


RESEARCH ARTICLE

Open Access



Genetic mapping for agronomic traits in a MAGIC population of common bean (*Phaseolus vulgaris* L.) under drought conditions

Santiago Diaz^{1†}, Daniel Ariza-Suarez^{1†}, Paulo Izquierdo^{1,2}, Juan David Lobaton^{1,3}, Juan Fernando de la Hoz^{1,4}, Fernando Acevedo^{1,5}, Jorge Duitama^{1,6}, Alberto F. Guerrero¹, Cesar Cajiao¹, Victor Mayor^{1,7}, Stephen E. Beebe¹ and Bodo Raatz^{1*} 

Abstract

Background: Common bean is an important staple crop in the tropics of Africa, Asia and the Americas. Particularly smallholder farmers rely on bean as a source for calories, protein and micronutrients. Drought is a major production constraint for common bean, a situation that will be aggravated with current climate change scenarios. In this context, new tools designed to understand the genetic basis governing the phenotypic responses to abiotic stress are required to improve transfer of desirable traits into cultivated beans.

Results: A multiparent advanced generation intercross (MAGIC) population of common bean was generated from eight Mesoamerican breeding lines representing the phenotypic and genotypic diversity of the CIAT Mesoamerican breeding program. This population was assessed under drought conditions in two field trials for yield, 100 seed weight, iron and zinc accumulation, phenology and pod harvest index.

Transgressive segregation was observed for most of these traits. Yield was positively correlated with yield components and pod harvest index (PHI), and negative correlations were found with phenology traits and micromineral contents. Founder haplotypes in the population were identified using Genotyping by Sequencing (GBS). No major population structure was observed in the population. Whole Genome Sequencing (WGS) data from the founder lines was used to impute genotyping data for GWAS. Genetic mapping was carried out with two methods, using association mapping with GWAS, and linkage mapping with haplotype-based interval screening. Thirteen high confidence QTL were identified using both methods and several QTL hotspots were found controlling multiple traits. A major QTL hotspot located on chromosome Pv01 for phenology traits and yield was identified. Further hotspots affecting several traits were observed on chromosomes Pv03 and Pv08. A major QTL for seed Fe content was contributed by MIB778, the founder line with highest micromineral accumulation. Based on imputed WGS data, candidate genes are reported for the identified major QTL, and sequence changes were

(Continued on next page)

* Correspondence: b.raatz@cgiar.org

[†]Santiago Diaz and Daniel Ariza-Suarez contributed equally to this work.

¹Bean Program, Agrobiodiversity Area, International Center for Tropical Agriculture (CIAT), Cali, Colombia

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

identified that could cause the phenotypic variation.

Conclusions: This work demonstrates the importance of this common bean MAGIC population for genetic mapping of agronomic traits, to identify trait associations for molecular breeding tool design and as a new genetic resource for the bean research community.

Keywords: Genotyping-by-sequencing (GBS), Genome-wide association study (GWAS), Multiparent advanced generation inter-crosses (MAGIC), Quantitative trait loci (QTL), Whole genome sequencing (WGS)

Background

Common bean (*Phaseolus vulgaris* L.) is one of the most important grain legumes for direct human consumption [1], and it is widely cultivated throughout the world, especially in tropical and subtropical countries of Africa and America [2]. Bean has been recognized as a highly valuable food for human nutrition, a rich and relatively inexpensive source of proteins, micronutrients such as iron and zinc, dietary fiber and essential vitamins [3] making it valuable food for over half a billion people, especially in developing countries [4]. Micronutrient malnutrition (MNM) is considered a major health threat affecting half the world's population, particularly women and children in developing countries [5]. Iron deficiency anemia, which is the most prevalent MNM in the world, can be alleviated by biofortification, particularly in legumes with high baseline Fe contents such as common bean [6].

Bean production is negatively affected by multiple abiotic and biotic constraints. Drought is a major factor to yield loss around the world and its incidence and duration is expected to increase due to climate change [4, 7]. It is estimated that approximately 60% of common bean production is affected by terminal or intermittent drought [8]. If current climate change trends continue, a large part of current common bean growing areas in southeastern Africa are predicted to become unsuitable for bean cultivation by 2050, generating reductions in yield for current varieties and potentially affecting the nutritional quality of the crop [9]. In that sense, drought represents a high-risk factor for bean farmers in the tropics, and improved varieties developed to withstand altered climatic conditions represent a valuable resource for future generations.

Genetic mapping is used to identify genomic regions that harbor variation responsible for altering phenotypic trait expression. This information is then applied in breeding by marker-assisted selection (MAS), whereby breeding lines with desirable trait attributes are identified through genetic screening. Different strategies are available for genetic mapping. Most molecular markers currently used in breeding were identified through linkage mapping, where a bi-parental population is generated for identifying the genomic regions that segregate with a trait. However, this strategy has low resolution

because of limited genetic recombinations and limited genetic variation, since only two haplotypes are observed per locus [10]. On the other hand, association mapping using Genome-wide association studies (GWAS) directly identifies marker-trait associations in diverse populations. This strategy also leverages low levels of linkage disequilibrium (LD) to achieve higher resolution. However, confounders such as population structure can produce spurious associations, and large populations are needed to overcome the lack of statistical power in the case of low allele frequencies [11].

A strategy to increase the number of investigated haplotypes while avoiding confounding population structure effects is to generate recombinant inbred lines (RILs) from multiple parents [12], where the genomes of the founders are first recombined through several rounds of mating and then advanced to generate a stable panel of inbred lines. The creation of Multiparent advanced generation intercross (MAGIC) populations holds great promise for genetic mapping, overcoming the main limitations of linkage and association mapping [13]. Using more than two parental accessions increases the allelic and phenotypic diversity compared to traditional bi-parental populations, raising the number of QTL that segregate in the population and the larger number of accumulated recombination events increases mapping accuracy [14]. Furthermore, MAGIC populations have been suggested to use breeding lines as founders, rather than conventional crosses between two phenotypically contrasting lines. Utilization of elite breeding germplasm in MAGIC population development strongly facilitates the transfer of identified QTL into breeding applications. In summary, a MAGIC population combines advantages of natural and synthetic populations with a better shuffling of the genome and increased genetic resolution (i.e. better QTLs) and better transferability to breeding applications [15].

Genetic mapping ideally leads to the identification of genes, whose functional alleles influence the observed phenotypic variation. Gene identification is most advanced in model crops like rice, where over 2000 genes controlling important agronomic traits have been cloned [16, 17]. There are few examples for cloned genes in common bean, including the *bc-3* gene conferring resistance to bean common mosaic virus, which was identified

as the *eIF4E* gene [18] or *Mrp1* gene, causative of *lpa1* mutant regulating low concentration of phytic acid in the bean seed [19]. In other studies, candidate genes related with agronomic traits in common bean have been identified based on annotations in the vicinity of significantly associated markers [20–23]. Further investigation of candidate genes is hindered by the scarcity of genetic resources and genetic transformation protocols in common bean [24], which would be required for validation experiments.

The main goal of this work was to develop a common bean MAGIC population from eight Mesoamerican elite breeding founder lines and to identify genomic regions associated with yield, micromineral accumulation, phenology and physiological traits under drought conditions.

Results

Phenotypic evaluation

A common bean MAGIC population was developed by inter-crossing of eight Mesoamerican elite breeding lines, including four founder lines derived from interspecific crosses with *P. coccineus*, *P. dumosus* and *P. acutifolius* (Table 1). Following the crossing scheme in Additional file 1, 996 RILs were generated. Field trials under drought stress were carried out in 2013 and 2014 (Additional file 2) with 636 and 599 RIL lines respectively. Agronomic performance was evaluated with respect to yield (Yd), days to flowering (DF), days to physiological maturity (DPM), seed weight (100SdW) and pod harvest index (PHI). Micronutrient contents such as iron and zinc in the seed (SdFe, SdZn) were evaluated in 2014 and again in non-stress conditions in 2016.

The phenotypic data display significant genetic variability within all traits, and transgressive segregation was observed for most of them (Fig. 1). The broad-sense heritability for phenological traits was high, with 0.98 for DF in both years and 0.96 and 0.79 respectively for DPM (Fig. 2). PHI had a heritability of 0.90. Likewise, high heritabilities of 0.96 and 0.85 were observed for 100SdW in respective years. In contrast, the heritability for yield was lower and not stable across both trials with 0.71 in

2013 and 0.31 in 2014. Heritabilities of iron and zinc content were intermediate, with 0.67 for SdFe and 0.47 for SdZn. In general, trait heritabilities in 2013 were higher compared to 2014 for all traits measured in both trials. In 2013, two-row plots were used compared to one-row plots in 2014, which may have caused more experimental noise, allowing more precise data in 2013. However, climatic conditions differed between seasons with stronger drought and less late rainfall in 2013 (Additional file 2), leading to a stronger yield reduction.

In spite of the seasonal variation, all traits measured in more than one trial were significantly positively correlated, ranging from 0.25 (Yd) to 0.66 (DF and 100SdW) (Fig. 2), indicating comparability of the data sets from the three trials for evaluated traits. Two trait clusters were identified according to the positive and generally significant correlations within the clusters and negative correlations between them. On the one hand, the phenology traits DF and DPM and micronutrient contents SdFe and SdZn form a group of traits that were largely positively correlated to each other. On the other hand, yield and 100SdW together with PHI show significant positive correlations among them, while being negatively correlated to traits of the first group (Fig. 2). The founders SCR2 and SCR9 performed best in terms of yield and related traits, while MIB778 showed delayed maturity, lowest yield components, but the highest Fe and Zn contents (Fig. 1 and Additional file 3). Taken together, the phenotypic datasets from the field trials show a good quality for subsequent analyses, with some lines showing transgressive segregation and outperforming elite founder lines.

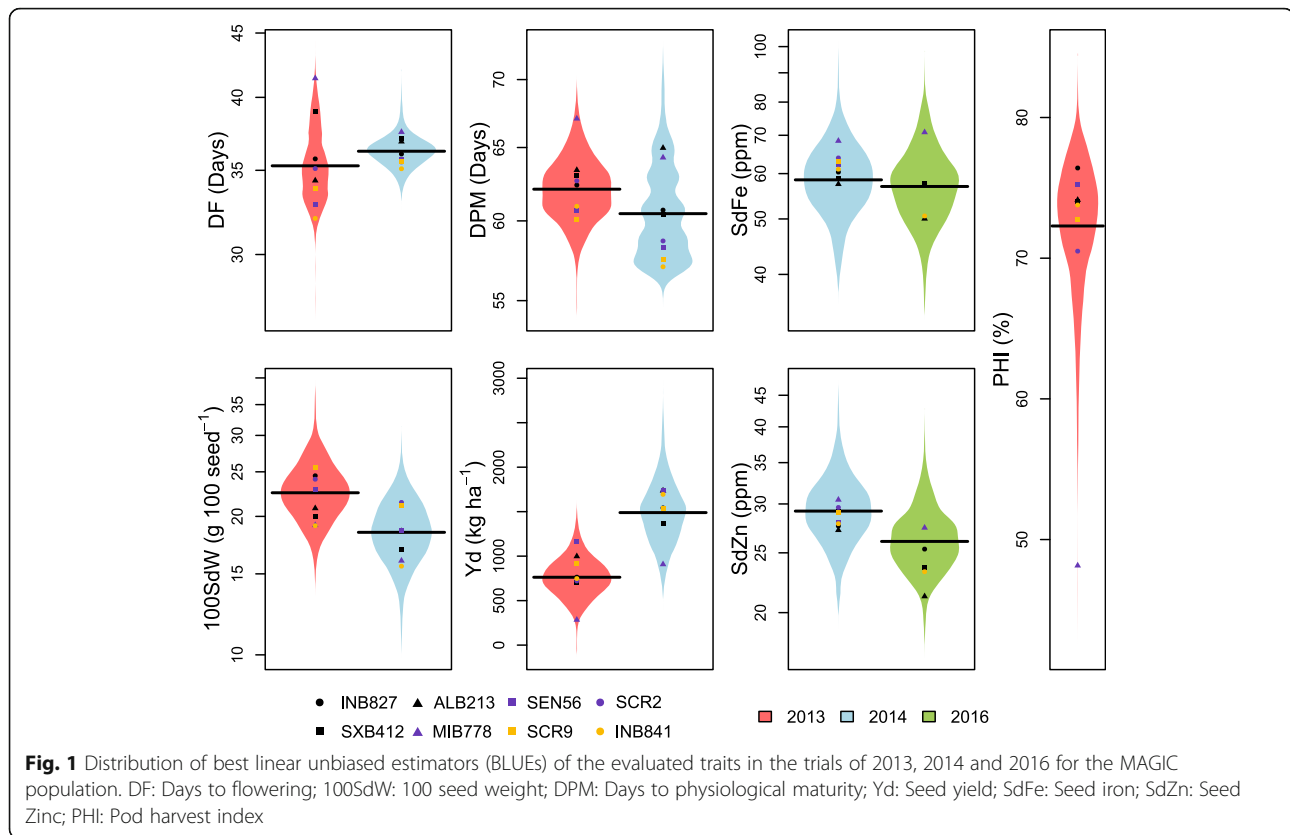
Genetic and population structure

The MAGIC population was genotyped by GBS, generating 20,615 polymorphic markers that were used to assess the population structure. A binning process to eliminate redundant markers was applied, resulting in 5738 non-redundant markers that were used to generate a genetic map and for QTL mapping. WGS data from the eight founder lines was used to impute 1,972,528 markers in the population for GWAS (Additional files 4 and 5). The

