Universidade Estadual de Londrina

JESSICA DELFINI DE PAULA IÁCONO

DIVERSIDADE GENÉTICA, ESTRUTURA POPULACIONAL E MAPEAMENTO ASSOCIATIVO PARA CARACTERÍSTICAS NUTRICIONAIS E AGRONÔMICAS EM UM PAINEL DE FEIJÃO MESOAMERICANO BRASILEIRO

Londrina
2020

JESSICA DELFINI DE PAULA IÁCONO

DIVERSIDADE GENÉTICA, ESTRUTURA POPULACIONAL E MAPEAMENTO ASSOCIATIVO PARA CARACTERÍSTICAS NUTRICIONAIS E AGRONÔMICAS EM UM PAINEL DE FEIJÃO MESOAMERICANO BRASILEIRO

Tese apresentada ao Programa de Pós-Graduação em Agronomia da Universidade Estadual de Londrina, como requisito para obtenção do título de doutor em Agronomia.

Orientador: Prof. Dr. Paulo Maurício Ruas.
Co-orientadora: Dra. Vânia Moda Cirino.
Co-orientador no exterior: Dr. Paul Gepts – University of California, Davis.

Londrina
2020

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

I11d Iácono, Jessica Delfini de Paula.
Diversidade genética, estrutura populacional e mapeamento associativo para características nutricionais e agronômicas em um painel de feijão mesoamericano brasileiro / Jessica Delfini de Paula Iácono. - Londrina, 2020.
117 f. : il.

Orientador: Paulo Maurício Ruas.
Coorientador: Vania Moda-Cirino.
Coorientador: Paul Gepts.
Tese (Doutorado em Agronomia) - Universidade Estadual de Londrina, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia, 2020.
Inclui bibliografia.

1. Agronomia - Tese. 2. Feijão comum - Tese. 3. Estudos de associação ampla do genoma - Tese. 4. Polimorfismos de nucleotídeo único - Tese. I. Ruas, Paulo Maurício. II. Moda-Cirino, Vania. III. Gepts, Paul. IV. Universidade Estadual de Londrina. Centro de Ciências Agrárias. Programa de Pós-Graduação em Agronomia. V. Título.

CDU 63

JESSICA DELFINI DE PAULA IÁCONO

DIVERSIDADE GENÉTICA, ESTRUTURA POPULACIONAL E MAPEAMENTO ASSOCIATIVO PARA CARACTERÍSTICAS NUTRICIONAIS E AGRONÔMICAS EM UM PAINEL DE FEIJÃO MESOAMERICANO BRASILEIRO

Tese apresentada ao Programa de Pós-Graduação em Agronomia da Universidade Estadual de Londrina, como requisito para obtenção do título de doutor em Agronomia.

BANCA EXAMINADORA

_____

Orientador: Prof. Dr. Paulo Maurício Ruas
Universidade Estadual de Londrina - UEL

_____

Co-Orientador: Dr.ª Vânia Moda Cirino
Instituto de Desenvolvimento Rural do Paraná – IDR-Paraná

_____

Prof. Dr. Leandro Simões Azeredo Gonçalves
Universidade Estadual de Londrina - UEL

_____

Dr. Luiz Filipe Protasio Pereira
Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA

_____

Dr. Gaspar Malone
GDM Seeds

_____

Dr. Leandro de Freitas Mendonça
GDM Seeds

Londrina, 27 de novembro de 2020.

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus por estar sempre presente em minha vida, me iluminando e me dando forças para enfrentar todos os obstáculos;

Ao meu orientador Prof. Dr. Paulo Maurício Ruas que me acompanha desde o início da graduação, juntamente com sua esposa Profa. Dra. Claudete de Fátima Ruas, e que juntos sempre me incentivaram e apoiaram meus estudos;

À minha co-orientadora Vânia Moda Cirino, um exemplo de mulher na pesquisa e que me inspira a seguir nessa jornada. Obrigada Dra. Vânia por sempre transmitir muito conhecimento, por me abrir muitas portas, pelo suporte, conselhos valorosos e amizade;

Ao Prof. Dr. Leandro Simões de Azeredo Gonçalves por todas as grandes contribuições, por idealizar esse trabalho, por todo o suporte oferecido, pelas oportunidades, conselhos e amizade;

Ao Dr. Paul Gepts que me recebeu durante o período de sanduíche na Universidade da Califórnia em Davis, por ter auxiliado e contribuído de maneira muito relevante;

Ao amigo e companheiro de trabalho durante todo o meu período de pós-graduação, José dos Santos Neto. O seu apoio foi fundamental na realização desse trabalho, obrigada por todo o tempo dedicado, conselhos e amizade;

A todos os servidores do Apoio de Feijão do IDR-Paraná e funcionários terceirizados, especialmente aos técnicos agrícolas Adair Morais e Paulo Lopez de Londrina, Eloir Myska de Guarapuava e Carlos Frederico de Oliveira de Ponta Grossa, que sempre me auxiliaram no desenvolvimento do trabalho a campo;

Aos estagiários que passaram pelo setor de Recursos Genéticos do IDR-Paraná que me ajudaram na montagem, condução e avaliação dos experimentos. Em especial Giovana, Camila, Gabriela, Guilherme, Leonardo, Débora, Gustavo, Deyvid e Abiézer.

Ao Laboratório de Solos e Tecidos Vegetais do IDR-Paraná, em especial à Rosinei Aparecida de Souza e sua equipe pelo apoio e condução das análises de caracterização nutricional;

Aos membros do Laboratório de Marcadores Moleculares e Citogenética de plantas da UEL, pela amizade, ensinamentos e auxílio;

Aos mestres e amigos Alison Fernando Nogueira e Luriam Aparecida Brandão pela amizade e auxílio na condução dos experimentos;

Ao Dr. Gustavo César Sant'Ana pelos conhecimentos transmitidos;

Ao Msc. Douglas Mariani Zeffa pelo auxílio e troca de conhecimentos;

A todos que de alguma maneira compartilharam conhecimento, experiências e me auxiliaram nas análises, em especial Dr. Nelson da Fonseca Júnior e Dra. Mariane Silva Felicio;

Aos membros da banca por terem aceitado o convite e disponibilizarem tempo para fazer contribuições;

À Universidade Estadual de Londrina, ao programa de Pós-Graduação em Agronomia e a todos professores que contribuíram com o meu conhecimento;

Ao Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná) pela infraestrutura disponibilizada para condução desse trabalho;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa no país e no exterior (Doutorado Sanduíche).

Aos meus pais Laercio e Lucinêa, e meu irmão Gabriel, por sempre me darem todo o apoio necessário e incentivo para realização dos meus sonhos, agradeço também meus avós que sempre torceram muito por mim;

Ao meu marido Mario Henrique, por ser meu maior incentivador e me apoiar em todas as minhas decisões, agradeço também a toda sua família pelo apoio;

Enfim, o sucesso na realização desse trabalho foi devido a colaboração de muitas pessoas, que de alguma forma me auxiliaram. Deixo aqui o meu agradecimento a todos que contribuíram!

MUITO OBRIGADA!

“Os atrasos, não são as negativas de Deus.
A razão porque as pessoas desistem tão rapidamente é porque elas tendem a olhar para quão distante estão do objetivo, ao invés de olhar para o quanto já se distanciaram do início.

Não importa quantos erros você cometeu ou quão lento progride, você ainda está na frente de todos que ainda não estão tentando.”

Tony Robins

IÁCONO, Jessica Delfini de Paula. **Diversidade genética, estrutura populacional e mapeamento associativo para características nutricionais e agronômicas em um painel de feijão mesoamericano brasileiro.** 2020. 117 f. Tese (Doutorado em Agronomia) – Universidade Estadual de Londrina, Londrina, 2020.

RESUMO

O feijão (*Phaseolus vulgaris* L.) é um dos alimentos básicos na dieta do brasileiro, bem como é considerado uma das leguminosas mais importantes no mundo. É reconhecidamente uma excelente fonte de proteínas, carboidratos e micronutrientes, e devido aos seus atributos nutricionais e importância na alimentação em países em desenvolvimento, se torna uma cultura adequada para biofortificação. A identificação de genes e regiões genômicas relacionados com características nutricionais e agronômicas, é de grande importância para o melhoramento genético de feijão, podendo acelerar o desenvolvimento de novas cultivares. Nesse contexto, o presente trabalho teve como objetivo acessar a diversidade genética, estrutura populacional e desequilíbrio de ligação em um painel de diversidade de feijão designado *Brazilian Diversity Panel (BDP)*, e em seguida identificar regiões genômicas associadas a características nutricionais e agronômicas. A genotipagem por sequenciamento (*Genotyping by sequencing – GBS*) de 219 acessos pertencentes ao BDP permitiu a identificação de 49.817 SNPs com MAF > 0,05. A análise bayesiana da estrutura populacional evidenciou a subdivisão do painel em 3 grupos, um formado por feijões de origem andina, outro formado predominantemente por feijões de origem mesoamericana pertencentes ao grupo comercial carioca, e outro por feijões de origem mesoamericana com cor da semente preta, creme, vermelho e outras. A análise do desequilíbrio de ligação evidenciou a existência de longos blocos de ligação e um baixo decaimento do LD em função da distância física entre os SNPs. Para os estudos de mapeamento associativo (*Genome wide association studies -GWAS*) foram utilizados 178 acessos de origem mesoamericana pertencentes ao BDP e 25.011 SNPs de boa qualidade. O painel foi fenotipado para nove nutrientes (fósforo, potássio, cálcio, magnésio, cobre, manganês, enxofre, zinco e ferro) e dez características agronômicas (altura de planta, altura de inserção da primeira vagem, número de nós, comprimento da vagem, número total de vagens por planta, número de locos por vagem, número de sementes por vagem, peso total de sementes por planta, peso de 100 sementes e rendimento de grãos) em três e quatro ambientes, respectivamente. Quatro métodos multi-locus de GWAS foram empregados nesse estudo (mrMLM, FASTmrMLM, pLARmEB e ISIS EM-BLASSO). Apenas *QTNs* (*Quantitative Trait Nucleotide*) detectados por diferentes métodos ou ambientes, foram considerados verdadeiramente significativos, resultando em 48 e 64 QTNs estáveis relacionados com o conteúdo nutricional e características agronômicas, respectivamente. *QTNs* pleiotrópicos e regiões genômicas sobrepostas em torno dos QTNs foram identificadas para entre diferentes características, demonstrando que os mecanismos relacionados a expressão dessas características podem estar associados. O acúmulo de alelos superiores revelou aumento gradual do teor de nutrientes no grão, bem como causaram um incremento positivo para a característica agronômicas. O *BDP* se mostrou eficiente para estudos de associação e a exploração de diferentes métodos e ambientes demonstrou a confiabilidade dos marcadores associados as características. Os lócus identificados no presente trabalho serão importantes para o melhoramento de feijões de origem mesoamericana e mais especificamente para os principais grupos consumidos no Brasil, preto e carioca.

Palavras-chave: *Phaseolus vulgaris* L.; *Genotyping by sequencing* (*GBS*); desequilibrio de ligação; *Single nucleotide polymorphism* (*SNP*); *Genome wide association studies* (GWAS).

IÁCONO, Jessica Delfini de Paula. **Genetic diversity, population structure and genome wide association studies for nutritional and agronomic traits in a Brazilian Mesoamerican bean panel.** 2020. 117 p. Thesis (Doctor's Degree in Agronomy) – Universidade Estadual de Londrina, Londrina, 2020.

ABSTRACT

The common bean (*Phaseolus vulgaris* L.) is one of the staple foods in the Brazilian diet, as well as being considered one of the most important legumes in the world. It is recognized as an excellent source of proteins, carbohydrates and micronutrients, and due to its nutritional attributes and importance in developing countries, it becomes a suitable crop for biofortification. The identification of genes and genomic regions related to nutritional and agronomic traits is of great importance for the genetic improvement of beans, being able to accelerate and provide more efficiency in the development of new cultivars. In this context, the present work aims to access the genetic diversity, population structure and the linkage disequilibrium in a common bean diversity panel called Brazilian Diversity Panel (BDP), and then identify genomic regions associated with nutritional and agronomic traits. Genotyping by sequencing (GBS) of 219 accessions belonging to the BDP allowed the identification of 49,817 SNPs with MAF> 0.05. The Bayesian analysis of the population structure showed that the panel was subdivided into 3 groups, one formed by common beans of Andean origin, another formed predominantly by Mesoamerican origin belonging to the carioca commercial group, and another by common beans of Mesoamerican origin with black seed color, cream, red and others. The analysis of the linkage disequilibrium showed the existence of long linkage blocks and a low LD decay due to the physical distance between the SNPs. For the genome wide association studies (GWAS) 178 accessions of Mesoamerican origin belonging to the BDP and 25,011 good quality SNPs were used. The panel was phenotyped for nine nutrients (phosphorus, potassium, calcium, magnesium, copper, manganese, sulfur, zinc and iron) and ten agronomic traits (plant height, first pod insertion height, number of nodes, pod length, total number of pods per plant, number of locules per pod, number of seeds per pod, total seed weight per plant, weight of 100 seeds and grain yield) in three and four environments, respectively. Four multi-locus GWAS methods were employed in this study (mrMLM, FASTmrMLM, pLARmEB and ISIS EM-BLASSO). Only QTNs (Quantitative Trait Nucleotide) detected by different methods or environments, were considered truly significant, resulting in 48 and 64 stable QTNs related to the mineral content and agronomic traits evaluated, respectively. Pleiotropic QTNs and overlapping of the genomic regions surrounding the QTNs were identified among different traits, demonstrating that the mechanisms related to expression between these characteristics may be associated. The accumulation of superior alleles revealed a gradual increase in the content of nutrients in the grain, as well as causing a positive increase for the agronomic traits. The BDP proved to be efficient for association studies and the exploration of different methods and environments demonstrated the reliability of the markers associated with the traits. The loci identified in the present study will be important for the breeding of common beans of Mesoamerican origin and more specifically for the main groups consumed in Brazil, black and carioca.

Keywords: *Phaseolus vulgaris* L.; *Genotyping by sequencing* (*GBS*); linkage desequilibrium; *Single nucleotide polymorphism* (*SNP*); *Genome wide association studies* (GWAS).

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

O feijão (*Phaseolus vulgaris* L.) é uma espécie de grande interesse agronômico, representando 50% do total de leguminosas consumidas pelo homem em todo o mundo. Essa leguminosa é uma importante fonte de proteína (aproximadamente 22%), vitaminas e minerais (Ca, Cu, Fe, Mg, Mn, Zn) na dieta humana, especialmente em países em desenvolvimento (BROUGHTON et al., 2003). O Brasil destaca-se no cenário internacional como um dos maiores consumidores e maior produtor mundial de feijão comum segundo estatística da FAO (2020). A cultura está amplamente distribuída em todo o território nacional e possui um alto valor socioeconômico, por ser produzida por pequenos, médios e grandes produtores em sistemas diversificados de produção.

O feijão, como muitas leguminosas, são ricos em ferro, zinco e outros microelementos que, geralmente, se encontram em baixas concentrações em cereais, e raízes e/ou tubérculos e, portanto, são bons candidatos para biofortificação (BLAIR, 2013). A biofortificação é o processo de melhoria de culturas básicas para o conteúdo mineral ou vitamínico como forma de enfrentar a desnutrição nos países em desenvolvimento. Sabendo-se do hábito da população brasileira de consumir diariamente o feijão, a biofortificação da cultura torna-se uma alternativa promissora, pois não altera o hábito do consumidor e não requer custos adicionais para administração da dieta.

O panorama da produtividade do feijão no Brasil foi significativamente otimizado nas últimas décadas, associado a melhoria da qualidade tecnológica e nutricional dos grãos. O incremento da produtividade deve-se a diversos fatores, como os relacionados as tecnologias para instalação e condução da lavoura, práticas culturais, insumos e de manejo e conservação de solos. Entretanto, o maior ganho em produtividade ocorreu principalmente em virtude do uso de cultivares melhoradas, com alto potencial de rendimento, estabilidade de produção conferida pela resistência ou tolerância a fatores bióticos e abióticos adversos e adaptadas a colheita mecanizada (RAMALHO; DIAS; CARVALHO, 2012; TSUTSUMI; BULEGON; PIANO, 2015).

Para o melhoramento de plantas, estudos de associação genômica ampla (*Genome-wide association studies* - *GWAS*) para características nutricionais e agronômicas são de grande importância, visando encontrar associações entre características e marcadores, o que pode agilizar a seleção de genótipos superiores. Essa técnica requer uma grande quantidade de marcadores para uma boa cobertura do genoma (EDWARDS; BATLEY, 2010; STAPLEY et al., 2010). As tecnologias de sequenciamento de próxima geração (*Next-generation sequencing* - *NGS*) estão

revolucionando estudos genéticos e desenvolvimento de marcadores moleculares, aumentando exponencialmente o número de variantes genéticas que podem ser descobertas em um único experimento (STAPLEY et al., 2010).

As *NGS* tem permitido que a técnica de genotipagem por sequenciamento (*genotyping-by-sequencing - GBS*) torne-se viável para estudos de espécies com genoma amplo e de alta diversidade (ELSHIRE et al., 2011). O GBS é uma técnica robusta, de alto rendimento, econômica e simples para obter milhares de marcadores de um grande número de indivíduos, no qual captura dados *SNPs* utilizando uma biblioteca de representação reduzida do genoma (ARIANI; BERNY MIER Y TERAN; GEPTS, 2016; SCHRÖDER et al., 2016).

O GWAS possibilita a identificação de múltiplos polimorfismos que ocorrem de forma natural dentro de uma espécie, já que é baseado em coleções de germoplasma com pouca estrutura genética, podendo ser empregado, preferencialmente, o uso de genótipos que possuem características de interesse para programas de melhoramento (KORTE; ASHLEY, 2013). Dessa forma, a formação do painel de diversidade a ser utilizado nesses estudos é de grande importância. Além de ser formado por acessos que apresentem alta variabilidade, esses devem representar os principais grupos comerciais alvo do melhoramento no país, bem como adaptados as condições climáticas do mesmo. No Brasil, há uma preferência por feijões de origem mesoamericana, sendo os grupos comerciais carioca e preto os mais consumidos.

Estudos de GWAS tem sido extensivamente empregados em diversas culturas de importância econômica. Em feijão, alguns estudos já foram realizados para diferentes características, entretanto direcionados principalmente para feijões de origem andina, ou à painéis compostos pelos dois centros de origem, andino e mesoamericano. Poucos estudos foram direcionados á painéis exclusivos de origem mesoamericana, bem como formados por feijões adaptados as condições climáticas do Brasil.

Nesse contexto, o presente trabalho tem como objetivo (i) acessar a diversidade genética, estrutura populacional e desequilíbrio de ligação em um painel de diversidade de feijão designado *Brazilian Diversity Panel* (*BDP*) composto por cultivares, variedades locais e linhagens melhoradas que representam a variabilidade da espécie presente do Brasil; e utilizando um subset do *BDP* composto apenas por acessos de origem andina, identificar regiões genômicas relacionadas (ii) a concentração de nutrientes nos grãos de feijão e (iii) a características morfo-agronômicas relacionadas a rendimento e seus componentes.

2 REVISÃO DE LITERATURA

2.1 CULTURA DO FEIJÃO

O feijão (*Phaseolus vulgaris* L.) é uma das cinco espécies cultivadas do gênero *Phaseolus* e uma das leguminosas mais consumidas em todo o mundo. Em muitos países é o grão mais importante para o consumo humano direto e a principal fonte de proteínas e micronutrientes (BROUGHTON et al., 2003).

A espécie *P. vulgaris* é uma planta diplóide, com 2n=22 cromossomos, e tem seu centro de origem localizado no México e América Central, os quais foram domesticados independentemente e originaram dois grupos gênicos (*gene pools)*, Andino e Mesoamericano, que são morfologicamente e geneticamente diferenciados, assim como evidenciado em diversos trabalhos (BITOCCHI et al., 2013; KWAK; GEPTS, 2009; SCHMUTZ et al., 2014; VLASOVA et al., 2016). Como resultado do processo de domesticação há uma grande divergência para características quantitativas morfoagronômicas, incluindo ciclo, tamanho e qualidade da semente, entre outras, e essas variações têm sido extensivamente utilizadas em programas de melhoramento (PÉREZ-VEGA et al., 2010).

O sistema de reprodução do feijão é predominantemente por autofecundação, em decorrência da estrutura da flor, na qual os estames e estigmas são bem protegidos pelas pétalas e ainda pelo fato de a deiscência das anteras ocorrer na fase de botão floral (VILHORDO; BURIN; GANDOLFI, 1988). O feijão é uma planta herbácea, de hábito de crescimento determinado ou indeterminado, quando determinado a planta é ereta, apresentando o eixo principal e os secundários sempre terminando em uma inflorescência, embora as flores também apareçam nas axilas das folhas, e quando indeterminado, o eixo principal nunca termina em inflorescência, as flores vão aparecendo na axila das folhas (VIEIRA, 1983). Uma das características mais importantes para classificar as variedades, do ponto de vista agronômico, é quanto ao hábito de crescimento, o Centro Internacional de Agricultura Tropical (CIAT), classifica o feijão em quatro categorias: I) de crescimento determinado, II) indeterminado de hastes curtas, III) indeterminado com hastes longas e IV) indeterminado trepador (VOYSEST, 2000).

Cultivares de porte ereto são desejadas para o cultivo do feijão, sendo assim grande parte das cultivares que atualmente estão em uso na agricultura possuem essa característica, sendo muitas delas de hábito indeterminado, porém de hastes curtas e em alguns casos hastes longas. Cultivares de hábito determinado também vêm sendo desenvolvidas e além de apresentarem

excelentes características agronômicas, como alto rendimento, facilitam a colheita mecanizada da cultura.

Os desafios enfrentados pela cultura do feijão são moldados pela sua história evolucionária, principalmente em ambientes da América Central e do Sul e também pelo conjunto de novos sistemas agroecológicos em que a cultura vem sendo implementada (ASSEFA et al., 2019). O feijão possui alta variabilidade morfológica, sendo uma cultura de ampla adaptação a diversos modos de cultivo e ambientes, podendo ser destinada a diferentes usos. O cultivo pode ser realizado em diversas regiões, localizadas desde o nível do mar até 3.000 metros de altitude, podendo ser usado em monocultivos, em associação ou em rotação (BROUGHTON et al., 2003). O ciclo (61 a 110 dias) e as características da planta favorecem os diferentes usos, sendo indicado desde sistemas agrícolas intensivos irrigados, altamente tecnificados, até aqueles com baixo uso tecnológico, como de subsistência (AIDAR, 2007).

Globalmente são produzidas em torno de 24 milhões de toneladas de grãos de feijão por ano (RAWAL; NAVARRO, 2019). O Brasil é o segundo maior produtor mundial, atrás apenas da Índia e seguido por Mianmar, China, Estados Unidos e México (FAO, 2020). Entre os principais países produtores de feijão, a produção do Brasil e do México é toda destinada para consumo nacional, enquanto os Estados Unidos, Canadá, Argentina e China são países exportadores. Em vários países em desenvolvimento da América Central, região Andina da América do Sul e nas regiões Sul e Leste da África, o feijão é cultivado para agricultura de subsistência e para mercados regionais, desempenhado um papel importante na segurança alimentar e geração de renda (BLAIR, 2013).

