

Genome-wide Association Study for Grain Mineral Content in a Brazilian Common Bean Diversity Panel

Jessica Delfini

Universidade Estadual de Londrina

Vânia Moda-Cirino

Instituto Agronômico do Paraná: Instituto Paranaense de Assistência Técnica e Extensão Rural

José dos Santos Neto

Instituto Agronômico do Paraná: Instituto Paranaense de Assistência Técnica e Extensão Rural

Douglas Mariani Zeffa

Universidade Estadual de Maringá

Alison Fernando Nogueira

Universidade Estadual de Londrina

Luriam Aparecida Brandão Ribeiro

Universidade Estadual de Londrina

Paulo Mauricio Ruas

Universidade Estadual de Londrina

Paul Gepts

UC Davis: University of California Davis

Leandro Simões Azeredo Gonçalves (✉ leandrosag@uel.br)

Universidade Estadual de Londrina <https://orcid.org/0000-0001-9700-9375>

Research Article

Keywords: Phaseolus vulgaris L., favorable alleles, biofortification, genotyping-by-sequencing (GBS)

Posted Date: March 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-288785/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Biofortification is one of the strategies developed to address malnutrition in developing countries, the aim of which is to improve the nutritional content of crops. The common bean (*Phaseolus vulgaris* L.), a staple food in several African and Latin American countries, has excellent nutritional attributes and is considered a strong candidate for biofortification. The objective of this study was to identify genomic regions associated with nutritional content in common bean grains using 178 Mesoamerican accessions belonging to a Brazilian Diversity Panel (BDP) and 25,011 good-quality single-nucleotide polymorphisms. The BDP was phenotyped in three environments for nine nutrients (phosphorus, potassium, calcium, magnesium, copper, manganese, sulfur, zinc, and iron) using four genome-wide association multi-locus methods. To obtain more accurate results, only Quantitative-Trait Nucleotides (QTNs) that showed repeatability (i.e., those detected at least twice using different methods or environments) were considered. Forty-eight QTNs detected for the nine minerals showed repeatability and were considered reliable. Pleiotropic QTNs and overlapping genomic regions surrounding the QTNs were identified, demonstrating the possible association between the deposition mechanisms of different nutrients in grains. The accumulation of favorable alleles in the same accession was associated with a gradually increasing nutrient content in the grain. The BDP proved to be a valuable source for association studies. The investigation of different methods and environments showed the reliability of markers associated with minerals. The loci identified in this study will potentially contribute to the improvement of Mesoamerican common beans, particularly carioca and black beans, the main groups consumed in Brazil.

Key Message

QTNs significantly associated to nine mineral content in grains of common bean were identified. The accumulation of favorable alleles was associated with a gradually increasing nutrient content in the grain.

Introduction

Of the leguminous plants consumed by humans, the common bean (*Phaseolus vulgaris* L.) has the highest worldwide demand and is widely cultivated, accounting for approximately 41.712 million hectares annually (RAWAL and NAVARRO 2019). It is considered a staple food for the populations of numerous countries, particularly in Latin America and Eastern and Southern Africa (Broughton et al. 2003). Per capita consumption varies depending on consumer preferences, countries, and regions, reaching up to 66 kg capita⁻¹ year⁻¹ in some African and American countries and ranging from 5 to greater than 10 kg capita⁻¹ year⁻¹ in the United States and Brazil, respectively (Blair 2013).

Common beans have high carbohydrate content and are known to be an excellent source of protein and micronutrients such as iron (Fe), zinc (Zn), thiamine, and folic acid (Petry et al. 2015). Although a cup of common beans can supply approximately 25% and 15% of the daily reference values of Fe and Zn, respectively, genetic improvement can increase the contents of these elements in common bean grains by two- or even three-fold (Cichy et al. 2009).

The process of improving nutrient contents in crops is referred to as biofortification, which is known as a sustainable and economic strategy to address malnutrition in developing countries, considering its focus on improving the staple foods consumed on a daily basis (Dwivedi et al. 2012). Biofortification can be performed using two strategies: genetic biofortification, which is based on traditional methods of plant breeding or genetic engineering, and agronomic biofortification, which is based on the optimized use of fertilizers (Cakmak 2008). Genetic biofortification is considered the most effective option in terms of costs and efficiency, given that as inbred lines with higher nutrient contents are developed, they can be used on a large scale without requiring further investment (Cu et al. 2020).

Several studies have shown a wide genetic variability in the nutritional characteristics of common beans, thereby indicating a considerable potential for increasing the content of Fe, Zn, and other minerals in this crop (Beebe et al. 2000; Islam et al. 2002; Pinheiro et al. 2010; Silva et al. 2012; Ribeiro et al. 2014; McClean et al. 2017; Delfini et al. 2020). Moreover, genetic mapping studies on biparental populations have identified certain quantitative trait locus (QTL) regions associated with different

nutritional characteristics, notably Fe and Zn contents (Blair et al. 2009, 2010; Cichy et al. 2009; Mukamuhirwa et al. 2015). However, few genome-wide association studies (GWAS) have analyzed this subject.

GWAS are an extension of QTL association mapping studies between characteristics of interest and molecular markers in large population samples. Moreover, statistical genotype and phenotype associations are identified in unrelated individuals and are only detected when the marker and QTL are in strong linkage disequilibrium (LD). This method requires an existing linkage map or a reference genome of the investigated species to facilitate marker positioning (Stapley et al. 2010). In addition, a large number of markers are required for good genome coverage, which has been made possible by the recent emergence of next-generation sequencing (NGS) (Edwards and Batley 2010).

Several techniques have been developed to generate large sets of single nucleotide polymorphisms (SNPs), such as high-density SNP arrays available for several crops, including common beans. However, these arrays are often designed based on a limited number of elite genotypes and can produce biased data when used to characterize non-elite genotypes (Rasheed et al. 2017; Raggi et al. 2019). The NGS technique is a promising alternative that produces a large number of SNPs, is not biased, and can be performed at a low cost. Genotyping-by-sequencing (GBS) is a technically direct and multiplexed approach that can be used to sequence subsets of a genome based on restriction enzymes for rapid, specific, and reproducible results, and the large number of SNPs generated provides a deeper understanding of population structure and genetic diversity, in addition to being suitable for GWAS and even for the discovery of candidate genes (Stansell et al. 2018).

Another important aspect of GWAS concerns the mathematical models used to associate the marker with the target trait. Advances in quantitative molecular genetics enabled the development of a large number of association mapping methods for the genetic dissection of complex plant characteristics. However, most of the previous studies on common beans used single-locus GWAS, such as general linear (GLM) and mixed linear (MLM) models. These models assess the significance of the marker-trait association considering one marker at a time, resulting in significant associations based on rigorous multiple test correction [e.g., Bonferroni and false discovery rate (FDR)]. Due to the high significance required, these methods detect QTNs that have a large effect only; they are unable to identify polygenes with small effects for complex characteristics (Lan et al. 2020).

To solve this problem, alternative multi-locus methods, including mrMLM (Wang et al. 2016), FASTmrMLM (Tamba and Zhang 2018), pLARmEB (Zhang et al. 2017), and ISIS-EM-BLASSO (Tamba et al. 2017), have been proposed, with the advantage that Bonferroni correction for multiple tests is not necessary. These methods differ from other multi-locus methods in that they comprise two steps. The first step considers the SNP effect as random, with all potentially associated markers being selected using a random-SNP-effect MLM with a modified Bonferroni correction for the significance test. In the second step, all markers are placed in a single model, and all effects other than zero are detected based on a likelihood ratio test to identify QTNs (Chang et al. 2018).

Given that the tools used in GWAS can identify, characterize, and develop molecular markers related to the allelic variation of nutritional characteristics of common beans and considering the socioeconomic importance of the crop, studies investigating the genetic architecture of these traits are fundamental (Katuuramu et al. 2018). Thus, the objective of the study was to identify regions related to the nutritional content of common bean grains using multi-locus methods in a Brazilian Diversity Panel (BDP) comprising Mesoamerican accessions adapted to tropical conditions. These findings will provide useful genetic information to improve the biofortification of common bean crops.

Material And Methods

Genetic material and experimental design

The Mesoamerican common bean diversity panel used in this study was established in order to group accessions adapted to tropical conditions and that represented the variability of the Mesoamerican domesticated gene pool of the species present in Brazil (Delfini et al. 2021). This panel includes cultivars, landraces, and inbred lines maintained in the germplasm bank of the

Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná). For the purposes of the present study, a BDP sub-set comprising 178 Mesoamerican common bean accessions was evaluated (Table S1).

The experiments were conducted during the 2018 rainy season (sowing between September and October) at different IDR-Paraná Research Stations in the cities of Londrina (LDA) (23°22'8"S; 51°9'48"O), Ponta Grossa (PG) (25°09'11"S; 50°09'22"O), and Guarapuava (GUA) (25°23'51"S; 51°32'36"O), all located in the state of Paraná, Brazil. The experiments were based on an incomplete block design with replicates in sets. Five sets with two replicates were used, each with 50 treatments (46 accessions and four controls). For each plot, common beans were planted in four 2.00-m-long rows, with a between-row spacing of 0.50 m and a density of 12 plants per linear meter. Fertilization and pest, disease, and invasive plant control were performed according to the technical recommendations for the crop.

Phenotyping for nutrient content in grains and statistical analysis

After physiological maturity, 100 g samples of disease-free seeds with no visible physical or insect-related damage from each experimental plot were collected. Before laboratory analyses, the samples were washed sequentially with running water, 0.01 M HCl solution, and distilled water to prevent contamination by soil particles. Subsequently, the seeds were oven-dried at 60°C for 24 h and then ground to a fine powder. The flour obtained was packed in plastic bags with hermetic insulation. Phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), copper (Cu), manganese (Mn), sulfur (S), Zn, and Fe contents in the flour were determined using the method described by Miyazawa (1999). Initially, 0.4 g samples of flour were subjected to nitroperchloric digestion with a 3:1 solution of HNO₃:HClO₄ in 80 mL digester tubes, and mineral concentrations were subsequently determined using an Inductively Coupled Plasma Optical Emission Spectrometer (ICP-OES, Optima 83000, Perkin Elmer, Waltham).

