wide association study (GWAS) [34–38]. GWAS is a powerful tool for forward genetics analyses (from phenotype-to-genotype-to-genome) in which population samples consisting of individuals with a broad genetic variance (diversity panel) are screened for phenotypes of interest. At least two different diversity panels, consisting of mainly American genotypes, have been established in common bean: the AN Diversity Panel [35] and the MA Diversity Panel [39]. These diversity panels have been used to investigate the genetic architecture of important morpho-agronomic and nutritional traits.

One of the most important factors to be considered in a GWAS is the constitution of the diversity panel, including sample size, linkage disequilibrium (LD), population structure, or genetic relationship between individuals [40,41]. In the current work, a common bean diversity panel representing most of the Spanish diversity for this species was set up, including the updated SCC, and several international breeding lines, as well as old and elite cultivars. The objective of the work described here was to assess the structure and diversity of this panel for its use in GWAS and to investigate the origin and diversity of the local germplasm in order to maximize its conservation and use.

## 2. Material and Methods

### 2.1. Plant Material

A group of 308 *P. vulgaris* materials selected on the basis of type of material (landrace or elite cultivar), the form in which the bean is consumed (dry or snap), and previous genetic knowledge, was assembled into a panel, referred to as the SDP (Spanish Diversity Panel; Table S1). Lines (Figure 1) were obtained by sowing one plant per accession in a greenhouse located at the SERIDA (Regional Service for Agri-Food Research and Development) station in Asturias, northern Spain (43°29′01″ N, 5°26′11″ W; elevation 6.5 m). Plants were watered and fertilized for normal growth and maintained under natural light, environmental relative humidity, and moderate temperature (18–25 °C) during the year 2016. The SDP included 220 landraces, most of them from the updated SCC [24,27], and 51 elite cultivars, most of them cultivated in Europe for snap bean consumption, with the remaining 37 lines derived from traditional old cultivars and well-known breeding lines. The two sequenced bean genotypes, G19833, of AN origin [33], and BAT93, of MA origin [42], were included in the panel as references for the common bean gene pool.



**Figure 1.** Phenotypic seed diversity included in the panel (one seed per line). Bar represents 3 cm.

### 2.2. DNA Isolation

Young leaves from one plant of each line were collected and DNA was isolated using the CTAB method [43] with modifications. Tissue was frozen in liquid nitrogen and pulverized. Concentrations of DNA were quantified photometrically (260–280 nm) using a Biomat 3 ultraviolet–visible spectrophotometer (Thermo Scientific, Waltham, MA, USA). The quality levels of the isolated DNA samples were verified in 1% agarose gels, stained with RedSafe (INtRON Biotechnology, Gyunggi-Do, Korea), and visualized under ultraviolet light. DNA samples were preserved at –80 °C.

### 2.3. Genotyping by Sequencing

Genotyping-by-sequencing, as described by Elshire et al. [32], was carried out at BGI-Tech (Copenhagen, Denmark) using the *ApeKI* restriction enzyme. A GBS sequencing library was prepared by ligating the digested DNA to unique nucleotide adapters (barcodes) followed by PCR with flow-cell attachment site tagged primers. Sequencing was performed using Illumina HiSeq4000 and 100x Paired-End. The sequencing reads from different genotypes were deconvoluted using the barcodes and aligned to the *Phaseolus vulgaris* L. v1 reference genome ([33]; Gene Bank Accession: GCF\_000499845.1), using the Burrow Wheelers Alignment tool [44]. Single nucleotide polymorphism markers were extracted using the GBS pipeline implemented in TASSEL 5.2.39 software [45]. Data were filtered considering missing values (<5%), physical distance (>500 bp), and minor allele frequency (MAF > 0.01). The distribution of the SNPs along chromosomes was calculated with the qqman package [46] of the R project for statistical computing [47]. In order to estimate the rate of mistake in the GBS analysis two duplicated DNA samples of the landrace BGE025740 derived from the same isolation process were included. Rate of mistake was calculated the ratio of the differences between duplicated samples by the total number of SNPs.

### 2.4. Linkage Disequilibrium

Linkage disequilibrium was estimated by calculating the square value of correlation coefficient ( $r^2$ ) between pairs of markers [48] using the TASSEL 5.2.39 software [45]. A threshold of  $r^2 \geq 0.5$  was considered to indicate LD. The level of LD was estimated for the entire panel and for the specific subgroups identified with Structure v2.3.4. Within these subgroups, LD was calculated considering only the polymorphic set of markers in each case.  $p$ -values for each  $r^2$  estimate were obtained with a two-tailed Fisher's exact probability test and a threshold of  $p < 0.0001$  was considered significant. Linkage disequilibrium patterns per chromosome were also calculated.

### 2.5. Data Analysis

Population structure was evaluated using Structure v2.3.4 [49] and Structure Plot v2 software [50]. The STRUCTURE parameters used were an admixture model with independent allele frequencies, a burn-in period of 1000 and 5000 Markov Chain Monte Carlo (MCMC) iterations, with 20 replications for each hypothetical number of subpopulations ( $K$ ) between 1 and 4. The optimum  $K$  value value was calculated according to Evanno et al. [51]. A new burn-in period of 10,000 and 30,000 MCMC iterations was conducted for the optimum  $K$  value to assign accessions to subpopulations.

Cluster analysis was conducted with the FactoMineR [52] package of the R project, considering the Euclidean distance and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) agglomeration method. The FactoMineR package was also used to compute a Principal Component Analysis (PCA). The contribution of each SNP in the final PCA plot was visualized with the function `fviz_contrib()` of factoextra [53] package.