O cultivo do feijão no Brasil é realizado em três safras na maioria dos estados brasileiros, proporcionando constante oferta do produto no mercado ao longo de todo o ano. As safras são conhecidas como safra das águas (1ª safra), safra da seca (2ª safra) e safra de inverno (3ª safra), sendo responsáveis por 33, 43 e 24% da produção total, respectivamente, na safra de 2018/2019 (CONAB, 2020). Na safra de 2018/2019, a produção brasileira foi de 3.017,7 mil toneladas com uma produtividade média de 1033 Kg ha^{-1}. Os principais estados produtores são Paraná, Minas Gerais, Goiás, Mato Grosso, São Paulo e Bahia.

Em diferentes partes do mundo há uma preferência de consumo em relação a grãos do tipo Andino e Mesoamericano. Grãos do tipo Mesoamericano são mais produzidos na América do Norte, América Central e em partes da América do Sul, enquanto que feijões do tipo Andino são mais consumidos em partes da África, Europa e partes da América do Sul (CICHY et al., 2009; OLADZAD et al., 2019). No Brasil há uma preferência por feijões do tipo Mesoamericano, e entre eles os mais consumidos são os pertencentes aos grupos comerciais

carioca e preto (PERSEGUINI et al., 2015; RIBEIRO et al., 2014). O feijão carioca é o mais produzido no Brasil, sendo responsável por cerca de 70% da produção nacional, enquanto o preto representa cerca de 15% da produção (MAPA, 2018).

2.2 DIVERSIDADE E MELHORAMENTO GENÉTICO DA CULTURA DO FEIJÃO

O feijão é cultivado principalmente por pequenos e médios produtores, caracterizados pela agricultura de baixa tecnologia, o que leva a uma maior vulnerabilidade a diversos estresses bióticos e abióticos. Enquanto que agricultores que utilizam de alta tecnologia tem maior controle sobre esses fatores, porém, independente do sistema, os estresses bióticos e abióticos são os principais fatores de redução da produtividade (MIKLAS et al., 2006).

O melhoramento genético de feijão tem como principal objetivo a melhoria em relação a estresses bióticos e abióticos, combinada com a necessidade de manter características particulares relacionadas a qualidade e classe comercial, essenciais para atender a preferência dos consumidores (ASSEFA et al., 2019). Além disso, outras características são foco do melhoramento de feijão, variando de acordo com a localização mundial, ou seja, de acordo com as necessidades de cada país, algumas delas são: conteúdo mineral (ferro e zinco), tempo de cozimento, qualidade para enlatamento, índice de colheita e classe comercial ou cor da semente. Mundialmente, o principal objetivo em comum dos programas de melhoramento é o incremento da produtividade, que é um grande desafio para a cultura já que é implementada em diversos sistemas de cultivo, em sua maior parte com recursos muito limitados (BROUGHTON et al., 2003).

O Brasil, por ser um dos principais produtores mundiais de feijão comum, possui diversas instituições que trabalham no desenvolvimento de novas cultivares de feijão, sendo a maioria dessas instituições empresas públicas. As principais instituições públicas são Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Instituto Agronômico de Campinas (IAC), Instituto de Desenvolvimento Rural do Paraná – IAPAR-EMATER (IDR-Paraná), Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG), Empresa de Pesquisa Agropecuária de Santa Catarina (EPAGRI), e também universidades como a Universidade Federal de Lavras (UFLA) e a de Viçosa (UFV). Algumas empresas privadas de melhoramento de feijão são FT-Pesquisa e Sementes, Terra Alta Agropecuária (TAA) e Sementes Agronorte.

Estudos de diversidade genética são de grande importância para programas de melhoramento genético, fornecendo informações valiosas para conservação e utilização efetiva do germoplasma disponível (CHENG et al., 2016). Esses estudos podem ter diversas

finalidades, como entender as relações de parentesco genético entre os acessos, identificar redundâncias e misturas no germoplasma e identificar pares de genitores com distância genética adequada de acordo com o objetivo do programa de melhoramento.

Há diversas maneira de explorar a variabilidade genética presente em uma espécie, as principais são: por meio de caracteristicas morfológicas, agronômicas e moleculares. Em relação as características morfológicas e agronômicas, essas podem variar de acordo com o objetivo do programa de melhoramento. Já em relação a marcadores moleculares, a eficiência do melhoramento de plantas teve grande avanço com a emergência dos mesmos. Alguns dos mais utilizados são: polimorfismos de DNA amplificados ao acaso (*Random Amplified Polymorphic DNA- RAPD*), polimorfismo de comprimento de fragmento amplificado (*Amplified Fragment Length Polymorphism – AFLP*), sequências simples repetidas (*Simple Sequence Repeats – SSR*) e polimorfismos de uma única base (*Single-Nucleotide Polymorphism - SNP*) (ZHANG et al., 2018b). Podendo ser acessado de diversas formas, uma extensiva fenotipagem e caracterização genética são necessárias para o descobrimento do potencial genético da cultura do feijão (RAGGI et al., 2019).

2.2.1 Aspectos nutricionais e a Biofortificação

O feijão é reconhecidamente, uma excelente fonte proteica, possui bom conteúdo de carboidratos, além de ser uma importante fonte de micronutrientes, como por exemplo, ferro, zinco, tiamina e ácido fólico (BLAIR, 2013a; PETRY et al., 2015). Considerando um consumo anual de 15 kg *per capita*, supre cerca de 12% da necessidade de um adulto em proteínas, 4% em amido, 8% em cálcio, 20% em fósforo, 16% em magnésio, 27% em ferro, 8% em zinco, 12% em cobre, 18% em manganês, e 12% em potássio, além de possuir altas quantidades de fibras e carboidratos (BROUGHTON et al., 2003).

Sabendo-se do hábito da população brasileira de consumir diariamente o feijão, a utilização de alimentos biofortificados torna-se uma alternativa promissora, pois não altera a rotina e não requer custos adicionais para administração da dieta. A biofortificação é o processo de melhoria do conteúdo de nutrientes em uma cultura, sendo considerada uma estratégia sustentável e econômica para enfrentar a desnutrição nos países em desenvolvimento, porque visa alimentos básicos que são consumidos diariamente (DWIVEDI et al., 2012).

A biofortificação pode ser realizada por meio do melhoramento genético de plantas (métodos convencionais ou engenharia genética) ou por meio de práticas agronômicas (biofortificação agronômica). A biofortificação agronômica é o processo de melhoria do teor

de micronutrientes em partes comestíveis das plantas por meio de aplicação de fertilizantes no solo ou nas folhas. No entanto, a biofortificação genética possui melhor relação custo-benefício à longo prazo (VALENÇA et al., 2017).

As principais culturas básicas que foram, até o momento, utilizadas na biofortificações de minerais à escala internacional, são, principalmente as culturas de grãos, como arroz, trigo, milho e feijão, que fazem parte do programa internacional *Harvest Plus challenge*, que tem como foco a melhoria de plantas em três micronutrientes, ferro, zinco e vitamina A (PFEIFFER; MCCLAFFERTY, 2007). Diversos estudos têm demonstrado a eficácia do melhoramento convencional no desenvolvimento de culturas biofortificadas, melhorando a ingestão e minimizando a deficiência de micronutrientes entre as populações alvo. Alguns exemplos são, incremento de vitamina-A nas culturas da batata-doce (LOW et al., 2007; VAN JAARSVELD et al., 2005), milho (GANNON et al., 2014) e mandioca (TALSMA et al., 2016), e de ferro para feijão (GLAHN et al., 2017; HAAS et al., 2016) e milheto (BEER et al., 2014; FINKELSTEIN et al., 2015). Cumulativamente, mais de 150 variedades biofortificadas de dez culturas têm sido oficialmente lançadas para produção em mais de 30 países da África, Ásia e América Latina e Caribe (BOUIS; SALTZMAN, 2017).

Levando-se em conta a grande importância do feijão na alimentação humana, a biofortificação tem sido implementada nos programas de melhoramento em diversas partes do mundo. Explorar a variabilidade existente na espécie é uma das etapas para se obter cultivares melhoradas. Para a cultura do feijão diversos estudos relacionados com a biofortificação já foram realizados em diferentes países, como por exemplo Brasil (DELFINI et al., 2020; RIBEIRO et al., 2014; SILVA et al., 2012), Colômbia (BEEBE; GONZALEZ; RENGIFO, 2000; ISLAM et al., 2002), Portugal (PINHEIRO et al., 2010) e também nos Estados Unidos (MCCLEAN et al., 2017).

Uma grande variabilidade é observada para concentração dos diferentes nutrientes no germoplasma de feijão, entre plantas silvestres e cultivadas, sendo assim estudos relacionados a essa variabilidade, são de grande contribuição para o melhoramento da cultura (RIBEIRO et al., 2013). A maioria dos trabalhos de biofortificação mineral tem sido efetuados por meio do melhoramento convencional com algumas tentativas de tecnologia transgênica (BLAIR, 2013a). O melhoramento direcionado ao aumento do conteúdo de micronutrientes requer melhorar concomitantemente uma série de características essenciais, para que as novas cultivares sejam aceitas pelos produtores, já que os níveis elevados de ferro e zinco não são visíveis (PFEIFFER; MCCLAFFERTY, 2007). A nova cultivar deve exibir simultaneamente alto rendimento agronômico, resistência a agentes patogênicos e outros estresses ambientais,

sendo tão lucrativa quanto as cultivares atuais (PETRY et al., 2015).

2.2.2 CARACTERÍSTICAS AGRONÔMICAS RELACIONADAS A PRODUÇÃO

Alcançar a segurança alimentar é um dos desafios mais importantes a enfrentar nas próximas três décadas, já que é esperado um grande crescimento populacional que irá aumentar a demanda por alimentos, especialmente em países em desenvolvimento (JENSEN et al., 2012; RAGGI et al., 2019). Nesse contexto, as leguminosas são geralmente consideradas como essenciais para melhorar a segurança alimentar, pois são uma fonte relativamente barata de aminoácidos e outros nutrientes importantes quando comparadas com outros alimentos (JENSEN et al., 2012).

O feijão é uma leguminosa fixadora de nitrogênio que possui baixa necessidade de fertilização, é importante em sistemas de rotação, quebrando ciclo de pragas e doenças de outras culturas e melhora a estrutura do solo (PARKER; PALKOVIC; GEPTS, 2020). A produtividade média em países em desenvolvimento (1035 $kg.ha^{-1}$) ainda é muito inferior as médias atingidas em países desenvolvidos (1944 $kg.ha^{-1}$), o que demonstra que o uso de novas técnicas e uma melhor exploração do germoplasma podem melhorar o desempenho da cultura (GEPTS et al., 2008). Sendo assim, o incremento da produtividade é um dos principais objetivo nos programas de melhoramento da cultura. O entendimento da arquitetura genética relacionada a produtividade, bem como a interação entre os componentes da produção é a base para o melhoramento da produção de grãos de feijão (KAMFWA; CICHY; KELLY, 2015a).

A produtividade é uma característica quantitativa e está relacionada com muitas outras características morfológicas, agronômicas e fisiológicas. O número de vagens por planta, número de sementes por vagem e peso de sementes são os componentes primários relacionados a produtividade, entretanto, outros fatores também influenciam a produtividade, como taxa de crescimento, capacidade das sementes em absorver fotossintatos, arquitetura da planta, entre outras características (ASSEFA et al., 2019; BEEBE et al., 2013; RESENDE et al., 2018a).

Características como dias para florescimento e maturidade e altura de planta também são características complexas e controladas por diversos fatores, e, impactam consideravelmente a adaptabilidade, biomassa e a produtividade em culturas agrícolas (ZHANG et al., 2015). A melhoria da arquitetura da planta e da tolerância ao acamamento por sua vez, afetam diretamente as perdas no momento da colheita, a ocorrência de doenças, o manejo da cultura e a colheita mecanizada (TEIXEIRA; RAMALHO; ABREU, 1999).

Produtividade e seus componentes são características quantitativas e altamente influenciadas pelo ambiente, sendo assim, entender a relação entre essas características é muito importante para poder direcionar as estratégias e esforços em programa de melhoramento genético (ASSEFA et al., 2019; SINGH; NODARI; GEPTS, 1991). Os métodos tradicionais de seleção no melhoramento genético requerem um trabalho intensivo de fenotipagem a campo, sendo necessário avaliações em vários ambientes e anos, e consequentemente possuem alto custo e consumo de tempo (IKRAM et al., 2020).

Por meio de marcadores moleculares, é possível identificar regiões genômicas que contribuem para a produtividade e seus componentes, o que os tornam uma ferramenta importante para auxiliar na seleção assistida por marcadores (KAMFWA; CICHY; KELLY, 2015a). Estudos de associação genômica ampla (*Genome Wide Association Studies – GWAS*) representam uma opção poderosa para a caracterização genética de características quantitativas, e têm sido amplamente utilizados em análises de características agronômicas em plantas (CUI; ZHANG; ZHOU, 2018).

2.3 GENOTIPAGEM POR SEQUENCIAMENTO (GBS)

As tecnologias de sequenciamento de próxima geração (*Next-generation sequencing - NGS*) estão revolucionando estudos genéticos e o desenvolvimento de marcadores moleculares, aumentando exponencialmente o número de variantes genéticas que podem ser descobertas em um único experimento (STAPLEY et al., 2010). A tecnologia de *NGS* tem permitido que a técnica de genotipagem por sequenciamento (*genotyping-by-sequencing - GBS*) torne-se viável para estudos de espécies com genoma amplo e de alta diversidade (ELSHIRE et al., 2011). Entre as *NGS* a técnica de *GBS* surgiu como uma nova abordagem para mitigar as restrições dos marcadores mais antigos (SIADJEU; MAYLAND-QUELLHORST; ALBACH, 2018). Essa técnica é baseada no uso de enzimas para redução da complexidade do genoma que em seguida é acoplado a adaptadores com código de barras (*barcodes*) para produzir bibliotecas multiplex de amostras prontas para sequenciamento *NGS* (POLAND et al., 2012).

A GBS é uma técnica robusta, de alto rendimento, econômica e simples para obter milhares de marcadores de um grande número de indivíduos, no qual captura dados de polimorfismos únicos (*Single nucleotide polymorphism - SNP*) utilizando uma biblioteca de representação reduzidas (ARIANI; BERNY MIER Y TERAN; GEPTS, 2016; SCHRÖDER et al., 2016). *SNPs* são as variações de sequência mais abundantes e universais em todos os

genomas, tornando-os marcadores de grande utilidade para análises genéticas em plantas (WANG et al., 2015).

Essa técnica é simples, rápida, extremamente específica, reproduzível e pode alcançar regiões importantes do genoma inacessíveis por outras técnicas de captura de sequenciamento. A correta escolha da enzima de restrição pode auxiliar na redução de regiões repetitivas do genoma, o que simplifica muito os problemas computacionais de alinhamento em espécies com altos níveis de diversidade genética (HE et al., 2014).

A GBS tem sido utilizado para o sequenciamento parcial do genoma de espécies e também para projetos de re-sequenciamento, onde o genoma de vários indivíduos da mesma espécies são sequenciados para descobrir um grande número de *SNPs* com a finalidade de explorar a diversidade dentro da espécie, construção de mapas e estudos de associação genômica ampla (*Genome-wide association - GWAS*) (ELSHIRE et al., 2011).

Outra vantagem é que os dados brutos obtidos por meio de GBS são dinâmicos, pois podem ser reanalisados descobrindo novas informações (ex. novos polimorfismos, genes anotados, etc.) à medida que as técnicas de bioinformática, genomas de referência e coleta de dados de sequenciamento avançam (POLAND et al., 2012).

A partir da conclusão do sequenciamento completo do genoma do feijão por Schmutz et al. (2014) tornou-se mais fácil a descoberta de *SNPs* e mapeamento, permitindo que a partir de leituras curtas de diferentes genótipos sejam montados mapas usando a sequência do genoma como um modelo, sendo mais confiável do que a montagem *de novo* e os *SNPs* podem ser automaticamente identificados em relação à sequência de referência (GUJARIA-VERMA et al., 2016).

Uma das aplicações mais utilizadas e eficiente da GBS é no melhoramento de plantas. É uma técnica rápida e de baixo custo que pode ser utilizada para genotipar populações em larga escala provenientes do melhoramento, e possibilita a implementação de muitos estudos como por exemplo GWAS, estudos de diversidade genética, análises de ligação, descobrimento de marcadores moleculares e seleção genômica (HE et al., 2014). O principal objetivo da técnica de GBS, portanto, não é apenas descobrir polimorfismos, mas descobrir simultaneamente polimorfismos e obter informações genotípicas da população de interesse (POLAND et al., 2012).

2.4 Estudos de Associação genômica ampla (GWAS)

Mapas genéticos fornecem informações úteis para diversos estudos, como por exemplo

a localização de regiões genômicas que controlam características fenotípicas, bem como podem ser utilizados para estudar os efeitos de caracteres quantitativos (CAMPOS et al., 2011). Os dois métodos mais utilizados para entender características complexas são os mapeamentos por análise de ligação e o mapeamento associativo (ZHU et al., 2008).

O mapeamento por análise de ligação, inicialmente, tinha como objetivo medir a proximidade genética dos loci entre si, para mapear características quantitativas, e mais tarde tornou-se possível realizar o mapeamento de *QTLs* (*Quantitative Trait Loci*). A identificação de locos que governam características complexas tem sido facilitada pelas novas técnicas de mapeamento de QTLs, no qual, convencionalmente, são utilizadas populações biparentais segregantes, ou seja, populações experimentais altamente estruturadas com pedigrees conhecidos, como por exemplo populações F_2 (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003). Melhoristas e geneticistas moleculares tem rotineiramente utilizado populações derivadas de cruzamentos biparentais para desenvolvimento de novas variedades e mapeamento de *QTLs* para características de interesse, entretanto a riqueza alélica e variação fenotípica em populações biparentais é um pouco limitada (ZHANG et al., 2018b).

Apesar de estudos de mapeamento de *QTLs* serem muito utilizados, o conhecimento dos genes que governam importantes características agronômicas se torna restrito, pois normalmente, os intervalos estimados se estendem por vários *cM*, que pode conter vários genes candidatos (SONAH et al., 2015). A análise de ligação, em plantas, geralmente localiza *QTLs* para intervalos de 10 a 20 *cM*, devido ao número limitado de eventos de recombinação que ocorrem durante a construção de populações para o mapeamento (DOERGE, 2002).

O mapeamento associativo, também conhecido como mapeamento por desequilíbrio de ligação (*Linkage Desequilibrium - LD*), surgiu como uma ferramenta para resolver a variação de características complexas a nível de sequência de DNA, a partir da exploração da recombinação de eventos históricos e evolucionários ao nível de população (NORDBORG et al., 2002). Em contraste com a análise de ligação tradicional, amplamente utilizada em plantas, o mapeamento associativo busca a variação funcional em um contexto de germoplasma muito mais amplo, permitindo que os pesquisadores usem tecnologias genômicas modernas para explorar a diversidade natural (ZHU et al., 2008).

Com base na escala e no foco do estudo, o mapeamento associativo, geralmente leva a duas categorias amplas, (i) mapeamento associativo de genes candidatos, que relaciona polimorfismos em genes selecionados, que são candidatos à controlar variáveis fenotípicas em características específicas; e (ii) mapeamento de associação genômica ampla, ou varredura do genoma, que examina a variação genética em todo o genoma e sinais de associação para várias

características complexas (RISCH; MERIKANGAS, 1996).

Os estudos de associação genômica ampla (*Genome-wide association - GWAS*) rapidamente se tornaram uma ferramenta poderosa para a identificação de regiões candidatas associadas a características quantitativas. O GWAS é uma extensão do mapeamento QTL, visa encontrar associações entre características e marcadores em indivíduos não relacionados e só são detectadas quando o marcador e QTL estão em forte desequilíbrio de ligação (LD) (STAPLEY et al., 2010). Em comparação ao mapeamento por ligação, na técnica de GWAS o desequilíbrio de ligação entre as marcas vizinhas é muito menor, pelo fato da utilização de populações de genótipos não aparentados, dessa forma é necessário um número muito maior de marcas genéticas, para que seja possível uma cobertura completa do genoma (HYTEN et al., 2007; SONAH et al., 2015).

Alguns pontos devem ser considerados para realização do mapeamento por associação: (1) seleção de uma coleção de germoplasma com alto nível de diversidade genética; (2) fenotipagem do germoplasma selecionado; (3) genotipagem dos indivíduos com alta densidade de marcadores; (4) quantificação da extensão do LD no genoma da população escolhida; (5) avaliação do nível de diferenciação genética entre grupos dentro dos indivíduos amostrados e o coeficiente de parentesco entre pares de indivíduos dentro de uma amostra; e (6) levar em consideração as informações obtidas através da quantificação de LD e da estrutura da população para a correlação de dados fenotípicos e genotípicos com a aplicação de uma abordagem estatística apropriada que revela "marcadores" posicionados nas proximidades da característica de interesse (ABDURAKHMONOV; ABDUKARIMOV, 2008).

A partir da publicação completa do genoma feijão (SCHMUTZ et al., 2014), e com as novas tecnologias de sequenciamento, ficou facilitada a implantação de novas tecnologias no melhoramento da cultura. Esses acontecimentos permitem que sejam obtidos conjuntos densos de marcadores *SNP* no genoma, que permitem uma exploração mais precisa das regiões genômicas responsáveis por características particulares em feijão (HOYOS-VILLEGAS; SONG; KELLY, 2017). O GWAS realizado com alta densidade de marcadores e populações de genótipos não recombinantes permite uma resolução maior de mapeamento em relação a tecnicas convencionais de mapeamento de *QTLs* baseadas em populações recombinantes, além de permitir predizer ou identificar o gene causal (ZHANG et al., 2015).

Até o momento existem três principais estratégias para realizar GWAS. A primeira delas é um modelo linear generalizado (*Generalized Linear Model – GLM*), que foi proposto para a análise genética de caracteristicas quantitativas no qual a estrutura populacional é ajustada como efeito fixo (PRICE et al., 2006), entretanto esse modelo não conseguiu controlar o efeito

poligênico. Em seguida um modelo linear misto (*Mixed Linear Model – MLM*) foi desenvolvido para levar em conta a estrutura populacional e o efeito poligênico usando a relação de parentesco (*kinship*) como efeito aleatório (YU et al., 2006; ZHANG et al., 2005). GLM e MLM são métodos conhecidos como *single-locus* por realizarem varreduras unidimensionais no genoma, ou seja, testam um marcador de cada vez, usando várias correções rigorosas de teste de significância para múltiplo-testes como por exemplo Bonferroni, sendo assim esses métodos tem um poder relativamente baixo para detectar poligenes de pequeno efeito que são resposáveis pela maioria das características quantitativas (HE et al., 2019a). Embora esses métodos tenham sido amplamente utilizados, características complexas governadas por múltiplos genes podem não ser eficientemente identificadas (LI et al., 2018b).