An analysis of variance (ANOVA) of the phenotypic data was performed using the PROC GLM function of the SAS software (SAS Institute 2000) according to the following statistical model: $Y_{ijkl} = \mu + A_i + S_j + AS_{ij} + R/AS_{kij} + G/Sl_j + AG/Sml_j + e_{ijklm}$, where μ is the mean; A_i is the fixed effect of the i -th environment; S_j is the effect of the j -th "set"; AS_{ij} is the effect of the interaction between environments and "sets"; R/AS_{kij} is the effect of the k -th repetition within the interaction between the i -th environment and the j -th set; G/Sl_j is the random effect of the l -th genotype within the j -th "set"; AG/Sml_j is the effect of the interaction of environments and accessions within the j -th "set"; and e_{ijklm} is the experimental error (Hallauer and Miranda Filho 1988). The means adjusted for each accession in each of the environments and across the environments were obtained using the LSmeans option in the GLM procedure. Heritability (h^2) was estimated using the equation $h^2 = \sigma^2_G / \sigma^2_P$, in which the genotypic (σ^2_G) and phenotypic (σ^2_P) variances were estimated using the following equations: $\sigma^2_G = (QM_G - QM_E) / ra$ and $\sigma^2_P = QM_G / ra$, where QM_G is the mean square of genotype within "sets," QME is the mean square of the error, r is the number of replicates, and a is the number of environments. Descriptive analysis was performed using the PROC UNIVARIATE function of the SAS software (SAS Institute 2000). A graphic representation of Pearson correlations was obtained using the *corrplot* package implemented in the R software (Wei and Simko 2017; R Core Team 2020).

Genotyping-by-sequencing

Genotyping was performed based on the GBS technique using the enzyme *Cvi*All (Ariani et al. 2016). Data processing was performed as described in Delfini et al. (2021), removing individuals containing less than 30% genotyped positions and SNPs with a minor allele frequency lower than 0.05. The data were imputed using the Beagle software v.5 (Browning et al. 2018), and only SNPs anchored to chromosomes in the common bean reference genome (Schmutz et al. 2014) were used.

Population structure and linkage disequilibrium

SNPs were filtered for LD using the *indep-pairwise* function of the PLINK software (Purcell et al. 2007), and only the SNPs with an LD value ≤ 0.2 were maintained. Analyses of population structure, principal components, and clustering were performed using the filtered data. The population structure was inferred using the Bayesian clustering algorithm in Structure v2.3.4 software (Pritchard et al. 2000). The model comprised an admixture with 100,000 burn-ins, 100,000 Monte-Carlo Markov chains, and 10 replicates for hypothetical numbers of subpopulations (K) between 1 and 10. The statistical parameter ΔK (Evanno et al. 2005) was used to determine the number of groups. Only accessions with a membership coefficient ≥ 0.6 were assigned to a genetic group, whereas those with coefficients < 0.6 were placed in a group designated as a mixture. Principal component

analysis (PCA) was performed using the *snpGdsPCA* function of the *SNPRelate* package (Zheng et al. 2012) of the R software. A neighbor-joining tree was constructed using the TASSEL 5.0 software (Bradbury et al. 2007).

The LD between SNPs was estimated using the *LdcorSV* package (Desrousseaux et al. 2017) of the R software, which can be used to correct LD biases caused by the population structure and relationship. In addition to the conventional r^2 , r^2 corrected for population structure (r^2_s), r^2 including relationship (r^2_v) and r^2 including population structure and relationship (r^2_{vs}) were calculated. The results obtained using the STRUCTURE software for $K = 2$ were used as population structure, and a kinship matrix was calculated for relationship using the *rrBLUP* package (Endelman 2011) of the R software. LD decay was calculated using a non-linear method (Hill and Weir 1988) and adjusted using the *n/s* function in the R software.

Genome-wide association studies

A total of four multi-locus GWAS implemented in the mrMLM.GUI 4.0 software (Ya-Wen et al. 2019) were used to detect significant QTNs for the target traits: mrMLM, FASTmrMLM, pLARmEB, and ISIS-EM-BLASSO. The population structure (Q) and the kinship matrix (K) were included in the model to reduce the occurrence of false positives and improve analytical power. The STRUCTURE software results ($K = 2$) were used for Q, and a kinship matrix was calculated using the mrMLM.GUI 4.0 software. For all methods, the parameters used were the standards, and a logarithm of the odd (LOD) score ≥ 3 was used as the critical value for significant associations. The phenotypic data used for GWAS were the adjusted means of each of the three environments and the overall adjusted mean (LDA, PG, GUA, and LSmeans). To obtain more accurate results, only QTNs showing repeatability (i.e., detected at least twice using different methods or in different environments) were considered significant and were used in searches for favorable alleles and candidate genes.

Identification of favorable alleles

For each QTN, all BDP accessions were initially divided into two groups based on the QTN genotype, and alleles associated with a positive effect on the phenotype (i.e., those associated with an increased mineral content) were identified. Subsequently, the number of favorable alleles for each accession was identified, and the association between the accumulation of these superior alleles in the same accession and nutritional contents was verified. Boxplots and a heatmap generated using the *heatmap* package (Kolde 2019) of the R software were produced for better result visualization.

Search for candidate genes

The search for potential candidate genes focused on the QTNs detected using multiple methods or in multiple environments. The search radius (physical distance) was determined according to LD half-decay corrected for the population structure and relationship (r^2_{vs}). Genes present in the association region were identified based on the annotation of the *Phaseolus vulgaris* v.2 common bean reference genome published on the Phytozome v10.3 website (<http://phytozome.net>). Subsequently, genes with known putative functions based on the Gene Ontology annotation (GO, <http://www.geneontology.org/>) related to the traits of interest were selected as candidate genes.

Results

Nutritional characterization

The contents of nine minerals in the grains of common bean accessions from the BDP cultivated in three environments were determined. The results of the ANOVA and related descriptive statistics (mean, maximum, minimum, standard deviation, asymmetry, and kurtosis) are shown in Table 1, and the frequency distributions of the contents of these minerals in each of the three environments are presented in Fig. 1.

Table 1

Analysis of variance and descriptive statistics for the contents of different minerals detected in common bean accessions from the Brazilian Diversity Panel (BDP) evaluated in three environments.

	p^a	K	Ca	Mg	Cu	Mn	S	Zn	Fe	
F_{env}^b	537,91 ^{***}	39,98 ^{***}	196,71 ^{***}	102,8 ^{***}	24,37 ^{***}	965,79 ^{***}	792,73 ^{***}	770,03 ^{***}	74,87 ^{***}	
F_{set}	46,76 ^{***}	82,08 ^{***}	12,28 ^{***}	52,51 ^{***}	55,77 ^{***}	52,94 ^{***}	148,25 ^{***}	13,11 ^{***}	48,75 ^{***}	
F_{env*set}	24,09 ^{***}	6,99 ^{***}	8,65 ^{***}	15,68 ^{***}	15,14 ^{***}	9,97 ^{***}	4,75 ^{***}	4,36 ^{***}	2,3 [*]	
F_{rep(env*set)}	58,17 ^{***}	25,35 ^{***}	11,23 ^{***}	33,36 ^{***}	17,15 ^{***}	10,37 ^{***}	59,52 ^{***}	28,56 ^{***}	10,45 ^{***}	
F_{treat(set)}	1,44 ^{**}	1,28 [*]	1,53 ^{***}	1,9 ^{***}	1,55 ^{***}	1,4 ^{**}	1,74 ^{***}	1,47 ^{***}	1,22 [*]	
F_{env*treat(set)}	0,8 ^{ns}	0,92 ^{ns}	0,7 ^{ns}	0,72 ^{ns}	0,79 ^{ns}	0,86 ^{ns}	0,66 ^{ns}	0,79 ^{ns}	0,89 ^{ns}	
CV(%)^c	9,06	9,32	18,27	8,88	15,52	15,02	10,64	10,08	17,07	
Heritability (h²)	0,31	0,22	0,35	0,47	0,35	0,28	0,42	0,32	0,18	
Mean	LDA	4,79	13,29	1,72	1,95	10,18	14,87	2,35	35,51	74,87
	PG	3,81	12,43	1,66	2,11	10,78	20,89	1,88	30,48	63,74
	GUA	4,35	12,83	1,33	2,14	10,91	13,23	1,74	26,25	70,25
Standard Deviation	LDA	0,399	0,866	0,261	0,128	1,924	2,162	0,273	2,709	12,363
	PG	0,375	1,290	0,234	0,225	1,224	2,748	0,188	2,213	7,899
	GUA	0,326	1,213	0,176	0,152	1,333	1,502	0,206	2,204	9,240
Minimum	LDA	3,76	11,01	1,11	1,64	6,37	9,92	1,68	28,31	50,75
	PG	3,07	8,78	1,14	1,69	7,96	15,12	1,49	25,34	46,31
	GUA	3,72	9,79	0,99	1,81	7,44	10,09	1,24	21,77	47,44
Maximum	LDA	5,64	15,09	2,37	2,26	16,72	22,19	3,02	42,20	117,81
	PG	4,99	16,08	2,30	2,90	14,66	28,54	2,46	38,15	87,86
	GUA	5,70	15,57	1,89	2,84	15,88	18,66	2,36	34,30	96,87
Skewness	LDA	-0,238	-0,351	0,112	-0,151	1,339	0,498	0,035	0,136	0,927
	PG	0,188	0,065	0,306	0,547	0,286	0,489	0,752	0,322	0,327
	GUA	0,949	-0,275	0,554	0,638	0,411	0,559	0,079	0,891	0,508
Kurtosis	LDA	-0,498	-0,374	-0,566	-0,529	2,290	0,242	-0,482	-0,297	0,964
	PG	-0,267	-0,128	-0,123	0,175	-0,202	-0,454	0,132	0,202	0,176
	GUA	1,840	-0,375	-0,014	2,100	1,265	0,856	-0,372	1,402	0,213

^aP, K, Ca, Mg, and S in g.kg⁻¹ and Cu, Mn, Zn, and Fe in mg.kg⁻¹; ^bF_{env}, F_{set}, F_{env*set}, F_{rep(env*set)}, F_{treat(set)} and F_{env*treat(set)} denote the values of F for environmental effects, "set," interaction between environment and set, repetition within environment and set, treatment within set, and interaction between environment and treatment within set, respectively; ^cCV (%) = coefficient of variation. *P < 0.01, **P < 0.001, *** P < 0.0001, and ^{ns}not significant.