Table 1 Description of the eight common bean (*P. vulgaris*) founder lines of the MAGIC population

Line	Interspecific introgressions from	Pedigree	Valuable traits
SXB412		A686/A774//NXB80/SEA15	Low fertility and drought tolerance
INB827	<i>P. acutifolius</i>	INB108/INB605	Drought tolerance
ALB213	<i>P. coccineus</i>	SER16/G35346-3Q//SER16	Drought and Al tolerance
SEN56		SXB123/DOR677//SEN34	Drought tolerance
SCR2		NCB226/RCB591	Drought tolerance and BCMV ^a resistance
MIB778	<i>P. dumosus</i>	FEB226/G35575-2P//FEB226	High Fe/Zn seed content
SCR9		SER176/RCB591	Drought tolerance and BCMV ^a resistance
INB841	<i>P. acutifolius</i>	INB108/INB605	Drought tolerance

^aBCMV bean common mosaic virus



population structure was assessed by constructing a NJ tree and performing a PCA. All MAGIC and founder lines are evenly separated from the center of the tree (Fig. 3a), except for MIB778 which showed a longer matching distance. There are no defined clusters that separate lines in the tree, indicating no significant population structure within the MAGIC population. In line with these results, each principal component explains only a small proportion of the variance, as the first two principal components account just for 4.73% (Fig. 3b). The two-dimensional space defined by the first and second principal components show a uniform dispersal of the genotypes, with MIB778 drawing more distant from the other founders. This result mirrors the actual pedigree, since MIB778 has the least number of common ancestors with other founder lines (Additional file 6). These results indicate that within the MAGIC lines the population structure has a low level of complexity.

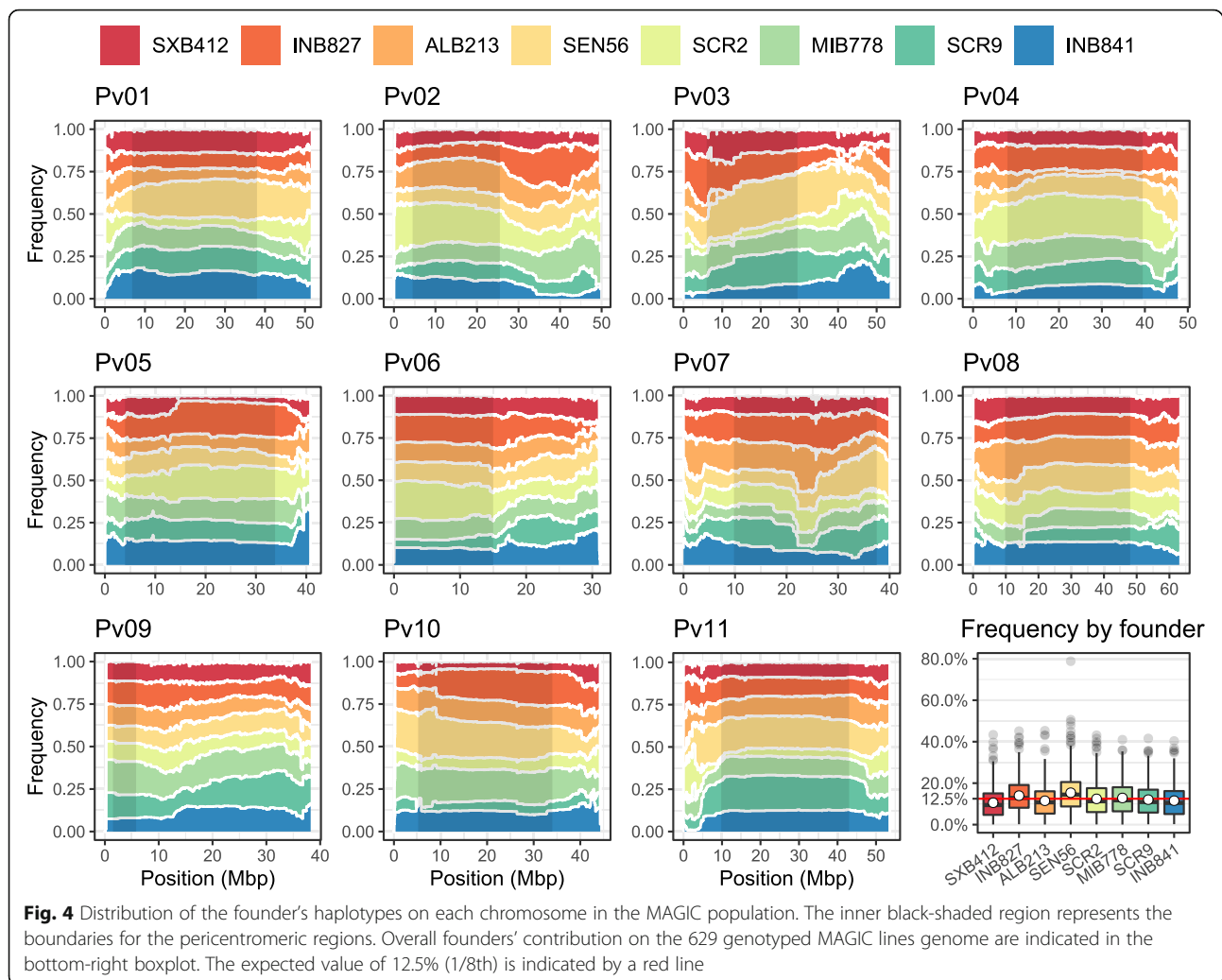
A haplotype analysis was performed to quantify the individual contribution from each founder line in the population. The average fraction of each founder on the MAGIC lines was similar, ranging between 10.6 and 15.4%. These values were close to the expected contribution of 1/8th (12.5%). However, this fraction was not constant along the chromosomes, especially in the euchromatic regions (Fig. 4), which may be attributed to arbitrary segregation after the early crosses that were

performed in a limited number of plants. The predicted haplotype composition for each RIL is presented in Additional file 7. Taken together, the reduced complexity level of population structure and an even representation of parental haplotypes in the MAGIC lines display the intended characteristics of the population to perform genetic analyses.

A genetic map for the MAGIC population was generated with a total length of 857 cM. The recombination rate was 2407 kbp/cM in the pericentromeric regions and 291 kbp/cM in the euchromatic regions of the chromosomes, with an overall rate of 603 kbp/cM. The MAGIC population had a mean of 3.22 recombination events per chromosome and a median of 44 recombination events per RIL in the entire genome (Additional files 7 and 8). The genome-wide rate of LD decay was 74 kbp at 0.19 r^2_V , about half of its initial predicted value. The rate of chromosome-specific LD decay ranged between 51 kbp (Pv08) and 154 kbp (Pv02) (Additional file 9). The high marker density and accuracy due to the large population size make this map suitable for the analysis of linkage and segregation patterns and for genetic mapping.

QTL analysis and genome-wide association study

Marker-trait associations were evaluated by both GWAS, using a Mixed Linear Model (MLM) approach, and QTL



mapping, using software designed for haplotype-based analysis of multi-parent populations. In total, 17 QTL for 7 traits were identified with significant peaks in the GWAS analysis (Additional file 10). GWAS peaks were selected with a significance greater than that established by the Bonferroni correction (2.42×10^{-6}) and selecting those regions that showed a clear peak above the background. QTL analysis resulted in 45 QTL for 7 traits exceeding the threshold of significance set by 1000 permutations ($\text{LOD} > 6.26$) (Additional file 11). High confidence QTL were defined as those showing significant marker-trait associations in both GWAS and QTL mapping. Following that rule, thirteen such “main QTL” were identified for 7 traits (Table 2). Founder haplotype assignment of phenotypic effects was generally the same between GWAS and QTL analysis (Additional files 10 and 11).

Most significant trait-associations were identified on chromosome Pv01 for phenology traits DF and DPM, in the region between 10 and 18 Mbp (Table 2). This

region showed additive effects of 0.1–1.7 days and 0.6–0.8 days for *DF1.2* and *DPM1.2* respectively. *DF1.2* and *DPM1.2* explained 35.8 and 5.5% of the phenotypic variance, respectively. A yield QTL was found in the same genomic region (*Yd1.1*), with phenotypic effects of up to 53 kg ha^{-1} . This genomic region represents a QTL hotspot with variation for several traits, as a significant GWAS association was identified also for SdZn (Additional file 10). Evaluating founder allele patterns at this QTL hotspot revealed that between 10.4 and 17.7 Mbp the alleles from SXB412, INB827 and MIB778 had a positive effect on phenology traits. In line with the negative correlation between Yd and phenology traits, alleles from these founder lines had a negative effect on Yd, (Additional file 10). QTL analysis results are mostly in accordance with GWAS results; SXB412, INB827 and MIB778 haplotypes always positively affected phenology traits at this locus. *DF1.2* evaluated in 2013 presented some deviations where SCR2 founder haplotype also had a strong positive effect (Additional file 11). *DF1.2/*

Table 2 Major QTL in the MAGIC population identified in both MLM-based GWAS and QTL analysis optimized for complex populations. QTL are listed that were significant in both analyses and show clear peaks in GWAS. A complete list of significant marker trait associations is available for GWAS (Additional file 10) and QTL mapping (Additional file 11)