Métodos alternativos *multi-locus* têm sido propostos para solucionar esse problema, com a grande vantagem de que não é utilizada a correção de Bonferroni, alguns desses métodos são mrMLM que atribui efeito aleatório aos SNPs (WANG et al., 2016), FASTmrMLM uma versão mais rápida e eficiente do método mrMLM (TAMBA; ZHANG, 2018), pLARmEB o qual integra *least angle regression* com Bayes empírico (ZHANG et al., 2017a), e ISIS EM-BLASSO que utiliza abordagem Bayesiana para identificação de associações (TAMBA; NI; ZHANG, 2017). Esses métodos diferem dos outros métodos multi-locus pelo fato de possuírem duas etapas, a primeira etapa considera o efeito do SNP como aleatório e todos os marcadores potencialmente associados são selecionados por um modelo MLM de efeito aleatório do SNP (random- SNP-effect MLM) com uma correção modificada de Bonferroni para o teste de significância, e na segunda etapa todos os marcadores são colocados em um único modelo e todos os efeitos diferentes de zero são detectados por um teste de razão de verossimilhança para identificação de QTNs (CHANG et al., 2018).

Muitos GWAS já foram realizados em genótipos de feijão para diversas características, como por exemplo características morfológicas, fisiológicas, bem como em resposta a estresses bióticos e abióticos. Em relação a fatores bióticos, marcas relacionadas a resistência a diversas doenças já foram encontradas. Algumas doenças já estudadas foram crestamento-bacteriano-comum (SHI et al., 2011), antracnose e mancha angular (PERSEGUINI et al., 2016), mosaico dourado do feijoeiro (HART; GRIFFITHS, 2015) e resistência de feijões andinos a antracnose (ZUIDERVEEN et al., 2016). Enquanto para características abióticas Hoyos-Villegas; Song; Kelly (2017) estudaram 96 genótipos de feijão-comum e encontraram 27 marcadores associados a características relacionadas a tolerância a seca a partir de dados *SNP*.

Alguns estudos para características agronômicas já foram realizados, sendo as principais: fibra na vagem, sementes por vagem, tipo de planta, hábito de crescimento, dias para

floração e maturação, biomassa e características de sementes (KAMFWA; CICHY; KELLY, 2015a; MACQUEEN et al., 2020; NEMLI et al., 2014; OLADZAD et al., 2019; RESENDE et al., 2018a; SOLTANI et al., 2016). Outras características já exploradas são: fixação simbiótica do nitrogênio (KAMFWA; CICHY; KELLY, 2015b) e tempo de cozimento (CICHY; WIESINGER; MENDOZA, 2015). Características nutricionais do feijoeiro ainda foram pouco exploradas nesse sentido. Katuuramu et al. (2018) realizou GWAS para características relacionadas a composição nutricional em feijão já cozido.

3 ARTIGO A - Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm

3.1 ABSTRACT

Brazil is the world's largest producer of common bean. Knowledge of the genetic diversity and relatedness of accessions adapted to Brazilian conditions is of great importance for the conservation of germplasm and for directing breeding programs aimed at the development of new cultivars. In this context, the objective of this study was to analyze the genetic diversity, population structure, and linkage disequilibrium (LD) of a diversity panel consisting of 219 common bean accessions, most of which belonging to the Mesoamerican gene pool. Genotyping by sequencing (GBS) of these accessions allowed the identification of 49,817 SNPs with minor allele frequency > 0.05. Of these, 17,149 and 12,876 were exclusive to the Mesoamerican and Andean pools, respectively, and 11,805 SNPs could differentiate the two gene pools. Further the separation according to the gene pool, Bayesian analysis of the population structure showed a subdivision of the Mesoamerican accessions based on the origin and color of the seed tegument. LD analysis revealed the occurrence of long linkage blocks and low LD decay with physical distance between SNPs (LD half decay in 249 Kb, corrected for population structure and relatedness). The GBS technique could effectively characterize the Brazilian common bean germplasms, and the diversity panel used in this study may be of great use in future genome-wide association studies.

Keywords: *Phaseolus vulgaris* L.; GBS; linkage disequilibrium

3.2 INTRODUCTION

The common bean (*Phaseolus vulgaris* L.) is one of the five cultivated species of the *Phaseolus* genus and is one of the most consumed legumes worldwide. It is the most important legume grain for direct human consumption and the main source of protein and micronutrients in several countries (BROUGHTON et al., 2003). Globally, around 31 million tons of bean grains are produced per year, with the Americas accounting for 32.4% of the total production. Brazil is the world's largest producer of common bean, and other countries that are among the largest producers are India, Myanmar, China, United States, and Mexico (FAO, 2020).

The common bean is known to have originated in Mexico and the Southern Andes, where it was domesticated independently to give rise to two gene pools, i.e., the Andean and Mesoamerican groups, which are morphologically and genetically different (BITOCCHI et al., 2013; KWAK; GEPTS, 2009; SCHMUTZ et al., 2014; VLASOVA et al., 2016). Different parts of the world prefer either the Andean or Mesoamerican grains. The Mesoamerican common beans are more common in North America, Central America, and the lowland part of South America, whereas the Andean common beans are preferred in parts of Africa, Europe, and Andean part of South America (CICHY et al., 2009; OLADZAD et al., 2019). In Brazil, Mesoamerican common beans are preferred, of which the carioca and black beans represent the most consumed commercial groups (PERSEGUINI et al., 2015; RIBEIRO et al., 2014). Carioca beans are the most widely produced in Brazil, accounting for approximately 70% of the national common bean production, whereas black beans represent about 15% of the total production (MAPA, 2018).

Genetic diversity studies are of great importance for breeding programs, as they provide valuable information for effective conservation and application of available germplasm (CHENG et al., 2016). Such studies facilitate the understanding of genetic relationships between accessions, identification of redundancies and admixtures in the germplasm, and determination of genitor pairs with adequate genetic distance.

Molecular markers have been widely used in plant breeding programs. Several different types of markers are available; however, their applications have been restricted in the past due to limitations such as low density, labor intensity, technical requirements, and high cost of large-scale analysis (BHATTARAI; SUBUDHI, 2018; HE et al., 2014). The advent of next-generation sequencing (NGS) technologies has resulted in an exponential increase in the number of genetic variants that can be discovered in a single experiment (STAPLEY et al., 2010). The publication of the complete genome sequence of the common bean (SCHMUTZ et

al., 2014) facilitated the discovery of single nucleotide polymorphisms (SNPs) and genetic mapping, further allowing the construction of maps from short reads of different genotypes using the genome sequence as a reference (GUJARIA-VERMA et al., 2016).

Among the NGS methods, the genotyping by sequencing (GBS) technique has emerged as a new approach to mitigate the constraints of previously employed markers (SIADJEU; MAYLAND-QUELLHORST; ALBACH, 2018). GBS is a robust, high-performance, cost-effective, and simple technique for obtaining thousands of markers from a large number of individuals, and allows the identification of SNPs using a reduced representation library (ARIANI; BERNY MIER Y TERAN; GEPTS, 2016; ELSHIRE et al., 2011; SCHRÖDER et al., 2016). SNPs are the most abundant and universal sequence variations in all genomes, which makes them very useful markers for genetic analyses in plants (WANG et al., 2015).

The GBS technique is often employed in plant breeding, and is frequently used in genetic diversity studies, mapping (linkage and association) studies, and genomic selection (GS) (HE et al., 2014). Genome-wide association studies (GWAS) are a powerful tool for identifying candidate genomic regions associated with traits of interest. Some of the most important parameters for successful GWAS are the representativity of the diversity panel, the size of the panel, the levels and genomic distribution of linkage disequilibrium (LD), and the population structure or genetic relationships among individuals (BURGHARDT; YOUNG; TIFFIN, 2017; KORTE; ASHLEY, 2013; ZHANG et al., 2010). The diversity panel should represent most of the available genetic and phenotypic diversity, and LD should be analyzed to determine the density of markers required for GWAS (NICOLAS et al., 2016).

Studies on genetic diversity and population structure have already been conducted for several crops, including wheat (ALIPOUR et al., 2017; BHATTA et al., 2018; ELTAHER et al., 2018), flaxseed (LUO et al., 2019), pepper (PEREIRA-DIAS et al., 2019) and rice (XU et al., 2016). Several diversity panels have also been developed for the common bean crop, including accessions from different regions of the world (CAMPA; MURUBE; FERREIRA, 2018; CICHY et al., 2015; LIOI et al., 2019; MOGHADDAM et al., 2016; PERSEGUINI et al., 2016; RAGGI et al., 2019). Based on these initial studies, several GWAS have further been conducted for different traits of interest, such as yield, plant architecture, nutritional content of grains, cooking time, resistance to diseases, and tolerance to abiotic factors (CICHY; WIESINGER; MENDOZA, 2015; KAMFWA; CICHY; KELLY, 2015b, 2015a; KATUURAMU et al., 2018; MOGHADDAM et al., 2016, 2018; OLADZAD et al., 2019; PERSEGUINI et al., 2016; RESENDE et al., 2018a; SOLTANI et al., 2017, 2018; TOCK et al., 2017; ZUIDERVEEN et al., 2016). Some GWAS have been conducted in Brazil

(PERSEGUINI et al., 2016; RESENDE et al., 2018b; VALDISSER et al., 2020), however, panels consisting of different genotypes, can contribute to a better understanding about the genetic diversity and relationships of the germplasm available for genetic breeding.

In view of the above, the objective of the present study was to analyze the genetic diversity, population structure, and LD of the Brazilian Diversity Panel (BDP), which is a common bean diversity panel representing a large proportion of the genetic diversity of Brazilian common bean populations. It is composed mainly of materials from the carioca and black bean commercial groups, which are the most consumed cultivars in the country, and is expected to be used for GWAS in the future.

3.3 MATERIAL AND METHODS

3.3.1 PLANT MATERIAL

The BDP, including 230 common bean accessions that represent a large component of the common bean genetic diversity in Brazil, was used in this study (Table S3.1). The diversity panel is composed of modern and old cultivars developed between 1968 and 2019 by different research institutions (Table 3.2), in addition to inbred lines and landraces, all of which belong to the germplasm bank of the Rural Development Institute of Paraná –IAPAR–EMATER (IAPAR). Among the CIAT accessions present in this panel, most of them are inbred lines from breeding programs directed to the needs of Brazil and/or are accessions that compose the genealogy of cultivars developed by Brazilian institutions. Most accessions in this panel are of Mesoamerican origin, which exhibit significant diversity in the color of the seed tegument and include different commercial classes, with the carioca and black bean groups being the most representative. In addition, 10 accessions of Andean origin were included in this study for comparison.

3.3.2 GENOTYPING-BY-SEQUENCING (GBS)

DNA extraction and the preparation of GBS libraries for sequencing was performed following the protocol developed for common bean by Ariani, Berny Mier y Teran and Gepts (2016). DNA was extracted from lyophilized leaves collected from a single plant of each accession grown in a green house. The extracted DNA was purified using the Genomic DNA Clean and Concentrator kit (Zymo Research, CA, USA), according to the manufacturer's

instructions. The DNA quality was checked using NanoDrop Lite (Thermo Fisher Scientific), and only samples with an absorbance ratio (A260/A280) greater than 1.7 were used for preparing the libraries. Genomic DNA was quantified using Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific), and 100 ng of the DNA from each genotype was used for preparing the libraries.

The genomic DNA was digested using the restriction enzyme *CviAII* (recognition site C'ATG); after the preparation process, the samples were multiplexed into two libraries with up to 144 accessions each, including as control a blank sample and the genotype of *P. vulgaris* used to construct the reference genome (G19833) in each of the two libraries (ARIANI; MIER TERAN; GEPTS, 2017). The presence of adapter dimers in the sequencing libraries was checked using DNA High Sensitivity Kit (Agilent 2100 Bionalyzer, Agilent Technologies).

The genomic libraries were sequenced using the Illumina HiSeq 4000 sequencer (Illumina, San Diego, CA, USA) with the 100-bp-single-end protocol, at the DNA Technologies & Expression Analysis Core Laboratory, located in the Genome Center, University of California, Davis, CA.

3.3.3 ANALYSIS OF SEQUENCING DATA

SNPs were called using the Tassel-5-GBS pipeline version 2 (GLAUBITZ et al., 2014), with the standard software settings, except for the minimum quality score (-mnQs 20) and minimum count (-c 10) parameters. The obtained sequences were aligned with the reference genome of *Phaseolus vulgaris* v2.0 obtained from the Phytozome website (https://phytozome.jgi.doe.gov, accessed on March 10, 2019), using the Burrows-Wheeler Alignment (BWA) (-aln option) tool version 0.7.10 (LI; DURBIN, 2010). Non-biallelic SNPs, and SNPs with indels, minor allele frequency (MAF) < 0.05, coefficient of inbreeding < 0.9, and those SNPs and accessions containing < 10% of genotyped positions were removed using VCFtools version 0.1.15 (DANECEK et al., 2011). Because common bean is an autogamous species, after the initial filtering, the occurrence of heterozygotes was insignificant, but heterozygous SNPs were treated as missing data, as they may indicate sequencing errors. After filtering, the SNPs were imputed using Beagle software version 5 (BROWNING; ZHOU; BROWNING, 2018), and only SNPs anchored to chromosomes in the common bean reference genome were used.

The SNPs were annotated according to the common bean genomic annotation (GFF3 file, version 2.1) available on the Phytozome website (https://phytozome.jgi.doe.gov, accessed

on January 10, 2019), using a custom R (R CORE TEAM, 2020) script developed by Hu et al. (2019) (https://github.com/zhenbinHU/Sorghum_SNP_dataset, accessed on June 17, 2019).

3.3.4 GENETIC DIVERSITY AND POPULATION STRUCTURE

The 219 accessions belonging to the BDP that passed by the quality control mentioned above were included in the initial analyses. The population structure was inferred using the Bayesian clustering algorithm in Structure v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000) software from the command line python program StrAuto (CHHATRE; EMERSON, 2017). The admixture model with 50,000 burn-ins, 200,000 MCMC, and 10 replications for hypothetical numbers of subpopulations (K) between 1 and 10 was used. The statistical parameter ΔK (EVANNO; REGNAUT; GOUDET, 2005) was used to determine the number of groups. Only the accessions with a membership coefficient equal or higher than 0.6 were assigned to a genetic group, and those with membership coefficient lower than 0.6 were clustered in the admixture group. The admixture model assumes that the markers are not strongly linked; hence, the SNPs were filtered based on LD, using the indep-pairwise option of the PLINK (PURCELL et al., 2007) software, and only SNPs with LD ≤ 0.2 were retained for population structure analysis. These data filtered for LD were also used for PCA, using the snpgdsPCA function of the SNPRelate (ZHENG et al., 2012) package in R.

After verifying the center of origin, only individuals of Mesoamerican origin were retained, and the SNPs were again filtered to exclude monomorphics, SNPs with MAF < 0.05 and LD ≥ 0.2, using VCFtools version 0.1.15 (DANECEK et al., 2011) and PLINK (PURCELL et al., 2007). These data were then used for PCA and population structure analysis, as previously described. In addition, phylogenetic inference was estimated using TASSEL v5 (BRADBURY et al., 2007), based on identity-by-state (IBS) distance and using Neighbor-Joining as the clustering method. The generated tree was customized using FigTree v1.4.4 (RAMBAUT, 2018).

To detect molecular differences in relation to the center of origin, color of the seed tegument, and institution of origin, new files were created from the initial file (including all SNPs) containing the different groups, and only polymorphic SNPs and those with MAF > 0.05 were retained. Subsequently, a Venn diagram was constructed to detect the differentiating SNPs for each of the three parameters using the JVENN tool (BARDOU et al., 2014). *Fst* index (WEIR; COCKERHAM, 1984), nucleotide diversity (π) and Tajima's D (TAJIMA, 1989),

were also calculated using VCFtools version 0.1.15 (DANECEK et al., 2011) and averaged on 100-kb genomic bins.

3.3.5 LINKAGE DISEQUILIBRIUM

LD between SNPs was estimated using the LDcorSV (DESROUSSEAUX et al., 2017) package in R. This package corrects for the bias due to population structure and relatedness while estimating LD. In addition to the conventional r^2, r^2 corrected for population structure (r^2_s), r^2 considering kinship (r^2_v), and r^2 including both population structure and kinship (r^2_{vs}) were calculated. Only individuals belonging to the Mesoamerican group were used for these calculations. The STRUCTURE result at K=2 for common beans of Mesoamerican origin was used as the population structure, and for relatedness the kinship matrix was calculated using the rrBLUP (ENDELMAN, 2011) package in R. LD decay was calculated using the nonlinear method proposed by Hill and Weir (HILL; WEIR, 1988), and adjusted with the nls function in R.

3.4 RESULTS

3.4.1 GENOTYPING BY SEQUENCING

Using the GBS method optimized for common beans (ARIANI; BERNY MIER Y TERAN; GEPTS, 2016), a total of 392,585,199 good barcoded reads were obtained from the sequenced accessions, of which 364,454,550 could be aligned with the Andean reference genome (G19833 (SCHMUTZ et al., 2014)), resulting in an average mapping rate of 93%. Initially, 461,199 SNPs were obtained, of which 49,817 SNPs were retained after filtering. Eleven accessions had a low rate of genotyping (less than 10% of genotyped positions) and were excluded from the BDP, for this reason 219 accessions were used in the subsequent analyses.

SNPs were unevenly distributed throughout the genome, and fewer SNPs were observed in regions near the centromere than in regions near the telomeres on the chromosome (Figure 3.1a). The mean number of SNPs per chromosome was 4,528, ranging from 3,361 to 5,910 SNPs on the Pv06 and Pv02 chromosomes, respectively (Table 3.1). Physical chromosome length was positively correlated with the number of SNPs ($r = 0.74$, $p < 0.01$).

Table 3.1 Number of SNPs in each of the 11 common bean chromosomes in the set of 219 accessions from the Brazilian Diversity Panel.

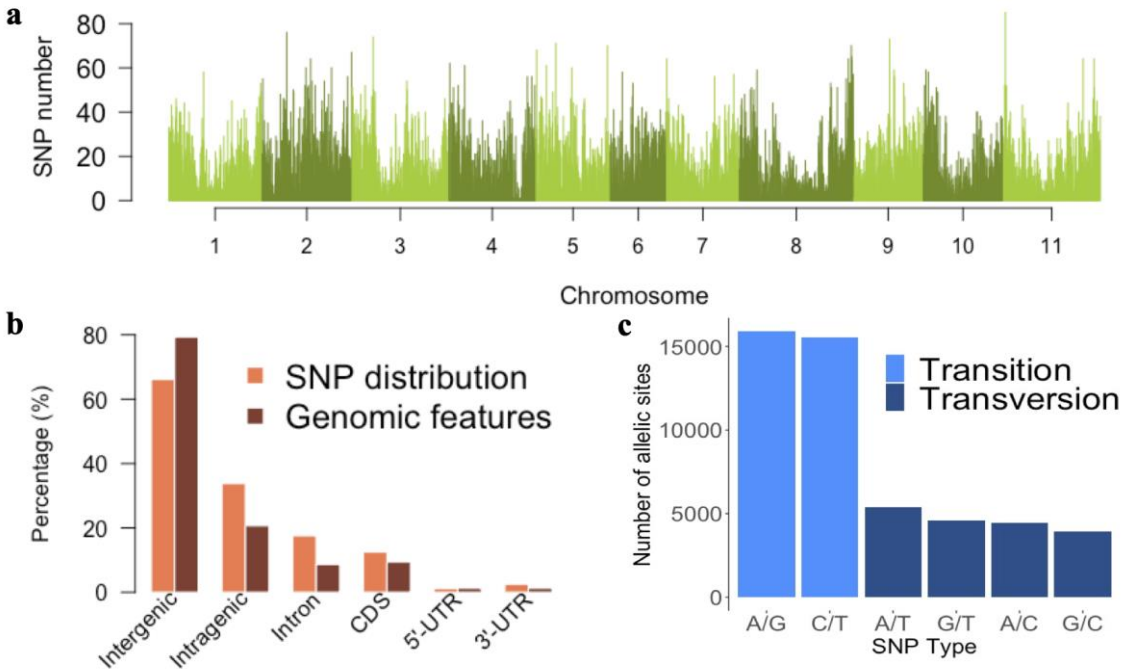
Chromosome	Physical Length (Mb)[1]	Number of Genes[1]	Total number of SNPs	Tagged genes
Pv01	52.20	2779	4718	787
Pv02	49.04	3435	5910	1111
Pv03	52.28	3058	4983	874
Pv04	45.96	1890	4650	596
Pv05	40.82	1928	4321	529
Pv06	31.97	2295	3361	721
Pv07	51.75	2895	3903	853
Pv08	59.66	3023	5635	914
Pv09	37.47	2719	3735	716
Pv10	43.27	1721	3939	512
Pv11	50.37	2253	4662	653
Total	-	27996	49817	8266

[1]Information obtained from the *EnsemblPlants* website (https://plants.ensembl.org)

Of the total SNPs obtained, 33.8% were located in intragenic regions (17.5% in intron and 16.3% in exons), 12.5% in coding DNA sequences, 1.2% in 5' UTR regions, and 2.6% in 3' UTR regions (Figure 3.1b). Thirty percent of the annotated genes in the reference genome of *Phaseolus vulgaris* v2.0 were tagged by at least one SNP (tagged genes) (Table 3.1). A positive correlation between the number of genes and tagged genes per chromosome was observed ($r = 0.96$, $p < 0.01$).

Of the different types of polymorphism, transitions (63.1%) were more frequent than transversions (36.9%), resulting in a transition/transversion rate of 1.71 (Figure 3.1c). The percentages of A/G and C/T transitions were very similar (32% and 31%, respectively), as were those of polymorphism due to A/T, A/C, G/T, and G/C transversions (11%, 9%, 9%, and 8%, respectively). Considering only the SNPs inside genes the transition/transversion rate was 1.27, smaller compared to the overall rate, for the reason that the percentage of transversions (44.1%) was greater than the overall.

Figure 3.1 Identification and annotation of 49,817 single nucleotide polymorphisms (SNPs) obtained from the genotyping of 219 common bean accessions. (a) Distribution of SNP density along the common bean genome in a 200 Kb sliding window. (b) Annotation of SNPs and proportion of genomic traits. (c) Transversion/transition ratio.



3.4.2 GENETIC DIVERSITY AND POPULATION STRUCTURE

The population structure of all accessions included in the BDP were analyzed using 819 SNPs that were retained after LD filtering ($r^2<0.2$). The results of the principal component analysis (PCA) showed that the accessions could be segregated into two distinct groups, based on the gene pools (Andean and Mesoamerican) (Figure 3.2a).

The two gene pools were also segregated in the Bayesian population structure analysis. However, based on the ΔK (EVANNO; REGNAUT; GOUDET, 2005) criterion, the number of groups (K) with the highest value of ΔK was three (K=3), which demonstrated a subdivision of the Mesoamerican group (Figure 3.3). Based on the membership coefficient (≥ 0.6), 90.9% of the accessions could be assigned to a specific group, and only 20 accessions were categorized as admixtures. The accessions of Andean origin formed a group, and the Mesoamerican accessions were divided into two distinct groups and the admixture group. In the two Mesoamerican groups formed solely by individuals with a membership coefficient ≥ 0.6, the accessions were distinguished by the color of the seed tegument; one group was composed

primarily of carioca-type grain accessions, whereas the other group included accessions with black, purple, red, cream, and other tegument colors.