ANOVA revealed a significant effect in all characteristics evaluated with regard to genotype and environmental variation; however, no Genotype × Environment (G×E) interactions for the evaluated traits were detected. Coefficients of variance (CV)

ranged from 8.88 to 18.27 for Mg and Fe contents, respectively, showing good experimental quality, whereas heritability values (h^2) ranged from 0.18 to 0.47 for K and Mg, respectively.

The mean P, K, Ca, S, Zn, and Fe contents were higher in common beans cultivated in LDA, whereas Mn and Mg were detected at higher concentrations in PG and GUA, respectively. Although the mean Cu content showed no substantial variation between the three locations, it was the mineral with the greatest variation between the minimum and maximum values among the accessions (up to 2.22-fold). Comparatively, the mean variation found for P, K, Mg, S, and Zn contents ranged between 1.55- and 1.78-fold, whereas Ca, Mn, and Fe contents had an approximately two-fold variation.

Correlation analysis showed significant effects for most minerals. A positive correlation was detected between Fe and Cu (0.67, $P \leq 0.05$); however, they were negatively correlated with P, K, Mg, and S. Similarly, positive correlations were detected among P, K, Ca, Mg, Mn, and S contents, ranging from 0.31 (Ca-P) to 0.79 (P-S) (Fig. S1).

Genotyping, population structure and linkage disequilibrium

After filtering the SNPs obtained by GBS, 25,011 high-quality SNPs were obtained. These SNPs were used for subsequent GWAS, and their distribution is shown in Fig. 2. After removing high-LD SNPs, a total of 707 remained for use in population structure studies. Population structure analysis performed using the ADMIXTURE method indicated a K value of 2, corroborating PCA and clustering analysis findings based on the neighbor-joining method (Fig. S2). Thus, two highly distinct groups were identified: one composed predominantly of carioca common bean commercial accessions and the other composed of black bean accessions. The third group included 12% of the accessions, for which the membership coefficient in either of the aforementioned two groups was less than 0.6, and was classified as a mixture group.

LD corrected for relationship (r^2_v) and population structure and relationship (r^2_{vs}) showed half-decay values well below those obtained for conventional r^2 and those corrected only for population structure (r^2_s). The LD half-decay values for r^2 , r^2_s , r^2_v , and r^2_{vs} were 1,414.41, 1,223.41, 296.76, and 296.54 kb, respectively. Thus, the search distance for candidate genes was set according to the distance determined based on r^2_{vs} (Fig. S3).

Genome-wide association studies

Altogether, 238 unique QTNs were significant for the nine minerals studied, of which 48 were detected at least twice by one of the four multi-locus methods used or in the different environments analyzed (LDA, PG, GUA, and LSmeans). These QTNs were considered promising and were subjected to a more thorough analysis (Fig. 3). The QTNs presenting repeatability are shown in Table 2. Results revealed differences in the numbers of QTNs associated with the different minerals; 7, 4, 4, 11, 2, 6, 5, 7, and 2 QTNs were detected for P, K, Ca, Mg, Cu, Mn, S, Zn, and Fe, respectively. Moreover, four pleiotropic QTNs were detected (i.e., they showed a significant association at least two times for more than one trait) with one QTN shared between P and Mg, one between Ca and Mn, and two between K and Mg. In addition to the pleiotropic QTNs, 12 QTNs overlapped considering the genomic region around the QTN defined according to LD (± 296 kb), with the two of them associated with Cu and P located on chromosomes Pv3 and Pv4, respectively, which overlapped with the two pleiotropic QTNs identified for K-Mg. The other overlapping QTNs were associated with Mg-Zn, Mg-S, Mn-Fe, P-Cu, and Zn-Zn located on chromosomes Pv1, Pv8, Pv9, Pv9, and Pv11, respectively.

Table 2

Quantitative trait nucleotides (QTNs) associated with the mineral contents of common beans from the Brazilian Diversity Panel (BDP), which were detected at least twice using different methods and/or in different environments.

Trait	SNP	Chr	Position (bp)	QTN effect ^a	LOD score ^b	PVE (%) ^c	MAF ^d	Gen	Env ^e	Methods ^f
P	S01_27883754	1	27883754	-0,17 ~ -0,14	3,33 ~ 5,34	2,4 ~ 2,56	0,0506	C	2	2, 4
	S04_275251	4	275251	-0,2 ~ -0,14	3,48 ~ 6,53	3,56 ~ 6,96	0,0674	G	2	1, 2, 4
	S04_45228381	4	45228381	0,08 ~ 0,08	3,75 ~ 3,9	2,06 ~ 3,75	0,2697	T	2	2, 4
	S06_14869772	6	14869772	0,12 ~ 0,14	4,06 ~ 5,49	5,1 ~ 7,08	0,1742	C	1	2, 4
	S06_5639061	6	5639061	-0,14 ~ -0,09	3,78 ~ 4,77	3,24 ~ 8,69	0,0795	G	4	2, 3, 4
	S08_39404548	8	39404548	0,07 ~ 0,08	3,91 ~ 4,2	2,5 ~ 3,2	0,3146	A	1	2, 4
	S09_36496109	9	36496109	-0,14 ~ -0,09	3,25 ~ 3,6	2,69 ~ 6,84	0,0625	T	4	3, 4
K	S03_552367	3	552367	-0,31 ~ -0,17	3,06 ~ 4,5	3,06 ~ 9,45	0,3011	G	2,4	1, 2, 3, 4
	S04_44875902	4	44875902	0,19 ~ 0,23	3,19 ~ 3,88	2,98 ~ 4,17	0,1798	A	1	1, 2, 4
	S05_9081162	5	9081162	-0,31 ~ -0,18	3,06 ~ 3,35	2,84 ~ 7,18	0,2045	C	4	1, 2, 3, 4
	S07_14454278	7	14454278	0,2 ~ 0,38	3,09 ~ 4,83	4,61 ~ 9,13	0,3466	A	2,4	1, 2, 3, 4
	Ca	S05_13646513	5	13646513	-0,05 ~ -0,05	3,38 ~ 4,71	2,93 ~ 3,73	0,3371	A	1
	S05_18968293	5	18968293	-0,05 ~ -0,05	3,64 ~ 4,52	5,07 ~ 5,87	0,1921	T	4	2, 4
	S07_35307332	7	35307332	-0,1 ~ -0,05	3,16 ~ 4,38	4,4 ~ 7,59	0,1186	G	2,4	1, 2
	S09_17096186	9	17096186	0,04 ~ 0,05	3,01 ~ 4,64	5,92 ~ 7,2	0,2825	C	4	1, 4
Mg	S01_46197108	1	46197108	-0,13 ~ -0,06	3,84 ~ 4,88	4,6 ~ 7,87	0,0682	C	2,4	1, 2, 3
	S02_2417813	2	2417813	0,06 ~ 0,1	3,31 ~ 4,7	3,4 ~ 7,98	0,0795	G	2,4	1, 2, 3, 4
	S03_552367	3	552367	-0,03 ~ 0,003	3,05 ~ 4,81	0,04 ~ 5,9	0,309	G	1,3,4	1, 2, 3, 4
	S04_275251	4	275251	-0,14 ~ -0,11	4,64 ~ 5,69	5,3 ~ 9,7	0,0674	G	2	1, 2, 3, 4
	S04_44875902	4	44875902	0,05 ~ 0,06	4,2 ~ 6,38	9,48 ~ 13,49	0,1798	A	1	1, 2, 3, 4

^aQuantitative trait nucleotide effect; ^bLOD value, the significant threshold for transformed P-value; ^cPVE (%): phenotypic variation explained; ^dMinor allele frequency; ^eEnvironment: 1-LDA, 2-PG, 3-GUA, 4-LSmeans; ^fMethods: 1-FASTmrMLM, 2-ISIS-EM-BLASSO, 3-mrMLM, 4-pLARmEB; pleiotropic QTNs are shown in bold.