QTL name	Trial	GWAS				QTL Mapping			
		Most significant marker ^a	p value	MAF	Effect ^b	Pos. (cM)	Nearest marker	LOD	PVE (%) ^c
<i>DF1.1</i>	2013	Pv2.1_01_4147947_C/G	5.46E-07	0.37 (G)	-0.46	32	Pv2.1_01_5474208_C/T	9.30	4.30
<i>DPM1.1</i>	2013	Pv2.1_01_763904_G/A	1.60E-06	0.27 (A)	0.43	36.5	Pv2.1_01_6773375_T/A	21.70	15.90
<i>DF1.2</i>	2013	Pv2.1_01_14672594_C/T	1.05E-35	0.34 (C)	1.63	38	Pv2.1_01_11238077_T/A	58.85	35.76
	2014	Pv2.1_01_17627372_A/G	2.36E-12	0.35 (G)	-0.28	40	Pv2.1_01_14582969_G/C	34.42	16.50
<i>DPM1.2</i>	2013	Pv2.1_01_14672594_C/T	2.77E-09	0.34 (C)	0.82	38.5	Pv2.1_01_13085941_C/A	8.19	5.47
	2014	Pv2.1_01_11567887_T/TA	2.13E-07	0.18 (T)	0.61				
<i>Yd1.1</i>	2013	Pv2.1_01_11250640_A/G	3.11E-08	0.16 (A)	-52.81	38	Pv2.1_01_11238077_T/A	8.76	2.03
	2014	Pv2.1_01_11221702_T/A	3.57E-07	0.21 (T)	-28.79				
<i>PHI2.1</i>	2013	Pv2.1_02_47566148_G/A	2.61E-08	0.37 (G)	-1.08	71	Pv2.1_02_47643879_G/T	10.73	7.62
<i>DF3.1</i>	2013	Pv2.1_03_39652029_G/A	5.23E-07	0.14 (A)	0.62	50	Pv2.1_03_39061112_G/C	7.50	3.43
	2014					50	Pv2.1_03_39061112_G/C	8.11	3.54
<i>DPM3.1</i>	2013	Pv2.1_03_39534987_C/A	2.08E-08	0.16 (C)	-0.75	48	Pv2.1_03_37341462_T/C	11.41	8.36
	2014					51.5	Pv2.1_03_39819948_G/A	10.24	5.11
<i>100SdW4.1</i>	2014	Pv2.1_04_1328790_G/A	4.08E-08	0.27 (A)	0.51	3	Pv2.1_04_522512_A/T	8.36	4.74
<i>SdFe6.1</i>	2014	Pv2.1_06_5311627_C/T	2.77E-07	0.17 (C)	2.09	2	Pv2.1_06_8725963_G/A	7.33	4.06
<i>SdFe6.2</i>	2016	Pv2.1_06_22844368_T/G	3.36E-07	0.08 (G)	-2.67	35	Pv2.1_06_23945432_T/C	6.46	8.68
<i>DPM8.1</i>	2013	Pv2.1_08_1070342_C/T	2.54E-09	0.11 (T)	-1.45	5	Pv2.1_08_1091562_T/G	7.92	5.56
	2014	Pv2.1_08_706882_A/G	3.58E-11	0.13 (G)	-1.77	4	Pv2.1_08_946243_C/T	18.15	9.09
<i>SdZn8.1</i>	2014	Pv2.1_08_60922896_C/G	4.39E-07	0.06 (C)	0.85	83	Pv2.1_08_60982396_C/G	6.81	2.65

^aMarker ID contains the physical position of the polymorphism in the reference version 2.1 and the genotype of the reference / alternative allele

^bEffect of alternative allele

^cPVE = Phenotypic variation explained

DPM1.2 showed seasonal stability being detected by GWAS and QTL analysis in both 2013 and 2014 trials. *Yd1.1* was detected at this locus in 2013; however, in 2014 a subtle QTL peak was visible but did not reach the significance threshold (Additional file 12).

Another QTL hotspot for maturity was observed on chromosome Pv03 at 37.3–39.9 Mbp (*DPM3.1* and *DF3.1*). In this region, the alleles of INB827 and INB841 had a negative allelic effect on DPM and DF. Similarly, on chromosome Pv06, between 4.4–9.1 Mbp, MIB778 alleles were associated with late maturity and low PHI. MIB778 was the latest flowering founder genotype; accordingly, it was involved in all three QTL on Pv01, Pv03 and Pv06.

A fourth QTL hotspot for maturity and PHI was identified on Pv08 (0–1.6 Mbp). The alleles from the founder line ALB213 were associated with later physiological maturity in both years and low PHI. An interspecific introgression in ALB213 from *P. coccineus* was reported in this same region on the intervals 881,865–884,647 and 1,181,015 to 1,234,070 bp [25], that may cause this phenotype (Table 2, Additional files 10 and 12). In all four QTL hotspots for maturity and yield component

traits, defined haplotypes increase maturity days and at the same time decrease yield-related traits, following the negative phenotypic correlations of these trait groups. An interesting QTL *PHI2.1* was found on chromosome Pv02. The alleles of the founder lines SCR2 and SCR9 have a negative effect of -1.08% on this trait (Additional file 10). This QTL was not linked to maturity traits, for this, it may be easier to employ *PHI2.1* in breeding.

Three main QTL for micromineral content were identified, *SdFe6.1*, *SdFe6.2* and *SdZn8.1* (Table 2). *SdFe6.2* at 21–24 Mbp was observed in two seasons (2014 by GWAS and 2016 by QTL analysis). MIB778 is the founder line with highest accumulation of Fe/Zn and provided the positive alleles for SdFe and SdZn, with an effect of 2.67 ppm in GWAS and 3.10 ppm in QTL analysis (Table 2, Additional files 10 and 11). *SdZn8.1* was located at 60–63 Mbp with an additive allelic effect of 0.85 ppm in GWAS and 0.92 ppm in QTL analysis, also contributed by MIB778 (Table 2). Micromineral levels were evaluated in drought and non-drought conditions to identify possible constitutive QTL. However, most QTL of Fe/Zn were not consistent between the two

years of evaluation, which may be due to specific genotype by environment (G×E) effects of drought stress. QTL analysis indicated several further independent QTL that were not detected by GWAS analysis. Among those, two major QTL for yield were located in Pv07 and Pv08 (*Yd7.1* and *Yd8.2*), which showed highly significant LOD scores (23.16 and 38.17 respectively). However, GWAS analysis showed no signal at these regions to validate those results (Additional files 10, 11 and 12).

Top 10 best performing lines were identified based on a weighted trait index (WTI). These lines show an enrichment of desirable haplotypes for major QTL, but no line holds positive alleles for all QTL as these don't explain the majority of phenotypic variation (Additional files 13 and 14). An outstanding line, MGC583 presents high yield in drought conditions in both trials (1100 and 1600 kg ha⁻¹) and a high accumulation of Fe and Zn (69 and 31 ppm respectively) (Additional file 13).

Candidate genes

GWAS and QTL analysis was performed using WGS data, based on imputation from re-sequenced founder lines. Hence, in principle all SNPs and small indels (< 20 bp) present in the population were evaluated in the GWAS. Close-ups of major QTL regions reveal that QTL analysis LOD curves and GWAS peaks often do not mark the exact same positions (Fig. 5 and Additional file 14); hence, candidate genes were considered based mainly on GWAS analysis data.

To identify candidate genes, only the polymorphisms in coding regions that had a moderate or high effect on the amino acid sequence (as defined by the SnpEff software, see Methods for details) were evaluated. Most candidate polymorphisms were found for the QTL hotspot on Pv01, which had the highest marker-trait associations, and covered the largest genomic region among the major QTL (Table 3). For *DFI.2*, 86 polymorphisms altering the protein sequences of 45 genes were identified and 8 SNPs in 4 genes are shared with *DPML.2* (Fig. 5 and Additional file 15).

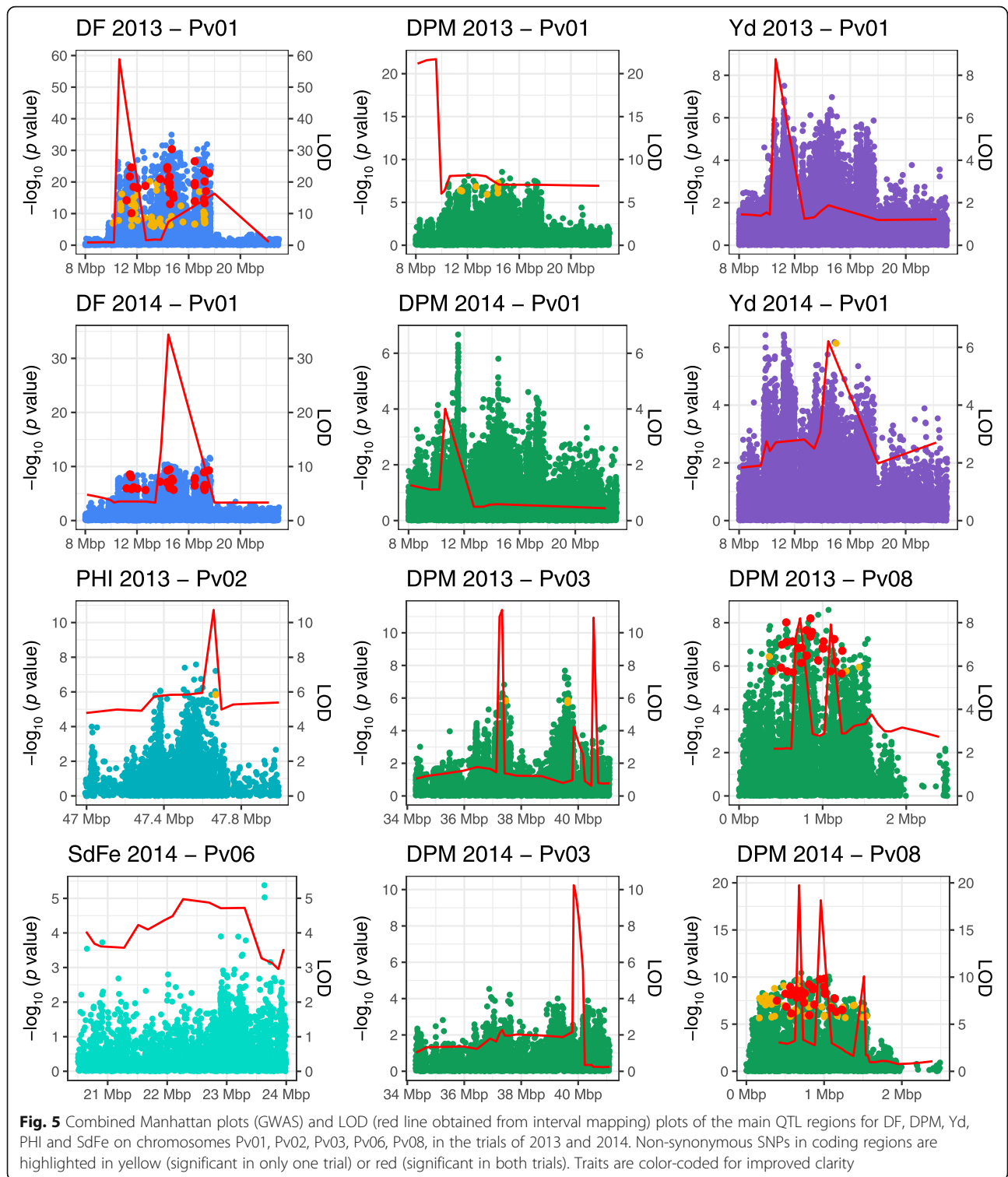
The QTL hotspot on Pv01 contains polymorphisms in genes directly or indirectly involved in flowering pathways. *Phvul.001G085200* encodes a Light Regulator Protein (*Lir1*), a homolog of the perennial ryegrass gene *LpLIR1*, that was reported to be involved in floral induction affecting the control of the circadian clock and vernalization [26]. *Phvul.001G087300* is a homolog of an Arabidopsis B-box zinc finger protein CONSTANS-like 9. CONSTANS and CONSTANS-like genes were reported in Arabidopsis as important regulators of flowering in response to inductive photoperiods, as a central component of photoperiodic floral control [27]. The BBX24 zinc finger transcription factor was also reported to be associated with flowering time and abiotic stress tolerance by repressing flowering and Gibberellin biosynthesis genes in

Chrysanthemum [28]. *Phvul.001G089900* is a homolog of the repressor of the gibberellin signaling pathway *RGL1*. This gene is involved in modulating floral development in Arabidopsis [29]. *Phvul.001G094300* is a homolog of Arabidopsis *BAF60*, a SWI/SNF subunit mediating ATP-dependent chromatin remodeling and directly regulating floral repressor FLOWERING LOCUS C [30]. One missense SNP for *Yd1.1* was found; however, annotation information is inconclusive as to regard it as a candidate gene. A short list of 10 most probable candidate genes for the QTL hotspot on Pv01 based on this evaluation is shown in Table 3. Given the large number of genes, a prime candidate cannot be pointed out.

In the QTL hotspot located on Pv03 for DPM and PHI, four polymorphisms in four genes were found. The annotations of these genes suggest their involvement in signaling or developmental processes, but a single candidate gene cannot be clearly attributed to the phenotype.

For the QTL hotspot for DPM and PHI on Pv08, 81 non-synonymous polymorphisms in 51 genes were found, two of which are shared between both traits. Several genes are involved in signaling or development, but the large number makes it difficult to determine a single clear candidate for phenology (Additional file 15). The most significant GWAS association in this QTL was found in *Phvul.008G012100*, next to three further non-synonymous SNPs found in that gene. This gene encodes a TATA BOX ASSOCIATED FACTOR II homolog, a group of genes reported to affect many regulatory processes [31]. A further candidate *Phvul.008G003500* encodes a protein of the NRT1/PTR family. These proteins were originally identified as nitrate or peptide transporters, and recently they have been reported to be transporters of auxins, ABA, and gibberellins [32]. *Phvul.008G014000* is a WUSCHEL-related homeobox 10-related (*WOX*) gene. *WOX* overexpressing lines have been reported with altered flowering times [33]. For *PHI2.1* two non-synonymous polymorphisms were identified in *Phvul.002G309200*, which encodes an AGAMOUS-like gene, a family reported to control development of flowers and fruits [34].