Functional annotations of specific chromosome regions were studied using Ensembl Plants resource [54]). To understand biological meaning behind a list of genes the Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 was used [55,56]).

### 2.6. Genome-Wide Association Study for Determinacy

Association mapping for determinacy was conducted to evaluate the utility of the SDP for GWAS. Phenotyping was performed at the same time as the lines were grown in the greenhouse. Growth habit was characterized as being either determinate (main stem ending in a terminal flower bud) or indeterminate (the flower bud was not terminal). Association studies were conducted using the generalized linear model (GLM) and the mixed linear model (MLM) implemented in TASSEL 5.2.39 software [45]. The GLM is appropriate for variables that are not normally distributed [41], and it is

based on  $P + Q$  matrices, where  $P$  is the phenotype matrix and  $Q$  is the population structure matrix from PCA. The MLM includes both fixed and random effects and it is based on the equation

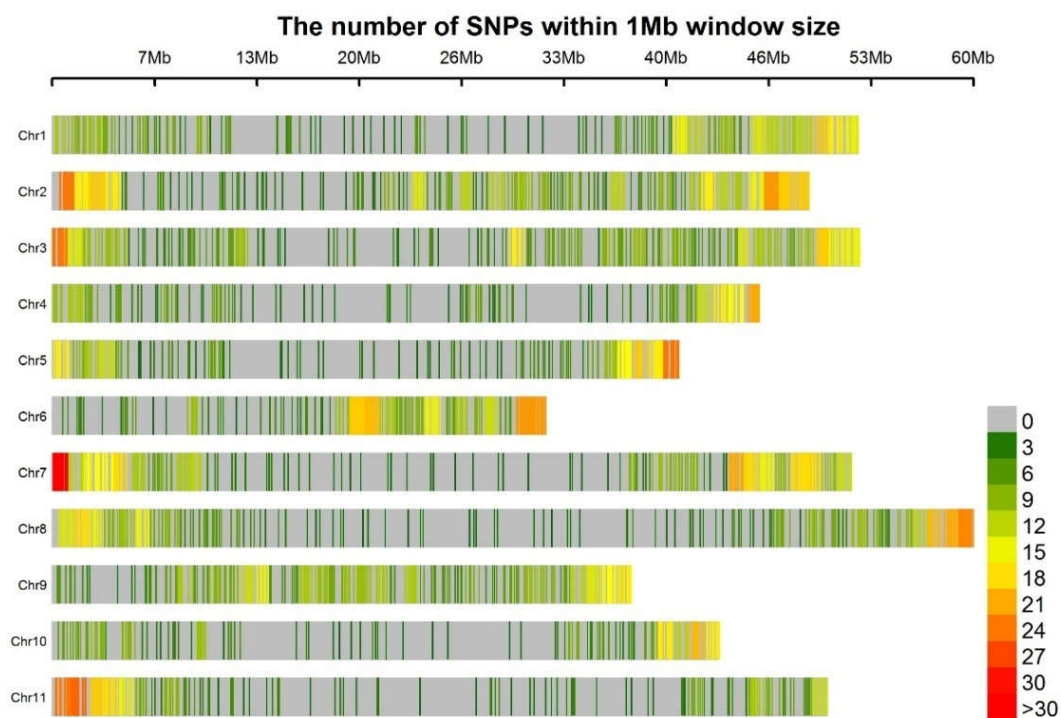
$$Y = X\alpha + P\beta + K\mu + e$$

where  $Y$  is phenotype,  $X$  is genotype,  $P$  is the PCA matrix, both  $X$  and  $P$  represent fixed effects,  $K$  is the relative kinship matrix value, and  $e$  is for residual effects. The Bonferroni correction for  $\alpha 0.001$  ( $-\text{Log}(p) = 6.5$ ) was used for the identification of significant SNP markers. Manhattan plots and QQ plots were developed using the qqman package in R [46,47].

### 3. Results

#### 3.1. Genotyping

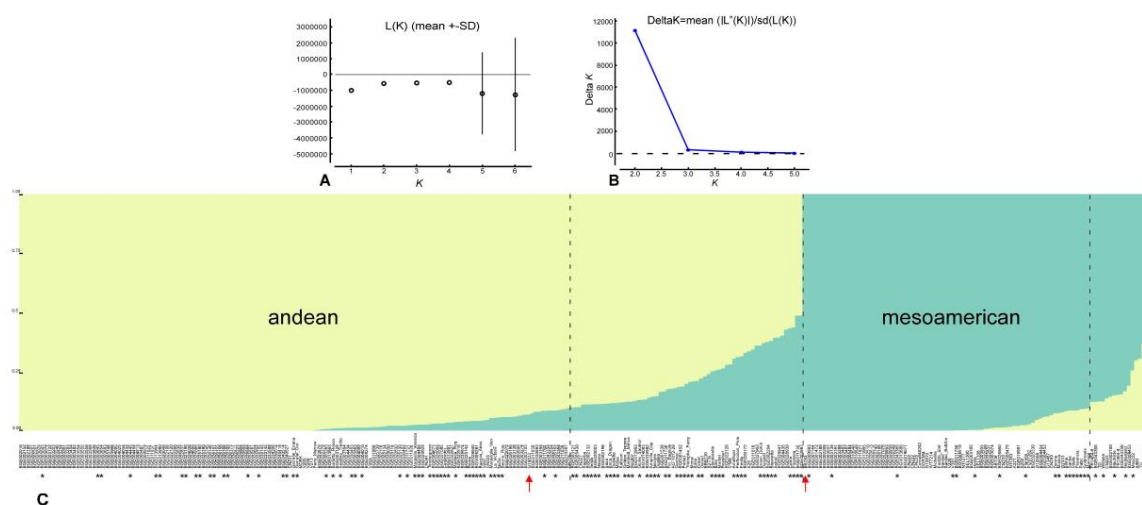
Sequencing of the GBS libraries yielded approximately 2.58 million reads per line, the Q20 value of each sample was above 98% and most of the sample mapping rate was >83%. A total of 9070 mapped SNPs were identified. The error rate of this analysis was estimated to be 0.13% by comparing the duplicated DNA samples from BGE025740. After filtering for missing values, physical distance, and minor allele frequency, a total of 3099 SNPs distributed among the eleven common bean chromosomes was selected (Figure 2; Table S2). The average number of SNPs per chromosome was 282, ranging from 179 on chromosome Pv10 to 399 on chromosome Pv02. For all chromosomes, fewer SNPs were identified in the regions around centromeres than in the regions around telomeres. The average distance between SNPs was 0.17 mega base pair (Mbp), with a minimum distance of 501 bp on chromosome Pv04 and a maximum distance of 3.99 Mb on chromosome Pv10.