Trait	SNP	Chr	Position (bp)	QTN effect ^a	LOD score ^b	PVE (%) ^c	MAF ^d	Gen	Env ^e	Methods ^f
	S08_3789382	8	3789382	-0,06 ~ 0,0004	3,37 ~ 4,52	0 ~ 9,83	0,1875	A	2,4	1, 3
	S08_59042075	8	59042075	-0,06 ~ -0,05	3,71 ~ 4,51	1,98 ~ 2,91	0,1348	G	2	1, 2, 4
	S09_27556168	9	27556168	-0,02 ~ -0,01	3,35 ~ 18,49	0,54 ~ 1,84	0,2809	T	3	1, 4
	S10_42297372	10	42297372	0,04 ~ 0,07	3,65 ~ 7,27	3,73 ~ 10,06	0,2443	A	2,4	1, 2, 3, 4
	S11_1201121	11	1201121	0,01 ~ 0,06	6,42 ~ 16,14	0,14 ~ 9,46	0,2247	A	3	2, 4
	S11_28734694	11	28734694	-0,04 ~ -0,02	4,55 ~ 34,13	0,94 ~ 5,4	0,2303	C	3	2, 4
Cu	S03_1013759	3	1013759	-0,46 ~ -0,33	3,24 ~ 5,66	3,06 ~ 5,96	0,1193	G	2	2, 3, 4
	S09_36738755	9	36738755	0,37 ~ 0,4	3,22 ~ 3,87	3,58 ~ 4,04	0,3989	T	1	2, 4
Mn	S02_32813039	2	32813039	-0,63 ~ -0,42	3,07 ~ 3,13	2,17 ~ 4,4	0,3708	C	2	2, 3
	S04_4882578	4	4882578	0,36 ~ 1,09	3,78 ~ 7,04	4,01 ~ 11,31	0,2841	G	2,4	1, 3, 4
	S05_11203900	5	11203900	-1,13 ~ -0,66	3,59 ~ 5,43	5,45 ~ 5,99	0,0966	G	2,4	2, 3, 4
	S09_13592607	9	13592607	-0,7 ~ -0,39	4,93 ~ 5,19	5,69 ~ 6,44	0,427	A	2,4	2, 4
	S09_16145160	9	16145160	0,58 ~ 0,87	3,17 ~ 3,37	4,85 ~ 7,47	0,1875	G	2,4	1, 3
	S09_17096186	9	17096186	0,43 ~ 0,75	5,04 ~ 5,33	5,54 ~ 6,03	0,2809	C	2,4	2, 4
S	S01_1213244	1	1213244	-0,09 ~ -0,09	3,87 ~ 3,99	4,62 ~ 5,81	0,0739	G	3,4	3, 4
	S04_33850915	4	33850915	0,0002 ~ 0,1	3,03 ~ 5,18	0 ~ 9,98	0,1348	C	1,4	1, 2, 3
	S08_59081232	8	59081232	-0,08 ~ -0,05	3,15 ~ 4,55	2,85 ~ 7,33	0,1573	G	4	1, 2, 3
	S09_18443083	9	18443083	-0,1 ~ -0,06	4 ~ 4,33	1,4 ~ 6,55	0,0787	T	1,4	1, 3
	S10_15778701	10	15778701	0,00002 ~ 0,08	3,18 ~ 4,03	0 ~ 5,41	0,0955	G	3	1, 2
Zn	S01_46657405	1	46657405	-0,65 ~ 0,00003	3,55 ~ 3,55	0 ~ 1,62	0,0787	C	3	1, 4
	S05_39601221	5	39601221	0,62 ~ 0,62	3,2 ~ 3,2	3,59 ~ 3,6	0,2191	G	1	2, 4

^aQuantitative trait nucleotide effect; ^bLOD value, the significant threshold for transformed P-value; ^cPVE (%): phenotypic variation explained; ^dMinor allele frequency; ^eEnvironment: 1-LDA, 2-PG, 3-GUA, 4-LSmeans; ^fMethods: 1-FASTmrMLM, 2-ISIS-EM-BLASSO, 3-mrMLM, 4-pLARmEB; pleiotropic QTNs are shown in bold.

Trait	SNP	Chr	Position (bp)	QTN effect ^a	LOD score ^b	PVE (%) ^c	MAF ^d	Gen	Env ^e	Methods ^f
	S09_29892463	9	29892463	-0,98 ~ -0,86	3,91 ~ 5,83	4,22 ~ 8,36	0,125	G	3	1, 2, 3, 4
	S10_818442	10	818442	-0,46 ~ -0,35	4,09 ~ 5,11	2,35 ~ 3,63	0,3708	T	3	1, 2, 4
	S11_18198447	11	18198447	-0,8 ~ -0,43	3,29 ~ 4,61	3,56 ~ 7,92	0,191	C	3,4	1, 2, 3, 4
	S11_50117148	11	50117148	0,51 ~ 0,65	3,74 ~ 4,81	5,17 ~ 7,98	0,3876	G	3	2, 3
	S11_50159956	11	50159956	0,4 ~ 0,48	4,52 ~ 4,52	2,93 ~ 3,21	0,4157	T	3	1, 4
Fe	S08_2828543	8	2828543	2,83 ~ 3,15	3,14 ~ 3,63	2,73 ~ 3,45	0,1573	G	1	2, 4
	S09_13611567	9	13611567	-4,78 ~ -2,95	3,15 ~ 4,55	3,76 ~ 9,43	0,2079	A	1	1, 2, 3

^aQuantitative trait nucleotide effect; ^bLOD value, the significant threshold for transformed P-value; ^cPVE (%): phenotypic variation explained; ^dMinor allele frequency; ^eEnvironment: 1-LDA, 2-PG, 3-GUA, 4-LSmeans; ^fMethods: 1-FASTmrMLM, 2-ISIS-EM-BLASSO, 3-mrMLM, 4-pLARmEB; pleiotropic QTNs are shown in bold.

QTNs were detected on all 11 common bean chromosomes, with the largest number located on Pv09 and Pv07, accounting for 10 and 7 QTNs, respectively. The phenotypic variation explained (PVE) by the different QTNs varied from very low values close to 0 to 13.49%. Although values close to zero were also observed for QTN effects, the effects varied from - 4.78 to 3.15. Different PVE and QTN effects results were obtained for the same QTN using different methods, with more significant results for PVE. For example, some QTNs detected using the FASTmrMLM method showed an effect and PVE close to zero.

For all QTNs detected using the four methods, the pLARmEB method identified the largest number of significant SNPs, followed by the ISIS-EM-BLASSO, mrMLM, and FASTmrMLM methods. However, considering only the QTNs that showed repeatability, the ISIS-EM-BLASSO method stood out, followed by FASTmrMLM, pLARmEB, and mrMLM. The FASTmrMLM method showed 100% efficiency (i.e., all SNPs detected using this method showed repeatability), whereas the mrMLM, ISIS-EM-BLASSO, and pLARmEB methods showed 73%, 57%, and 19% efficiency, respectively. Although the pLARmEB method detected the highest initial number of SNPs, many of these SNPs were discarded at subsequent stages of the study. The GUA environment detected the highest initial number of significant SNPs, followed by PG, LSmeans, and LDA. However, PG presented more stable SNPs, followed by LSmeans, LDA, and GUA. Some environments showed no significant SNPs for certain characteristics (Fig. 3).

Favorable alleles

Favorable alleles were identified for each of the QTNs that showed repeatability (i.e., alleles associated with increased mineral content), followed by an analysis of whether the accumulation of these alleles reflected an increasing mineral content in the genotypes (Fig. 4). Results showed a gradual increase in P, K, Ca, Mg, Mn, S, Zn, and Fe contents in common bean grains according to the number of favorable alleles present in the genotype.

Figure 5 shows a clustering of BDP accessions according to the contents of the nine minerals studied, with lower nutrient content groups presenting fewer favorable alleles (< 20) and higher nutrient content groups presenting more favorable alleles (> 25). Forty-three accessions had 50% or more favorable alleles, of which 25 accessions were common beans from the carioca commercial group and 11 were from the black commercial group; the IAPAR 16 cultivar was characterized by the highest percentage (62%) of favorable alleles. The number of favorable alleles in the 31 inbred lines developed by IDR-Paraná ranged between 13 and 28, with nine showing a favorable allele percentage of at least 50%.

Potential candidate genes

Initially, a search was performed for genes in the genomic regions identified around QTNs. Using GO annotation, the genes were grouped into three functional categories related to cellular components, biological processes, and molecular function. In the cellular component category, the main functions identified were related to the membrane and its integral components, whereas in the biological processes category, the functions were related to oxido-reduction processes, protein phosphorylation, transcription regulation, metabolic processes, transmembrane transport, and metal ion transport. In the molecular function category, protein binding, ATP binding, DNA binding, nucleic acid binding, heme binding, zinc binding, iron binding, protein kinase activity, and metal ion binding were some of the main functions of the identified genes.

With the exception of K and Mn, all minerals had at least one candidate gene with a nutrient-related function identified (Table 3). The largest number of candidate genes were identified for P and Zn (13 and 10, respectively), whereas only a single candidate gene was associated with S. For some of the identified candidate genes, the QTN was located within the gene itself, whereas others were detected at distances of between 6.6 and 287 kbp from the QTN.

Table 3

List of potential candidate genes located in genomic regions underlying the quantitative trait nucleotides (QTNs) associated with nutritional content variation in common beans.

Trait	QTN	Candidate gene	Functional Annotation
P	S04_275251	Phvul.004G004800	Protein phosphorylation
		Phvul.004G006600	Protein phosphorylation
	S04_45228381	Phvul.004G150500	Phosphoethanolamine N-methyltransferase activity
		Phvul.004G149200	Protein phosphorylation
		Phvul.004G150100	Phosphogluconate dehydrogenase (decarboxylating) activity
		Phvul.004G151200	Protein phosphorylation
		Phvul.004G151400	Protein phosphorylation
		Phvul.004G152000	Protein phosphorylation
	S06_14869772	Phvul.006G043700	Protein phosphorylation
		Phvul.006G046700	GDP-D-glucose phosphorylase activity
S06_5639061	Phvul.006G012900	Phospholipid binding	
S09_36496109	Phvul.009G243200	Pyridoxal phosphate binding	
	Phvul.009G244600	Protein phosphorylation	
Ca	S07_35307332	Phvul.007G231400	Calcium ion binding
		Phvul.007G232100	Calcium-dependent lipid binding domain
Mg	S04_275251	Phvul.004G000800	Magnesium ion binding
	S08_59042075	Phvul.008G243200	Magnesium ion binding
	S11_1201121	Phvul.011G013600	Magnesium ion binding
Cu	S03_1013759	Phvul.003G012600	Electron carrier activity - Cupredoxins - blue copper proteins
	S09_36738755	Phvul.009G247600	Copper transport protein ATOX1-Related
S	S01_1213244	Phvul.001G014532	Iron-sulfur cluster binding
Zn	S01_46657405	Phvul.001G208400	Zinc ion binding
		Phvul.001G212400	Zinc ion binding
	S05_39601221	Phvul.005G166300	Zinc ion binding
	S09_29892463	Phvul.009G197800	Zinc ion binding
		Phvul.009G198300	Zinc ion binding
	S10_818442	Phvul.010G005300	Zinc ion binding
	S11_18198447	Phvul.011G115800	zinc ion binding
		Phvul.011G115900	zinc ion binding
	S11_50117148	Phvul.011G187200	Zinc ion binding
	S11_50159956	Phvul.011G190200	Zinc ion binding
Fe	S08_2828543	Phvul.008G034400	Iron ion binding

Trait	QTN	Candidate gene	Functional Annotation
		Phvul.008G034500	Iron ion binding
	S09_13611567	Phvul.009G081633	Iron ion binding
		Phvul.009G081700	Iron ion binding

Discussion

The discovery of genes or genomic regions associated with the nutritional contents of common bean grains may accelerate the development of new biofortified cultivars. In this context, common bean germplasms preserved worldwide represent an invaluable genetic resource for identifying genomic regions associated with nutrient accumulation in grains based on associative mapping techniques. In the present study, GWAS were conducted for nine nutrients based on the screening of a Mesoamerican common bean diversity panel, including accessions adapted to Brazilian climatic conditions and that represent the most widely consumed commercial groups in the country, the black and carioca groups.