For *SdFe6.2* located on Pv06, four genes were identified with non-synonymous polymorphisms, including *Phvul.006G114800*, which encodes a MYB domain protein 4 homolog (Table 3). MYB-like transcription factors have been shown to affect iron deficiency in *Malus xiaojinensis* [35] or *Chlamydomonas reinhardtii* [36], among multiple other metabolic and developmental processes [37]. The ectopic expression of the *DwMYB2* transgene in *A. thaliana*, a distant homolog of *Phvul.006G114800*, regulated the expression of genes related to iron transporters and homeostasis in root and shoot [38]. Taken together, the evaluation of GWAS results from WGS imputed markers resulted in several plausible candidate genes that can be further evaluated with other methods.



Discussion

This work reports the first MAGIC population in common bean, constructed from eight Mesoamerican breeding lines. MAGIC populations have been developed recently in several plant species, such as Arabidopsis

[10], maize [39], rice [15], wheat [40], tomato [41] and *Vicia faba* [42], among others. MAGIC populations are designed to capture more genetic variability and deliver more precise genetic mapping compared to commonly used bi-parental RIL populations. Utilization of breeding

Table 3 Candidate genes for the major QTL identified in the MAGIC population. Genes are shown that harbor non-synonymous polymorphisms which had significant associations in GWAS. Most likely candidate genes for major QLT are shown, a complete list of candidate polymorphisms and genes in Additional file 15

Gene	Chr	QTL	Position (bp) ^a	Variant (# polymorphisms)	Gene annotation
<i>Phvul.001G077300</i>	Pv01	<i>DF1.2 - DPM1.2</i>	11,441,607	missense variant (1)	26S proteasome regulatory subunit N6
<i>Phvul.001G077700</i>	Pv01	<i>DF1.2</i>	11,571,295	splice region (2)	L-arabinokinase
<i>Phvul.001G085200</i>	Pv01	<i>DF1.2</i>	13,124,249	missense variant (1)	Light regulated protein Lir1
<i>Phvul.001G087100</i>	Pv01	<i>DF1.2 - DPM1.2</i>	13,513,228	missense variant (1)	GRAM domain (GRAM)
<i>Phvul.001G087300</i>	Pv01	<i>DF1.2</i>	13,548,191	missense variant (1)	Zinc finger protein CONSTANTS-like 14-related
<i>Phvul.001G089400</i>	Pv01	<i>DF1.2 - DPM1.2</i>	14,352,260	missense variant (5)	PPR repeat (PPR)
<i>Phvul.001G089900</i>	Pv01	<i>DF1.2</i>	14,661,652	missense variant (1)	GRAS domain family (GRAS)
<i>Phvul.001G090100</i>	Pv01	<i>DF1.2 - DPM1.2</i>	14,694,391	missense variant (1)	Exocyst complex protein EXO70
<i>Phvul.001G091400</i>	Pv01	<i>DF1.2</i>	15,553,393	missense variant (1)	Squamosa promoter-binding protein-like (SBP domain)
<i>Phvul.001G094300</i>	Pv01	<i>DF1.2</i>	16,832,664	missense variant (1)	SWIB/MDM2 domain
<i>Phvul.001G093700</i>	Pv01	<i>DF1.2</i>	17,339,464	missense variant (1)	Copine - DCD domain protein
<i>Phvul.002G309200</i>	Pv02	<i>PHI2.1</i>	47,667,592	missense variant (2)	AGAMOUS-like 29
<i>Phvul.003G158500</i>	Pv03	<i>DPM3.1</i>	37,464,971	missense variant (1)	Chitinase-like protein-related
<i>Phvul.006G114800</i>	Pv06	<i>SdFe6.2</i>	22,340,161	missense variant (1)	MYB-like DNA-binding protein MYB4
<i>Phvul.008G003500</i>	Pv08	<i>DPM8.1</i>	325,238	missense variant (3)	Protein NRT1/PTR family
<i>Phvul.008G005500</i>	Pv08	<i>DPM8.1</i>	512,002	missense variant (2)	LOB domain-containing protein 40
<i>Phvul.008G006800</i>	Pv08	<i>DPM8.1</i>	640,209	premature start codon variant (1)	Transducin family protein / WD-40
<i>Phvul.008G010800</i>	Pv08	<i>DPM8.1</i>	946,243	splice donor variant (1)	Dolichyl-phosphate beta-glucosyltransferase
<i>Phvul.008G012100</i>	Pv08	<i>DPM8.1</i>	1,001,427	missense variant (4)	transcription initiation factor TFIID subunit 6
<i>Phvul.008G013300</i>	Pv08	<i>DPM8.1</i>	1,091,953	missense variant (1)	subtilase family protein
<i>Phvul.008G014000</i>	Pv08	<i>DPM8.1</i>	1,154,068	missense variant (1)	WUSCHEL-related homeobox 10-related

material produces results that are theoretically more easily transferable to breeding programs [15]. In this work, founder lines were selected to represent breeding variability in the CIAT Mesoamerican breeding program. These cover several market classes, an important aspect in common bean breeding, and furthermore, several important agronomic traits such as tolerance to drought, low fertility and aluminum, resistance to virus and high accumulation of Fe/Zn. Hence, these lines represent relevant allelic variation, applicable for several breeding efforts.

Recombinations are the basis of genetic mapping and generate the genetic variability for breeders' selection. Due to several rounds of crossing, MAGIC populations offer increased recombinations and higher mapping precision than bi-parental populations. The MAGIC population presented here had on average 3.22 recombination events per chromosome. This value is expectedly higher than the averages observed for bi-parental RIL populations of common bean, ranging between 1.01 and 1.28 (DOR365xG19833, IJRxAFR298 and RCB593xINB841 populations, unpublished observations). In tomato, mean values of 2.49 and 1.46 recombination events were reported for MAGIC and bi-parental populations, respectively [41]. The rate of LD

decay observed for this MAGIC population was 74 kbp at $0.19 r_V^2$. This rate is faster than the 428 kbp reported for a natural population of cultivars/lines of the Mesoamerican genepool [43], Likewise, Blair et al. [44], describe a faster LD decay in an Andean population of common bean due to reduced population structure compared to a Mesoamerican population. Similar results have been reported for other MAGIC populations of winter-wheat [40] and cotton [45], and indicate that the consecutive rounds of mating may have broken linkage blocks, consequently reducing LD [45].

In this work, a high-density genetic map that incorporated multiple recombination events from eight founder lines using GBS SNP data was produced. This is the first map that incorporates such higher recombination frequency for common bean. It can be useful for improving the performance of some genotype imputation algorithms, incorporating recombination information with more precision [46]. In addition, the sigmoidal curves depicting recombination rates (Additional file 8) resembled those of previously published maps [44, 47]. However, the pericentromeric boundaries in the Mesoamerican founders' genomes deviate from those defined in the reference genome of G19833 (Additional file 8). These results can be useful to

study differences of local recombination events between the Mesoamerican and Andean gene pools. Furthermore, four founder lines were reported to bear interspecific introgressions from *P. coccineus* and *P. acutifolius* in chromosomes Pv07, Pv08, Pv09 and Pv11 [25]. Some of these regions display minor deviations on the proportional contribution (Fig. 4), but there is no clear signal of segregation distortion due to interspecific introgressions.

Breeding applications of the MAGIC population

Breeding for drought tolerance has been a long-standing effort at CIAT's and other common bean breeding programs. Yield has been used as a key selection criterion to advance the phenotypic responses under drought conditions.

One of the goals of the MAGIC population was to obtain lines with good agronomic performance and desirable QTL that are identified and tagged with markers to be used as resources for breeding programs. The average yield of the population was 763 and 1499 kg ha⁻¹ in 2013 and 2014 respectively, which is comparable to other populations phenotyped under drought conditions such as 1047 kg ha⁻¹ [8], 1030 kg ha⁻¹ [48], or 1173 kg ha⁻¹ [49]. Several RILs in the population showed superior performance compared to the founder lines. In this study, the top ten agronomically performing RILs were identified (Additional file 13). Some lines combine good yield performance with micro mineral levels. This is an important result because obtaining varieties with high yield and high accumulation of micronutrients has been a key challenge in the area of bean biofortification [6], due to the negative correlation between these two traits. Expectedly, the haplotypes of the major QTL that these lines carry do not completely explain performance, as many genetic factors that control these traits are not discovered. Yet, positive haplotypes of major QTL are enriched in the top lines. This suggests that MAS for major QTL and phenotypic selection should be combined to optimize genetic gain in selection.

Significant effort has been invested in the past in studies of physiology and genetics to obtain a deeper understanding of the agronomic responses under stress conditions beyond yield per se. The positive correlations between Yd and Yd components have previously been reported under drought conditions [48, 50–52]. In the same way, the observed negative correlations between Yd and phenology traits have been reported to be influenced by hot and dry growing conditions [53], revealing a strong dependence on the environmental factors for the relationship of these traits. Previous studies have used PHI as an effective selection criterion for identifying genotypes with improved drought tolerance [50]. Also, in this study, PHI had a positive phenotypic correlation with Yd. In general, these results show that

superior yielding lines under drought stress are characterized by full commitment to reproductive growth, effective seed filling and resource translocation, leading to large seed and high PHI, as well as early maturity for drought avoidance.

QTL for agronomic traits have been widely studied in common bean using QTL mapping and GWAS studies [54]. DF is a key trait in the adaptation of common bean and QTL for DF and DPM have been previously reported in all chromosomes of the bean genome, except for Pv10. In this study, two major QTL in Pv01 for maturity were found, *DF1.1/DPM1.1* and *DF1.2/DPM1.2* in intervals 5–7 and 9–17 Mbp respectively. Other studies report QTL for DF in the same region also under drought stress, and showed that Mesoamerican genotypes contributed the additive allelic effect [22, 51, 53, 55, 56]. These data suggest there is an important locus on the start of chromosome Pv01 that controls maturity, next to the known *fin* locus at the lower arm of that chromosome in the interval 47–52 Mbp [57]. Likewise, other studies have reported QTL for DF and Yd located near to *DF3.1/DPM3.1* by linkage analysis and GWAS [57–59] and *DPM8.1* [56, 60].

QTL for seed weight have been previously reported in all chromosomes [54]. The main QTL found in this study for seed weight (*100SdW4.1*) represents new allelic variation for this oligogenic trait. The QTL hotspots (Pv01, Pv03, Pv06 and Pv08) identified in this work control both phenology and yield related traits, where the same alleles increase DF/DPM and decrease Yd and related traits. Hence, breeders may not be able to utilize these QTL for yield improvement, as primary consideration may have to be given to the maturity requirements for a certain target region and maturity is often negatively correlated with yield [53].

In this work, four PHI QTL were found on chromosomes Pv01, Pv02, Pv06 and Pv08 (Additional file 10). Mukeshimana et al. [51] reported two QTL under drought stress on Pv01 and one of these QTL was in the same region of *PHI1.1* (around 2 Mbp). Previous studies reported PHI QTL on chromosomes Pv06 and Pv08, at alternative chromosomal regions as identified here [53, 61]. *PHI6.1* could be useful in breeding in some genetic backgrounds, to avoid the negative allele from founder line MIB778, which was the lowest yielder in all trials. The main QTL found in this study for PHI (*PHI2.1*) has not been reported previously in other studies. *PHI2.1* appears to be independent from phenology traits and could be used for marker development, being associated with raising PHI values by 1.08%.