**Figure 2.** Distribution along the eleven bean chromosomes of the 3099 single nucleotide polymorphisms (SNPs).

### 3.2. Population Structure

A hypothetical number of subpopulations between two and four were tested with Structure v2.3.4 (Figure S1). The  $\Delta K$  value indicated an optimal number of subpopulations of two (Figure 3A,B). At  $K = 2$  and a threshold of 0.9 for Q statistics, two main groups are identified (Figure 3C): a group of 216 lines closely related to G19833, of AN origin, and a group of 92 lines closely related to BAT93, of MA origin. However, a total of 82 lines showed recombination between the two main gene pools. Lines with potential use as snap beans were observed in both groups, but principally they showed recombination. Only three lines derived from the elite snap bean cultivars did not show recombination between the two gene pools: ‘Garrafal Oro’ and ‘Garrafal Enana’ assigned to the AN group, and ‘Helda’ assigned to MA. With regard to lines derived from the SCC, they were assigned to the two main groups, MA and AN, and 30 of them (approximately 14%) showed recombination.



**Figure 3.** (A) Mean  $L(K)$  ( $\pm$ SD) over 20 runs for each  $K$  value. (B) Plot of  $\Delta K$ . (C) Plot of ancestry estimated for  $K = 2$ . Bars represent the estimated membership coefficients for each accession in each population (represented by different colors) using a threshold value of 0.9 for the Q statistic. Arrows indicate the Andean cultivar G19833 and the Mesoamerican BAT93. Asterisks indicate lines that can be used for snap consumption.

### 3.3. Linkage Disequilibrium

Linkage disequilibrium level was calculated for the entire panel and separately within the subgroups of 148 AN lines, 78 MA lines, and 82 recombinant lines (Table 1). Only 10.11% of the total panel showed significant LD. LD levels were very low, less than 1% in the AN, MA, and recombinant subgroups. For this reason, LD patterns were only calculated for the total panel (Table 2). Chromosomes Pv01, Pv03, Pv09, and Pv11 showed the highest percent of intrachromosomal LD (>20%), coinciding with regions around centromeres (Figure S2). Chromosome Pv09 showed the largest interchromosomal LD, with more than 20% LD with chromosomes Pv01, Pv03, Pv07 and Pv11.

**Table 1.** Pairwise linkage disequilibrium (LD) analysis for polymorphic loci calculated in the whole Panel, and in the subgroups Andean, Mesoamerican, and recombinants. A threshold of  $r^2 \geq 0.5$  was considered as LD.

Group	n	No Polymorphic SNPs	Pairwise LD		
			Average $r^2$	%LD	%Significant LD
Total Panel	308	3099	0.18	10.11	100
Andean	148	2152	0.03	0.95	92.61
Mesoamerican	78	2596	0.05	0.92	90.2
Recombinants	82	2994	0.06	1.08	89.2

**Table 2.** Linkage disequilibrium pattern in the Spanish Diversity Panel (SDP). Average value for  $r^2$  is indicated. %LD indicates the percentage of SNP pairs in linkage disequilibrium ( $r^2 \geq 0.5$ ).

		Chr Pv01	Chr Pv02	Chr Pv03	Chr Pv04	Chr Pv05	Chr Pv06	Chr Pv07	Chr Pv08	Chr Pv09	Chr Pv10	Chr Pv11
Chr Pv01	$r^2$	0.3										
	%LD	29.4%										
Chr Pv02	$r^2$	0.15	0.18									
	%LD	4%	9.30%									
Chr Pv03	$r^2$	0.23	0.15	0.24								
	%LD	18.7%	3.3%	20.3%								
Chr Pv04	$r^2$	0.15	0.13	0.15	0.16							
	%LD	6.4%	3.3%	5.7%	10%							
Chr Pv05	$r^2$	0.19	0.13	0.17	0.13	0.19						
	%LD	12.3%	2.7%	8.2	3.1%	11.7%						
Chr Pv06	$r^2$	0.21	0.14	0.19	0.13	0.17	0.24					
	%LD	13.3%	2%	10.9%	3.4%	6.2%	16.4%					
Chr Pv07	$r^2$	0.2	0.14	0.19	0.14	0.15	0.18	0.21				
	%LD	15.5%	3.1%	13.4%	4.5%	6.7%	10.3	16%				
Chr Pv08	$r^2$	0.18	0.14	0.18	0.14	0.15	0.17	0.17	0.19			
	%LD	11.7%	4.10%	10.1%	6.4%	6.5%	7.50%	8.7%	13.1%			
Chr Pv09	$r^2$	0.26	0.18	0.24	0.17	0.19	0.22	0.24	0.21	0.35		
	%LD	24.7%	6.4%	21.6%	6.7%	10.4%	16.4%	21.2%	14%	35.4%		
Chr Pv10	$r^2$	0.13	0.10	0.14	0.09	0.11	0.13	0.13	0.11	0.17	0.14	
	%LD	5.5%	0.9%	5.9%	0.01%	0.02%	3.5%	4.8%	2.2%	8.4%	7%	
Chr Pv11	$r^2$	0.22	0.15	0.20	0.15	0.17	0.18	0.19	0.18	0.24	0.13	0.25
	%LD	19.4%	3.8%	15.1%	6.2%	9.5%	10.6%	13.7%	11.7%	22.1%	4.2%	21.4%