The admixture group comprised accessions that had resulted from hybridization between the previous two groups. The accessions of commercial groups other than black and carioca (i.e., purple, red, cream, and others) were predominantly grouped with the black commercial group; however, there was a tendency to cluster according to the color of the flower, which is purple in the black group, white in the carioca group, and variable (white, pink, and purple) in other accessions. Three accessions initially identified as Mesoamerican were assigned to the Andean group in these analyses and were therefore treated as Andean in subsequent analyses.

Removal of accessions of Andean origin from the panel left 207 accessions of the Mesoamerican origin. Among these 207 accessions, 25,136 SNPs with MAF > 0.05 could be identified, i.e., the number of SNPs per chromosome was reduced on average by 50% relative to the number of SNPs identified when the Andean accessions were included in the panel. The chromosomes exhibiting the greatest reduction in the number of SNPs were Pv05 and Pv11, whereas Pv09 and Pv01 presented the smallest reduction.

Figure 3.2 Genetic differentiation between Andean and Mesoamerican gene pools. (a) Principal component analysis of 219 accessions of Andean and Mesoamerican origin including different commercial groups (black, carioca, cream, red, etc.). (b) Venn diagram of the total set of SNPs and SNPs belonging to the Andean and Mesoamerican groups. (c) Distribution of the *Fst* values of each SNP (colored according to the population in which they occur). (d) Total number of differentiating SNPs on each chromosome and number of differentiating SNPs located within genes. (e) Distribution of the 11,805 differentiating SNPs of the Andean and Mesoamerican groups along the common bean genome in a 200 Kb sliding window.

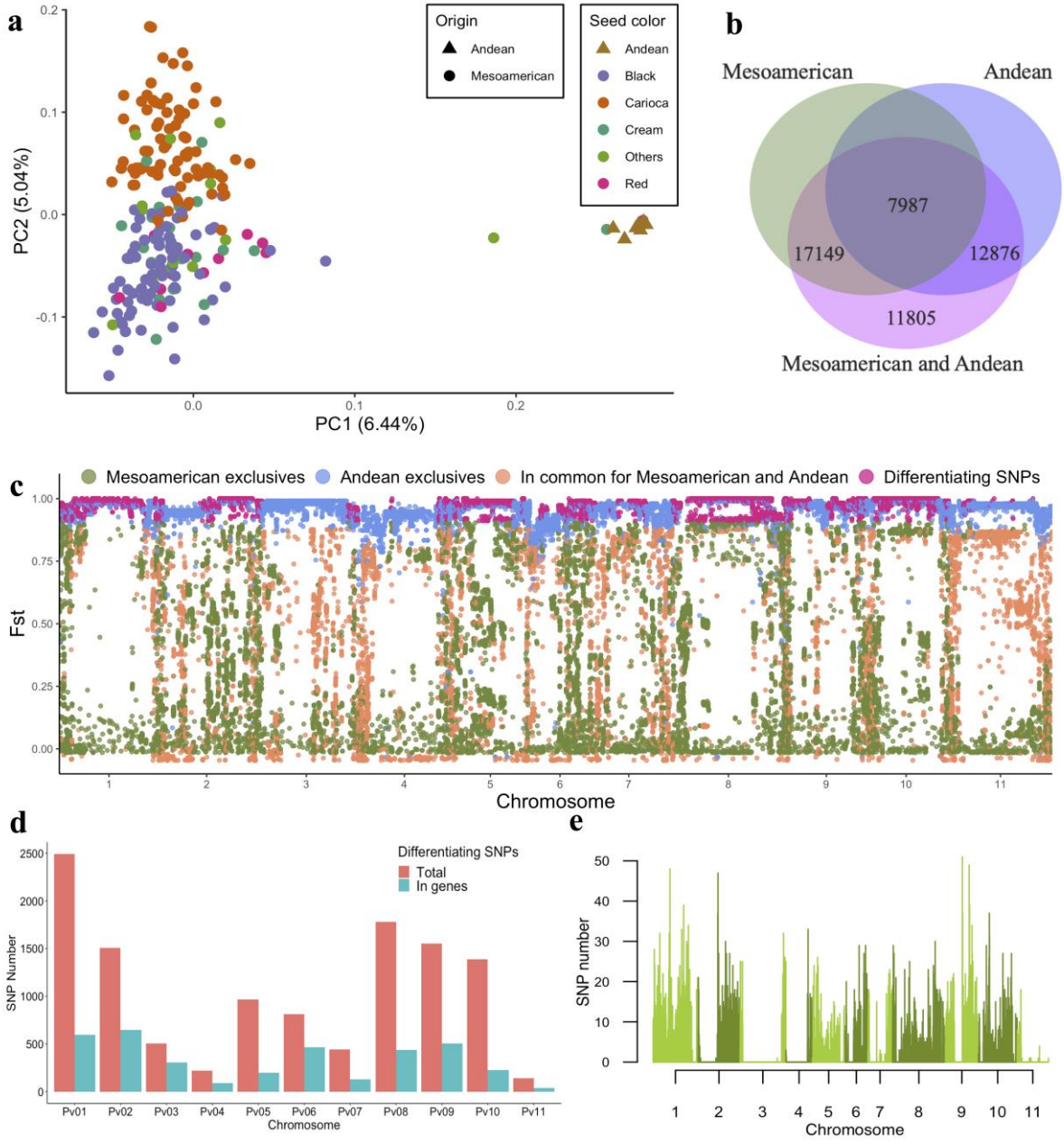
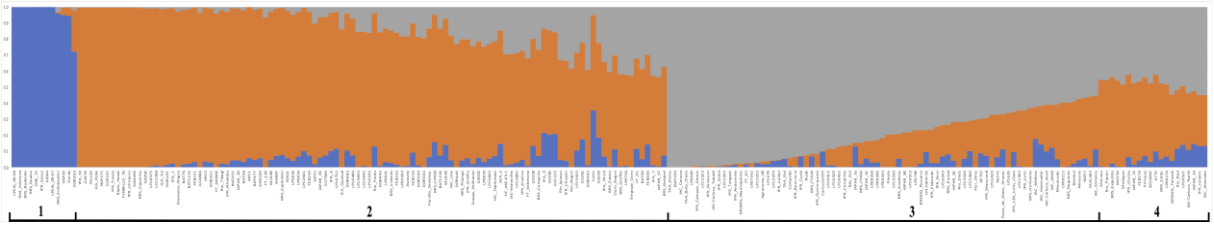


Figure 3.3 Analysis of the population structure using 219 accessions belonging to the Brazilian common bean diversity panel with K = 3: (1) corresponds to the group of common beans of Andean origin; (2) mostly formed by Mesoamerican accessions of black, cream, red, and other seed tegument colors; (3) mostly formed by Mesoamerican accessions from the carioca commercial group; and (4) mostly formed by Mesoamerican accessions with membership coefficient < 0.6 for the previous groups.



3.4.3 GENETIC DIFFERENTIATION BETWEEN ANDEAN AND MESOAMERICAN GENE POOLS

The two gene pools shared 7,987 SNPs, whereas 17,149 and 12,876 SNPs were unique to the Mesoamerican and Andean groups (Figure 3.2b), respectively. The mean pairwise fixation index (*Fst*) for each of these SNP groups was 0.39, 0.34, and 0.94, respectively (Figure 3.2c). A total of 11,805 SNPs differentiating the Andean and Mesoamerican groups were detected, with a mean *Fst* of 0.97. The mean *Fst* between the Andean and Mesoamerican pools was 0.77 when all the SNPs were included. The Mesoamerican group showed greater mean nucleotide diversity (π = 0.31) than the Andean group (π = 0.22). Regarding Tajima's D, the Mesoamerican gene pool showed a positive value (D = 1.50), while the Andean gene pool showed a negative value (D = -0.50) (Table 3.2).

Most SNPs that differentiate the Andean and Mesoamerican pools were located on chromosomes Pv01 (2,492 SNPs), Pv08 (1,781 SNPs), Pv09 (1,554 SNPs), Pv02 (1,506 SNPs), and Pv10 (1,387 SNPs) (Figure 3.2d and e), and 30.8% were located within genes, with 2,187 genes including at least one differentiating SNP. Most of these SNPs inside genes were located on chromosomes Pv02 (648 SNPs), Pv01 (595 SNPs), Pv09 (504 SNPs), Pv06 (476 SNPs), and Pv08 (435 SNPs) chromosomes (Figure 3.2d). Among the SNPs located in coding regions, 26% were synonymous SNPs and 74% were non-synonymous (being 68% missense variants). Of the genes containing the differentiating SNPs, 279 were putative candidates for domestication, of which 179 are known to be involved in the domestication of the Mesoamerican group, 91 in that of the Andean group, and 9 in the domestication of both these groups (SCHMUTZ et al., 2014).

Table 3.2 Nucleotide diversity (π), Tajima's D and weighted *Fst* estimated in the Brazilian common bean diversity panel in relation to different centers of origin, seed colors, and institutions of origin.

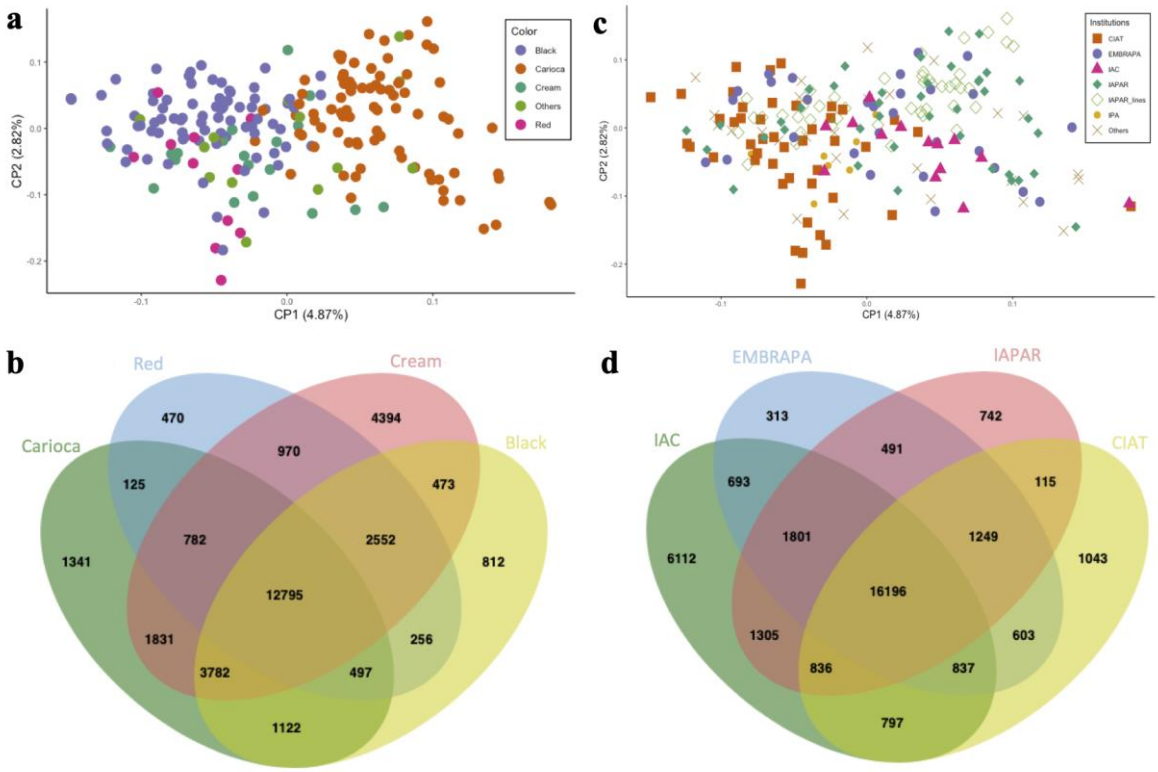
	N	SNPs	π	D	*Fst*		
Origin					Andean		
Mesoamerican	207	25,136	0.31	1.50	0.77		
Andean	12	20,863	0.22	-0.50			
Seed tegument color					Black	Red	Cream
Carioca	85	22,275	0.32	1.32	0.12	0.22	0.12
Black	78	22,289	0.31	1.23		0.10	0.03
Red	11	18,447	0.35	0.69			0.08
Cream	19	27,579	0.29	0.40			
Institutions of origin[2]					EMBRAPA	IAC	IAPAR
CIAT	45	21,676	0.34	1.28	0.06	0.12	0.09
EMBRAPA	29	22,183	0.34	1.08		0.09	0.01
IAC	14	28,577	0.30	0.44			0.06
IAPAR	84	22,735	0.32	1.35			

N = number of accessions, SNPs = number of SNPs, π = nucleotide diversity, D = Tajima's D statistics. [1]Weir and Cockerham, 1984. [2]CIAT = International Center for Tropical Agriculture (Centro Internacional de Agricultura Tropical), EMBRAPA = Brazilian Agricultural Research Corporation (Empresa Brasileira de Pesquisa Agropecuária), IAC = Agronomic Institute of Campinas (Instituto Agronômico de Campinas), IAPAR = Rural Development Institute of Paraná – IAPAR – EMATER (Instituto de Desenvolvimento Rural do Paraná).

3.4.4 Genetic differentiation between the Mesoamerican accessions

As seen in the PCA and Bayesian analysis of population structure, the accessions of Mesoamerican origin were also segregated into two main groups in the phylogenetic tree, based on the tegument color, with one group consisting of the carioca commercial group and the other including the accessions with black, cream, red, white and purple tegument (Figure 3.4a and Figure 3.5).

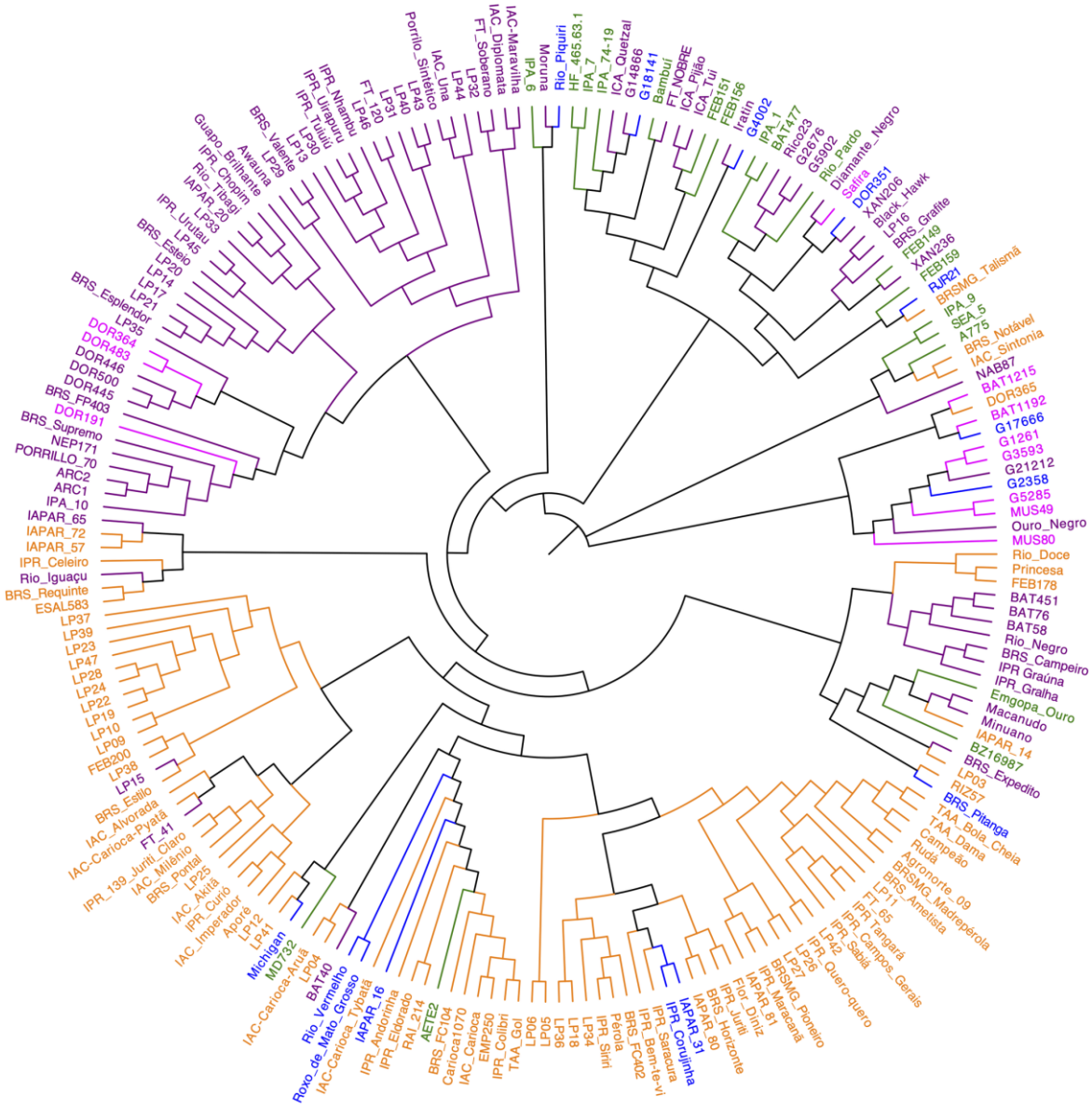
Figure 3.4 Principal component analyses and Venn diagrams. (a) Principal component analysis of 207 accessions of common beans of Mesoamerican origin with different seed tegument colors. (b) Venn diagram for the different sets of SNPs related to seed tegument color. (c) Principal component analysis of 207 accessions of common beans from different research institutions. (d) Venn diagram for the different sets of SNPs related to the institutions of origin.



The separation of Mesoamerican individuals by seed color showed that each group had a variable number of SNPs, and only 12,795 SNPs were common to all these color groups (Figure 3.4b). The cream-colored accessions exhibited the highest number of SNPs (27,579) and the lowest π (0.29) value. The red-colored group had the highest π (0.35), whereas π values of the carioca and black groups were similar (0.32 and 0.31, respectively) (Table 3.2). According to *Fst*, the carioca and red groups were the most different, with an *Fst* value of 0.22, whereas comparisons between the other colors yielded low *Fst* values. The Tajima's D values were all positive in relation to the seed tegument color as well as for the institution of origin (Table 3.2).

Figure 3.5 Dendrogram showing the genetic relatedness among 207 common bean accessions belonging to the Brazilian Diversity Panel. The different colors identify the accessions

according to the color of the seed tegument. Purple = black tegument, Orange = carioca-type tegument, Green = cream tegument, Pink = red tegument, and Blue = others.



Regarding the institution of origin, a clustering trend was observed for the accessions of the International Center for Tropical Agriculture (CIAT), Agronomic Institute of Pernambuco (IPA), and the more recent inbred lines of the Rural Development Institute of Paraná – IAPAR–EMATER (IAPAR) (Figure 3.4c). The number of SNPs was variable for each institution; however, the π was similar. A total of 16,196 SNPs was shared in the accessions of all institutions, whereas 8,210 were exclusive, i.e., belonged to only one institution (Figure 3.4d). The Agronomic Institute of Campinas (IAC) accessions presented the highest number of exclusive markers (6,112), whereas the Brazilian Agricultural Research Corporation

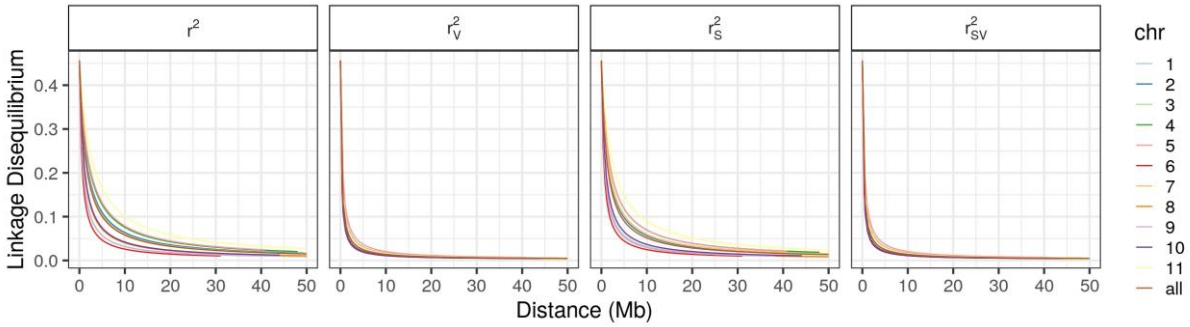
(EMBRAPA) accessions included only 313 unique markers. Comparison of the accessions from different institutions did not yield high differentiation indexes (*Fst*), with the highest value being observed between CIAT and IAC (0.12) and the lowest value between IAPAR and EMBRAPA (0.01) (Table 3.2).

3.4.5 LINKAGE DISEQUILIBRIUM

LD decay and half-decay distances were calculated for individual chromosomes and for the whole genome. In both cases, the differences between conventional r^2 and population structure-corrected r^2 (r^2_s) were small. Considering the whole genome (all chromosomes), the half-decay distance was 1,361 Kb and 1,180 Kb for r^2 and r^2_s, respectively. The r^2 was remarkably different when compared with r^2_v (r^2 corrected for relatedness) and r^2_{vs} (r^2 corrected for population structure and relatedness) (Figure 3.6). The latter two measures exhibited very similar decay values, with half-decay occurring at 249 Kb.

In the analysis of r^2_v and r^2_{vs} of individual chromosomes, the half-decay distance ranged from 183 to 397. The highest decay values were noted for chromosomes Pv10, Pv08, Pv01, and Pv06 chromosomes (183, 187, 193 and 198 Kb, respectively), whereas the lowest decay values were presented by chromosomes Pv05, Pv04, Pv07, and Pv09 (397, 322, 317 and 310, respectively) (Figure 3.6).

Figure 3.6 Analysis of linkage disequilibrium (LD) decay as a function of physical distance without correction (r^2), and after correcting for population structure (r^2_s), relatedness (r^2_v), and for both population structure and relatedness (r^2_{vs}).



3.5 DISCUSSION

The common bean is a very important crop in Brazil and is cultivated in all states of the country, mainly by family farmers. Considering the history of common bean cultivation in the country, domesticated common beans are highly diverse, although Brazil is not a primary center of diversity (BURLE et al., 2010; VALDISSER et al., 2017). In this context, the present study was developed to understand the genetics and population structure of a newly created common bean diversity panel that includes a large part of the diversity of the most consumed common bean types in Brazil. These results will assist future GWAS for determining genomic regions or genes associated with several economically important traits.

The GBS methodology proposed by Ariani, Berny Mier y Teran and Gepts (2016) was used in this study, which could effectively detect numerous SNPs in the analyzed accessions. These authors found that the *CviAII* enzyme was more effective than the commonly used *ApeKI* enzyme. As a methylation-insensitive enzyme, *CviAII* exhibited a higher number of restriction sites and acted preferentially on non-repetitive parts of the genome, allowing the identification of thousands of markers spaced unevenly throughout the common bean genome, with a density distribution resembling that of the distribution of genes.