The ANOVA conducted to assess the relationships between common bean accessions and environments with regard to the nine target minerals showed that the BDP can be used in GWAS and that different QTNs can be identified in different environments. The heritability values (h^2) obtained tend to indicate a high influence of the environment on these traits. However, heritability and genetic models may vary among different studies, as relevant calculations depend on the study population, experimental design, and environmental conditions (Lynch and Walsh 1998).

The distinction of two groups broadly corresponding to the commercial groups of black and carioca common beans was already used in several studies (Valdisser et al. 2016; Gioia et al. 2019; Delfini et al. 2021). In contrast to conventional QTL mapping, association mapping is based on unstructured populations; consequently, it is fundamental to consider the population structure and relationship between individuals to prevent false associations due to the confounding effects of population mixture (Oraguzie et al. 2007). This consideration may be particularly applicable in the case of diversity panels, which are assembled from germplasm collections, inbred lines obtained from breeding, and cultivars already released on the market. Thus, selecting the appropriate association method is important and must consider the population structure and the relationship between individuals (Shi et al. 2011).

Regarding to the distribution of the SNPs obtained by GBS, it was observed a higher density towards the ends of the chromosomes, and similar results was obtained in other bean materials (Ariani et al. 2016; Garcia et al. 2021). Mapping resolution and statistical power are the main aspects considered in GWAS, with the former being strongly influenced by LD. Under high-LD conditions, a lower density of markers is necessary in target regions with a high potential for detecting markers strongly associated with polymorphisms of the genes of interest, even if physically distant (Shi et al. 2011). In this regard, a high LD is observed in predominantly autogamous crops (Li et al. 2016) which enables GWAS in common bean populations. In the present study, the LD half-decay corrected for the population structure and kinship (r^2_{vs}) obtained was 296.5 kb, which is relatively higher than the one previously obtained (249 kb) in studies conducted with all Mesoamerican BDP individuals (Delfini et al. 2021), thereby indicating that LD can vary depending on the study population. Considering the LD decay distance in the study population, regions within a distance of 296.5 kb on either side of the detected QTNs were searched for candidate genes.

A comparison of the four multi-locus methods used indicated that the ISIS-EM-BLASSO method identified the highest number of co-detected QTNs, corroborating the findings of previous studies (Cui et al. 2018; Ma et al. 2018; Misra et al. 2018; Zhang et al. 2018b; Fang et al. 2020). Although using the pLARmEB method initially detected the greatest number of QTNs, few of these contributed significantly to heritability, as previously reported by Li et al. (2018) and Lü et al. (2018). Moreover, this method was the least efficient for detecting reliable QTNs. Furthermore, although the FASTmrMLM method presented 100% efficiency (all detected SNPs showed repeatability), the associated PVE and QTN effects were close to zero. Although all four assessed methods involve a combined two-step approach, the different number of QTNs identified may be due to different screening and

estimation models associated with each method (Zhang et al. 2018a). Therefore, to obtain more reliable results, only the QTNs identified with more than one method or in more than one cultivation environment were considered.

Several QTNs detected at least twice using different methods or in different environments were identified and were thus considered reliable. The co-detection of these QTNs indicates the reliability and complementarity of using different methods and environments and also reveals that certain QTNs show stable behavior in different environments. Moreover, the combined use of multiple statistical methods has the advantage of facilitating the identification of small-effect QTNs for characteristics with a complex genetic basis and low heritability (He et al. 2019a).

The detection of pleiotropic QTNs or overlapping genomic regions for different nutrients reinforces the results obtained based on correlation analysis, which revealed positive correlations between most minerals, thereby indicating that the accumulation or transport of different nutrients in grains can be associated with common mechanisms and/or genetic factors. These findings corroborate the findings of previous studies that reported positive correlations among minerals in other crops, including wheat (Cu et al. 2020) and millet (Jaiswal et al. 2019), and in common beans (Delfini et al. 2020).

The large number of QTNs identified may confirm the hypothesis that nutrient accumulation and transport in grains are complex characteristics controlled by many genes or gene families that each have a small effect and that can be detected by the multi-locus models used. Moreover, the results indicate that it is possible to detect QTNs in diversity panels comprising accessions that are routinely developed and used in breeding programs, thereby facilitating the incorporation of these QTNs during the improvement stages (Shi et al. 2011).

In addition to being useful for the identification of genes, the implementation of GWAS can also facilitate the identification of genomic regions associated with the traits of interest, and marker information can be used in breeding programs to efficiently incorporate specific loci in elite germplasms (Collard et al. 2005). The QTNs identified in the present study could play a role in this regard, since the accumulation of favorable alleles is associated with a gradual increase in content of the nutrients studied. Results show that these favorable alleles probably have an additive effect on nutrient accumulation in grains (Zhang et al. 2018b). Increased Fe and Cu contents were not detected, which could be attributed to the fact that few associated QTNs were identified for these elements. Silva, Abreu, and Ramalho (2013) reported that for Zn, additive allele interaction alone explains the content variation, whereas for Fe the occurrence of dominance is also important, and the same may be true for Cu. Thus, it seems that the nutrient content of grains can be improved by increasing the number of favorable alleles in inbred lines and cultivars that already show a good agronomic performance, and marker-assisted selection (MAS) can serve as a useful tool for this purpose.

Although some of the QTNs identified in the present study were characterized as having a small effect, these may prove useful for genomic selection (GS) (He et al. 2019a). In GS models, instead of using a complete panel of randomized SNPs, the use of markers associated with traits of interest may decrease the number of markers and, consequently, the costs of genotyping large breeding populations (Lan et al. 2020). Moreover, the use of specific trait-associated markers in GS models can improve prediction accuracy by reducing background noise in model construction (He et al. 2019b; Ali et al. 2020).

The availability of a genome sequence and gene annotation for common beans enabled the search for genes located in genomic regions around the QTNs that may be associated with nutrient content variation in grains. Potential candidate genes were identified for all nutrients, except K and Mn, with more than one gene identified in most cases and with genes distributed in more than one chromosome for each trait. These results reinforce the quantitative inheritance of these traits, which is well documented for Fe and Zn, elements that tend to be the main focus of studies investigating nutrient contents in common beans (Blair et al. 2009).

Many of the genes detected in the present study were identified as having the molecular function of protein binding, which may be associated with transduction signaling and the transcription factors that modulate gene expression (Villordo-Pineda et al. 2015). Moreover, other genes not directly related to the nutrient with which the QTN was associated were identified, possibly indicating correlations in the mechanisms of absorption or transport between the minerals analyzed. Examples of detected functions include calmodium binding, phospholipid binding, Cu ion transmembrane transport activity, and 4 Fe-4 S cluster

binding. With respect to K, for which none of the detected QTNs were associated with genes directly related to the nutrient, two genes associated with potassium ion transmembrane transport and potassium ion binding functions were identified on chromosomes Pv9 and Pv4, in genomic regions near the QTNs detected for P and Mg, respectively.

Although common beans are rich in multiple minerals, Fe and Zn are the main nutrients studied in this crop, for which numerous QTLs were already described. However, most previous studies focused on common beans of Andean origin or inter-gene-pool populations obtained by crossing Andean and Mesoamerican genotypes; to date, few studies have exclusively focused on panels of Mesoamerican origin. In a QTL study of a Mesoamerican population, Blair et al. (2010) identified new sites related to Fe and Zn that had not been detected in previous studies on Andean or intergenic populations. Similarly, the sites identified in the present study differ from those found previously by Katuuramu et al. (2018), who conducted GWAS for protein, Zn, and Ca contents and the bioavailability of Fe using an Andean diversity panel, as well as from those identified by Erdogmus et al. (2020), who evaluated an intergenic panel with regards to Ca and Mn contents. Thus, the loci identified in the present study will make an important contribution to the improvement of Mesoamerican common beans and more specifically, the black and carioca common beans, the primary types consumed in Brazil.

This study shows that the BDP is efficient for association studies and enables the identification of high variability between different quantified minerals. Moreover, the evaluation of different multi-locus methods and cultivation environments established the reliability of markers associated with minerals. Further studies should validate the importance and functionality of candidate genes associated with the accumulation of nutrients in grains. However, the identified QTNs showed promising potential to be used in breeding programs focused on the pyramiding of favorable alleles, which could be monitored through MAS. In addition, these loci can be incorporated into SNP panels related to traits of interest in GS.