### 3.4. Unweighted Pair Group Method with Arithmetic Mean Clustering

To characterize the relatedness among the 308 bean lines, a dendrogram was constructed using the UPGMA method and 3099 SNPs. Figure 4 shows the circular phylogenetic tree obtained. Two main groups were observed, one including the MA line BAT93 and the other the AN line G19833. Most lines (70%) were clustered within the AN gene pool.

Lines showing the same genetic profile for the 3099 SNPs were observed in both groups. The MA cluster had three groups of materials showing the same profile: lines derived from accessions BGE027076-BGE039249, lines derived from Sanilac-SanilacBC6\_Are, and lines derived from elite cultivar Bilma-Sacha. Within the AN gene pool three groups of materials also showed the same profile:

- (i) Lines derived from commercial elite cultivars for snap bean consumption: ‘Perfeccion Negra Polo’ and ‘Emilia’, ‘Manteca de los Mercados’, and ‘Rocdor’;
- (ii) lines derived from accessions included in the SCC: BG026222-BGE004813, BGE026158-BGE002207, BGE040514-BGE002152, BGE003550-BGE003274, and BGE029629-BGE004489-BGE027085, BGE022494-V208;
- (iii) lines derived from commercial elite cultivars for snap bean consumption were included in the SCC, ‘Garrafal enana’-BGE026151, ‘Garrafal Oro’-BGE025180-BGE013964-BGE022837, and ‘Buenos Aires Roja’-BGE028940-BGE025142-BGV008281.

### 3.5. Principal Component Analysis

Figure 5 shows the two-dimensional plot obtained in the PCA analysis. The first component (Dim1) accounted for 40% of the variance and distinguished between the two main groups, AN and MA. The recombinant lines identified with STRUCTURE clustered at the intersection between these two main groups. The second principal component (Dim2) accounted for only 4.9% of the variance but revealed more diversity within the MA group than within the AN, in which many accessions occupied a similar position in the plot.







'Buenos Aires Roja' and the accessions BGE028940, BGE025142, and BGV008281. This finding suggests that some accessions which were considered to be landraces probably derived from commercial/elite cultivars after several years of maintenance by farmers, and reflects the difficulty in differentiating between the two types of materials. In fact, both cultivars 'Garrafal Oro' and 'Buenos Aires Roja' are old cultivars that have already been described in the Spanish fields in 1960 [26]. This result is to be expected, because the presence of elite cultivars, mainly for snap bean consumption, derived from bean breeding programs in Europe which began at the end of the 19th century [58]. Moreover, *P. vulgaris* is a highly self-pollinated species so local farmers frequently use their own seed for planting, with the maintenance of cultivars being quite straightforward. This work also reflects the important genetic diversity present within elite cultivars, because they are an important source of genes for a species.

#### 4.2. Population Structure

Structure, cluster, and PCA analysis based on the 3099 SNPs showed the existence of two main groups of germplasm corresponding to the AN and MA gene pools, although most lines (70%) were attributable to the AN gene pool. This result agrees with previous works in which the AN gene pool was prevalent within the European material [20–23,25] and also within the Spanish germplasm [24]. However, the MA group of materials analyzed in this work showed a greater genetic diversity than the AN group, as had been reported in previous studies [12,33,35,59,60]. A higher level of diversity in the MA gene pool than in the AN gene pool was also found in the wild forms [10,61]. Based on sequence data, Bitocchi et al. [4] proposed that, before the domestication of the common bean, there was a severe genetic bottleneck in the AN wild populations, which could explain the narrower AN diversity.

Even though two main subpopulations were identified in the structure analysis, 82 lines showed introgression between the two gene pools. This finding agreed with the high proportion (approximately 44%) of European common bean germplasm that is estimated to be derived from hybridization between the two gene pools (Angioi et al. 2010). In the present work, approximately 14% of the Spanish landraces showed recombination between the two gene pools, a finding which is in agreement with the results of Angioi et al. [25], who observed an uneven distribution of hybrids around Europe, with low frequencies in Spain and Italy. Interestingly, most of those recombinant lines are cultivated for snap bean consumption (see Table S1), which suggests that it is likely both gene pools were used in the breeding of snap bean cultivars. Lines for snap consumption which showed no recombination between the two gene pools were also detected. Of particular interest are the old snap bean cultivars 'Garrafal Oro' and 'Garrafal Enana' assigned to the AN group, and 'Helda' to the MA one. The origin of cultivars for snap bean consumption is not clear. They are thought to have been mainly derived from dry beans after introgression in Europe, where they were rapidly consolidated as a new crop [62]. Accordingly to Brown et al. [63] and Gepts et al. [5] they are predominantly of AN origin but Blair et al. [64] suggests a MA origin. It is important to note that genetic diversity is a concept that depends on the type and number of molecular markers used and it is influenced by sampling effects, so comparison between results from different studies can be difficult.