Initially, 461,199 SNPs were identified. However, 89% of the markers did not satisfy the filtering criteria (Non-biallelic, indels, MAF < 0.05, coefficient of inbreeding < 0.9 and less than 10% of genotyped positions) and were not used in subsequent analyses. Polymorphisms were widely distributed across the 11 chromosomes and were highly correlated with the length and number of genes on each chromosome. The transition/transversion rate was consistent with that observed in other studies on common bean and other species (ALIPOUR et al., 2017; GAUR et al., 2015; LIOI et al., 2019; PAVAN et al., 2017). Transitions are usually more frequent than transversions in several species, which indicates that the former are better tolerated during natural selection, which may be due to the fact that they are synonymous mutations in protein-coding sequences (GUO et al., 2017; LUO et al., 2019).

Because LD may affect the inference of the population structure, an LD filter was further applied, which resulted in a decrease of the number of SNPs. This is due to the fact that the common bean is an autogamous plant with very long blocks of markers in LD (BERNY MIER Y TERAN et al., 2018; BLAIR et al., 2018; LIOI et al., 2019).

Genetic differentiation between common bean accessions based on the gene pool has been well documented in several previous studies (CAMPA; MURUBE; FERREIRA, 2018; DINIZ et al., 2019; KWAK; GEPTS, 2009; LIOI et al., 2019; PIPAN; MEGLIČ, 2019; RAATZ et al., 2019; VALENTINI et al., 2018). The relationship between the genetic similarity of the Mesoamerican accessions and the color of the seed tegument was also observed by Valdisser

et al. (2016) and Gioia et al. (2019). In Brazil, breeding programs for the carioca and black commercial groups have different objectives (DELFINI et al., 2017). Moreover, genetic breeding of the carioca group is much more advanced than that of the black group, because of its greater importance in the country due to consumers and market preferences. Efforts to improve the carioca bean are directed towards the grain size traits, to satisfy the consumers' preference for larger grains. However, the grain size is negatively correlated with yield, in case of the black group, selection is based mainly on yield, resulting in cultivars with smaller grains (DELFINI et al., 2018).

Several SNPs exclusive to either of the gene pools were observed, in addition to the differentiating SNPs between the two pools. Other authors have also reported that the proportion of polymorphic loci tends to be higher in populations composed of accessions from the two centers of origin, and it tends to decrease when they are studied separately (BLAIR et al., 2013; CORTÉS; CHAVARRO; BLAIR, 2011; VALDISSER et al., 2016). The two gene pools differ in both phenotypic and molecular characteristics, which is supported by the high rates of genetic differentiation obtained in the present analysis and in other studies (ARIANI; MIER TERAN; GEPTS, 2017; BURLE et al., 2010; VALDISSER et al., 2016, 2017). In addition, the Mesoamerican gene pool exhibits higher nucleotide diversity than the Andean, possibly because a strong bottleneck occurred during the dispersal of Southern Andean common beans from Mesoamerica, which drastically reduced its nucleotide diversity (ARIANI; MIER TERAN; GEPTS, 2017; BITOCCHI et al., 2013; CAMPA; MURUBE; FERREIRA, 2018; CICHY et al., 2015; SCHMUTZ et al., 2014; VALENTINI et al., 2018).

To identify genomic signatures of selection between the Andean and Mesoamerican pools, the *Fst* was estimated for each SNP. The *Fst* of nucleotide positions that were polymorphic only when the two gene pools were studied together was close to 1, and these SNPs were therefore highly discriminating between the gene pools. The *Fst* of SNPs present only in the Andean group was also high, similar to that of the discriminating markers. This may be due to the small number of Andean accessions included in this study.

There is significant evidence supporting the independent domestication of the Andean and Mesoamerican gene pools. Schmutz et al. (2014) identified 1,835 candidate genes for domestication in the Mesoamerican group and 748 candidate genes in the Andean group. Of these, only 59 genes were common to both groups. These genes are mainly located on chromosomes Pv01, Pv02, Pv07, Pv09, and Pv10. In the present study, 11% of all the candidate genes for domestication harbored differentiating SNPs. These genes have also been identified in other studies aimed at finding selection signatures between the Andean and Mesoamerican

(CAMPA; MURUBE; FERREIRA, 2018; LIOI et al., 2019) accessions. The candidate genes for domestication are directly or indirectly associated with the main characteristics that distinguish the two gene pools, such as flowering time, plant size, and seed size.

The low rate of differentiation of the accessions based on the institution of origin may be related to the protocols of the breeding programs. Breeding programs tend to be conservative and almost always employ the Mesoamerican germplasm, with little exploration of exotic germplasms; in addition, they use a selected group of elite parents, which further narrows the genetic base (DELFINI et al., 2017; DINIZ et al., 2019; GIOIA et al., 2019; VALDISSER et al., 2016). Because of significant exchange of germplasm between institutions, there was no formation of well-defined groups among the accessions from the different institutions of origin (DINIZ et al., 2019), only a trend for clustering was observed for accessions belonging to CIAT, IPA, and the more recent inbred lines of IAPAR.

LD measurement is very important in association mapping studies for identifying loci associated with quantitative traits. Importantly, the population structure and relatedness between the analyzed accessions may cause a bias in LD estimation. Frequent selection, admixture of populations, and crossing of a small number of cultivars in breeding programs reduces genetic diversity and affects LD patterns (CONTRERAS-SOTO et al., 2017a). These factors can affect different genomic regions in several ways, which can introduce heterogeneity of LD through the genome. This makes the resolution and power achieved in GWAS dependent on the species and the population under study. LD decay is slower in autogamous species, such as common bean and soybean, in which recombination is less effective than in allogamous species (CONTRERAS-SOTO et al., 2017a; FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003).

In this study, the LD corrected for population structure (r^2_s) was not significantly different from the conventional r^2. However, r^2_v and r^2_{vs} exhibited a faster LD decay when compared with the conventional r^2. The fact that r^2_v was considerably lower than conventional r^2 demonstrates the need to remove the effect of relatedness to reduce the overestimation of LD. The similarity between the estimated r^2 and r^2_s (LD half-decay with 296 Kb) shows that the BDP is not highly structured, which is consistent with the results of other studies on common bean diversity panels (DINIZ et al., 2019; RESENDE et al., 2018a; VALDISSER et al., 2017). As observed by Diniz et al. (2019) in panels composed mainly of improved genotypes, the degree of relatedness between individuals was very high.

The present study demonstrated that GBS is a powerful approach for analyzing the population structure and genetic diversity in common bean. The newly developed diversity

panel, which represents a large proportion of the Brazilian common bean diversity, exhibited high genetic diversity, and was shown to be adequate for future studies to identify genomic regions related to traits of interest (GWAS).

ADDITIONAL INFORMATION

Table S3.1 List of accessions constituting the Brazilian Diversity Panel (BDP).

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
BRS Ártico	Andean	Cultivar	EMBRAPA	White
BRS Embaixador	Andean	Cultivar	EMBRAPA	Red
BRS Radiante	Andean	Cultivar	EMBRAPA	Red (Others)
BRS Realce	Andean	Cultivar	EMBRAPA	Red (Others)
DRK 15	Andean	Cultivar	CIAT	Red
G6416	Andean	Breeding line	CIAT	Red
IPR Garça	Andean	Cultivar	IAPAR	White
KID44	Andean	Breeding line	CIAT	Red
LP01	Andean	Breeding line	IAPAR	Carioca
LP02	Andean	Breeding line	IAPAR	Carioca
Diamante Negro	Mesoamerican	Cultivar	EMBRAPA	Black
Aporé	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS Ametista	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS Campeiro	Mesoamerican	Cultivar	EMBRAPA	Black
BRS Esplendor	Mesoamerican	Cultivar	EMBRAPA	Black
BRS Esteio	Mesoamerican	Cultivar	EMBRAPA	Black
BRS Estilo	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS Expedito	Mesoamerican	Cultivar	EMBRAPA	Black
BRS FC104	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS FC402	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS FP403	Mesoamerican	Cultivar	EMBRAPA	Black
BRS Grafite	Mesoamerican	Cultivar	EMBRAPA	Black
BRS Horizonte	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS Notável	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS Pitanga	Mesoamerican	Cultivar	EMBRAPA	Purple
BRS Pontal	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS Requinte	Mesoamerican	Cultivar	EMBRAPA	Carioca

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
BRS Supremo	Mesoamerican	Cultivar	EMBRAPA	Black
BRS Valente	Mesoamerican	Cultivar	EMBRAPA	Black
BRS MGMadrepérola	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS MGPioneiro	Mesoamerican	Cultivar	EMBRAPA	Carioca
BRS MGTalismã	Mesoamerican	Cultivar	EMBRAPA	Carioca
Bambuí	Mesoamerican	Cultivar	EMBRAPA	Cream
Guapo Brilhante	Mesoamerican	Cultivar	EMBRAPA	Black
Macanudo	Mesoamerican	Cultivar	EMBRAPA	Black
Minuano	Mesoamerican	Cultivar	EMBRAPA	Black
Ouro Negro	Mesoamerican	Cultivar	UFV/EPAMIG	Black
Pérola	Mesoamerican	Cultivar	EMBRAPA	Carioca
Rudá	Mesoamerican	Cultivar	EMBRAPA	Carioca
Safira	Mesoamerican	Cultivar	EMBRAPA	Red
FT120	Mesoamerican	Cultivar	FT Sementes	Black
FT NOBRE	Mesoamerican	Cultivar	FT Sementes	Black
FT 41	Mesoamerican	Cultivar	FT Sementes	Black
FT 65	Mesoamerican	Cultivar	FT Sementes	Carioca
FT Soberano	Mesoamerican	Cultivar	FT Sementes	Black
IAC Akitã	Mesoamerican	Cultivar	IAC	Carioca
IAC Alvorada	Mesoamerican	Cultivar	IAC	Carioca
IAC Carioca	Mesoamerican	Cultivar	IAC	Carioca
IAC Diplomata	Mesoamerican	Cultivar	IAC	Black
IAC Formoso	Mesoamerican	Cultivar	IAC	Carioca
IAC Imperador	Mesoamerican	Cultivar	IAC	Carioca
IAC Milênio	Mesoamerican	Cultivar	IAC	Carioca
IAC Sintonia	Mesoamerican	Cultivar	IAC	Carioca
IAC Una	Mesoamerican	Cultivar	IAC	Black
IAC CariocaAruã	Mesoamerican	Cultivar	IAC	Carioca
IAC CariocaPyatã	Mesoamerican	Cultivar	IAC	Carioca
IAC CariocaTybatã	Mesoamerican	Cultivar	IAC	Carioca

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
IAC Maravilha	Mesoamerican	Cultivar	IAC	Black
Moruna	Mesoamerican	Cultivar	IAC	Black
IAPAR 16	Mesoamerican	Cultivar	IAPAR	Carioca (Others)
RAI214	Mesoamerican	Cultivar	IAPAR	Carioca
IAPAR 57	Mesoamerican	Cultivar	IAPAR	Carioca
IAPAR 65	Mesoamerican	Cultivar	IAPAR	Black
IAPAR 14	Mesoamerican	Cultivar	IAPAR	Carioca
IAPAR 20	Mesoamerican	Cultivar	IAPAR	Black
IAPAR 31	Mesoamerican	Cultivar	IAPAR	Carioca (Others)
IAPAR 72	Mesoamerican	Cultivar	IAPAR	Carioca
IAPAR 80	Mesoamerican	Cultivar	IAPAR	Carioca
IAPAR 81	Mesoamerican	Cultivar	IAPAR	Carioca
IPR 139 - JuritiClaro	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Andorinha	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Bem-te-vi	Mesoamerican	Cultivar	IAPAR	Carioca
IPR CamposGerais	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Celeiro	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Chopim	Mesoamerican	Cultivar	IAPAR	Black
IPR Colibri	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Corujinha	Mesoamerican	Cultivar	IAPAR	Carioca (Others)
IPR Curió	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Eldorado	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Gralha	Mesoamerican	Cultivar	IAPAR	Black
IPR Graúna	Mesoamerican	Cultivar	IAPAR	Black
IPR Juriti	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Maracanã	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Nhambu	Mesoamerican	Cultivar	IAPAR	Black
IPR Quero-quero	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Sabiá	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Saracura	Mesoamerican	Cultivar	IAPAR	Carioca

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
IPR Siriri	Mesoamerican	Cultivar	IAPAR	Carioca
IPR Tangará	Mesoamerican	Cultivar	IAPAR	Carioca
FEB200	Mesoamerican	Breeding line	CIAT	Carioca
IPR Tuiuiú	Mesoamerican	Cultivar	IAPAR	Black
IPR Uirapuru	Mesoamerican	Cultivar	IAPAR	Black
IPR Urutau	Mesoamerican	Cultivar	IAPAR	Black
Rio Iguaçu	Mesoamerican	Cultivar	IAPAR	Black
Rio Negro	Mesoamerican	Cultivar	IAPAR	Black
Rio Pardo	Mesoamerican	Cultivar	IAPAR	Cream
Rio Piquiri	Mesoamerican	Cultivar	IAPAR	Brown
Rio Tibagi	Mesoamerican	Cultivar	IAPAR	Black
Rio Vermelho	Mesoamerican	Cultivar	IAPAR	Purple
Rio Doce	Mesoamerican	Cultivar	IAPAR	Carioca
Gordo	Mesoamerican	Cultivar	IPA	Cream
HF465.63.1	Mesoamerican	Cultivar	IPA	Cream
IPA1	Mesoamerican	Cultivar	IPA	Cream
IPA10	Mesoamerican	Cultivar	IPA	Black
IPA6	Mesoamerican	Cultivar	IPA	Cream
IPA7	Mesoamerican	Cultivar	IPA	Cream
IPA74-19	Mesoamerican	Cultivar	IPA	Cream
IPA9	Mesoamerican	Cultivar	IPA	Cream
Princesa	Mesoamerican	Cultivar	IPA	Carioca
TAA Bola Cheia	Mesoamerican	Cultivar	TAA	Carioca
TAA Dama	Mesoamerican	Cultivar	TAA	Carioca
TAA Gol	Mesoamerican	Cultivar	TAA	Carioca
Awauna	Mesoamerican	Cultivar	UEM	Black
Flor Diniz	Mesoamerican	Cultivar	UEM	Carioca
Rico23	Mesoamerican	Cultivar	UFV	Black
Campeão	Mesoamerican	Cultivar	Agristar	Carioca
Agronorte09	Mesoamerican	Cultivar	Agronorte	Carioca

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
ICA Pijão	Mesoamerican	Cultivar	ICA	Black
ICA Quetzal	Mesoamerican	Cultivar	ICA	Black
ICA Tui	Mesoamerican	Cultivar	ICA	Black
Iratin	Mesoamerican	Landrace	-	Black
Emgopa Ouro	Mesoamerican	Cultivar	Incaper	Cream
A775	Mesoamerican	Breeding line	CIAT	Cream
A779	Mesoamerican	Breeding line	CIAT	Cream
AETE2	Mesoamerican	Cultivar	IAC	Cream
ARC1	Mesoamerican	Breeding line	CIAT	Black
ARC2	Mesoamerican	Breeding line	CIAT	Black
BAT1215	Mesoamerican	Breeding line	CIAT	Red
BAT40	Mesoamerican	Breeding line	CIAT	Black
BAT41	Mesoamerican	Breeding line	CIAT	Black
BAT451	Mesoamerican	Breeding line	CIAT	Black
BAT58	Mesoamerican	Breeding line	CIAT	Black
BAT76	Mesoamerican	Breeding line	CIAT	Black
BAT1192	Mesoamerican	Breeding line	CIAT	Red
BAT477	Mesoamerican	Breeding line	CIAT	Cream
Black Hawk	Mesoamerican	Cultivar	MSU	Black
BZ16987	Mesoamerican	Breeding line	-	Cream
Carioca1070	Mesoamerican	Breeding line	CENA/USP	Carioca
DOR191	Mesoamerican	Breeding line	CIAT	Red
DOR351	Mesoamerican	Breeding line	CIAT	Purple
DOR365	Mesoamerican	Breeding line	CIAT	Carioca
DOR445	Mesoamerican	Breeding line	CIAT	Black
DOR446	Mesoamerican	Breeding line	CIAT	Black
DOR483	Mesoamerican	Breeding line	CIAT	Red
DOR500	Mesoamerican	Breeding line	CIAT	Black
DOR303	Mesoamerican	Breeding line	CIAT	Carioca (Others)
DOR364	Mesoamerican	Breeding line	CIAT	Red

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
EMP250	Mesoamerican	Breeding line	CIAT	Carioca
ESAL583	Mesoamerican	Breeding line	ESALQ	Carioca
FEB149	Mesoamerican	Breeding line	CIAT	Cream
FEB151	Mesoamerican	Breeding line	CIAT	Cream
FEB156	Mesoamerican	Breeding line	CIAT	Cream
FEB159	Mesoamerican	Breeding line	CIAT	Cream
G1261	Mesoamerican	Landrace	CIAT	Red
G14866	Mesoamerican	Landrace	CIAT	Black
G17666	Mesoamerican	Landrace	CIAT	Yellow
G18141	Mesoamerican	Cultivar	CIAT	Others
G2358	Mesoamerican	Landrace	CIAT	White
G2676	Mesoamerican	Cultivar	CIAT	Black
G3593	Mesoamerican	Landrace	CIAT	Red
G4002	Mesoamerican	Landrace	CIAT	Carioca (Others)
G4825	Mesoamerican	Landrace	CIAT	Carioca
G5285	Mesoamerican	Cultivar	CIAT	Red
G5902	Mesoamerican	Landrace	CIAT	Black
FEB178	Mesoamerican	Breeding line	CIAT	Carioca
G21212	Mesoamerican	Breeding line	CIAT	Black
MD732	Mesoamerican	Breeding line	IAPAR	Cream
Michigan	Mesoamerican	Cultivar	MSU	White
MUS49	Mesoamerican	Breeding line	CIAT	Red
MUS80	Mesoamerican	Breeding line	CIAT	Red
NAB87	Mesoamerican	Breeding line	CIAT	Black
NEP171	Mesoamerican	Breeding line	IICA	Black
PORRILLO70	Mesoamerican	Cultivar	CIAT	Black
PorrilloSintético	Mesoamerican	Cultivar	CIAT	Black
RIZ57	Mesoamerican	Breeding line	CIAT	Carioca
RJR21	Mesoamerican	Breeding line	-	White
RosinhaG1	Mesoamerican	-	IAC	Rosinha

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
RoxinhoIvaí	Mesoamerican	Landrace	-	Black
RoxodeMatoGrosso	Mesoamerican	Landrace	-	Purple
RoxodeMinas	Mesoamerican	Landrace	-	Purple
SEA5	Mesoamerican	Breeding line	CIAT	Cream
Vermelho Imbituva	Mesoamerican	Landrace	-	Red
XAN206	Mesoamerican	Breeding line	CIAT	Black
XAN236	Mesoamerican	Breeding line	CIAT	Black
LP03	Mesoamerican	Breeding line	IAPAR	Carioca
LP04	Mesoamerican	Breeding line	IAPAR	Carioca
LP05	Mesoamerican	Breeding line	IAPAR	Carioca
LP06	Mesoamerican	Breeding line	IAPAR	Carioca
LP07	Mesoamerican	Breeding line	IAPAR	Carioca
LP08	Mesoamerican	Breeding line	IAPAR	Carioca
LP09	Mesoamerican	Breeding line	IAPAR	Carioca
LP10	Mesoamerican	Breeding line	IAPAR	Carioca
LP11	Mesoamerican	Breeding line	IAPAR	Carioca
LP12	Mesoamerican	Breeding line	IAPAR	Carioca
LP13	Mesoamerican	Breeding line	IAPAR	Black
LP14	Mesoamerican	Breeding line	IAPAR	Black
LP15	Mesoamerican	Breeding line	IAPAR	Black
LP16	Mesoamerican	Breeding line	IAPAR	Black
LP17	Mesoamerican	Breeding line	IAPAR	Black
LP18	Mesoamerican	Breeding line	IAPAR	Carioca
LP19	Mesoamerican	Breeding line	IAPAR	Carioca
LP20	Mesoamerican	Breeding line	IAPAR	Black
LP21	Mesoamerican	Breeding line	IAPAR	Black
LP22	Mesoamerican	Breeding line	IAPAR	Carioca
LP23	Mesoamerican	Breeding line	IAPAR	Carioca
LP24	Mesoamerican	Breeding line	IAPAR	Carioca
LP25	Mesoamerican	Breeding line	IAPAR	Carioca

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color
LP26	Mesoamerican	Breeding line	IAPAR	Carioca
LP27	Mesoamerican	Breeding line	IAPAR	Carioca
LP28	Mesoamerican	Breeding line	IAPAR	Carioca
LP29	Mesoamerican	Breeding line	IAPAR	Black
LP30	Mesoamerican	Breeding line	IAPAR	Black
LP31	Mesoamerican	Breeding line	IAPAR	Black
LP32	Mesoamerican	Breeding line	IAPAR	Black
LP33	Mesoamerican	Breeding line	IAPAR	Black
LP34	Mesoamerican	Breeding line	IAPAR	Carioca
LP35	Mesoamerican	Breeding line	IAPAR	Black
LP36	Mesoamerican	Breeding line	IAPAR	Carioca
LP37	Mesoamerican	Breeding line	IAPAR	Carioca
LP38	Mesoamerican	Breeding line	IAPAR	Carioca
LP39	Mesoamerican	Breeding line	IAPAR	Carioca
LP40	Mesoamerican	Breeding line	IAPAR	Black
LP41	Mesoamerican	Breeding line	IAPAR	Carioca
LP42	Mesoamerican	Breeding line	IAPAR	Carioca
LP43	Mesoamerican	Breeding line	IAPAR	Black
LP44	Mesoamerican	Breeding line	IAPAR	Black
LP45	Mesoamerican	Breeding line	IAPAR	Black
LP46	Mesoamerican	Breeding line	IAPAR	Black
LP47	Mesoamerican	Breeding line	IAPAR	Carioca
LP48	Mesoamerican	Breeding line	IAPAR	Carioca

4 ARTIGO B – Genome wide association studies in a Brazilian common bean panel for detection of loci related to nutritional content

4.1 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 (mrMLM, FASTmrMLM, pLARmEB, and ISIS-EM-BLASSO). To obtain more accurate results, only Quantitative-Trait Nucleotides (QTNs) that showed repeatability (i.e., those that could be detected at least twice using different methods or in different environments) were considered significant and were used to locate favorable alleles and candidate genes. A total of 238 unique QTNs were detected for the nine minerals studied; of these, 48 showed repeatability and were considered reliable. The number of QTNs obtained varied for 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. Pleiotropic QTNs and overlapping genomic regions surrounding the QTNs were identified for different minerals, 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 for carioca and black beans, the main groups consumed in Brazil

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

4.2 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; 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; GONZALEZ; RENGIFO, 2000; DELFINI et al., 2020; ISLAM et al., 2002; MCCLEAN et al., 2017; PINHEIRO et al., 2010; RIBEIRO et al., 2014; SILVA et al., 2012). 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; TUSIIME; MUKANKUSI, 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; 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 (RAGGI et al., 2019; RASHEED et al., 2017). 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; ZHANG, 2018), pLARmEB (ZHANG et al., 2017a), and ISIS-EM-BLASSO (TAMBA; NI; ZHANG, 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.

4.3 MATERIAL AND METHODS

4.3.1 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 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 S4.1).