Declarations

Funding

Not applicable

Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

Not applicable

Authors' contributions

J.D., V.M.C. and L.S.A.G. conceived and designed the study; J.D. collected plant material, extracted DNA and performed the genotyping; J.D., J.S.N., A.F.N., L.A.B.R., D.M.Z. performed the phenotyping; J.D. and L.S.A.G. performed bioinformatics and statistical analyses; J.D. drafted the manuscript; J.D., V.M.C., J.S.N., D.M.Z., P.M.R, P.G., and L.S.A.G. edited and revised the final manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank the Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná) and the University of California, Davis (through the Gepts' Lab) for support this research and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) for the scholarship to JD in Brazil and abroad (Finance Code 001).

References

1. Ali M, Zhang Y, Rasheed A, et al (2020) Genomic prediction for grain yield and yield-related traits in chinese winter wheat. *Int J Mol Sci* 21:1–18. doi: 10.3390/ijms21041342
2. Ariani A, Berny Mier y Teran JC, Gepts P (2016) Genome-wide identification of SNPs and copy number variation in common bean (*Phaseolus vulgaris* L.) using genotyping-by-sequencing (GBS). *Mol Breed*. doi: 10.1007/s11032-016-0512-9
3. Beebe S, Gonzalez AV, Rengifo J (2000) Research on trace minerals in the common bean. *Food Nutr Bull* 21:387–391.
4. Blair MW (2013) Mineral biofortification strategies for food staples: The example of common bean. *J Agric Food Chem* 61:8287–8294. doi: 10.1021/jf400774y
5. Blair MW, Astudillo C, Grusak MA, et al (2009) Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol Breed* 23:197–207. doi: 10.1007/s11032-008-9225-z
6. Blair MW, Medina JI, Astudillo C, et al (2010) QTL for seed iron and zinc concentration and content in a Mesoamerican common bean (*Phaseolus vulgaris* L.) population. *Theor Appl Genet* 121:1059–1070. doi: 10.1007/s00122-010-1371-0
7. Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi: 10.1093/bioinformatics/btm308
8. Broughton WJ, Hernandez G, Blair M, et al (2003) Beans (*Phaseolus* spp.) - model food legumes. *Plant Soil* 252:55–128. doi: 10.1023/A:1024146710611
9. Browning BL, Zhou Y, Browning SR (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* 103:338–348. doi: 10.1016/j.ajhg.2018.07.015
10. Cakmak I (2008) Enrichment of cereal grains with zinc: Agronomic or genetic biofortification? *Plant Soil* 302:1–17. doi: 10.1007/s11104-007-9466-3
11. Chang F, Guo C, Sun F, et al (2018) Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. *Front Plant Sci* 9:1–13. doi: 10.3389/fpls.2018.01184
12. Cichy KA, Caldas G V., Snapp SS, Blair MW (2009) QTL analysis of seed iron, zinc, and phosphorus levels in an andean bean population. *Crop Sci* 49:1742–1750. doi: 10.2135/cropsci2008.10.0605
13. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142:169–196. doi: 10.1007/s10681-005-1681-5
14. Cu ST, Guild G, Nicolson A, et al (2020) Genetic dissection of zinc, iron, copper, manganese and phosphorus in wheat (*Triticum aestivum* L.) grain and rachis at two developmental stages. *Plant Sci* 291:110338. doi: 10.1016/j.plantsci.2019.110338
15. Cui Y, Zhang F, Zhou Y (2018) The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. *Front Plant Sci* 9:1–9. doi: 10.3389/fpls.2018.01464
16. Delfini J, Moda-Cirino V, dos Santos Neto J, et al (2021) Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. *Sci Rep*. doi: 10.1038/s41598-021-82437-4
17. Delfini J, Moda-Cirino V, dos Santos Neto J, et al (2020) Diversity of nutritional content in seeds of Brazilian common bean germplasm. *PLoS One* 1–13. doi: 10.1371/journal.pone.0239263
18. Desrousseaux D, Sandron F, Siberchicot A, et al (2017) LDcorSV: Linkage disequilibrium corrected by the structure and the relatedness. R package version 1.3.2.
19. Dwivedi SL, Sahrawat KL, Rai KN, et al (2012) Nutritionally Enhanced Staple Food Crops. *Plant Breed Rev* 36:169–291. doi: 10.1002/9781118358566

20. Edwards D, Batley J (2010) Plant genome sequencing: Applications for crop improvement. *Plant Biotechnol J* 8:2–9. doi: 10.1111/j.1467-7652.2009.00459.x
21. Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome J* 4:250. doi: 10.3835/plantgenome2011.08.0024
22. Erdogmus S, Ates D, Nemli S, et al (2020) Genome-wide association studies of Ca and Mn in the seeds of the common bean (*Phaseolus vulgaris* L.). *Genomics* 112:4536–4546. doi: 10.1016/j.ygeno.2020.03.030
23. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14:2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
24. Fang Y, Liu S, Dong Q, et al (2020) Linkage Analysis and Multi-Locus Genome-Wide Association Studies Identify QTNs Controlling Soybean Plant Height. *Front Plant Sci* 11:1–14. doi: 10.3389/fpls.2020.00009
25. Garcia T, Duitama J, Zullo SS, et al (2021) Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. *Nat Commun* 12:1–17. doi: 10.1038/s41467-021-20921-1
26. Gioia T, Logozzo G, Marzario S, et al (2019) Evolution of SSR diversity from wild types to U.S. Advanced cultivars in the Andean and Mesoamerican domestications of common bean (*Phaseolus vulgaris*). *PLoS One* 14:1–21. doi: 10.1371/journal.pone.0211342
27. Hallauer AR, Miranda Filho JB (1988) Quantitative genetics in maize breeding. Iowa State University Press, Ames
28. He L, Xiao J, Rashid KY, et al (2019a) Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.). *Front Plant Sci* 9:1–15. doi: 10.3389/fpls.2018.01982
29. He L, Xiao J, Rashid KY, et al (2019b) Evaluation of genomic prediction for PasmO resistance in flax. *Int J Mol Sci*. doi: 10.3390/ijms20020359
30. Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol* 33:54–78. doi: 10.1016/0040-5809(88)90004-4
31. Islam FMA, Basford KE, Jara C, et al (2002) Seed compositional and disease resistance differences among gene pools in cultivated common bean. *Genet Resour Crop Evol* 49:285–293. doi: 10.1023/A:1015510428026
32. Jaiswal V, Bandyopadhyay T, Gahlaut V, et al (2019) Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (*Setaria italica* L.). *J Cereal Sci* 85:48–55. doi: 10.1016/j.jcs.2018.11.006
33. Katuramu DN, Hart JP, Porch TG, et al (2018) Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (*Phaseolus vulgaris* L.). *Mol Breed*. doi: 10.1007/s11032-018-0798-x
34. Kolde R (2019) pheatmap: Pretty Heatmaps. R package version 1.0.12.
35. Lan S, Zheng C, Hauck K, et al (2020) Genomic prediction accuracy of seven breeding selection traits improved by QTL identification in flax. *Int J Mol Sci*. doi: 10.3390/ijms21051577
36. Li J, Tang W, Zhang YW, et al (2018) Genome-wide association studies for five forage quality-related traits in sorghum (*sorghum bicolor* L.). *Front Plant Sci* 9:1–8. doi: 10.3389/fpls.2018.01146
37. Li L, Guo N, Niu J, et al (2016) Loci and candidate gene identification for resistance to *Phytophthora sojae* via association analysis in soybean [*Glycine max* (L.) Merr.]. *Mol Genet Genomics* 291:1–9. doi: 10.1007/s00438-015-1164-x
38. Lü H, Yang Y, Li H, et al (2018) Genome-wide association studies of photosynthetic traits related to phosphorus efficiency in soybean. *Front Plant Sci*. doi: 10.3389/fpls.2018.01226
39. Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland
40. Ma L, Liu M, Yan Y, et al (2018) Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. *Front Plant Sci* 9:1–15. doi: 10.3389/fpls.2018.00561
41. McClean PE, Moghaddam SM, López-Millán A-F, et al (2017) Phenotypic diversity for seed mineral concentration in North American dry bean germplasm of middle american ancestry. *Crop Sci* 57:3129–3144. doi: 10.2135/cropsci2017.04.0244
42. Misra G, Badoni S, Domingo CJ, et al (2018) Deciphering the genetic architecture of cooked rice texture. *Front Plant Sci*. doi: 10.3389/fpls.2018.01405