The identification of duplicated materials in germplasm resources was not cost-effective until the development of the sequencing technologies. In this work, redundant material has been identified based on the 3099 SNPs. This information can be useful for the optimization of the SCC, as redundancy is one of the main problems facing germplasm collections, which consumes gene bank resources. On the other hand, an optimized SDP without the presence of redundant genotypes should be considered for future GWAS.

#### 4.3. Linkage Disequilibrium

Linkage disequilibrium is the nonrandom association of alleles at distinct loci in the genome of a sampled population [48] and constitutes the basis for association mapping approaches. Linkage disequilibrium is highly population-specific and can determine the utility of a panel for GWAS [41]. Based on the 3099 SNPs used in this work, only 10% of the pairwise LD comparisons were in

disequilibrium, even when a very restrictive threshold of  $r^2 \geq 0.5$  was taken into account. Concerning the distribution of LD patterns, chromosomes Pv01, Pv03, Pv09, and Pv11 showed the highest percentage of intrachromosomal LD (>20%), coinciding with highly conserved centromeric regions. This is to be expected as LD has been shown to be noticeably elevated (~5 Mb) in centromeres and other heterochromatic regions, as well as in duplicated regions of the genome [65]. For interchromosomal LD, chromosome Pv09 showed the highest percentage of LD (>20% LD) with chromosomes Pv01, Pv03, Pv07, and Pv11. Different LD patterns could also be related to the independent domestication events for the MA and AN gene pools, in which different chromosomes regions were indirectly selected. According to Schmutz et al. [33], chromosomes Pv02, Pv07, and, in particular, chromosome Pv09 showed signatures of selection in the MA population, whereas the Andean domestication event primarily involved chromosomes Pv01, Pv02, and Pv10. Results of the PCA analysis support the proposal that chromosome Pv09 plays an important role in the differentiation of the two main common bean gene pools. Among the 15 SNP markers that showed the greatest contribution to the differentiation between the MA and the AN groups of the SDP, five are located on chromosome Pv09, involving regions in which 50 candidate genes associated with domestication have been described ([33], see Table S3).

#### 4.4. Genome-Wide Association Study

To evaluate the utility of the SDP, an association mapping analysis for the well-known morphological trait determinacy was conducted. The *fin* gene is involved in the genetic control of this trait, with recessive genotypes controlling the determinate growth habit. This locus was mapped to the end of chromosome Pv01 and a candidate gene (*PvTFL1y*, PHAVU\_001G189200g; Pv01:45,561,512..45,563,326) has been reported from homology with the *TFL1y* gene of *Arabidopsis thaliana* L. [66,67]. In the GLM-GWAS a significant determinacy-associated region was identified on chromosome Pv01, from 6 to 45 Mbp. The candidate gene for determinacy, *Phvul.001G189200*, was identified in the GLM-GWAS, although the chromosome region identified spread to 6 Mbp of chromosome Pv01. This could be explained by the strong LD block identified in this region of chromosome Pv01, which evince the importance of considering the distribution of LD in each panel for the interpretation of GWAS results. Concerning MLM-GWAS, a significant determinacy-associated SNP was only identified on chromosome Pv01, in the position 37 Mbp. The region of 45 Mbp, in which the candidate gene, *PHAVU\_001G189200g*, has been located, was not identified in the MLM-GWAS. Determinacy is a complex trait that has been selected during domestication of common bean, and leguminous crops in general [68]. Multiple origins have been proposed for determinacy in common bean based on the broad mutational spectrum observed in *PvTFL1y*, including retrotransposon insertion and deletion [69], so the possibility of other genes apart from *PvTFL1y* involved in the genetic control of determinacy cannot be discarded.

## 5. Conclusions

In this work, a diversity panel of 308 common bean lines (SDP) was established and genotyped with 3099 SNPs obtained through GBS. Broad genetic and morphological diversity was observed in the SDP. Most of the landraces included were derived from the Spanish common bean core collection, so this panel can be considered to be representative of the local Spanish diversity for this species. SDP also contains snap bean elite cultivars obtained from breeding programs, so it is appropriate for the study of pod morphological traits related to snap bean consumption. Some groups of accessions with the same profile from the 3099 SNPs were identified, suggesting the possibility of removing some duplicate accessions in order to maximize panel diversity. Information concerning redundant accessions can be useful for the management of the Spanish local diversity maintained in gene banks. Close relationships between lines derived from landraces and old cultivars were identified, revealing the difficulty of differentiating between both types of materials. The usefulness of SDP for future GWAS was validated through the association mapping of determinacy.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/9/11/518/s1>. Table S1: List of accessions included in the Spanish Diversity Panel, Table S2: Distribution of the 3099 SNPs, Table S3: Tag sequences, Table S4: Regions of 100-kb surrounding each one of the most influent SNPs, Table S5: Genes annotated in the Regions of 100-kb surrounding each one of the most influent SNPs, Table S6: Determinacy-associated SNPs, Figure S1: Plot of ancestry, Figure S2: Linkage disequilibrium plots, Figure S3: PCA Plot, Figure S4: Manhattan plots.