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), Ponta Grossa (PG), and Guarapuava (GUA), 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.

4.3.2 PHENOTYPING FOR NUTRIENT CONTENTS 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 HNO_3:$HClO_4$ 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: $Yijkl = \mu + Ai + Sj + ASij + R/ASkij + G/Slj + AG/Smlj + eijklm$, where μ is the mean; *Ai* is the fixed effect of the i-th environment; *Sj* is the effect of the j-th "set"; *ASij* is the effect of the interaction between environments and "sets"; *R/ASkij* is the effect of the k-th repetition within the interaction between the i-th environment and the j-th set; *G/Slj* is the random effect of the l-th genotype within the j-th "set"; *AG/Smlj* is the effect of the interaction of environments and accessions within the j-th "set"; and *eijklm* is the experimental error (HALLAUER; 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 (R CORE TEAM, 2020; WEI; SIMKO, 2017).

4.3.3 GENOTYPING-BY-SEQUENCING

Genotyping was performed based on the GBS technique using the enzyme *Cvi*AII (ARIANI; BERNY MIER Y TERAN; GEPTS, 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; ZHOU; BROWNING, 2018), and only SNPs anchored to chromosomes in the common bean reference genome were used.

4.3.4 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; STEPHENS; DONNELLY, 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; REGNAUT; GOUDET, 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; WEIR, 1988) and adjusted using the *nls* function in the R software.

4.3.5 GENOME-WIDE ASSOCIATION STUDIES

A total of four multi-locus GWAS implemented in the mrMLM.GUI 4.0 software (YA-WEN; PEI; YUAN-MING, 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.

4.3.6 IDENTIFICACATION OF FAVORABLES 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 higher nutritional contents was verified. Boxplots and a heatmap generated using the *pheatmap* package (KOLDE, 2019) of the R software were produced for better result visualization.

4.3.7 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.

4.4 RESULTS

4.4.1 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 4.1, and the frequency distributions of the contents of these minerals in each of the three environments are presented in Figure 4.1.

ANOVA revealed a significant effect in all characteristics evaluated with regard to genotype and environmental variation; however, no Genotype × Environment (GxE) 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 (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.

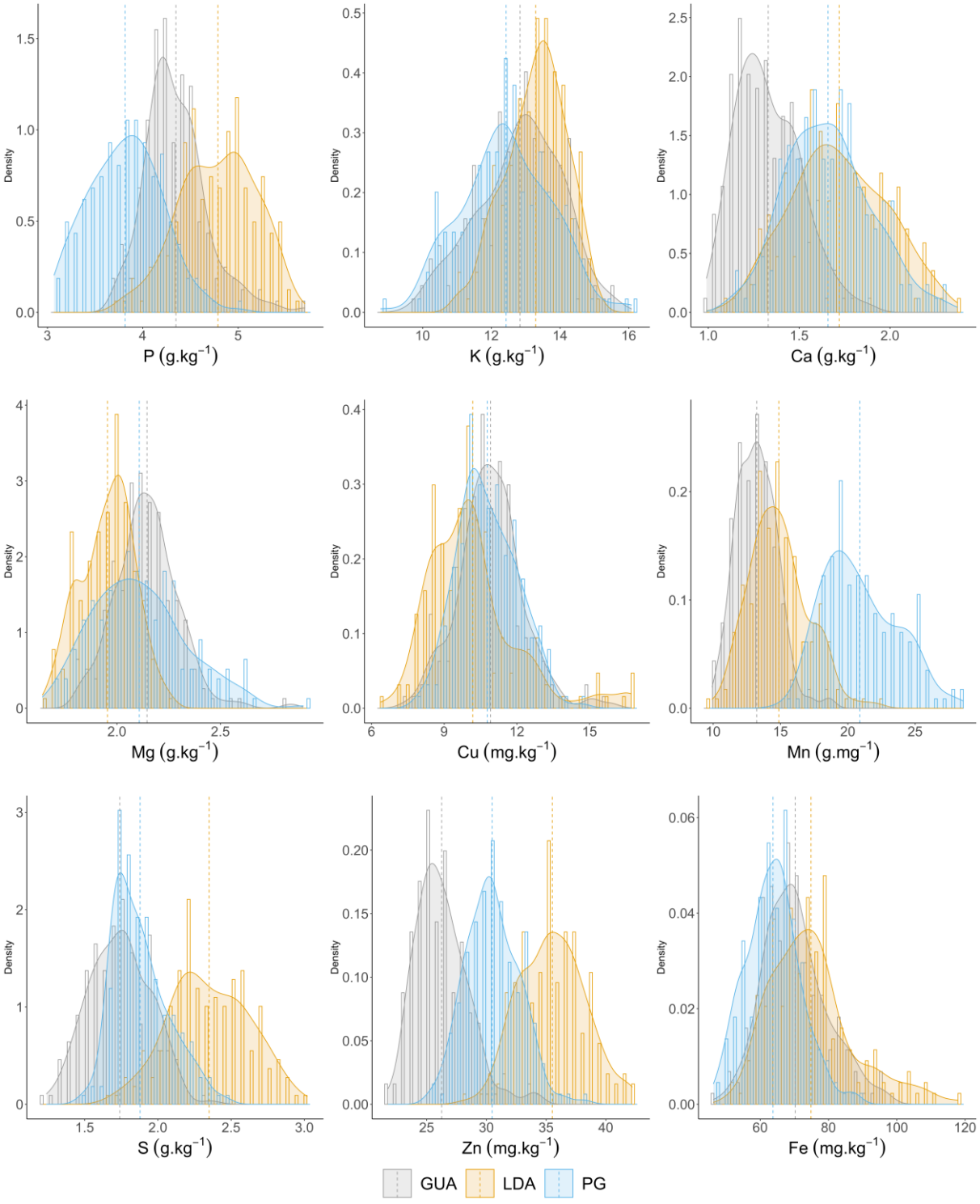
Correlation analysis showed significant effects for most minerals. A positive correlation was detected between Fe and Cu (0.67, $P \leq 0.05$); however, they were negatively correlated with P, K, Mg, and S. Similarly, positive correlations were detected among P, K, Ca, Mg, Mn, and S contents, ranging from 0.31 (Ca-P) to 0.79 (P-S) (Figure S4.1).

Table 4.1 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[1]	K	Ca	Mg	Cu	Mn	S	Zn	Fe
F_{env}[2]		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*
$F_{env*treat(set)}$		0,8ns	0,92ns	0,7ns	0,72ns	0,79ns	0,86ns	0,66ns	0,79ns	0,89ns
CV(%)[3]		9,06	9,32	18,27	8,88	15,52	15,02	10,64	10,08	17,07
Heritability (h^2)		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

[1]P, K, Ca, Mg, and S in $g.kg^{-1}$ and Cu, Mn, Zn, and Fe in $mg.kg^{-1}$; [2]F_{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; [3]CV (%) = coefficient of variation. *P < 0.01, **P < 0.001, *** P < 0.0001, and nsnot significant.

Figure 4.1 Frequency 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).

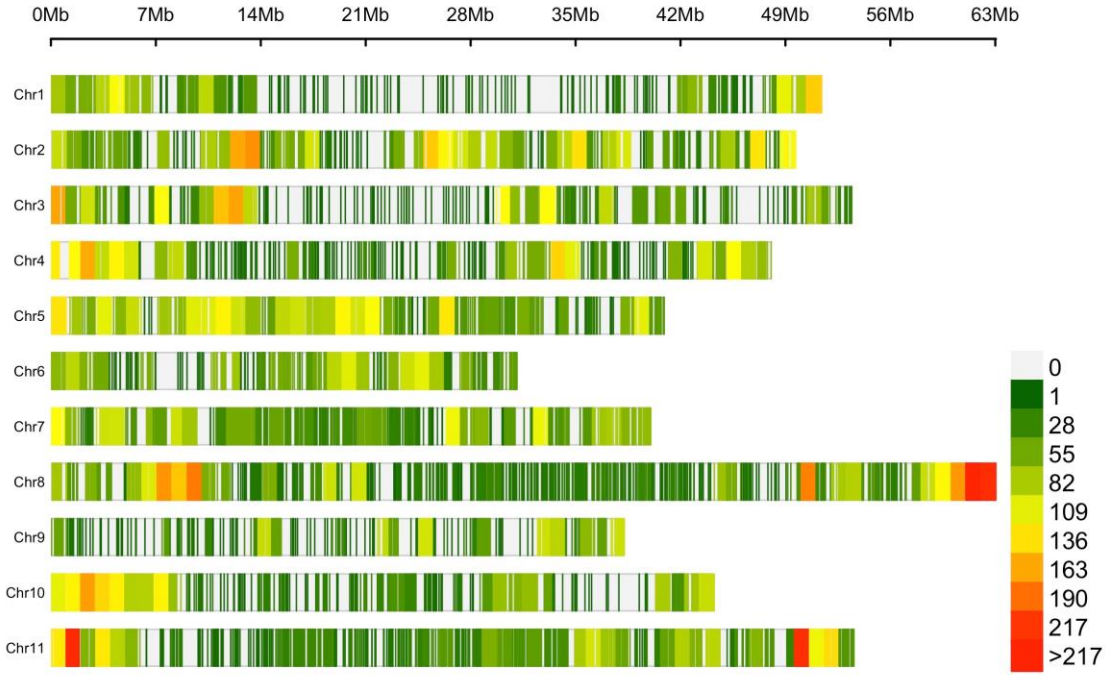


4.4.2 GENOTYPING, POPULATION STRUCTURE, AND LINKAGE DISEQUILIBRIUM

After filtering the SNPs obtained based on GBS, 25,011 high-quality SNPs were obtained. These SNPs were used for subsequent GWAS, and their distribution is shown in Figure 4.2. 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 (Figure S4.2). 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} (Figure S4.3).

Figure 4.2 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.



4.4.3 GENOME-WIDE ASSOCIATION STUDIES

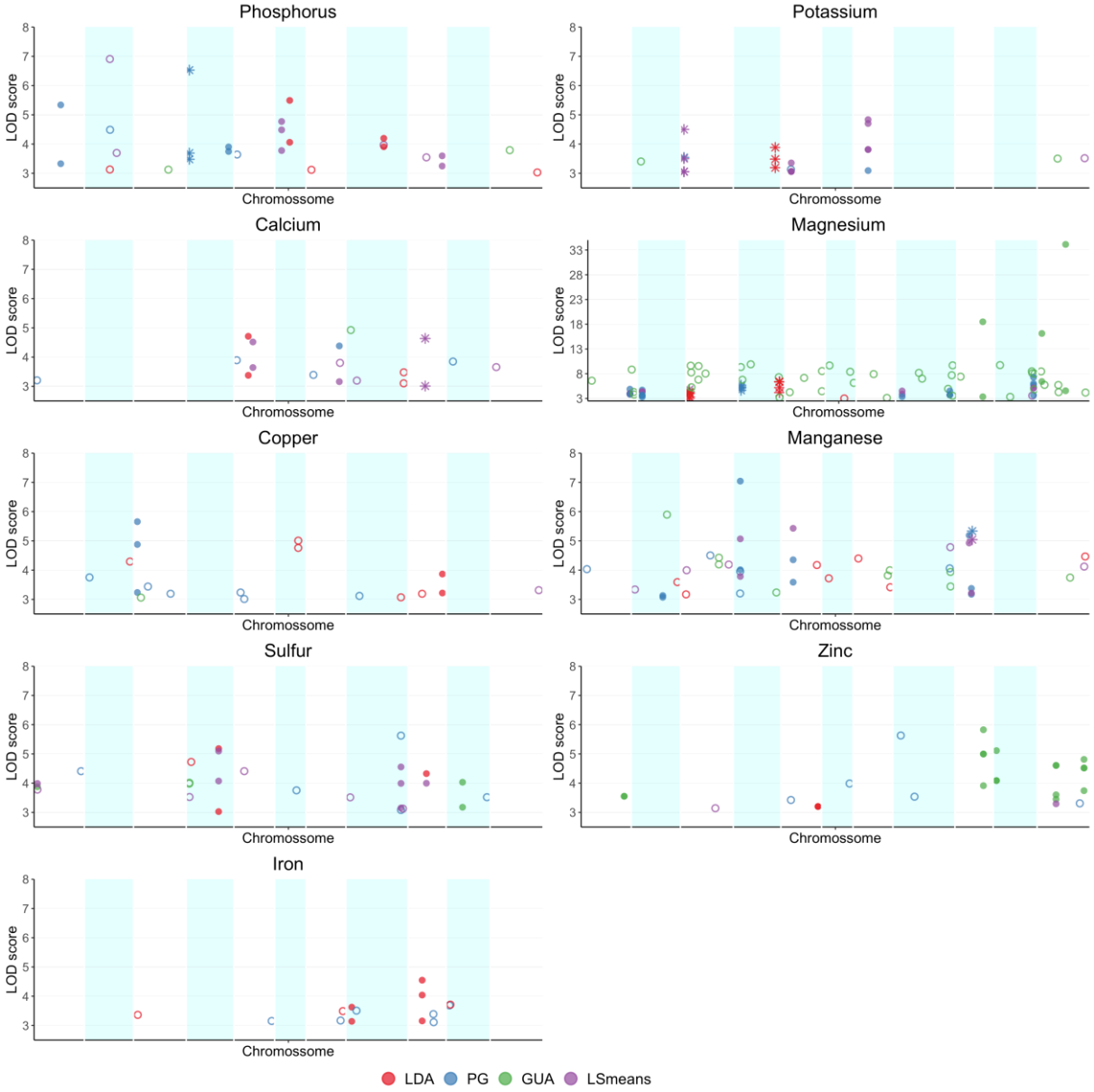
Altogether, 238 unique QTNs were significant for the nine minerals studied, of which 48 were detected at least twice by one of the four multi-locus methods used or in the different environments analyzed (LDA, PG, GUA, and LSmeans). These QTNs were considered promising and were subjected to a more thorough analysis (Figure 4.3). The QTNs presenting repeatability are shown in Table 4.2. 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.

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 (Figure 4.3).

Figure 4.3 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.

Table 4.2 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[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]
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	2, 4
	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
	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

(Continue)

(Continuation)

Trait	SNP	Chr	Position (bp)	QTN effect[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]
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
	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

[1]Quantitative trait nucleotide effect; [2]LOD value, the significant threshold for transformed P-value; [3]PVE (%): phenotypic variation explained; [4]Minor allele frequency; [5]Environment: 1-LDA, 2-PG, 3-GUA, 4-LSmeans; [6]Methods: 1-FASTmrMLM, 2-ISIS-EM-BLASSO, 3-mrMLM, 4-pLARmEB; pleiotropic QTNs are shown in bold font.

4.4.4 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 (Figure 4.4). 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 4.5 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%.

Figure 4.4 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)**.**

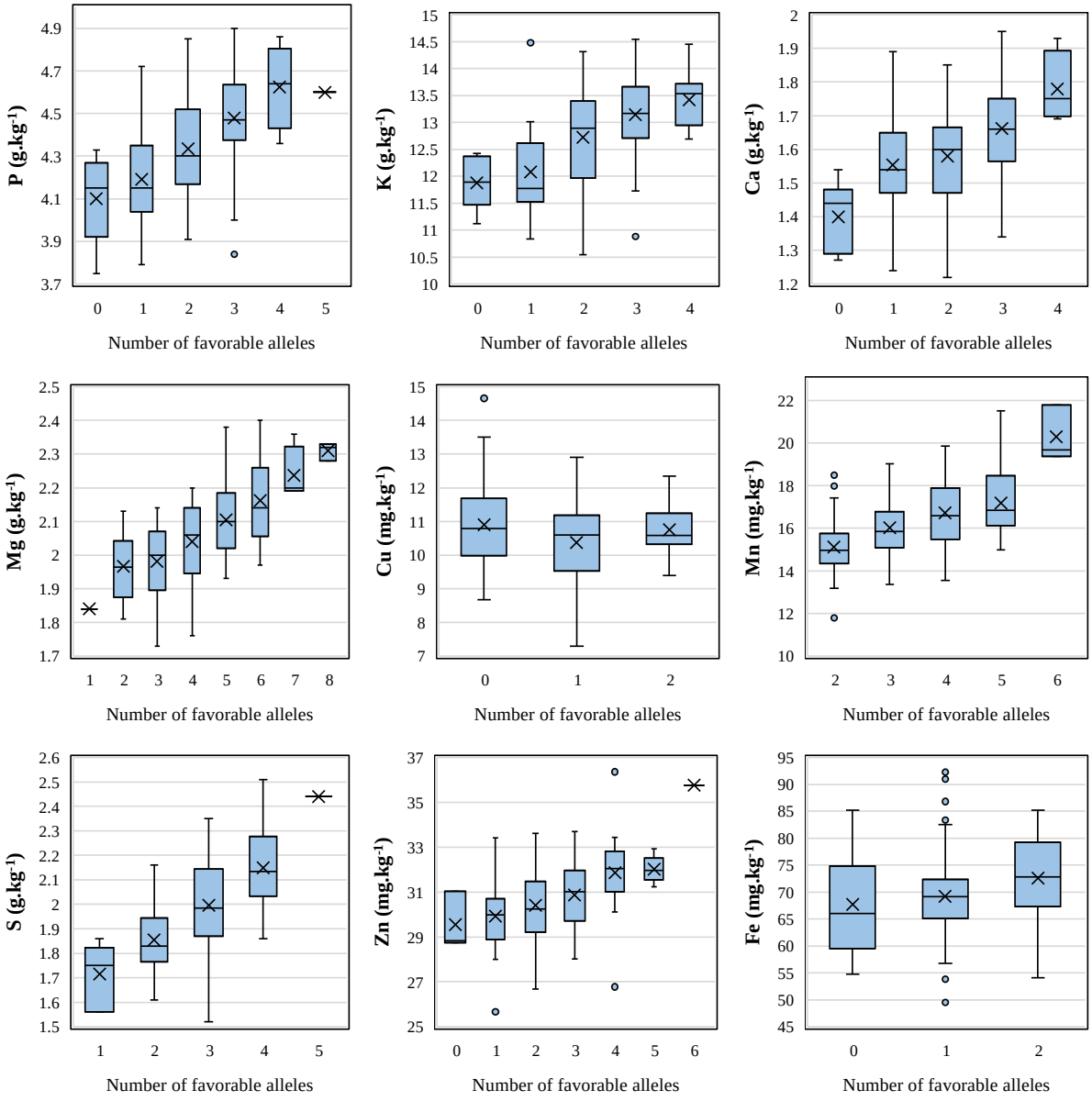
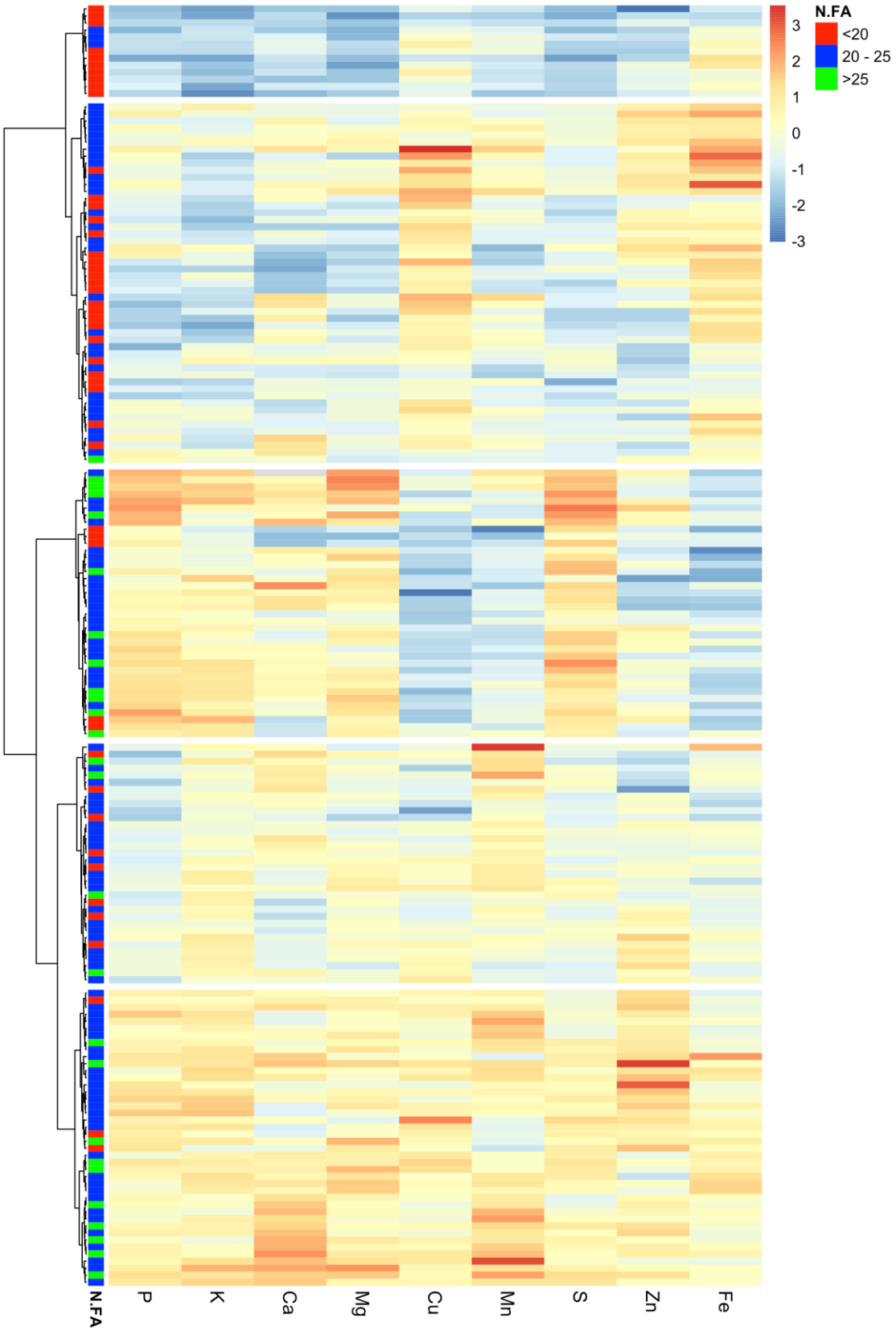


Figure 4.5 Number of favorable alleles (N.FA) and Ward's hierarchical clustering based on Euclidian distance associated with heatmap for different mineral contents quantified on common bean accessions from the Brazilian Diversity Panel (BDP)**.**



4.4.5 POTENCIAL 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 biding, 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 4.3). 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.

4.5 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.

Table 4.3 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	Candidat 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
		Phvul.008G034500	Iron ion binding
	S09_13611567	Phvul.009G081633	Iron ion binding
		Phvul.009G081700	Iron ion binding

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; 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 (DELFINI et al., 2021; GIOIA et al., 2019; VALDISSER et al., 2016). 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).

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; ZHANG; ZHOU, 2018; FANG et al., 2020; MA et al., 2018; MISRA et al., 2018; ZHANG et al., 2018b). 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 (ALI et al., 2020; HE et al., 2019b).

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).