43. Miyazawa M, Pavan MA, Muraoka T, et al (1999) Análises químicas de tecido vegetal. In: Silva FC (ed) Manual de análises químicas de solos, plantas e fertilizantes. Embrapa, Brasília, pp 171–223
44. Mukamuhirwa F, Tusiime G, Mukankusi MC (2015) Inheritance of high iron and zinc concentration in selected bean varieties. *Euphytica* 205:349–360. doi: 10.1007/s10681-015-1385-4
45. Oraguzie N, Rikkerink E, Gardiner S, Silva H (2007) Association mapping in plants. Springer, New York
46. Petry N, Boy E, Wirth JP, Hurrell RF (2015) Review: The potential of the common bean (*Phaseolus vulgaris*) as a vehicle for iron biofortification. *Nutrients* 7:1144–1173. doi: 10.3390/nu7021144
47. Pinheiro C, Baeta JP, Pereira AM, et al (2010) Diversity of seed mineral composition of *Phaseolus vulgaris* L. germplasm. *J Food Compos Anal* 23:319–325. doi: 10.1016/j.jfca.2010.01.005
48. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
49. Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. doi: 10.1086/519795
50. R Core Team (2020) R: A language and environment for statistical computing.
51. Raggi L, Caproni L, Carboni A, Negri V (2019) Genome-wide association study reveals candidate genes for flowering time variation in common bean (*Phaseolus vulgaris* L.). *Front Plant Sci* 10:1–14. doi: 10.3389/fpls.2019.00962
52. Rasheed A, Hao Y, Xia X, et al (2017) Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. *Mol Plant* 10:1047–1064. doi: 10.1016/j.molp.2017.06.008
53. RAWAL V, NAVARRO DK (2019) The Global Economy of Pulses. FAO
54. Ribeiro ND, Rodrigues J de A, Prigol M, et al (2014) Evaluation of special grains bean lines for grain yield, cooking time and mineral concentrations. *Crop Breed Appl Biotechnol* 14:15–22. doi: 10.1590/s1984-70332014000100003
55. SAS Institute (2000) JMP IN 4.0.3.
56. Schmutz J, McClean PE, Mamidi S, et al (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet* 46:707–13. doi: 10.1038/ng.3008
57. Shi C, Navabi A, Yu K, et al (2011) Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. *BMC Plant Biol* 11:52. doi: 10.1186/1471-2229-11-52
58. Silva CA, Abreu Â de FB, Ramalho MAP (2013) Genetic control of zinc and iron concentration in common bean seeds. *African J Agric Res* 8:1001–1008. doi: 10.5897/AJAR12.2172
59. Silva CA, Abreu Â de FB, Ramalho MAP, Maia LGS (2012) Chemical composition as related to seed color of common bean. *Crop Breed Appl Biotechnol* 12:132–137.
60. Stansell Z, Hyma K, Fresnedo-Ramírez J, et al (2018) Genotyping-by-sequencing of *Brassica oleracea* vegetables reveals unique phylogenetic patterns, population structure and domestication footprints. *Hortic Res*. doi: 10.1038/s41438-018-0040-3
61. Stapley J, Reger J, Feulner PGD, et al (2010) Adaptation genomics: The next generation. *Trends Ecol Evol* 25:705–712. doi: 10.1016/j.tree.2010.09.002
62. Tamba CL, Ni YL, Zhang YM (2017) Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput Biol* 13:1–20. doi: 10.1371/journal.pcbi.1005357
63. Tamba CL, Zhang Y-M (2018) A fast mrMLM algorithm for multi-locus genome-wide association studies. 341784. doi: 10.1101/341784
64. Valdisser PAMR, Pappas GJ, de Menezes IPP, et al (2016) SNP discovery in common bean by restriction-associated DNA (RAD) sequencing for genetic diversity and population structure analysis. *Mol Genet Genomics* 291:1277–1291. doi: 10.1007/s00438-016-1182-3
65. Villordo-Pineda E, González-Chavira MM, Giraldo-Carbajo P, et al (2015) Identification of novel drought-tolerant-associated SNPs in common bean (*Phaseolus vulgaris*). *Front Plant Sci* 6:1–9. doi: 10.3389/fpls.2015.00546

66. Wang SB, Feng JY, Ren WL, et al (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci Rep* 6:1–10. doi: 10.1038/srep19444
67. Wei T, Simko V (2017) R package “corrplot”: Visualization of a Correlation Matrix (Version 0.84).
68. Ya-Wen Z, Pei L, Yuan-Ming Z (2019) mrMLM.GUI: Multi-Locus Random-SNP-Effect Mixed Linear Model Tools for Genome-Wide Association Study with Graphical User Interface.R package version 4.0.
69. Zhang J, Feng JY, Ni YL, et al (2017) PLARmEB: Integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity (Edinb)* 118:517–524. doi: 10.1038/hdy.2017.8
70. Zhang K, Liu S, Li W, et al (2018a) Identification of QTNs controlling seed protein content in soybean using multi-locus genome-wide association studies. *Front Plant Sci* 871:1–14. doi: 10.3389/fpls.2018.01690
71. Zhang Y, Liu P, Zhang X, et al (2018b) Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front Plant Sci*. doi: 10.3389/fpls.2018.00611
72. Zheng X, Levine D, Shen J, et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328. doi: 10.1093/bioinformatics/bts606

Figures

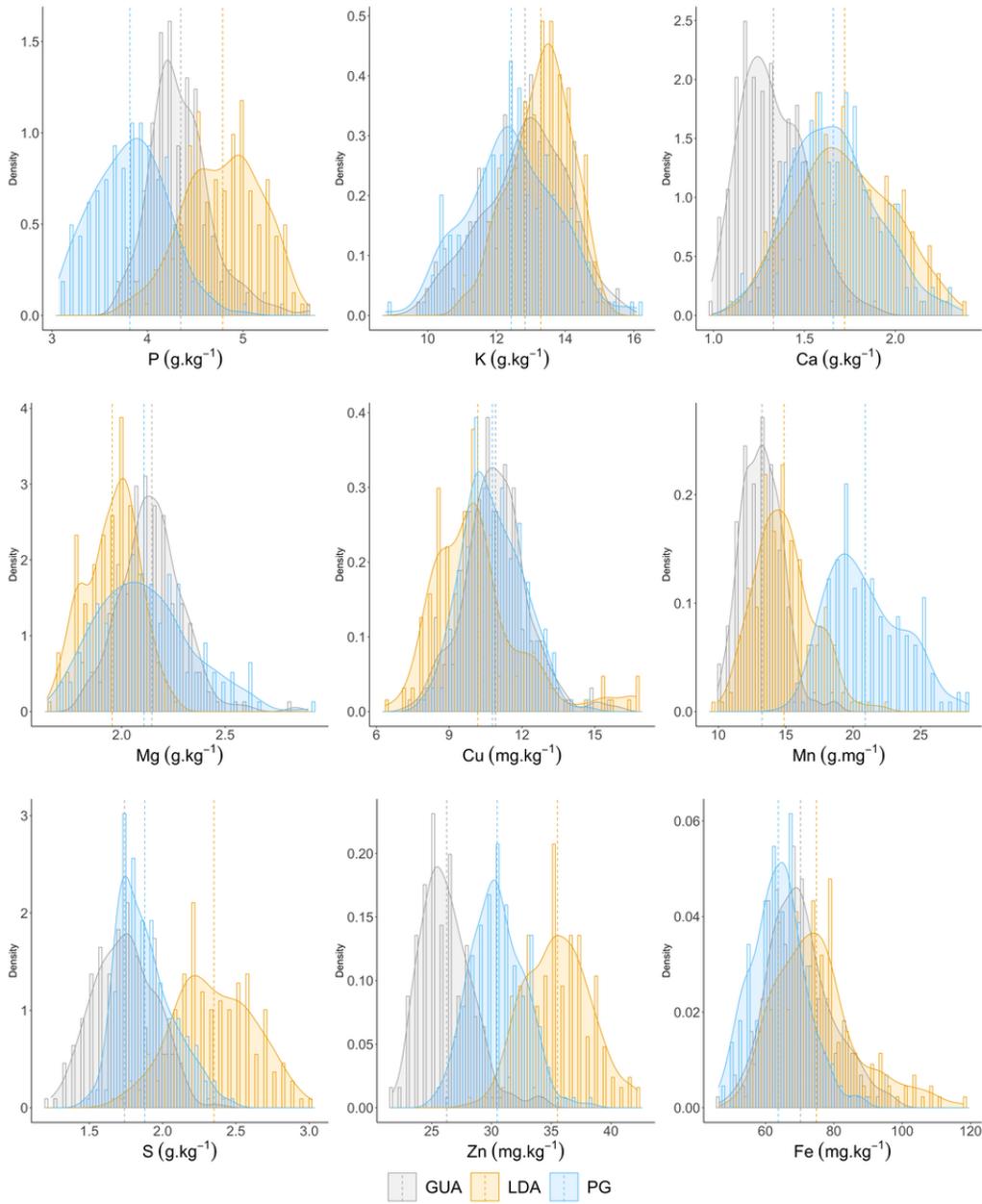


Figure 1

Histogram and density distribution of nine nutrients evaluated in three environments, Londrina (LDA), Ponta Grossa (PG), and Guarapuava (GUA), in common bean accessions from the Brazilian Diversity Panel (BDP). Dashed lines indicate the mean value.



Figure 2

Density distribution of single nucleotide polymorphisms (SNPs) identified by genotyping-by-sequencing (GBS) from the Brazilian Diversity Panel (BDP) along the common bean genome at a 1-Mb window size.