**Author Contributions:** Formal Analysis, A.C. and E.M.; Investigation, A.C. and E.M.; Writing—Original Draft Preparation, A.C. and E.M.; Writing—Review & Editing, A.C., E.M., and J.J.F.; Project Administration, J.J.F. and A.C.; Funding Acquisition, J.J.F. and A.C.

**Funding:** This work was supported in part by grant AGL2017-87050-R of the Spanish Government. E. Murube (FPI-INIA) and A Campa (DR13-0222) are recipients of a salary from the Instituto de Investigación y Tecnología Agraria y Alimentaria (INIA, Spain) cofunding with European Regional Development's Funds (FEDER).

**Acknowledgments:** The authors thank M. Bueno, J.A. Poladura, and F. Díaz for their technical assistance.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Food and Agriculture Organization of the United Nations. *FAOSTAT Statistics Database*; FAO: Rome, Italy, 1998.
2. Toro, O.; Tohme, J.; Debouck, D.G. *Wild Bean (Phaseolus vulgaris L.): Description and Distribution*; IBPGR and CIAT: Cali, Colombia, 1990; ISBN 958-9183-22-0.
3. Rossi, M.; Bitocchi, E.; Bellucci, E.; Nanni, L.; Rau, D.; Attene, G.; Papa, R. Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol. Appl.* **2009**, *2*, 504–522. [[CrossRef](#)] [[PubMed](#)]
4. Bitocchi, E.; Nanni, L.; Bellucci, E.; Rossi, M.; Giardini, A.; Zeuli, P.S.; Logozzo, G.; Stougaard, J.; McClean, P.; Attene, G.; et al. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci USA* **2012**, *109*, 788–796. [[CrossRef](#)] [[PubMed](#)]
5. Gepts, P.; Osborne, T.C.; Rashka, K.; Bliss, F.A. Phaseolin protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris* L.): Evidence for multiple centers of domestication. *Econ. Bot.* **1986**, *40*, 451–468. [[CrossRef](#)]
6. Beebe, S.; Skroch, P.; Tohme, J.; Duque, M.; Pedraza, F.; Nienhuis, J. Structure of genetic diversity among common bean landraces of middle-American origin based on correspondence analysis of RAPD. *Crop Sci.* **2000**, *40*, 264–273. [[CrossRef](#)]
7. Papa, R.; Gepts, P. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor. Appl. Genet.* **2003**, *106*, 239–250. [[CrossRef](#)] [[PubMed](#)]
8. Chacón, S.M.I.; Pickersgill, B.; Debouck, D.G. Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theor. Appl. Genet.* **2005**, *110*, 432–444. [[CrossRef](#)] [[PubMed](#)]
9. Papa, R.; Acosta, J.; Delgado-Salinas, A.; Gepts, P. A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theor. Appl. Genet.* **2005**, *111*, 1147–1158. [[CrossRef](#)] [[PubMed](#)]
10. Kwak, M.; Gepts, P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor. Appl. Genet.* **2009**, *118*, 979–992. [[CrossRef](#)] [[PubMed](#)]
11. Mamidi, S.; Rossi, M.; Annam, D.; Moghaddam, S.; Lee, R.; Papa, R.; McClean, P. Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant. Biol.* **2011**, *38*, 953–967. [[CrossRef](#)]
12. Bitocchi, E.; Bellucci, E.; Giardini, A.; Rau, D.; Rodriguez, M.; Biagetti, E.; Santilocchi, R.; Spagnoletti Zeuli, P.; Gioia, T.; Logozzo, G.; et al. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol.* **2013**, *197*, 300–313. [[CrossRef](#)] [[PubMed](#)]
13. Delgado Salinas, A.O.; Bonet, A.; Gepts, P. The wild relative of *Phaseolus vulgaris* in Middle America. In *Genetic Resources in Phaseolus Beans*; Gepts, P., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1988; pp. 163–184, ISBN 978-94-010-7753-8.