For seed Fe content, two main QTL located in Pv06 (*SdFe6.1* and *SdFe6.2*) are reported, contributed by MIB778, the founder with highest Fe and Zn accumulation. *SdFe6.1* and *SdFe6.2* have allelic effects of 2.1 and

2.7 ppm and could be employed in biofortification breeding. The allelic effects of the markers are moderate, also observed by Izquierdo et al. [62] and Ponce et al. [63]. This is according to expectation, as quantitative traits in elite breeding lines are not likely to vary in major effect loci, such as those reported for disease traits. Other QTL have been reported near to *SdFe6.1* and *SdFe6.2* in Andean populations [64, 65], a Mesoamerican population [66] and an inter-genepool population [67]. This data suggests that allelic variation at these loci exists in both Andean and Mesoamerican genepools. GWAS showed that positive alleles from MIB778 are mostly identical with the reference genome (Andean genotype G19833) alleles. Accordingly, this founder was reported to have a large Andean introgression on chromosome Pv06 in the same region [25]. The main QTL for seed Zn content (*SdZn8.1*) in this study was also based on the haplotype from MIB778. Previously reported QTL for Zn in Pv08 were located at an interval of ~ 8 Mbp from *SdZn8.1* [65, 66]. The founder line MIB778 was added to the MAGIC founders as a source for its high micronutrient levels. It supplied the strongest QTL, which supports the semi-quantitative inheritance of this oligo genetic trait.

The responses to drought stress include morphological, physiological, biochemical and genetic mechanisms closely related, such as early flowering and maturity (escape mechanisms) [51], or the control of the root system architecture [68]. Some QTL related to this trait have been reported near the QTL hotspot identified in this study. A QTL controlling the number of root branches and root length diameter are located near to *DF1.1/DPM1.1* and *DPM8.1* respectively [69, 70]. It has been suggested that QTL affecting root traits in common beans are based on the constitutive expression of genes, and drought tolerance based on characteristics of the roots can be used in molecular breeding [71].

Previous studies on QTL for agronomic traits in common beans employed crosses between genetically contrasting parents, following the basic logic of QTL population development and mapping. This approach has been very successful in identifying and transferring disease resistance genes for, e.g. common bacterial blight (CBB) [72] or angular leaf spot resistance (ALS) [73]. It has also been used to analyze abiotic stress-related traits like drought tolerance [51, 56, 60] or tolerance to low levels of phosphorus [53]. However, the usefulness of this approach has been questioned for quantitative traits, because evaluating contrasting lines means to include alleles with undesirable agronomic effects for the investigated trait, which are likely to have been selected out of breeding populations. Strongly contrasting crosses are usually not used in breeding because it would result in much undesirable variation, particularly to regenerate

commercial grain quality. For this reason, data from contrasting crosses is often not easily transferable to breeding applications. In contrast, genetic variation in a MAGIC population is analyzed in elite genetic backgrounds without detrimental exotic alleles; hence, data can be directly transferable to breeding populations.

Comparison of genetic mapping methods: GWAS and QTL mapping

In this study, genotype-phenotype associations were simultaneously evaluated using linkage mapping with a method designed for eight-way MAGIC populations and using association mapping with the MLM approach. Using the QTL mapping strategy, the QTL *DF1.1*, *Yd1.1* and *DPM3.1* were identified in the two evaluations in 2013 and 2014. However, close-ups of these regions show that the significant LOD peaks do not overlap with each other, suggesting two different adjacent QTL that affect the trait independently in two seasons. Similarly, QTL mapping software IciMap [74] from the same author group was previously suggested to delimit QTL regions too much, leading to comparable problems of QTL proliferation in narrow regions, which does not appear biologically sensible [53]. In addition, the linkage strategy occasionally produced LOD peaks with very large significance scores that had no corresponding GWAS signal. Given that the interval QTL mapping software can utilize the known population and haplotype structure it should theoretically constitute a more powerful tool for genetic mapping, but the results presented here leave doubts if the data is completely reliable. On the other hand, GWAS results varied by applying different modifications of the GLM and MLM approaches (data not shown). Whereas major marker-trait associations are usually retained, significantly different results for intermediate and minor association regions can be obtained depending on the chosen analysis method, while no gold standard has been established in the literature for a correct analysis. For this reason, we reported reliable major QTL regions in this study that were supported by the significant signals produced by both the linkage and association strategies, providing enough evidence to delimit the associated regions, whereas only the GWAS results seem to be suitable to search for candidate genes.

Value of WGS and imputed GWAS/QTL for candidate genes

Previous studies in common bean have identified candidate genes relying on either a priori candidates based on the published literature or comparing the functional annotation of gene models within a window, centered around significant SNPs markers e.g. [19, 22, 75–77]. The availability of WGS data for GWAS allows to largely

delimit the list of potential candidate genes by removing those that do not show polymorphisms in the evaluated population. In this work, we performed GWAS using imputed WGS data from the founder lines. In principle, this could take in all polymorphisms in the population compared to the reference genome, including those that cause the phenotypic variation. This can lead to the accurate identification of candidate genes by tracing the polymorphism with the most significant association.

The strategy implemented in this study for gene mapping harbors some limitations. In the first place, the Andean reference genome of G19833 used here leaves out of the association analysis the unique Mesoamerican regions present in the founder lines. These regions would be included using a Mesoamerican reference, yet the reference genome of BAT93 still presented some contiguity and completeness issues, making it unsuitable to be used in this study. Furthermore, the founder lines likely contain introgressions from sister species, harboring unique coding regions that are not traceable and thus are left out. For example, *DPM8.1* is based on an allele from ALB213 where a *P. coccineus* introgression was reported previously [25]. Despite that, the mapping rates of the founders' sequences were all close to 90%, which shows that most genomic regions that are common between both gene pools have a low divergence, suitable for identifying and testing variants in the population. In addition, the polymorphism calling did not consider long structural variants (> 20 bp). Finally, we only evaluated modifications to the protein coding sequence of the reference genome to identify candidate genes; hence, any type of promoter mutations or non-annotated genes would be missed. This can be an important limitation and should be considered for further studies, because larger variants (insertions/deletions > 20 pb) can be found in promoter regions of the genes, causing functional changes such as those reported for the *GSE5* promoter in rice [78], *stiff1* promoter in maize [79] or *PHYA3* promoter in common bean [80].

Following the approach of candidate gene identification using imputed WGS data, several candidate genes were identified for the evaluated agronomic quantitative traits. For the major phenology QTL on Pv01 and Pv08, many associated non-synonymous polymorphisms were found. Even though several plausible candidates are listed, no primary candidate stands out, which suggests that the genetic resolution, albeit quite large in this MAGIC population, is not enough to narrow down the candidate genes to a very small number. The strategy implemented in this study using resequencing of parental lines can be employed in other available RILs or multi-parental populations as a low-cost strategy for candidate gene identification. A similar strategy has been effective in the identification of QTLs for traits of

variable genetic complexity in MAGIC populations of tomato and maize [39, 41] and the identification of candidate genes in rice [81].

All methods of candidate gene identification should be followed up by direct candidate gene validation. Some strategies of virus induced silencing or genetic transformation have been tested in common bean with a few successful reported cases [82–88]. It shows that candidate gene validation is still a complex task for the species and therefore not widely adopted up to date [24]. In that sense, EcoTILLING approaches could also be pursued to identify further functional alleles of a candidate gene in order to validate the gene function and identify further variability for breeding [83]. Currently only few WGS data sets have been published [25, 89, 90], but ongoing projects will make a much larger set of WGS data available in the near future that could be mined for allelic variation in genes of interest.

Conclusions

This study presented the first common bean MAGIC population of the Mesoamerican gene pool. A genetic map comprising multiple recombination events between the founder lines was generated. To our knowledge, this map represents the largest and most dense genetic map available in common bean. The results presented here demonstrate that GWAS and haplotype-based interval mapping are successful tools in this population, identifying QTL for quantitative agronomic traits under drought conditions. Major QTL were identified to be controlling more than one trait, even in different seasons. This result is in line with the phenotypic correlations observed between some phenology and agronomic traits, suggesting there is extensive genetic correlation among them. Information on QTL can be used for molecular marker design for molecular breeding. Candidate genes for major QTL were identified using imputed WGS data from founder lines for GWAS. This method can be employed in RIL and MAGIC studies in common bean and other crops. Hence, this project provides data for applications in breeding and breeding tool development, especially for drought tolerance. This will support efforts to develop climate resilient germplasm, as well as information for basic research questions aiming to uncover the genetic basis of important agronomic traits.

Methods

Population development

Eight Mesoamerican common bean elite lines were selected from the breeding program at CIAT as founders of an 8-way MAGIC population: SXB412 (A), INB827 (B), ALB213 (C), SEN56 (D), SCR2 (E), MIB778 (F), SCR9 (G) and INB841 (H). These lines were selected based on genetic diversity (introgressions from *Phaseolus*

acutifolius, *Phaseolus dumosus* and *Phaseolus coccineus*), phenotypic diversity for abiotic tolerance, micromineral concentration, disease resistance, and agronomic performance (Table 1).

The breeding scheme of the MAGIC population used in this study is shown in Additional File 1. In brief, the eight lines were crossed in four pairs to create F_1 seed of SXB412 x INB827 (AxB), ALB213 x SEN56 (CxD), SCR2 x MIB778 (ExF) and SCR9 x INB841 (GxH). F_1 plants were then intercrossed (AxB x CxD) and (ExF x GxH) to generate 323 “4-way” (ABCD) and 272 “4-way” (EFGH) F_1 seed. These F_1 plants were intercrossed once again (ABCD x EFGH) to generate 728 “8-way” (ABCDEFGH) F_1 seed. 500 lines were randomly selected to advance to F_2 . Two plants per F_2 family were selected to assure that the variability of segregation in the F_2 would not be lost, and advanced to F_5 through single seed descent, to obtain 996 RILs. DNA was collected from F_5 individual plants. A first field trial was carried out with 636 RILs from the bulk harvested $F_{4,6}$, in 2013. Trial sizes were limited by available field space; entries from the complete population were randomly selected. Individual F_5 selections were advanced and 599 $F_{5,7}$ RILs were phenotyped in the 2014 trial. Due to a communication error between programs, the two genotype sets were not identical, so that both trials share 437 RILs evaluated in the two seasons (Additional file 1).

Field trial design and phenotyping

The MAGIC RILs and eight parents were planted at the International Center for Tropical Agriculture (CIAT) in Palmira, Colombia (with an altitude of 1000 m.a.s.l., latitude of 3° 32' N and longitude of 76° 18' W) in 2013 and 2014. The field experimental design for both trials was an alpha-lattice incomplete-block design with three replicates. Each genotype was laid out in two-row plots in 2013 and one-row plots in 2014 of 2.22 m² each. Around 10% of the plots in 2013 were used for planting 5 different check lines evenly distributed across the field, and 7.5% of the plots in 2014 were used for founder lines randomly distributed across each replicate. Border plots surrounding the trial plots were planted in both trials. These trials were carried out in the dry season (precipitations: 96.3 mm in 2013 and 250 mm in 2014 throughout the crop cycle, average temperature: 24.7 °C in both seasons, see Additional file 2). Three irrigations were applied, the first three days before sowing and the two others at 10 and 21 days after sowing. Standard field practices were applied over the plant growing seasons across years including the application of fungicide seed treatment and foliar insecticides.

The number of days to flowering (DF) was measured from planting to the day when 50% of the plants in the plot had at least one open flower. Days to physiological

maturity (DPM) was measured as the number of days from planting until 50% of plants had at least one pod losing its green pigmentation [50]. Yield (Yd, kg ha⁻¹) was obtained per plot and corrected for the percentage of moisture of the seed (seed moisture of 14%). Seed weight (100SdW, g 100 seeds⁻¹) was obtained from 100 seeds. These four traits were measured in both 2013 and 2014 trials. At the time of the harvest, 0.3 m² per plot were collected separately to measure pod harvest index (PHI, %) defined as the ratio between seed weight to pod weight. PHI was measured only in 2013. The samples to evaluate iron and zinc concentration in the seed (SdFe and SdZn, ppm) were prepared according to the method described by Stangoulis and Sison [91] and quantified by X-ray fluorescence method using an Energy dispersive X-ray fluorescence (EDXRF) instrument X-Supreme 8000 (Oxford Instruments, UK) [92]. Micromineral concentration was evaluated in three replicates in 2014 and in a non-replicated trial in 2016. Additional information on description of phenotypic traits can be found in “Trait Dictionaries for Fieldbook Development” (www.cropontology.org).