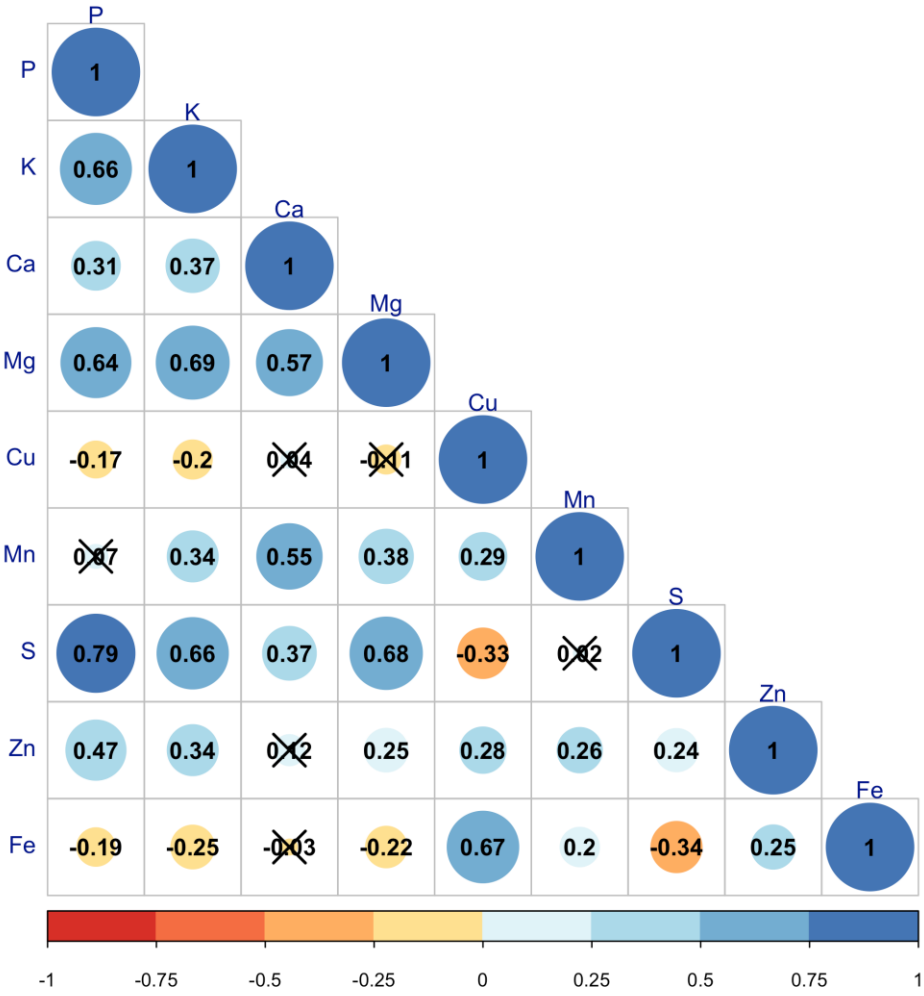
Many of the genes detected in the present study were identified as having the molecular function of protein binding, which may be associated with transduction signaling and the transcription factors that modulate gene expression (VILLORDO-PINEDA et al., 2015). Moreover, other genes not directly related 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 intergenic 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.

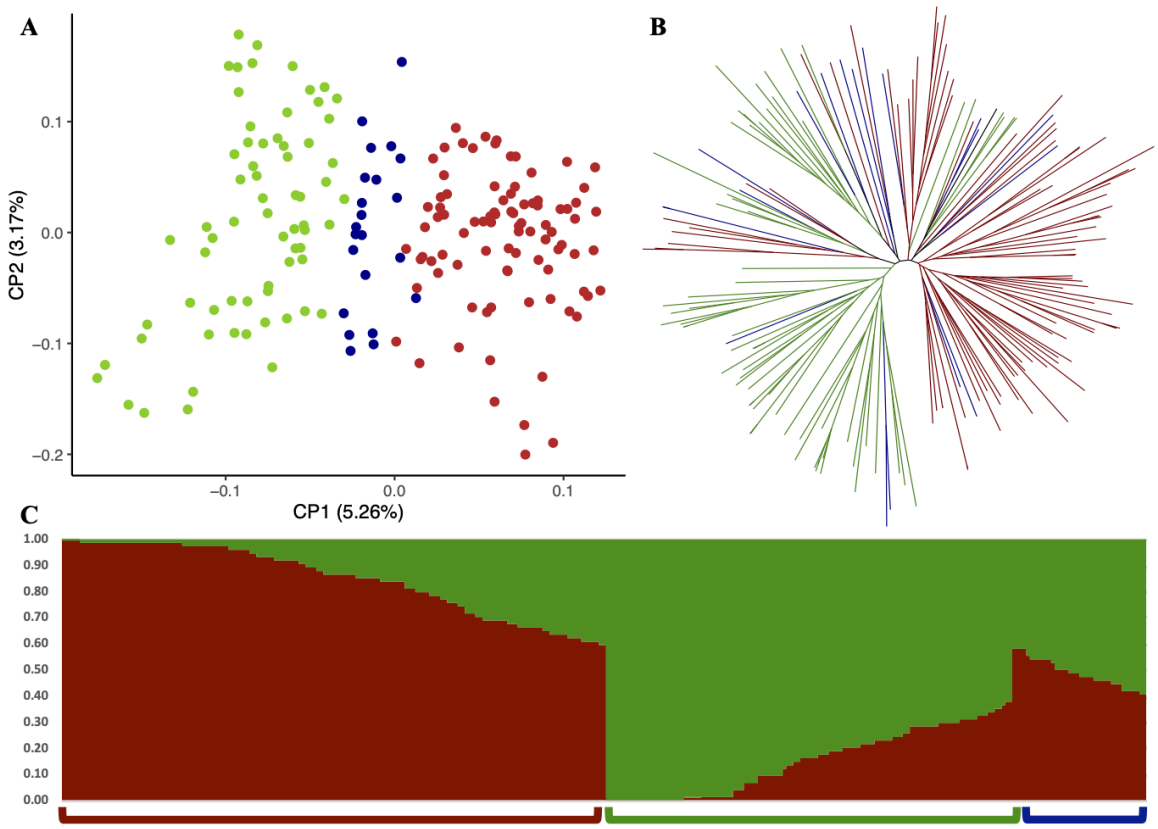
ADDITIONAL INFORMATION

Figure S4.1 Pearson correlation analysis for mineral contents detected in common bean accessions from the Brazilian Diversity Panel (BDP). (X) = not significant at 5% probability.



(X) = not significant at 5% probability

Figure S4.2 Population structure of 178 common bean accessions from the Brazilian Diversity Panel (BDP). **(A)** Principal component analysis (PCA). **(B)** Neighbor-joining tree. (**C)** Population structure.



The three colors represent the three groups formed: (1) red indicates the group formed predominantly by accessions of the black common bean commercial group; green indicates the group formed predominantly by accessions of the carioca common bean commercial group; and (3) blue indicates the group of accessions that did not reach a membership coefficient > 0.6 in either of the previous groups.

Figure S4.3 Analysis of linkage disequilibrium decay (LD decay) according to the physical distance without correction (r^2) (A) and corrected for relationship (r^2_v) (B), population structure (r^2_s) (C), and population structure and relationship (r^2_{sv}) (D).

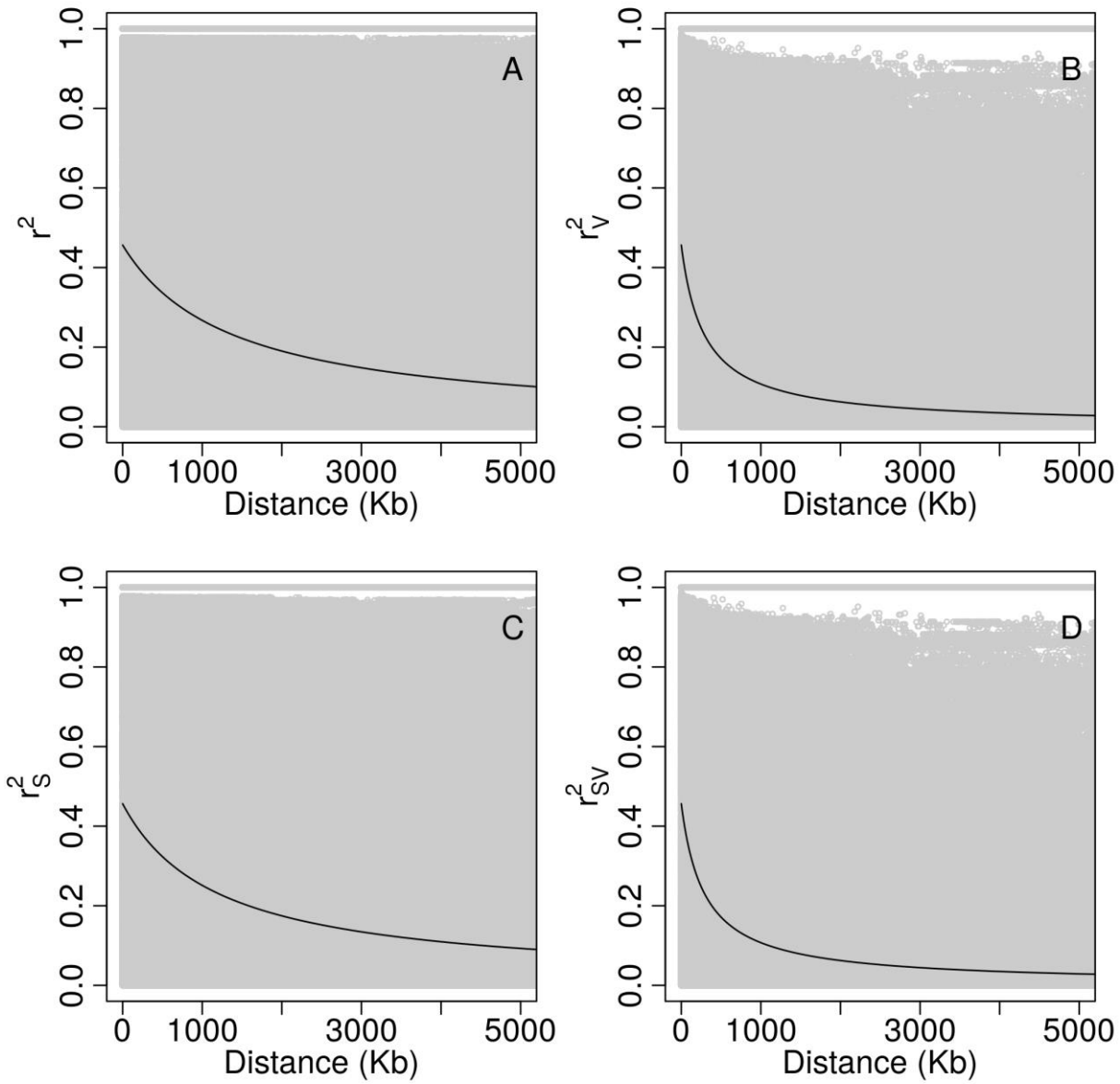


Table S4.1 Mesoamerican accessions from the Brazilian Diversity Panel (BDP) used in the genome-wide association study (GWAS) and structure group to which each access belongs.

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color	Group[1]
Diamante Negro	Mesoamerican	Cultivar	EMBRAPA	Black	1
Aporé	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
BRS Ametista	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
BRS Campeiro	Mesoamerican	Cultivar	EMBRAPA	Black	1
BRS Esplendor	Mesoamerican	Cultivar	EMBRAPA	Black	1
BRS Expedito	Mesoamerican	Cultivar	EMBRAPA	Black	3
BRS FC104	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
BRS FC402	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
BRS FP403	Mesoamerican	Cultivar	EMBRAPA	Black	1
BRS Grafite	Mesoamerican	Cultivar	EMBRAPA	Black	1
BRS Horizonte	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
BRS Notável	Mesoamerican	Cultivar	EMBRAPA	Carioca	1
BRS Pitanga	Mesoamerican	Cultivar	EMBRAPA	Purple	1
BRS Pontal	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
BRS Requinte	Mesoamerican	Cultivar	EMBRAPA	Carioca	3
BRS Supremo	Mesoamerican	Cultivar	EMBRAPA	Black	1
BRS Valente	Mesoamerican	Cultivar	EMBRAPA	Black	1
BRS MGMadrepérola	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
BRS MGPioneiro	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
Bambuí	Mesoamerican	Cultivar	EMBRAPA	Cream	3
Guapo Brilhante	Mesoamerican	Cultivar	EMBRAPA	Black	1
Macanudo	Mesoamerican	Cultivar	EMBRAPA	Black	2
Minuano	Mesoamerican	Cultivar	EMBRAPA	Black	3
Ouro Negro	Mesoamerican	Cultivar	UFV/EPAMIG	Black	1
Pérola	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
Rudá	Mesoamerican	Cultivar	EMBRAPA	Carioca	2
Safira	Mesoamerican	Cultivar	EMBRAPA	Red	1
FT120	Mesoamerican	Cultivar	FT Sementes	Black	1
FT NOBRE	Mesoamerican	Cultivar	FT Sementes	Black	1
FT 41	Mesoamerican	Cultivar	FT Sementes	Black	1
FT 65	Mesoamerican	Cultivar	FT Sementes	Carioca	2
FT Soberano	Mesoamerican	Cultivar	FT Sementes	Black	1
IAC Akitã	Mesoamerican	Cultivar	IAC	Carioca	2

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color	Group[1]
IAC Alvorada	Mesoamerican	Cultivar	IAC	Carioca	3
IAC Carioca	Mesoamerican	Cultivar	IAC	Carioca	2
IAC Diplomata	Mesoamerican	Cultivar	IAC	Black	1
IAC Imperador	Mesoamerican	Cultivar	IAC	Carioca	2
IAC Milênio	Mesoamerican	Cultivar	IAC	Carioca	2
IAC Sintonia	Mesoamerican	Cultivar	IAC	Carioca	2
IAC Una	Mesoamerican	Cultivar	IAC	Black	1
IAC CariocaAruã	Mesoamerican	Cultivar	IAC	Carioca	2
IAC CariocaPyatã	Mesoamerican	Cultivar	IAC	Carioca	3
IAC Maravilha	Mesoamerican	Cultivar	IAC	Black	1
Moruna	Mesoamerican	Cultivar	IAC	Black	3
IAPAR 16	Mesoamerican	Cultivar	IAPAR	Pintado	2
RAI214	Mesoamerican	Cultivar	IAPAR	Carioca	2
IAPAR 57	Mesoamerican	Cultivar	IAPAR	Carioca	3
IAPAR 65	Mesoamerican	Cultivar	IAPAR	Black	1
IAPAR 14	Mesoamerican	Cultivar	IAPAR	Carioca	2
IAPAR 20	Mesoamerican	Cultivar	IAPAR	Black	1
IAPAR 31	Mesoamerican	Cultivar	IAPAR	Pintado	2
IAPAR 72	Mesoamerican	Cultivar	IAPAR	Carioca	3
IAPAR 80	Mesoamerican	Cultivar	IAPAR	Carioca	2
IAPAR 81	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR 139 - JuritiClaro	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Andorinha	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Bem-te-vi	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR CamposGerais	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Celeiro	Mesoamerican	Cultivar	IAPAR	Carioca	3
IPR Chopim	Mesoamerican	Cultivar	IAPAR	Black	1
IPR Colibri	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Corujinha	Mesoamerican	Cultivar	IAPAR	Pintado	2
IPR Curió	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Eldorado	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Gralha	Mesoamerican	Cultivar	IAPAR	Black	1
IPR Graúna	Mesoamerican	Cultivar	IAPAR	Black	1
IPR Juriti	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Maracanã	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Nhambu	Mesoamerican	Cultivar	IAPAR	Black	1

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color	Group[1]
IPR Quero-quero	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Sabiá	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Saracura	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Siriri	Mesoamerican	Cultivar	IAPAR	Carioca	2
IPR Tangará	Mesoamerican	Cultivar	IAPAR	Carioca	2
FEB200	Mesoamerican	Breeding Line	CIAT	Carioca	1
IPR Tuiuiú	Mesoamerican	Cultivar	IAPAR	Black	1
IPR Uirapuru	Mesoamerican	Cultivar	IAPAR	Black	1
IPR Urutau	Mesoamerican	Cultivar	IAPAR	Black	3
Rio Iguaçu	Mesoamerican	Cultivar	IAPAR	Black	3
Rio Negro	Mesoamerican	Cultivar	IAPAR	Black	3
Rio Pardo	Mesoamerican	Cultivar	IAPAR	Cream	1
Rio Tibagi	Mesoamerican	Cultivar	IAPAR	Black	1
Rio Red	Mesoamerican	Cultivar	IAPAR	Purple	3
Rio Doce	Mesoamerican	Cultivar	IAPAR	Carioca	2
HF465.63.1	Mesoamerican	Cultivar	IPA	Cream	1
IPA1	Mesoamerican	Cultivar	IPA	Cream	1
IPA10	Mesoamerican	Cultivar	IPA	Black	1
IPA7	Mesoamerican	Cultivar	IPA	Cream	3
IPA74-19	Mesoamerican	Cultivar	IPA	Cream	1
IPA9	Mesoamerican	Cultivar	IPA	Cream	1
Princesa	Mesoamerican	Cultivar	IPA	Carioca	3
TAA Bola Cheia	Mesoamerican	Cultivar	TAA	Carioca	2
TAA Dama	Mesoamerican	Cultivar	TAA	Carioca	2
TAA Gol	Mesoamerican	Cultivar	TAA	Carioca	2
Awauna	Mesoamerican	Cultivar	UEM	Black	1
Flor Diniz	Mesoamerican	Cultivar	UEM	Carioca	2
Rico23	Mesoamerican	Cultivar	UFV	Black	1
Campeão	Mesoamerican	Cultivar	Agristar	Carioca	2
Agronorte09	Mesoamerican	Cultivar	Agronorte	Carioca	2
ICA Pijão	Mesoamerican	Cultivar	ICA	Black	1
ICA Quetzal	Mesoamerican	Cultivar	ICA	Black	1
ICA Tui	Mesoamerican	Cultivar	ICA	Black	1
Iratin	Mesoamerican	Landrace	-	Black	1
Emgopa Ouro	Mesoamerican	Cultivar	Incaper	Cream	3
A775	Mesoamerican	Breeding Line	CIAT	Cream	3

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color	Group[1]
AETE2	Mesoamerican	Cultivar	IAC	Cream	2
ARC1	Mesoamerican	Breeding Line	CIAT	Black	1
ARC2	Mesoamerican	Breeding Line	CIAT	Black	1
BAT1215	Mesoamerican	Breeding Line	CIAT	Red	1
BAT451	Mesoamerican	Breeding Line	CIAT	Black	1
BAT58	Mesoamerican	Breeding Line	CIAT	Black	1
BAT76	Mesoamerican	Breeding Line	CIAT	Black	1
BAT1192	Mesoamerican	Breeding Line	CIAT	Red	1
BAT477	Mesoamerican	Breeding Line	CIAT	Cream	1
Black Hawk	Mesoamerican	Cultivar	MSU	Black	1
Carioca1070	Mesoamerican	Breeding Line	CENA/USP	Carioca	2
DOR191	Mesoamerican	Breeding Line	CIAT	Red	1
DOR351	Mesoamerican	Breeding Line	CIAT	Purple	1
DOR365	Mesoamerican	Breeding Line	CIAT	Carioca	1
DOR445	Mesoamerican	Breeding Line	CIAT	Black	1
DOR483	Mesoamerican	Breeding Line	CIAT	Red	1
EMP250	Mesoamerican	Breeding Line	CIAT	Carioca	2
ESAL583	Mesoamerican	Breeding Line	ESALQ	Carioca	3
FEB149	Mesoamerican	Breeding Line	CIAT	Cream	1
FEB151	Mesoamerican	Breeding Line	CIAT	Cream	1
FEB156	Mesoamerican	Breeding Line	CIAT	Cream	1
FEB159	Mesoamerican	Breeding Line	CIAT	Cream	1
G1261	Mesoamerican	Landrace	CIAT	Red	1
G14866	Mesoamerican	Landrace	CIAT	Black	1
G17666	Mesoamerican	Landrace	CIAT	Mostarda	1
G18141	Mesoamerican	Cultivar	CIAT	Rajado	1
G2358	Mesoamerican	Landrace	CIAT	branco	1
G2676	Mesoamerican	Cultivar	CIAT	Black	1
G3593	Mesoamerican	Landrace	CIAT	Red	1
G4002	Mesoamerican	Landrace	CIAT	Carioca	1
G5285	Mesoamerican	Cultivar	CIAT	Red	1
G5902	Mesoamerican	Landrace	CIAT	Black	1
FEB178	Mesoamerican	Breeding Line	CIAT	Carioca	3
MUS49	Mesoamerican	Breeding Line	CIAT	Red	1
MUS80	Mesoamerican	Breeding Line	CIAT	Red	1
NAB87	Mesoamerican	Breeding Line	CIAT	Black	1

(Continue)

(Continuation)

Access name	Origin	Genetic Material	Developing institution[1]	Seed Color	Group[1]
NEP171	Mesoamerican	Breeding Line	IICA	Black	1
PorrilloSintético	Mesoamerican	Cultivar	CIAT	Black	1
RIZ57	Mesoamerican	Breeding Line	CIAT	Carioca	2
PurpledeMatoGrosso	Mesoamerican	Landrace	-	Purple	2
XAN206	Mesoamerican	Breeding Line	CIAT	Black	1
XAN236	Mesoamerican	Breeding Line	CIAT	Black	1
LP03	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP04	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP11	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP15	Mesoamerican	Breeding Line	IAPAR	Black	3
LP16	Mesoamerican	Breeding Line	IAPAR	Black	1
LP17	Mesoamerican	Breeding Line	IAPAR	Black	1
LP18	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP19	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP20	Mesoamerican	Breeding Line	IAPAR	Black	1
LP22	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP23	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP24	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP25	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP26	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP27	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP28	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP29	Mesoamerican	Breeding Line	IAPAR	Black	1
LP30	Mesoamerican	Breeding Line	IAPAR	Black	1
LP31	Mesoamerican	Breeding Line	IAPAR	Black	1
LP32	Mesoamerican	Breeding Line	IAPAR	Black	1
LP33	Mesoamerican	Breeding Line	IAPAR	Black	1
LP35	Mesoamerican	Breeding Line	IAPAR	Black	1
LP36	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP37	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP38	Mesoamerican	Breeding Line	IAPAR	Carioca	3
LP40	Mesoamerican	Breeding Line	IAPAR	Black	1
LP42	Mesoamerican	Breeding Line	IAPAR	Carioca	2
LP43	Mesoamerican	Breeding Line	IAPAR	Black	1
LP45	Mesoamerican	Breeding Line	IAPAR	Black	1
LP46	Mesoamerican	Breeding Line	IAPAR	Black	1
LP47	Mesoamerican	Breeding Line	IAPAR	Carioca	2

5 ARTIGO C – Genome-wide association study identifies genomic regions for important morpho-agronomic traits in Mesoamerican common bean

5.1 ABSTRACT

The population growth trend in recent decades has resulted in continuing efforts to guarantee food security in which leguminous plants, such as the common beans (*Phaseolus vulgaris* L.), play a particularly important role as they are relatively cheap, and have high nutritional value. To meet this demand for food, the main target for genetic improvement programs is to increase productivity, which is a complex quantitative trait influenced by many other traits. This research aims to identify Quantitative Trait Nucleotides (QTNs) associated with productivity and its components using multi-locus genome-wide association studies. Ten morpho-agronomic traits [plant height (PH), first pod insertion height (FPIH), number of nodules (NN), pod length (PL), total number of pods per plant (NPP), number of locules per pod (LP), number of seeds per pod (SP), total seed weight per plant (TSW), 100-seed weight (W100), and grain yield (YLD)] were evaluated in four environments for 178 Mesoamerican common bean accessions belonging to the Brazilian Diversity Panel. In order to identify stable QTNs, only those identified by multiple methods or in multiple environments were selected. In the identified QTNs, 64 were detected at least thrice by different methods or in different environments, and 39 showed significant phenotypic differences between their corresponding alleles. The alleles that positively increased the corresponding traits, except PH (for which lower values are desired), were considered favorable alleles. PH was the traits most influenced by the accumulation of favorable alleles, showing a 51.7% reduction, while NN, TSW, YLD, FPIH, and NPP increased between 18% and 34%. Identifying QTNs in several environments and by multiple methods reinforces the reliability of the associations obtained and the importance of conducting these studies in multiple environments. Using these QTNs through molecular techniques for genetic improvement, such as marker-assisted selection or genomic selection, can be a strategy to increase common bean production.