14. Debouck, D.G.; Tohme, J. Implications for bean breeders of studies on the origins of common beans. *Phaseolus vulgaris* L. In *Current Topics in Breeding of Common Bean*; Working Document No. 47; Beebe, S., Ed.; Centro Internacional de Agricultura Tropical: Cali, Colombia, 1989; pp. 3–42.
15. Gepts, P.; Debouck, D.G. Origin, domestication and evolution of common bean, *Phaseolus vulgaris*. In *Common Beans: Research for Crop Improvement*; Schoonhoven, A., Voyses, O., Eds.; CAB International: Wallingford, UK, 1991; pp. 4–54, ISBN 085198679X, ISBN 9780851986791.
16. Singh, S.P.; Nodari, R.; Gepts, P. Genetic diversity in cultivated common bean: I. Allozymes. *Crop Sci.* **1991**, *31*, 19–23. [[CrossRef](#)]
17. Singh, S.P.; Gepts, P.; Debouck, D.G. Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* **1991**, *45*, 379–396. [[CrossRef](#)]
18. Koenig, R.L.; Gepts, P. Allozyme diversity in wild *Phaseolus vulgaris*: Further evidence for two major centers of genetic diversity. *Theor. Appl. Genet.* **1989**, *78*, 809–817. [[CrossRef](#)] [[PubMed](#)]
19. Ortwin-Sauer, C. *The Early Spanish Man*; University of California Press: Berkeley/Los Angeles, CA, USA, 1966; pp. 51–298, ISBN 978-0521088480.
20. Gepts, P.; Bliss, F.A. Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. II Europe and Africa. *Econ. Bot.* **1988**, *42*, 86–104. [[CrossRef](#)]
21. Lioi, L. Geographical variation of phaseolin patterns in an Old World collection of *Phaseolus vulgaris*. *Seed Sci. Technol.* **1989**, *17*, 317–324.
22. Piergiovanni, A.R.; Taranto, G.; Losavio, F.P.; Pignone, D. Common bean (*Phaseolus vulgaris* L.) landraces from Abruzzo and Lazio regions (Central Italy). *Genet. Res. Crop Evol.* **2006**, *53*, 313–322. [[CrossRef](#)]
23. Logozzo, G.; Donnoli, R.; Macaluso, L.; Papa, R.; Knupffer, H.; Spagnoletti Zeuli, P.L. Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. *Genet. Resour. Crop Evol.* **2007**, *54*, 1763–1779. [[CrossRef](#)]
24. Pérez-Vega, E.; Campa, A.; De la Rosa, L.; Giraldez, R.; Ferreira, J.J. Genetic diversity in a core collection established from the main bean Genebank in Spain. *Crop Sci.* **2009**, *49*, 1377–1386. [[CrossRef](#)]
25. Angioi, S.A.; Rau, D.; Attene, G.; Nanni, L.; Belluci, E.; Logozzo, G.; Negri, V.; Spagnoletti Zeuli, P.L.; Papa, R. Beans in Europe: Origin and structure of the European landraces of *Phaseolus vulgaris* L. *Theor. Appl. Genet.* **2010**, *121*, 829–843. [[CrossRef](#)] [[PubMed](#)]
26. Puerta Romero, J. *Varietades de Judías Cultivadas en España*; Monogr. Inst. Nacional de Investigación y Tecnología Agraria y Alimentaria, Ministerio de Agricultura: Madrid, Spain, 1961; p. 11.
27. De la Rosa, L.; Lázaro, A.; Varela, F. Racionalización de la colección española de *Phaseolus vulgaris* L. In *II Seminario de Judía de la Península Ibérica*; Asociación Española de Leguminosas: Villaviciosa, Spain, 2000; pp. 55–62, ISBN 84-7847-532-X.
28. Colección Nuclear de Judías Española. Available online: <http://www.crf.inia.es/judias/> (accessed on 1 July 2018).
29. Pascual, A.; Campa, A.; Pérez-Vega, E.; Giraldez, R.; Miklas, P.N.; Ferreira, J.J. Screening common bean for resistance to four *Sclerotinia sclerotiorum* isolates collected in northern Spain. *Plant Dis.* **2010**, *94*, 885–890. [[CrossRef](#)]
30. Trabanco, N.; Perez-Vega, E.; Campa, A.; Rubiales, D.; Ferreira, J.J. Genetic resistance to powdery mildew in common bean. *Euphytica* **2012**, *186*, 875–882. [[CrossRef](#)]
31. Ferreira, J.J.; Campa, A.; Perez-Vega, E. Variation in the response to ascochyta blight in common bean germplasm. *Eur. J. Plant. Pathol.* **2016**, *146*, 977–985. [[CrossRef](#)]
32. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)] [[PubMed](#)]
33. Schmutz, J.; McClean, P.E.; Mamidi, S.; Wu, G.A.; Cannon, S.B.; Grimwood, J.; Jenkins, J.; Shu, S.; Song, Q.; Chavarro, C.; et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **2014**, *46*, 707–713. [[CrossRef](#)] [[PubMed](#)]
34. Hart, J.P.; Griffiths, P.D. Genotyping-by-sequencing enabled mapping and marker development for the *By-2* potyvirus resistance allele in common bean. *Plant Genome* **2015**, *8*, 1–14. [[CrossRef](#)]