Phenotypic data analysis

The field map of the plots was used to assign row and column coordinates. The phenotypic data of each trial was analyzed by fitting a linear mixed model with random effects for rows and columns using the functions ‘SpATS’ and ‘PSANOVA’ [nseg = c (180,24) and nseg = c (36,54) for the trials in 2013 and 2014 respectively] of the R package SpATS (v1.0–9) [93]. To model the spatial variability in the field, this model contains a smooth bivariate surface composed of a parametric and a smoothing component. The parametric component includes the intercept, fixed linear trends along rows and columns and their linear interaction trend. The smoothing component models the deviation from the previous compound linear trend using one-dimensional and tensor product P-splines. The effect of each line on the phenotype was fitted as fixed and random to obtain the best linear unbiased estimators (BLUES) and best linear unbiased predictors (BLUPs), respectively. The heritability was calculated using the function ‘getHeritability’ of SpATS, which uses the effective and nominal dimensions of the genotypic component calculated in the model [93]. Hence, the heritability was only obtained when the genotypic term was taken as random. BLUES were used to calculate Pearson correlation coefficients among each trait-trial combination assessed in this study, and their significance was tested using a two-tailed *t*-test.

To select the top ten agronomically performing RILs, a weighted trait index (WTI) was constructed for each trial using BLUPs. This WTI included the traits Yd, DF

and DPM with a weight of 40, 15 and 15% respectively. The remaining 30% was assigned to PHI in WTI of 2013 and SdFe in WTI of 2014. To construct the WTI, every trait was scaled using the Z transformation, assigning positive and negative scaled values to good and poor performing lines respectively (early maturity is interpreted as good performance). These values were combined using their corresponding weights for each trial separately. Then, the weighted averages for each trial were summed up and the top ten lines with highest values were selected.

Genotyping

DNA extraction and sequencing of the founder lines are described in detail by Lobaton et al. [25]. In summary, the DNA was extracted from young leaves using liquid N_2 and the urea buffer-based extraction miniprep protocol. The libraries were prepared using the Tru-Seq DNA PCR-Free library preparation kit, and the sequencing was performed by the HudsonAlpha Institute for Biotechnology, generating paired-end reads, yielding a raw sequencing depth ranging between 7x and 10x. The DNA extraction and genotyping by sequencing (GBS) of a subset of 629 MAGIC lines is described in detail by Perea et al. [94], following the same DNA extraction protocol described above. The DNA was sent to the Cornell sequencing facility for the GBS library preparation using the restriction enzyme *ApeKI* [95]. Each plate of 96-wells was sequenced in two lanes of an Illumina HiSeq platform using single-end reads. On average, each sample had a raw sequencing depth of 0.4x.

The mapping and variant calling processes for the founder lines is described in detail by Lobaton et al. [25]. Multi-allelic sites in the identified variants were split into bi-allelic sites using the module 'norm' with the options -m and -f from bcftools (v1.8) [96], yielding a total of 6,284,436 bi-allelic variants from the founders' sequencing data (Additional file 4). The GBS reads were demultiplexed using NGSEP (v3.1.2) [97]. Adapters and low-quality bases from the raw sequencing data were trimmed using Trimmomatic (v0.36) [98], and the processed reads were aligned to the reference genome of *P. vulgaris* accession G19833 v2.1 [46]. using Bowtie2 (v2.2.30) [99] with default parameters. The variant calling process was performed using NGSEP following recommended parameters for GBS data [94]. The list of variants identified previously in the founder lines was used as the variants to be genotyped in the MAGIC lines. The resulting genotype matrix was filtered for variants with a genotype quality above 40, MAF above 0.05, and at least 260 individuals genotyped per site. The final genotype matrix from GBS data contained 20,615 variants with ~25% missing genotype calls in the whole matrix (Additional files 4 and 5). The GBS matrix was

used to assess the population structure in the MAGIC lines by performing a principal component analysis (PCA) using GAPIT (v3.0) [100], and constructing an unrooted neighbor-joining (NJ) tree using SplitsTree (v4.14.16) [101]. The GBS matrix was also used to calculate pairwise measures of LD in sliding windows of 100 markers for each chromosome. The LD measures were corrected for kinship relationships in the population (r_V^2) as implemented in the R package LDcorSV (v1.3.2) [102]. The LD decay was estimated regressing the pairwise r_V^2 values on the physical distance of their markers using the locally estimated scatterplot smoothing implemented in the R function 'loess' (v3.6.3), with a span value of 0.5.

To quantify the proportional contribution of genomic information from the founders' genomes to the MAGIC lines, the parental haplotype blocks in the population were estimated using the method proposed for haplotype prediction in an *Arabidopsis thaliana* MAGIC population [10] (<http://mtweb.cs.ucl.ac.uk/mus/www/19genomes/magic.html>). This was performed using the whole set of founders and GBS markers, masking out indel variants and repetitive regions in the reference genome [25]. This method infers the breakpoints in the haplotypes by a dynamic programming algorithm, akin to the Viterbi path from a hidden Markov model (HMM). To construct a genetic linkage map, the inferred haplotype blocks and the GBS matrix were used to calculate genetic distances among markers using the Kosambi mapping function implemented in the integrated genetic analysis software for multi-parental pure-line populations (GAPL v1.2) [103]. The GBS marker set in this matrix was reduced by a binning procedure based on their recombination frequency as implemented in GAPL, generating a subset of 5738 non-redundant markers (Additional file 4). To impute the founders' WGS variants in the RILs, the founders' markers and the GBS markers were merged into a single matrix of 6,284,436 variants and 637 individuals. The missing data in this matrix was imputed with Beagle (v5.0) [46], providing the genetic linkage map (5738 non-redundant markers) and setting the effective population size to 8. The final imputed matrix was filtered for variants with MAF below 0.05, producing a matrix of 1,972,528 markers with genotype calls for all 637 samples (Additional files 4 and 5) that was used thereafter for the GWAS analyses.

QTL analysis and GWAS

QTL analysis was conducted using the genetic map described above and the genotypic BLUPs from the trials separately. Detection of QTL and estimation of the genetic parameters for each trait evaluated were performed using the composite interval mapping with the procedure for

additive effects (ICIM-ADD) of the software GAPL (v1.2) [103], employing the forward and backward regression model, with a 5 cM window size and a 0.5 cM sliding window. A significant QTL was declared if the logarithm of odds (LOD) was greater than the significance threshold of 6.26, which was obtained by performing a permutation test 1000 times with a p value of 0.05 to minimize the experimental type-I error rate. To identify the best allele for each QTL, a Tukey's multiple comparison test ($\alpha = 0.05$) was performed using the founders' haplotypes, assessing the effect of each haplotype separately.

Based on the results from the population structure assessment, different modifications to the general linear model (GLM) and the mixed linear model (MLM) approaches were tested for GWAS. This test included variable number of principal components as covariates, different methods to calculate the genetic relatedness (kinship) matrix, and different model implementations in the software packages Tassel (v5.2.44), GAPIT (v3.0) [100] and GENESIS (v2.8.1) [104] (data not shown). The selection criteria for the tested models was based on the calculation of the mean squared difference (MSD) between the observed and expected p values [105]. The model with the lowest MSD was obtained with the R package GENESIS (<https://github.com/UW-GAC/GENESIS>). This model accounts for population structure using the top five principal components (described previously) as fixed effects. It also accounts for random polygenic effects with a kinship matrix as variance-covariance structure, calculated using the EMMA algorithm implemented in GAPIT (v3.0) [100]. Significant associations with the trait of interest were declared when the p value was equal to or smaller than the Bonferroni threshold calculated with the GBS markers (2.42×10^{-6} with 20,615 markers). The association and linkage analyses were performed using a subset of 444 lines for the 2013 trial and 602 lines for the 2014 and 2016 trials, for which genotypic and phenotypic data was available.

Candidate gene identification

To detect putative candidate genes and candidate polymorphisms affecting the phenotypic variation, the major QTL regions identified by both QTL analysis and GWAS were selected. This search was restricted to the significant variants that had a high or moderate effect in the coding regions of the reference genome as defined by the annotation using SnpEff (v4.3) [106]. In addition, the gene expression data reported by O'Rourke et al. [107] and Phytozome (v12.1.6) (<https://phytozome.jgi.doe.gov>) was used to check if the selected genes had relevant expression levels in the tissue of interest. A gene was declared a candidate if its gene ontology description in Phytozome (v12.1.6) included a function related to the trait evaluated.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-020-07213-6>.

Additional file 1. (a) Crossing and selection scheme of the common bean MAGIC population. (b) Venn diagram showing the number of RILs used for genotyping and phenotyping in each trial.

Additional file 2. Precipitation, maximum and minimum temperatures during trials at Palmira, Colombia.

Additional file 3. Phenotypic variability, least significant difference (LSD) and broad-sense heritability (H^2) for best linear unbiased predictors (BLUPs) of the evaluated traits in the trials of 2013, 2014 and 2016 of the MAGIC population.

Additional file 4. Distribution of markers per chromosome obtained from WGS of the eight founder lines. Markers from GBS of the whole population and the resulting thinned markers used to construct the genetic map for QTL analysis are listed.

Additional file 5. Heat map of the density of markers called from WGS and GBS along the eleven chromosomes of the *P. vulgaris* reference genome. Each color band represents a region of 250 kbp. The inner black lines represent the boundaries of the pericentromeric regions as defined by Schmutz et al. [47].

Additional file 6. Pedigree tree for six of the eight founder lines of the MAGIC population. The founders are highlighted in dark gray at the lower tips of the tree. Exact pedigrees of INB lines are not currently available. This tree was generated using Helium (v1.18.03.15).

Additional file 7. Haplotype composition of the MAGIC lines. The parental source, the length and genomic construction of each haplotype in each MAGIC line is shown, using the method proposed by Kover et al. [10].

Additional file 8. Comparison between the physical location and the recombination frequency (genetic map position) based on thinned GBS markers (5,738 markers). The dashed horizontal lines represent the boundaries of the pericentromeric regions as defined by Schmutz et al. [47].

Additional file 9. Pattern of linkage disequilibrium (LD) decay calculated genome-wide (black-dashed line) and for each chromosome separately (colored lines) in a MAGIC population of common beans. The pairwise measures of LD were calculated in sliding windows of 100 markers and corrected for kinship relationships in the population ($rV2$). Each line corresponds to a locally estimated scatterplot smoothing (LOESS) regression on the LD measures.

Additional file 10. Significant markers identified in genome wide association studies, genetic and physical position, p value, allele frequency, phenotypic effect and founder genotypes associated with 9 traits in the MAGIC population evaluated in 2013, 2014 and 2016. Favorable alleles are colored in green.

Additional file 11. Details of QTL identified by interval mapping, genetic and physical position, LOD, phenotypic variation explained, and founders' allelic effects mapped for 7 traits in the MAGIC population evaluated in 2013, 2014 and 2016.

Additional file 12. Manhattan, quantile-quantile and LOD plots of the association and linkage mapping for each of the evaluated traits. The Bonferroni correction threshold ($p = 0.05$) using the WGS (1,972,528) and the GBS (20,615) markers are depicted as red and green horizontal dashed lines, respectively, in the Manhattan plot. The significance threshold for the QTL mapping analysis is depicted as the blue dashed line in the LOD plot.

Additional file 13. Phenotypic performance of top 10 RILs of the common bean MAGIC population and their genotypes at 12 major QTL. Phenotypic values for each of the top ten MAGIC RIL lines are shown for each trial, color coded from desirable (green) to undesirable (red) values. For each RIL line founder haplotypes for the top 12 QTL are shown together with the haplotype effects color coded from desirable (blue) to undesirable (red) QTL effects. The data mean phenotype for each parental haplotype in the main QTL and maximum and minimum trait

values or haplotype effects are shown below. Letters indicate significant differences using Tukey test ($\alpha = 0.05$).

Additional file 14. Haplotype composition per chromosome of the top 10 RILs of the common bean MAGIC population.

Additional file 15. Significant polymorphisms in major QTL identified by both methods MLM GWAS and QTL mapping based on haplotypes that affect protein coding regions of candidate genes with moderate or high effect as defined by SnpEff (v4.3). Gene expression data added from O'Rourke et al. [105] and [Phytozome.org](https://www.phytozome.org) (v12.1.6).