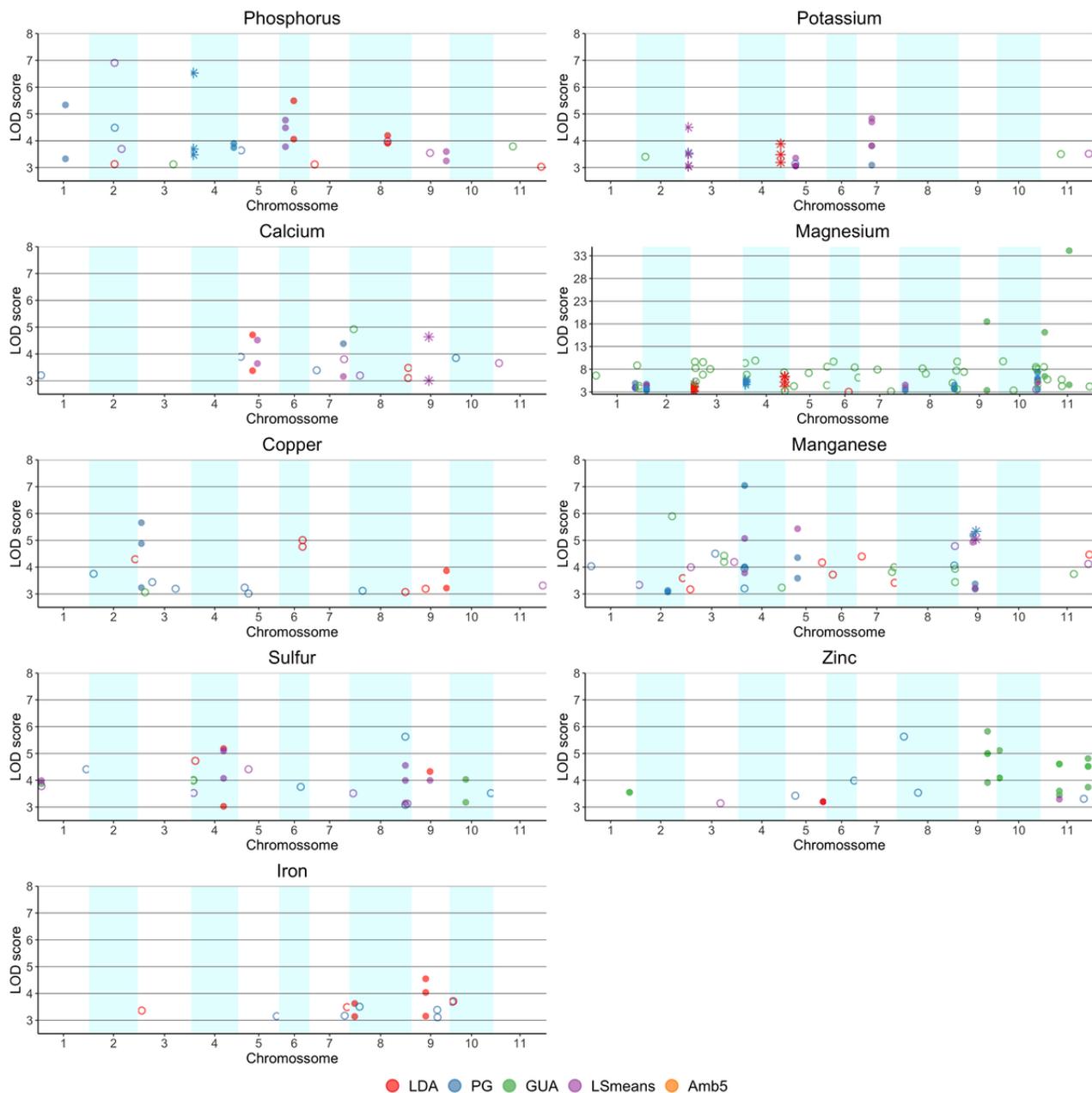


Figure 3

Quantitative trait nucleotides (QTNs) associated with mineral contents in common beans from the Brazilian Diversity Panel (BDP), detected using different methods and in different environments. LDA = Londrina, PG = Ponta Grossa, GUA = Guarapuava, LSmeans = overall mean. Empty points correspond to QTNs detected only once, full points indicate QTNs detected at least twice, and asterisks indicate QTNs detected at least twice and for more than one mineral.

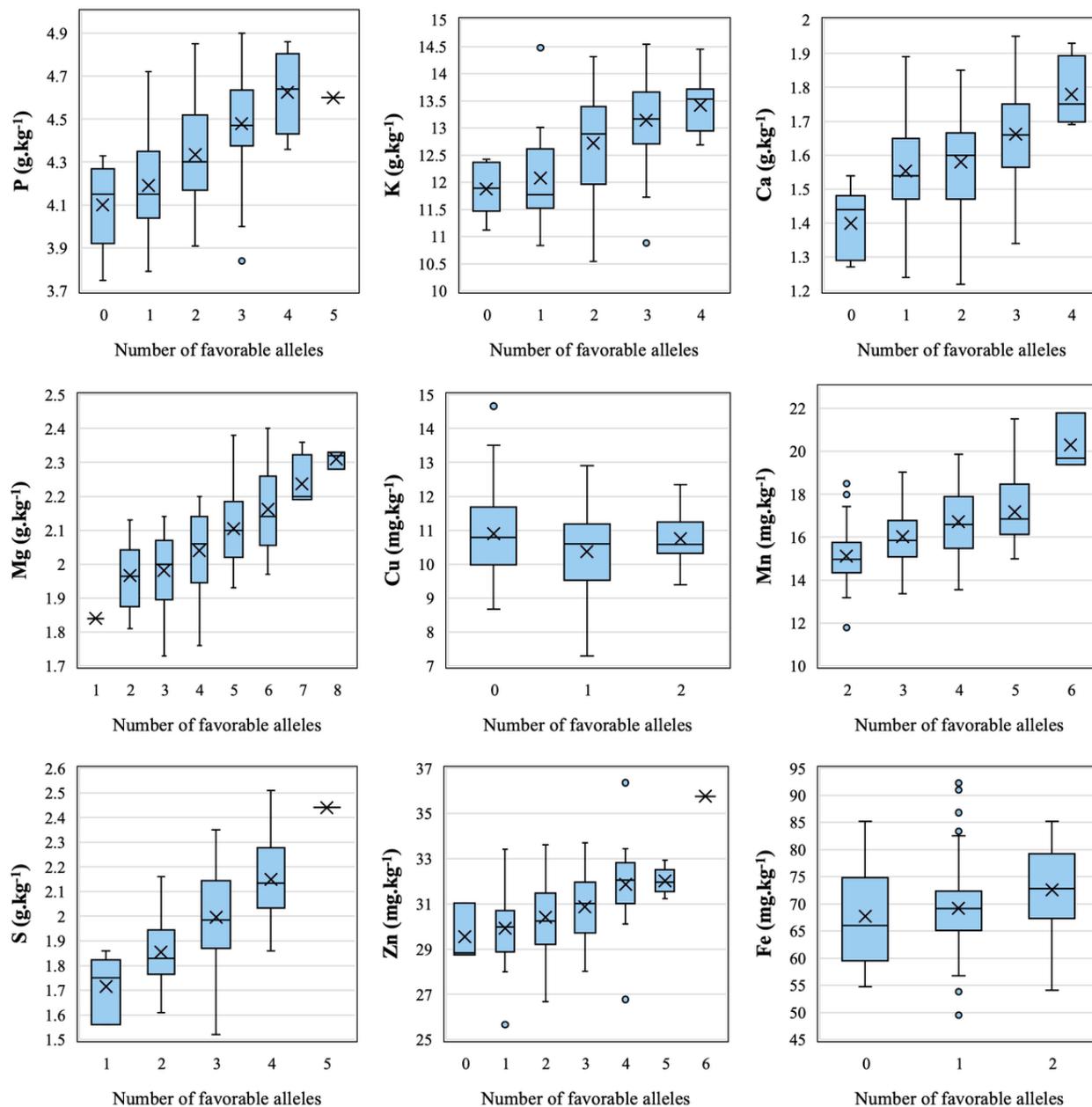


Figure 4

Accumulation of favorable alleles in relation to overall adjusted means (LSmeans) for each of the nine minerals present in common bean accessions from the Brazilian Diversity Panel (BDP).

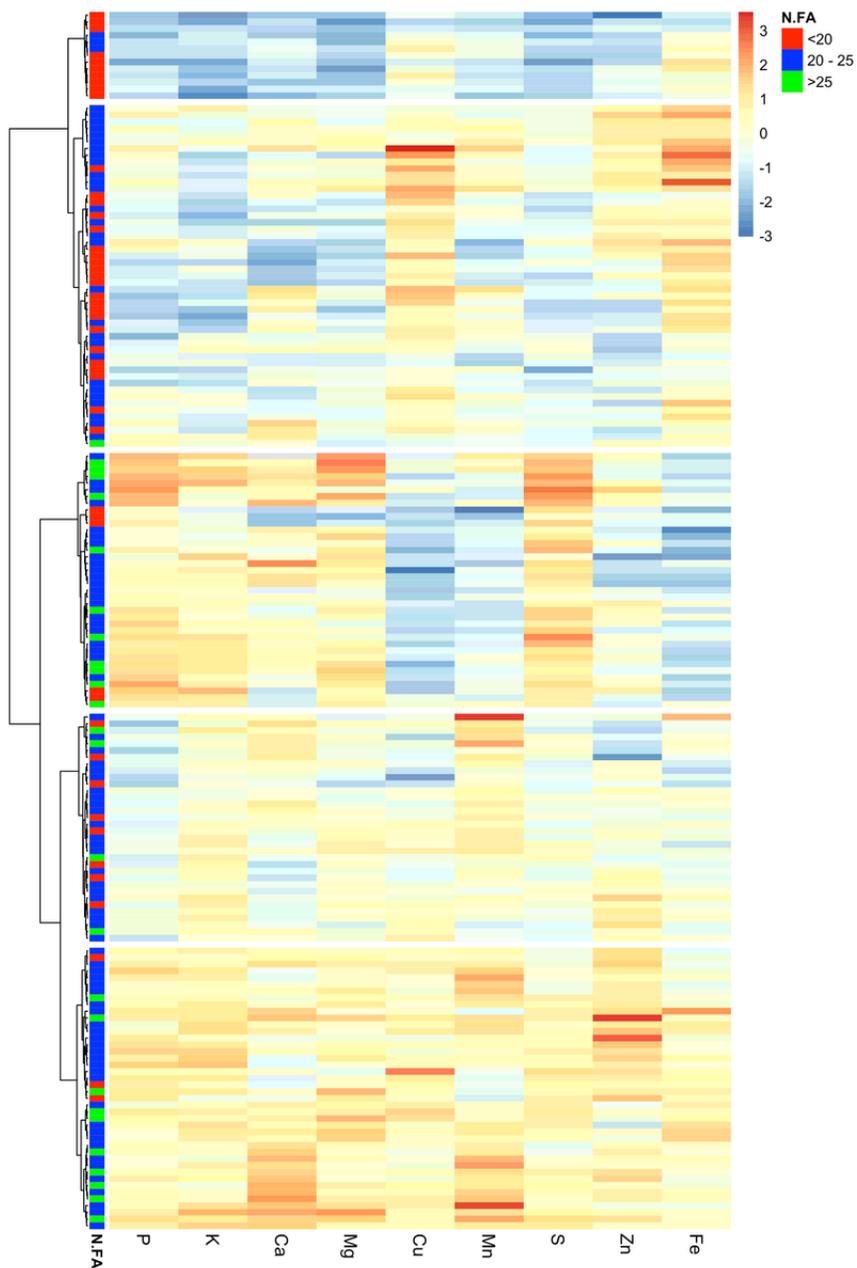


Figure 5

Number of favorable alleles (N.F.A) and Ward's hierarchical clustering based on Euclidian distance associated with heatmap for different mineral contents quantified on common bean accessions from the Brazilian Diversity Panel (BDP).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile.pdf](#)