35. Cichy, K.A.; Porch, T.G.; Beaver, J.S.; Cregan, P.; Fourie, D.; Glahn, R.P.; Grusak, M.A.; Kamfwa, K.; Katuutamu, D.N.; McClean, P.; et al. A *Phaseolus vulgaris* diversity Panel for Andean bean improvement. *Crop Sci.* **2015**, *55*, 2149–2160. [[CrossRef](#)]
36. Ariani, A.; Teran, J.C.B.M.; Gepts, P. Genome-wide identification of SNPs and copy number variation in common bean (*Phaseolus vulgaris* L.) using genotyping-by-sequencing (GBS). *Mol. Breed.* **2016**, *36*, 87. [[CrossRef](#)]
37. Ferreira, J.J.; Murube, E.; Campa, A. Introgressed genomic regions in a set of near-isogenic lines of common bean revealed by genotyping-by-sequencing. *Plant Genome* **2017**, *10*. [[CrossRef](#)] [[PubMed](#)]
38. Katuramu, D.N.; Hart, J.P.; Porch, T.G.; Grusak, M.A.; Glahn, R.P.; Cichy, K.A. Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (*Phaseolus vulgaris* L.). *Mol. Breed.* **2018**, *38*, 44. [[CrossRef](#)]
39. Moghaddam, S.M.; Mamidi, S.; Osorno, J.M.; Lee, R.; Brick, M.; Kelly, J.; Miklas, P.; Urrea, C.; Song, Q.; Cregan, P.; et al. Genome-Wide association study identifies candidate loci underlying agronomic traits in a middle American diversity panel of common bean. *Plant Genome* **2016**, *9*, 1–21. [[CrossRef](#)] [[PubMed](#)]
40. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **2013**, *9*, 29. [[CrossRef](#)] [[PubMed](#)]
41. Burghardt, L.T.; Young, N.D.; Tiffin, P. A guide to genome-wide association mapping in plants. *Curr. Prot. Plant. Biol.* **2017**, *2*, 22–38. [[CrossRef](#)]
42. Vlasova, A.; Capella-Gutiérrez, S.; Rendón-Anaya, M.; Hernández-Oñate, M.; Minoche, A.E.; Erb, I.; Câmara, F.; Prieto-Barja, P.; Corvelo, A.; Sanseverino, W.; et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biol.* **2016**, *17*, 32. [[CrossRef](#)] [[PubMed](#)]
43. Doyle, J.J.; Doyle, L.H. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.
44. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
45. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)] [[PubMed](#)]
46. Turner, S.D. qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots. 2014. Available online: <https://cran.r-project.org/web/packages/qqman/index.html> (accessed on 1 July 2018).
47. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016; ISBN 3-900051-07-0. Available online: <http://www.R-project.org> (accessed on 1 July 2018).
48. Hill, W.G.; Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **1968**, *38*, 226–231. [[CrossRef](#)] [[PubMed](#)]
49. Pritchard, J.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [[PubMed](#)]
50. Ramasamy, R.K.; Ramasamy, S.; Bindroo, B.B.; Naik, V.G. STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springplus* **2014**, *13*, 431. [[CrossRef](#)] [[PubMed](#)]
51. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)] [[PubMed](#)]
52. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate. *J Stat. Softw.* **2008**, *25*, 1–18. [[CrossRef](#)]
53. Kassambara, A.; Mundt, F. Extract and Visualize the Results of Multivariate Data Analyses, v 1.0.5. 2017. Available online: <http://www.sthda.com/english/rpkgs/factoextra> (accessed on 1 July 2018).
54. Bolser, D.; Staines, D.M.; Pritchard, E.; Kersey, P. Ensembl Plants: Integrating tools for visualizing, mining, and analyzing plant genomics data. *Methods Mol. Biol.* **2016**, *1374*, 115–140. [[CrossRef](#)] [[PubMed](#)]
55. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009**, *37*, 1–13. [[CrossRef](#)] [[PubMed](#)]
56. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)] [[PubMed](#)]
57. Brodie, A.; Azaria, J.R.; Ofran, Y. How far from the SNP may the causative genes be? *Nucleic Acids Res.* **2016**, *44*, 6046–6054. [[CrossRef](#)] [[PubMed](#)]

58. Doré, C.; Varoquaux, F. Le haricot: Histoire des variétés cultivées en France. In *Histoire et amélioration de cinquante plantes cultivées*; Collection Savoir-Faire; Institut Scientifique de Recherche Agronomique: Paris, France, 2006; pp. 338–339.
59. McClean, P.; Terpstra, J.; McConnell, M.; White, C.; Lee, R.; Mamidi, S. Population structure and genetic differentiation among the USDA common bean (*Phaseolus vulgaris* L.) core collection. *Genet. Resour. Crop Evol.* **2012**, *59*, 499–515. [[CrossRef](#)]
60. Mamidi, S.; Rossi, M.; Moghaddam, S.M.; Annam, D.; Lee, R.; Papa, R.; McClean, P. Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity* **2013**, *110*, 267–276. [[CrossRef](#)] [[PubMed](#)]
61. Bitocchi, E.; Rau, D.; Belluci, E.; Rodríguez, M.; Murgia, M.L.; Gioia, T.; Santo, D.; Nanni, L.; Attene, G.; Papa, R. Beans (*Phaseolus* spp.) as a model for understanding crop evolution. *Front. Plant Sci.* **2017**, *8*, 722. [[CrossRef](#)] [[PubMed](#)]
62. Myers, J.R.; Baggett, J.R. Improvement of snap bean. In *Common Bean Improvement in the Twenty-First Century*; Sing, S., Ed.; Springer: Dordrecht, The Netherlands, 1999; Volume 7, pp. 289–329, ISBN 978-90-481-5293-3.
63. Brown, J.W.S.; McFerson, J.R.; Bliss, F.A.; Hall, T.C. Genetic divergence among commercial classes of *Phaseolus vulgaris* in relation to phaseolin pattern. *HortScience* **1982**, *17*, 752–754.
64. Blair, W.B.; Chaves, A.; Tofiño, A.; Calderón, J.F.; Palacio, J.D. Extensive diversity and inter-pool introgression in a world-wide collection of indeterminate snap bean accessions. *Theor. Appl. Genet.* **2010**, *120*, 1381–1391. [[CrossRef](#)] [[PubMed](#)]
65. Smith, A.V.; Thomas, D.J.; Munro, H.M.; Abecasis, G.R. Sequence features in regions of weak and strong linkage disequilibrium. *Genome Res.* **2005**, *15*, 1519–1534. [[CrossRef](#)] [[PubMed](#)]
66. Kwak, M.; Velasco, D.; Gepts, P. Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). *J. Hered.* **2008**, *99*, 283–291. [[CrossRef](#)] [[PubMed](#)]
67. Repinski, S.L.; Kwak, M.; Gepts, P. The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis TFL1*. *Theor. Appl. Genet.* **2012**, *124*, 1539–1547. [[CrossRef](#)] [[PubMed](#)]
68. Li, S.; Ding, Y.; Zhang, D.; Wang, X.; Tang, X.; Dai, D.; Jin, H.; Lee, S.H.; Cai, C.; Ma, J. Parallel domestication with a broad mutational spectrum of determinate stem growth habit in leguminous crops. *Plant J.* **2018**. [[CrossRef](#)] [[PubMed](#)]
69. Kwak, M.; Toro, O.; Debouck, D.G.; Gepts, P. Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). *Ann. Bot.* **2012**, *110*, 1573–1580. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).