Abbreviations

MAGIC: Multiparent advanced generation intercross; GBS: Genotyping by sequencing; WGS: Whole genome sequencing; QTL: Quantitative trait loci; GWAS: Genome-wide association study; MNM: Micronutrient malnutrition; MAS: Marker-assisted selection; LD: Linkage disequilibrium; RIL: Recombinant inbred line; Yd: Yield; DF: Days to flowering; DPM: Days to physiological maturity; 100SdW: Seed weight; PHI: Pod harvest index; SdFe: Iron accumulation in seed; SdZn: Zinc accumulation in seed; MLM: Mixed linear model; GxE: Genotype x environment; WTI: Weighted trait index; CBB: Common bacterial blight; ALS: Angular leaf spot; EDXRF: Energy dispersive X-ray fluorescence; BLUEs: Best linear unbiased estimators; BLUPs: Best linear unbiased predictors; PCA: Principal component analysis; NJ: Neighbor-joining; HMM: Hidden Markov model; LOD: Logarithm of odds; GLM: General linear model

Acknowledgments

We would like to thank the contributions and technical support made by the bean breeding team at the International Center for Tropical Agriculture (CIAT) for phenotyping work and population development. We are also thankful to HarvestPlus for providing support on the micromineral analysis work and the Data management team at CIAT for making the datasets used in this work available with open access.

Authors' contributions

SB conceived the study and designed the crossing scheme for the population. CC, VM carried out the breeding population advancement and field trial evaluation. DAS, AFG, JFDLH did the phenotypic data analysis. DAS, JDL, JFDLH, FA, JD performed the genotypic data analysis. SD, DAS, PI did the association and linkage analyses. SD, BR contributed the candidate gene identification. SD, DAS and BR wrote the manuscript. SD and DAS should be regarded as Joint First Authors. All authors read, contributed to and approved the final version of the manuscript.

Funding

We would like to acknowledge funding support from the Tropical Legumes I - Improving tropical legume productivity for marginal environments in sub-Saharan Africa (Grant No: OPPGD 1392) and Tropical Legumes III - Improving Livelihoods for Smallholder Farmers: Enhanced Grain Legume Productivity and Production in Sub-Saharan Africa and South Asia (OPP1114827) projects funded by the Bill & Melinda Gates Foundation. This project was supported by the CGIAR Research Program on Grain Legumes and we would like to thank USAID for their contributions, particularly in the form of the "Global Hunger and Food Security Research Strategy: Climate Resilience, Nutrition, and Policy - Feed the Future Innovation Lab for Climate Resilience in Beans (CRIB)" Project (prime award number AID-OAA-A-13-00077, sub award number 4943-CIDT-AID-0077).

Availability of data and materials

The raw sequencing data used in this study is available at the NCBI sequence read archive (SRA) database with the bioproject accession numbers PRJNA294602 and PRJNA289910 as published by Perea et al. [82] and Lobaton et al. [25]. The variants for the eight founder lines used in this study are available for downloading at dryad (<https://doi.org/10.5061/dryad.46pk7>) and for browsing at the European Variation Archive database of the European Bioinformatics Institute with the accession number PRJEB18671 and biosample accession numbers SAMN07312849, SAMN07312850, SAMN07312851, SAMN07312852, SAMN07312853, SAMN07312854, SAMN07312855 and SAMN07312856. The GBS matrix and the genetic map, as well as raw and modeled phenotypic data used in this study are available for download at Harvard Dataverse (<https://doi.org/10.7910/DVN/JR4X4C>).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Bean Program, Agrobiodiversity Area, International Center for Tropical Agriculture (CIAT), Cali, Colombia. ²Present Address: Department of Plant Soil and Microbial Sciences, Michigan State University, East Lansing, MI, USA. ³Present Address: School of Environmental and Rural Sciences, University of New England, Armidale, SA, Australia. ⁴Present Address: Bioinformatics Interdepartmental Ph.D. Program, University of California, Los Angeles, Los Angeles, CA, USA. ⁵Departamento de Agronomía, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, Colombia. ⁶Present Address: Systems and Computing Engineering Department, Universidad de los Andes, Bogotá, Colombia. ⁷Present Address: Progeny Breeding, Madrid, Colombia.

Received: 30 March 2020 Accepted: 5 November 2020

Published online: 16 November 2020

References

- Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J. Beans (*Phaseolus* spp.) - model food legumes. *Plant Soil*. 2003;252(1):55–128. <https://doi.org/10.1023/A:1024146710611>.
- Joshi PK, Rao PP. Global pulses scenario: status and outlook. *Ann N Y Acad Sci*. 2017;1392(1):6–17.
- Winham D, Webb D, Barr A. Beans and good health. *Nutr Today*. 2008;43(5):201–9.
- Beebe SE. Common bean breeding in the tropics. In: *Plant Breeding Reviews*; 2012. p. 357–426.
- Bouis HE, Welch RM. Biofortification—a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. *Crop Sci*. 2010;50(5):20–32.
- Raatz B. Biofortification of grain legumes. In: Sivasankar S, Bergvinson D, Gaur P, Kumar S, Beebe SE, Tamò M, editors. *Achieving sustainable cultivation of grain legumes*. Cambridge: Burleigh Dodds; 2018. p. 1–24.
- Lauer JG, Bijl CG, Grusak MA, Baenziger PS, Boote K, Lingle S, et al. The scientific grand challenges of the 21st century for the crop science Society of America. *Crop Sci*. 2012;52(3):1003–10.
- Beebe SE, Rao IM, Cajiao C, Grajales M. Selection for drought resistance in common bean also improves yield in phosphorus limited and favorable environments. *Crop Sci*. 2008;48(2):582–92.
- Hummel M, Hallahan BF, Brychkova G, Ramirez-Villegas J, Guwela V, Chataika B, et al. Reduction in nutritional quality and growing area suitability of common bean under climate change induced drought stress in Africa. *Sci Rep*. 2018;8(16187):1–11.
- Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, et al. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet*. 2009;5(7):e1000551.
- Ingvarsson PK, Street NR. Association genetics of complex traits in plants. *New Phytol*. 2011;189(4):909–22.
- Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The collaborative cross, a community resource for the genetic analysis of complex traits. *Nat Genet*. 2004;36(11):1133.
- Cavanagh C, Morell M, Mackay I, Powell W. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol*. 2008;11(2):215–21.
- Valdar W, Flint J, Mott R. Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics*. 2006;172:1783–97.
- Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, et al. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice*. 2013; 6(11):1–15.
- Li Y, Xiao J, Chen L, Huang X, Cheng Z, Han B, et al. Rice functional genomics research: past decade and future. *Mol Plant*. 2018;11(3):359–80.

17. Liu H-J, Yan J. Crop genome-wide association study : a harvest of biological relevance. *Plant J.* 2019;97(1):8–18.
18. Naderpour M, Lund OS, Larsen R, Johansen E. Potyviral resistance derived from cultivars of *Phaseolus vulgaris* carrying *bc-3* is associated with the homozygotic presence of a mutated eIF4E allele. *Mol Plant Pathol.* 2010; 11(2):255–63.
19. Panzeri D, Cassani E, Doria E, Tagliabue G, Forti L, Campion B, et al. A defective ABC transporter of the MRP family, responsible for the bean *lpa1* mutation, affects the regulation of the phytic acid pathway, reduces seed myo-inositol and alters ABA sensitivity. *New Phytol.* 2011; 191(1):70–83.
20. Cichy KA, Wiesinger JA, Mendoza FA. Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theor Appl Genet.* 2015;128(8):1555–67.
21. Kamfwa K, Cichy KA, Kelly JD. Genome-wide association study of agronomic traits in common bean. *Plant Genome.* 2015;8(2):1–12.
22. Moghaddam SM, Mamidi S, Osorno JM, Lee R, Brick M, Kelly J, 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(3):1–21.
23. Katuramu DN, Hart JP, Porch TG, Grusak MA, Glahn RP, Cichy KA. Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (*Phaseolus vulgaris* L.). *Mol Breed.* 2018; 38(44):1–18.
24. Hnatuszko-Konka K, Kowalczyk T, Gerszberg A, Wiktorrek-Smagur A, Kononowicz A. *Phaseolus vulgaris* — recalcitrant potential. *Biotechnol Adv.* 2014;32(7):1205–15.
25. Lobaton JD, Miller T, Gil J, Ariza D, de la Hoz JF, Soler A, et al. Resequencing of common bean identifies regions of inter-Gene Pool introgression and provides comprehensive resources for molecular breeding. *Plant Genome.* 2018;11(2):1–21.
26. Ciannamea S, Jensen CS, Agerskov H, Petersen K, Lenk I, Didon T, et al. A new member of the LIR gene family from perennial ryegrass is cold-responsive , and promotes vegetative growth in *Arabidopsis*. *Plant Sci.* 2007; 172(2):221–7.
27. Tiwari SB, Shen Y, Chang H, Hou Y, Harris A, Ma SF, et al. The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *New Phytol.* 2010;187(1):57–66.
28. Yang Y, Xu Y, Wei Q, Imtiaz M, Lan H, Gao S, et al. A zinc finger protein regulates flowering time and abiotic stress tolerance in *Chrysanthemum* by modulating gibberellin biosynthesis. *Plant Cell.* 2014;26:2038–54.
29. Tyler L, Thomas SG, Hu J, Dill A, Alonso JM, Ecker JR, et al. DELLA proteins and gibberellin-regulated seed germination and floral development in *Arabidopsis*. *Plant Physiol.* 2004;135(2):1008–19.
30. Jégu T, Latrasse D, Delarue M, Hirt H, Domenichini S, Ariel F, et al. The BAF60 subunit of the SWI/SNF chromatin-remodeling complex directly controls the formation of a gene loop at FLOWERING LOCUS C in *Arabidopsis*. *Plant Cell.* 2014;26:538–51.
31. Chowdhury ZS, Sato K, Yamamoto D. The core-promoter factor TRF2 mediates a fruitless action to masculinize neurobehavioral traits in *Drosophila*. *Nat Commun.* 2017;8(1):1–10.
32. Chiba Y, Shimizu T, Miyakawa S, Kanno Y, Koshiba T, Kamiya Y, et al. Identification of *Arabidopsis thaliana* NRT1/PTR FAMILY (NPF) proteins capable of transporting plant hormones. *J Plant Res.* 2015;128(4):679–86.
33. Ikeda M, Mitsuda N, Ohme-Takagi M. *Arabidopsis* WUSCHEL is a Bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. *Plant Cell.* 2009;21(11):3493–505.
34. Gramzow L, Theissen G. A hitchhiker's guide to the MADS world of plants. *Genome Biol.* 2010;11(6):214.
35. Shen J, Xu X, Li T, Cao D, Han Z. An MYB transcription factor from *Malus xiaojinensis* has a potential role in Iron nutrition. *J Integr Plant Biol.* 2008; 50(10):1300–6.
36. Urzica E, Casero D, Yamasaki H, Hsieh SI, Adler LN, Karpowicz SJ, et al. Systems and trans-system level analysis identifies conserved Iron deficiency responses in the plant lineage. *Plant Cell.* 2012;24(October):3921–48.
37. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 2010;15(10): 573–81.
38. Chen Y-H, Wu X-M, Ling H-Q, Yang W-C. Transgenic expression of DwmMYB2 impairs iron transport from root to shoot in *Arabidopsis thaliana*. *Cell Res.* 2006;16(10):830–40.
39. Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, et al. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol.* 2015;16(1):167.
40. Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, et al. An Eight-Parent Multiparent Advanced Generation Inter-Cross Population for Winter-Sown Wheat : Creation, Properties, and Validation. *G3 Genes Genomes Genet.* 2014;4(9):1603–10.
41. Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, Bouchet J-P, et al. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol J.* 2015;13(4):565–77.
42. Sallam A, Martsch R. Association mapping for frost tolerance using multiparent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica.* 2015;143(4):501–14.
43. Valdisser P, Pereira W, Almeida J, Muller B, Coelho GP, de Menezes I, et al. In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics.* 2017;18(423):1–19.
44. Blair MW, Cortés AJ, Farmer AD, Huang W, Ambachew D, Varma Penmetasa R, et al. Uneven recombination rate and linkage disequilibrium across a reference SNP map for common bean (*Phaseolus vulgaris* L.). *PLoS One.* 2018;13(3):1–21.
45. Islam S, Thyssen G, Jenkins J, Zeng L, Delhom C, McCarty J, et al. A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton. *BMC Genomics.* 2016;17(903):1–17.
46. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet.* 2018;103(6):338–48.
47. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet.* 2014;46(7):707–13. <https://doi.org/10.1038/ng.3008>.
48. Polania JA, Poschenrieder C, Beebe SE, Rao IM. Effective use of water and increased dry matter partitioned to grain contribute to yield of common bean improved for drought resistance. *Front Plant Sci.* 2016;7:1–10.
49. Blair MW, Galeano CH, Tovar E, Torres MCM, Castrillón AV, Beebe SE, et al. Development of a Mesoamerican intra-genepool genetic map for quantitative trait loci detection in a drought tolerant × susceptible common bean (*Phaseolus vulgaris* L.) cross. *Mol Breed.* 2012;29(1):71–88.
50. Assefa T, Beebe SE, Rao IM, Cuasquer JB, Duque MC, Rivera M, et al. Pod harvest index as a selection criterion to improve drought resistance in white pea bean. *F Crop Res.* 2013;148:24–33.
51. Mukeshimana G, Butare L, Cregan PB, Blair MW, Kelly JD. Quantitative trait loci associated with drought tolerance in common bean. *Crop Sci.* 2014; 54(3):923–38.
52. Polania JA, Rao IM, Cajiao C, Rivera M, Raatz B, Beebe SE. Physiological traits associated with drought resistance in Andean and Mesoamerican genotypes of common bean (*Phaseolus vulgaris* L.). *Euphytica.* 2016;210(1): 17–29.
53. Diaz LM, Ricaurte J, Tovar E, Cajiao C, Teran H, Grajales M, et al. QTL analyses for tolerance to abiotic stresses in a common bean (*Phaseolus vulgaris* L.) population. *PLoS One.* 2018;13(8):1–26.
54. Gonzalez AM, Yuste-Lisbona F, Fernandez-Lozano A, Lozano R, Santalla M. Genetic mapping and QTL analysis in common bean. In: *The Common Bean Genome*; 2017. p. 69–108.
55. Blair MW, Iriarte G, Beebe SE. QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris* L.) cross. *Theor Appl Genet.* 2006;112(6):1149–63.
56. Hoyos-Villegas V, Song Q, Wright EM, Beebe SE, Kelly JD. Joint linkage QTL mapping for yield and agronomic traits in a composite map of three common bean RIL populations. *Crop Sci.* 2016;56(5):2546–63.
57. Chavarro MC, Blair MW. QTL analysis and effect of the fin locus on tropical adaptation in an inter-Gene Pool common bean population. *Trop Plant Biol.* 2010;3(4):204–18.
58. Wright EM, Kelly JD. Mapping QTL For seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). *Euphytica.* 2011; 179(3):471–84.
59. Trapp JJ, Urrea CA, Cregan PB, Miklas PN. Quantitative trait loci for yield under multiple stress and drought conditions in a dry bean population. *Crop Sci.* 2015;55(4):1596–607.
60. Oladad A, Porch T, Rosas J, Moghaddam S, Beaver J, Beebe S, et al. Single and multi-trait GWAS identify genetic factors associated with production

- traits in common bean under abiotic stress environments. *Genes Genomes Genetics*. 2019;9:1881–92.
61. Asfaw A, Blair MW, Struik PC. Multi-environment quantitative trait loci analysis for Photosynthate acquisition, accumulation, and remobilization traits in common bean under drought stress. *G3 genes, Genomes, Genet*. 2012;2(5):579–95.
 62. Izquierdo P, Astudillo C, Blair MW, Iqbal AM, Raatz B, Cichy KA. Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet*. 2018;131(8):1645–58.
 63. Ponce KS, Ye G, Zhao X. QTL Identification for Cooking and Eating Quality in indica Rice Using Multi-Parent Advanced Generation Intercross (MAGIC) Population. *Front Plant Sci*. 2018;9(868):1–9.
 64. Cichy KA, Caldas GV, Snapp SS, Blair MW. QTL analysis of seed iron, zinc, and phosphorus levels in an Andean bean population. *Crop Sci*. 2009;49(5):1742–50.
 65. Blair MW, Astudillo C, Rengifo J, Beebe SE, Graham R. QTL analyses for seed iron and zinc concentrations in an intra-gene pool population of Andean common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet*. 2011;122(3):511–21.
 66. Blair MW, Medina JI, Astudillo C, Rengifo J, Beebe SE, Machado G, et al. QTL for seed iron and zinc concentration and content in a Mesoamerican common bean (*Phaseolus vulgaris* L.) population. *Theor Appl Genet*. 2010; 121(6):1059–70.
 67. Blair MW, Astudillo C, Grusak MA, Graham R, Beebe SE. Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol Breed*. 2009;23(2):197–207.
 68. Rao I. Role of physiology in improving crop adaptation to abiotic stress in the tropics: the case of common bean and tropical forages. In: *Handbook Plant and Crop Physiology*; 2001. p. 583–613.
 69. Singh J, Gezan S, Vallejos E. Developmental Pleiotropy shaped the roots of the domesticated common bean (*Phaseolus vulgaris*). *Plant Physiol*. 2019; 180:1467–79.
 70. Nakedde T, Ibarra-Perez F, Mukankusi C, Waines J, Kelly J. Mapping of QTL associated with *Fusarium* root rot resistance and root architecture traits in black beans. *Euphytica*. 2016;212:51–63.
 71. Asfaw A, Blair M. Quantitative trait loci for rooting pattern traits of common bean grown under drought stress versus non-stress conditions. *Mol Breeding*. 2012;30:681–95.
 72. Singh SP, Miklas PN. Breeding common bean for resistance to common blight: a review. *Crop Sci*. 2015;55(3):971–84.
 73. Nay MM, Souza TLPO, Raatz B, Mukankusi CM, Pastor-Corrales MA, Abreu AFB, et al. A review of angular leaf spot resistance in common bean. *Crop Sci*. 2019;59(4):1376–91.
 74. Meng L, Li H, Zhang L, Wang J. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J*. 2015;3:269–83. <https://doi.org/10.1016/j.cj.2015.01.001>.
 75. Hart JP, Griffiths PD. A series of eIF4E alleles at the Bc-3 locus are associated with recessive resistance to clover yellow vein virus in common bean. *Theor Appl Genet*. 2013;126(11):2849–63.
 76. Zuiderveen GH, Padder BA, Kamfwa K, Song Q, Kelly JD. Genome-wide association study of anthracnose resistance in Andean beans (*Phaseolus vulgaris*). *PLoS One*. 2016;11(6):1–17.
 77. Soltani A, Mafi Moghaddam S, Oladzad-Abbasabadi A, Walter K, Kearns PJ, Vasquez-Guzman J, et al. Genetic Analysis of Flooding Tolerance in an Andean Diversity Panel of Dry Bean (*Phaseolus vulgaris* L.). *Front Plant Sci*. 2018;9(767):1–15.
 78. Duan P, Xu J, Zeng D, Zhang B, Geng M, Zhang G, et al. Natural variation in the promoter *GSE5* contributes to grain size diversity in Rice. *Mol Plant*. 2017;10:685–94.
 79. Zhang Z, Zhang X, Lin Z, Wang J, Liu H, Zhou L, et al. A large transposon insertion in the *stiff1* promoter increases stalk strength in maize. *Plant Cell*. 2020;32:152–65.
 80. Weller J, Vander Schoor J, Perez-Wright E, Hecht V, Gonzales A, Capel C, et al. Parallel origins of photoperiod adaptation following dual domestications of common bean. *J Exp Bot*. 2019;70(4):1209–19.
 81. Ogawa D, Yamamoto E, Ohtani T, Kanno N, Tsunematsu H, Nonoue Y, et al. Haplotype-based allele mining in the Japan-MAGIC rice population. *Sci Rep*. 2018;8(4379):1–11.
 82. Liu Z, Park B-J, Kanno A, Kameya T. The novel use of a combination of sonication and vacuum infiltration in agrobacterium-mediated transformation of kidney bean (*Phaseolus vulgaris* L.) with *lea* gene. *Mol Breed*. 2005;16(3):189–97.
 83. Bonfim K, Faria JC, Nogueira EOPL, Mendes ÉA, Aragão FJL. RNAi-mediated resistance to bean golden mosaic virus in genetically engineered common bean (*Phaseolus vulgaris*). *Mol Plant-Microbe Interact*. 2007;20(6):717–26.
 84. Pflieger S, Blanchet S, Meziadi C, Richard MMS, Thureau V, Mary F, et al. The “one-step” Bean pod mottle virus (BPMV) - derived vector is a functional genomics tool for efficient overexpression of heterologous protein, virus-induced gene silencing and genetic mapping of BPMV R-gene in common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol*. 2014;14(1):232.
 85. Collado R, Bermúdez-Carballoso I, García LR, Veitia N, Torres D, Romero C, et al. Agrobacterium-mediated transformation of *Phaseolus vulgaris* L. using indirect organogenesis. *Sci Hortic*. 2015;195:89–100.
 86. McClean PE, Bett KE, Stonehouse R, Lee R, Pflieger S, Moghaddam SM, et al. White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytol*. 2018;219:1112–23.
 87. Marroni F, Pinosio S, Di Centa E, Jurman I, Boerjan W, Felice N, et al. Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. *Plant J*. 2011;67(4):736–45.
 88. Song G, Han X, Wiersma A, Zong X, Awale H, Kelly J. Induction of competent cells for *Agrobacterium tumefaciens*-mediated stable transformation of common bean (*Phaseolus vulgaris* L.). *PLoS ONE*. 2020; 15(3):e0229909.
 89. Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, Vlasova A, Capella-Gutiérrez S, Ordaz-Ortiz JJ, et al. Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol*. 2017;18(1):60.
 90. Wu J, Wang L, Fu J, Chen J, Wei S, Zhang S, et al. Resequencing of 683 common bean genotypes identifies yield component trait associations across a north-south cline. *Nat Genet*. 2020;25:118–25 81.
 91. Stangoulis J, Sison C. Crop sampling protocols for micronutrient analysis. *Harvest Plus Tech Monogr Ser*. 2008;7:1–20.
 92. Guild GE, Paltridge NG, Andersson MS, Stangoulis JCR. An energy-dispersive X-ray fluorescence method for analysing Fe and Zn in common bean, maize and cowpea biofortification programs. *Plant Soil*. 2017;419:457–66.
 93. Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat Stat*. 2018;23:52–71.
 94. Perea C, De La Hoz JF, Cruz DF, Lobaton JD, Izquierdo P, Quintero JC, et al. Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC Genomics*. 2016;17(S5):539–69.
 95. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PlosOne*. 2011;6(5):1–10.
 96. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–93.
 97. Tello D, Gil J, Loaiza CD, Riascos JJ, Cardozo N, Duitama J. NGSEP3: accurate variant calling across species and sequencing protocols. *Bioinformatics*. 2019;35(22):4716–23.
 98. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequencing data. *Bioinformatics*. 2014;30(15):2114–20.
 99. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9(4):357–9.
 100. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, et al. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. *Plant Genome*. 2016;9(2):1–9.
 101. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2005;23(2):254–67.
 102. Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Hereditas*. 2012;108(3):285–91.
 103. Zhang L, Meng L, Wang J. Linkage analysis and integrated software GAPL for pure-line populations derived from four-way and eight-way crosses. *Crop J*. 2019;7(3):283–93.
 104. Conomos MP, Thornton T. GENetic ESTimation and inference in structured samples (GENESIS): statistical methods for analyzing genetic data from samples with population structure and/or relatedness. *R Package version*. 2016;2(0.1).
 105. Stich B, Möhring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE. Comparison of mixed-model approaches for association mapping. *Genetics*. 2008;178(3):1745–54.
 106. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide

polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80–92.

107. O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, et al. An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics*. 2014;866(15):1–16.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