Keywords: *Phaseolus vulgaris* L.; *GWAS*; yield; plant breeding

5.2 INTRODUCTION

More than 24 million tons of common beans (*Phaseolus vulgaris* L.) are produced per year worldwide, and the main producing countries are located in Asia and the Americas (RAWAL; NAVARRO, 2019). This crop is mainly grown by small producers, often in low fertility areas with low-level technology, resulting in low mean productivity (BROUGHTON et al., 2003).

Increasing productivity is one of the main objectives of breeding programs. In this context, understanding the genetic constitution related to the productivity and of the production components are the basis for improvement (KAMFWA; CICHY; KELLY, 2015a). In the cultivation of common beans, productivity is related to several morphological, agronomic, and physiological characteristics. The number of pods per plant (NPP), number of seeds per pod (SP), and seed weight are the primary components related to productivity, but other characteristics are also influential, such as growth rate, the capacity of the seeds to absorb photosynthates and plant architecture (ASSEFA et al., 2019; BEEBE et al., 2013; RESENDE et al., 2018a). Physiological characteristics such as days to flowering and maturity also significantly impact adaptability, biomass, and productivity (ZHANG et al., 2015).

Productivity and its components are quantitative traits and are highly influenced by the environment. Thus, understanding the relationship between these traits is very important for directing strategies and efforts in genetic improvement programs (ASSEFA et al., 2019; SINGH; NODARI; GEPTS, 1991). Traditional selection methods in plant breeding require intensive phenotyping fieldwork, with evaluations in several environments and years, resulting in a high cost and time-consuming process (IKRAM et al., 2020). The use of molecular markers can increase efficiency and reduce the costs of phenotyping in plant breeding programs. Using molecular tools makes it possible to identify genomic regions related to the productivity and its components, which can be used in marker-assisted selection (KAMFWA; CICHY; KELLY, 2015a). Genome-Wide Association Studies (GWAS) represent a powerful option for the genetic characterization of quantitative traits and have been widely used to analyze agronomic characteristics in plants (CONTRERAS-SOTO et al., 2017b; CUI; ZHANG; ZHOU, 2018; IKRAM et al., 2020; WANG et al., 2012; WARD et al., 2019; ZHANG et al., 2017b).

GWAS enables the study of different regions of the genome simultaneously using high-resolution mapping. It is a powerful tool in the identification of multiple polymorphisms that occur naturally within a species, being based on germplasm collections with little genetic structure, and genotypes that have traits of interest for breeding programs can be preferably

used (KORTE; ASHLEY, 2013). These studies use linkage disequilibrium (LD) with the population structure's knowledge to discover important genetic factors. Large and highly diverse association panels have unique recombination histories, allowing the detection of small and large genetic effects associated with a particular trait (OLADZAD et al., 2019).

Although the statistical power in the detection of Quantitative Trait Nucleotides (QTNs) improves after controlling the polygenic background of the population under study, most of the small effects associated with complex traits are still not captured by the GWAS single-locus methods (CUI; ZHANG; ZHOU, 2018). Single-locus methods perform one-dimensional scan of the genome; that is, they test one marker at a time, using several rigorous significance test corrections for multiple tests, such as the Bonferroni and False Discovery Rate (FDR) tests. However, these methods can be very conservative in eliminating true QTNs (HE et al., 2019a). Multi-locus models are being developed to solve this problem. These models involve a multi-dimensional scanning of the genome, in which the effects of all markers are simultaneously estimated (CUI; ZHANG; ZHOU, 2018). The advantage of these models is that it is unnecessary to perform multiple test corrections; therefore, more markers associated with traits of interest are identified (LI et al., 2018a).

Several studies seeking to identify allelic variations responsible for traits directly or indirectly related to productivity have already been conducted for common beans (KAMFWA; CICHY; KELLY, 2015a; LEI et al., 2020; MOGHADDAM et al., 2016; NASCIMENTO et al., 2018; NEMLI et al., 2014; RESENDE et al., 2018a; SOLTANI et al., 2016; WU et al., 2020). Knowing that abiotic factors, like drought and high temperatures, directly influence production, studies on plant behavior under stress conditions were also conducted (HOYOS-VILLEGAS; SONG; KELLY, 2017; KELLER et al., 2020; OLADZAD et al., 2019).

Common beans of Mesoamerican origin are the most consumed in Brazil, with a preference for the Carioca and Black commercial classes. Few GWAS studies were directed towards exclusive panels of Mesoamerican origin and plants adapted to the Brazilian climatic conditions. Studies of genetic variation in accessions adapted to the target habitats are a powerful and effective approach to investigate the genetic architecture of complex traits, and later these natural allelic variations can be directly employed in breeding programs (NAKANO; KOBAYASHI, 2020). Moreover, multi-locus methods were little explored in the cultivation of common beans. In this context, the present study's objective was to identify genomic regions related to morpho-agronomic traits in Mesoamerican common beans belonging to the Brazilian Diversity Panel (BDP) using the GWAS multi-locus methods.

5.3 MATERIAL AND METHODS

5.3.1 GENETIC MATERIAL, FIELD EXPERIMENTS, AND PHENOTYPING

In all, 178 Mesoamerican common bean accessions belonging to the BDP were evaluated (DELFINI et al., 2021). This panel is constituted by accessions that represent a large part of the variability present in Brazil and are adapted to tropical growing conditions. The phenotyping was conducted at the research stations of the Instituto de Desenvolvimento Rural do Paraná (IDR–Paraná), located in the state of Paraná, Brazil. The experiments were conducted in two seasons: the 2018 rainy season in the cities of Londrina (LDA_A18), Ponta Grossa (PG_A18), and Guarapuava (GUA_18); and the 2018/2019 dry season in Ponta Grossa (PG_S19), totaling four environments. The experiment used an incomplete block design with replication in sets. Five sets with two repetitions were used, and each set covered 50 treatments, that is, 46 accessions and four checks. Each set consisted of four 2 m long rows, spaced 0.50 m between rows and with a density of 12 plants per linear meter. Management and treatments were conducted according to the technical recommendations for the plant's cultivation.

Seven plants from the two lateral lines of the plot were used to evaluate traits such as plant height (PH, in cm), first pod insertion height (FPIH, in cm), number of nodules (NN, count variable), pod length (PL, in cm), total number of pods per plant (NPP, count variable), number of locules per pod (LP, count variable), number of seeds per pod (SP, count variable), total seed weight per plant (TSW, in g) and 100-seed weight (W100, in g). Grain yield (YLD, in kg ha^{-1} and 13% moisture) was estimated by harvesting the two central lines.

5.3.2 STATISTICAL ANALYSIS OF PHENOTYPIC DATA

An analysis of variance (ANOVA) was conducted using the PROC GLM function in the SAS software (SAS INSTITUTE, 2000). The following mathematical model was used: $Y_{ijkl} = \mu + A_i + S_j + AS_{ij} + R/AS_{kij} + G/S_{lj} + AG/S_{mlj} + e_{ijklm}$, where μ is the general mean, *Ai* 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/S_{lj} is the random effect of the l-th genotype within the j-th set; AG/S_{mlj} is the effect of the interaction of environments and accessions within the j-th set, and e_{ijklm} is the experimental error (HALLAUER; MIRANDA FILHO, 1988).

The means adjusted for each accession in each of the environments as well as the overall mean of all environments were obtained through the LSmeans option of the GLM procedure. The heritability (h^2) was estimated by the equation: $h^2 = \sigma_G^2 \ / \ \sigma_P^2$, where genotypic ($\sigma^2_G$) and phenotypic ($\sigma^2_P$) variances were estimated by the following equations: $\sigma_G^2 = (QM_G - QM_E)$ and $\sigma_P^2 = (QM_G \ / \ ra)$ where *QMG* is the mean square of genotype within sets; QM_E is the mean square of error, r is the number of repetitions, and a is the number of environments. The descriptive analysis was determined by the means adjusted for the two repetitions of each traits in each environment using the PROC UNIVARIATE function in the SAS software. Pearson's simple linear correlations were calculated and graphically presented using the R software (https://www.r-project.org/) using the 'corrplot' package (WEI; SIMKO, 2017).

5.3.3 GENOTYPING AND GENOME WIDE ASSOCIATION STUDY (GWAS)

The genotyping-by-sequencing (GBS) technique was used to obtain the SNPs. The methodology used, as well as the results of the population structure and linkage disequilibrium (LD) analyses, are detailed in previous work (DELFINI et al., 2021). In summary, genotyping was conducted using the restriction enzyme *CviAII* and the data were imputed using Beagle software version 5 (BROWNING; BROWNING, 2016). After quality control, 25,011 SNPs (MAF > 0.05) were used to perform the GWAS analyses.

For conducting the GWAS, mixed multi-locus models were used with the mrMLM.GUI software version 4.0 0 (YA-WEN; PEI; YUAN-MING, 2019). Four different methods were used: mrMLM (WANG et al., 2016), FASTmrMLM (TAMBA; ZHANG, 2018), pLARmEB [(ZHANG et al., 2017a), and ISIS EM-BLASSO [(TAMBA; NI; ZHANG, 2017). The critical values for significant associations were LOD ≥ 3 for all methods. Population structure and the kinship matrix were included in these models to minimize the identification of false positive associations and increase the statistical power of the analyses. The result of $K = 2$ was obtained by the Structure v2.3.4 software (PRITCHARD; STEPHENS; DONNELLY, 2000)100,000 burn-in, 100,000 MCMC, and ten repetitions for hypothetical numbers of subpopulations (K) between 1 and 10], while the kinship matrix was obtained using the mrMLM.GUI software version 4.0.

The phenotypic values used were the adjusted means for each of the four environments and the overall adjusted mean (LDA_18, PG_18, GUA_18, PG_19, and LSmeans). In order to obtain more accurate results, only QTNs that presented repeatability, that is, detected at least

three times by different methods or environments, were considered truly significant and used in the search for favorable alleles and candidate genes.

5.3.4 FAVORABLE ALLELES AND SEARCH FOR CANDIDATE GENES

For each QTN, all accessions were divided into two groups based on the QTN genotype. A t-test was then conducted to test if there was a significant difference in phenotypic mean between the two groups. Only the statistically stable QTNs between the environments, i.e., those that showed significant difference ($P \leq 0.05$) in the phenotypes in at least three of the five environments (LDA_18, PG_18, GUA_18, PG_19, and LSmeans), were used as favorable alleles. The favorable genotype of each QTN was then selected, i.e., the genotype that causes the desired effect according to each trait, and these effects can be positive or negative in the case of PH. Then, the number of favorable alleles was identified at each accession and visualized in a boxplot if the accumulation of these favorable alleles results in phenotypes with more desirable traits.

The identification of candidate genes was conducted at a physical distance of 296 kbp above and below the SNP associated with each of the assessed trait. This distance is the point at which the half decay of the LD calculated with correction by population structure and relatedness (r^2_{vs}) occurs (ARTIGO B). The genes present in the association region with known putative functions according to the GeneOntology (GO, http:// www.geneontology.org/) were identified based on the reference genome annotation of *Phaseolus vulgaris v*.2 published on the Phytozome v10.3 website (http://phytozome.net).

5.4 RESULTS

5.4.1 ANALYSIS OF VARIANCE (ANOVA), HERITABILITY, AND ENVIRONMENTAL EFFECT

The analysis of variance showed a significant effect ($P \leq 0.01$) in the accessions and environments for all traits evaluated (Table 5.1). Significant effects were also observed ($P \leq 0.01$) in the interaction between Genotype x Environment (GE) for the traits PH, LP, SP, W100, and YLD. The coefficients of variation (CV) varied between 4.79 (PL) and 25.62% (TSW). As for heritability estimates (h^2), the TSW, NN, NPP, and FPIH traits presented moderate values, between 0.54 and 0.68, while high values were detected for the other traits. YLD, SP, and LP

had h^2 values between 0.71 and 0.77, and the PH, PL, and W100 traits had the highest values, 0.88, 0.94, and 0.94, respectively.

Comparing the mean performance of the accessions in each of the environments (Figure 5.1), the GUA_18 environment showed the highest general averages for NPP (21.13) and YLD (4,128.8 kg.ha^{-1}), while PG_18 showed the lowest values for traits related to plant morphology: PH (53.46 cm), FPIH (15.19 cm), NN (10.86) and PL (8.98 cm). The LDA_18 environment showed the lowest values for the production components TSW (16.63 cm), W100 (20.95 g), and YLD (2368.78 kg.ha^{-1}). PL (7.13–12.49 cm), LP (5.06–7.92), SP (4.49–7.55), and W100 (14.69–33.48) did not show significant variations in the minimum, maximum, and mean values for each environment.

5.4.2 CORRELATION BETWEEN TRAITS

Significant and positive correlations ($P \leq 0.05$) were observed between PH, FPIH, and NN traits. The PL, LP, and SP traits also correlated positively with each other (Figure 5.2). Positive correlations were also observed between YLD and the primary components TSW and W100 (r = 0.31 and 0.32, respectively). In addition, YLD also correlated positively with the LP (r = 0.24) and SP (r = 0.27) traits, while NPP correlated positively with TSW (r = 0.65). On the other hand, negative correlations were observed between PPN × FPIH (r =−0.33), PPN × PL (r =-0.30), PPN × W100 (r =-0.31) and SP × W100 (r =-0.15).

5.4.3 QTNS IDENTIFIED BY ML-GWAS

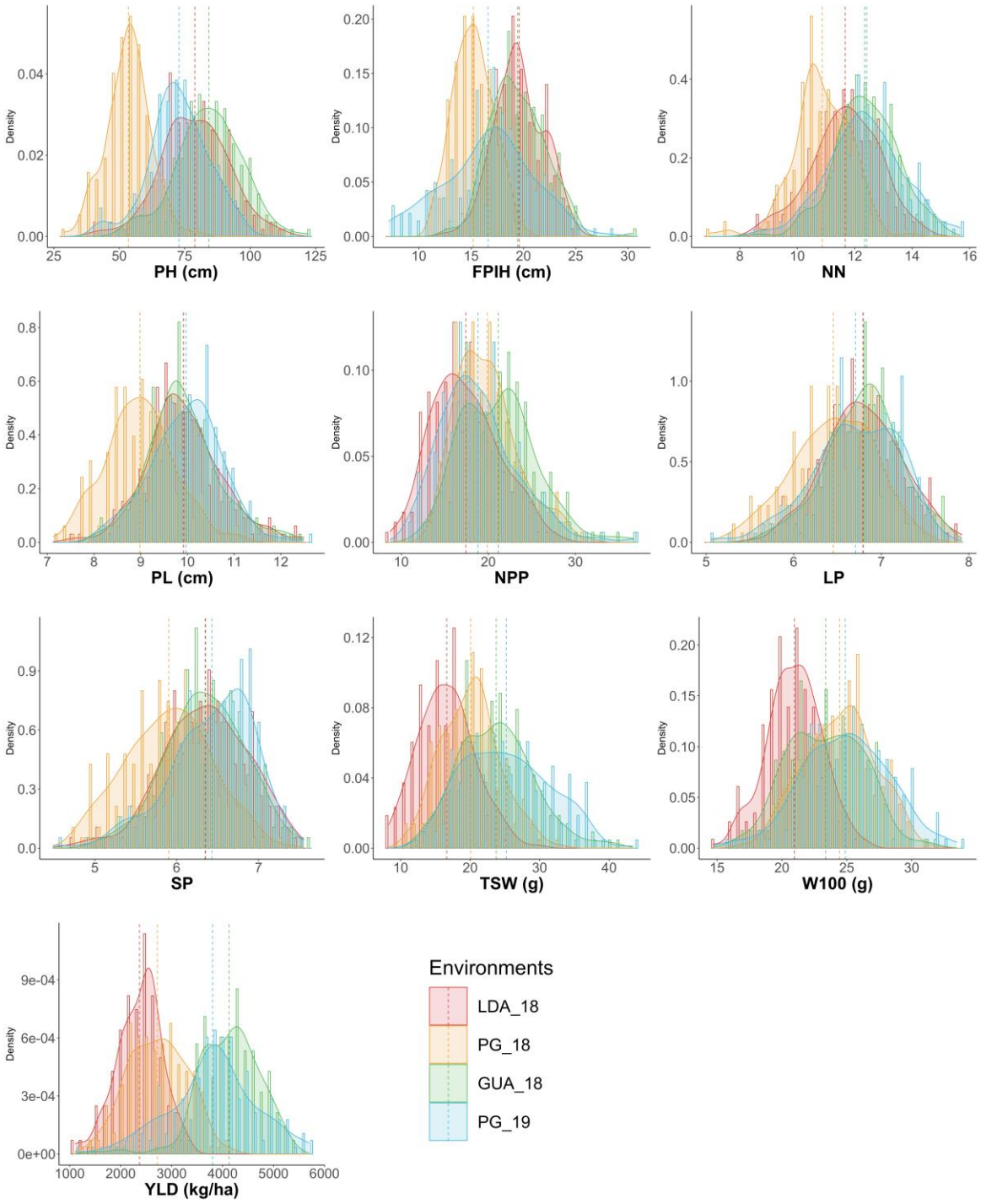
The four ML-GWAS methods identified 297 QTNs associated with the ten morpho-agronomic traits evaluated. Among these, 131 QTNs were detected at least twice by two methods and/or two different environments (Table S5.1), while 64 QTNs were detected at least three times by multiple tests and/or multiple environments (Table 5.2). The highest number of QTNs was observed on the Pv01 and Pv08 chromosomes (nine significant QTNs each), followed by Pv02 and Pv11 (eight QTNs each). Only the 64 QTNs that presented repeatability at least three times were considered reliable and followed in the study.

Table 5.1 Analysis of variance and descriptive statistics for morpho-agronomic traits evaluated in common bean accessions belonging to the Brazilian Diversity Panel (BDP) evaluated in four environments.

		PH[1]	FPIH	NN	PL	NPP	LP	SP	TSW	W100	YLD
F_{env}[2]		827.46***	193.95***	120.58***	369.85***	51.05***	53.95***	88.47***	165.96***	373.06***	627.45***
F_{set}		13.14***	17.57***	9.99***	47.87***	0.85ns	34.55***	24.52***	6.82***	1.42ns	19.15***
$F_{env*set}$		5.75***	9.78***	11.99***	14.01***	8.84***	10.79***	15.13***	11.58***	8.08***	4.89***
$F_{rep(env*set)}$		13.91***	11.04***	7.21***	6.64***	10.25***	10.27***	9.51***	8.69***	5.71***	14.55***
$F_{treat(set)}$		8.17***	3.13***	2.34***	15.78***	2.77***	4.34***	4.06***	2.18***	16***	3.45***
$F_{env*treat(set)}$		1.44***	1.11ns	1.07ns	1.11ns	1.09ns	1.23**	1.26**	1.13ns	1.61***	1.29***
CV(%)[3]		11.77	16.20	10.49	4.79	22.67	6.13	7.38	25.62	7.16	19.10
Heritability (h^2)		0.88	0.68	0.57	0.94	0.64	0.77	0.75	0.54	0.94	0.71
Mean	LDA_18	78.84	19.58	11.67	9.91	17.40	6.79	6.35	16.63	20.95	2368.78
	PG_18	53.46	15.19	10.86	8.98	19.87	6.45	5.90	20.05	24.43	2718.92
	GUA_18	84.26	19.44	12.41	9.91	21.13	6.79	6.35	23.74	23.37	4128.83
	PG_19	72.84	16.61	12.34	9.97	18.80	6.70	6.43	25.18	24.88	3801.88
Minimum	LDA_18	12.81	2.34	1.18	0.81	3.87	0.44	0.52	4.03	2.11	437.90
	PG_18	8.38	1.88	1.00	0.74	3.56	0.46	0.52	4.16	2.73	575.14
	GUA_18	12.55	2.65	1.09	0.77	4.30	0.42	0.48	5.32	2.93	605.74
	PG_19	11.33	4.21	1.32	0.77	4.42	0.49	0.52	6.18	3.31	852.26
Maximum	LDA_18	41.11	12.46	8.61	7.52	8.69	5.55	4.73	8.31	14.69	1110.91
	PG_18	28.84	11.23	6.95	7.13	13.18	5.28	4.75	10.74	17.15	1189.14
	GUA_18	50.48	12.54	8.70	8.10	12.09	5.60	5.01	12.31	16.12	1480.22
	PG_19	41.03	7.13	8.49	8.03	10.45	5.06	4.49	13.63	15.30	1114.61
Skewness	LDA_18	113.26	25.14	14.43	12.47	27.82	7.92	7.48	29.63	26.35	3358.46
	PG_18	81.06	20.18	13.90	11.10	31.37	7.47	7.33	32.59	31.26	4118.57
	GUA_18	122.00	30.44	15.15	12.30	36.43	7.71	7.55	40.73	31.44	5336.69
	PG_19	99.65	28.18	15.67	12.49	36.70	7.68	7.45	43.38	33.48	5691.76
Standard Deviation	LDA_18	-0.072	-0.032	-0.269	0.290	0.363	0.001	-0.286	0.323	-0.116	-0.360
	PG_18	0.035	0.150	-0.486	0.207	0.720	-0.246	0.020	0.242	-0.089	-0.054
	GUA_18	-0.128	0.432	-0.104	0.471	0.567	-0.212	-0.149	0.461	0.071	-0.832
	PG_19	-0.246	-0.147	-0.049	-0.035	0.859	-0.553	-0.776	0.289	-0.082	-0.245
Curtose	LDA_18	0.270	-0.220	-0.070	0.728	-0.348	-0.143	0.045	-0.031	0.022	0.098
	PG_18	0.489	-0.455	1.521	0.101	0.260	-0.383	-0.413	-0.083	-0.465	-0.473
	GUA_18	0.272	0.858	0.300	0.624	0.548	0.064	-0.101	0.324	-0.401	2.382
	PG_19	0.351	-0.244	0.174	0.288	1.252	0.156	0.690	-0.626	-0.013	0.184

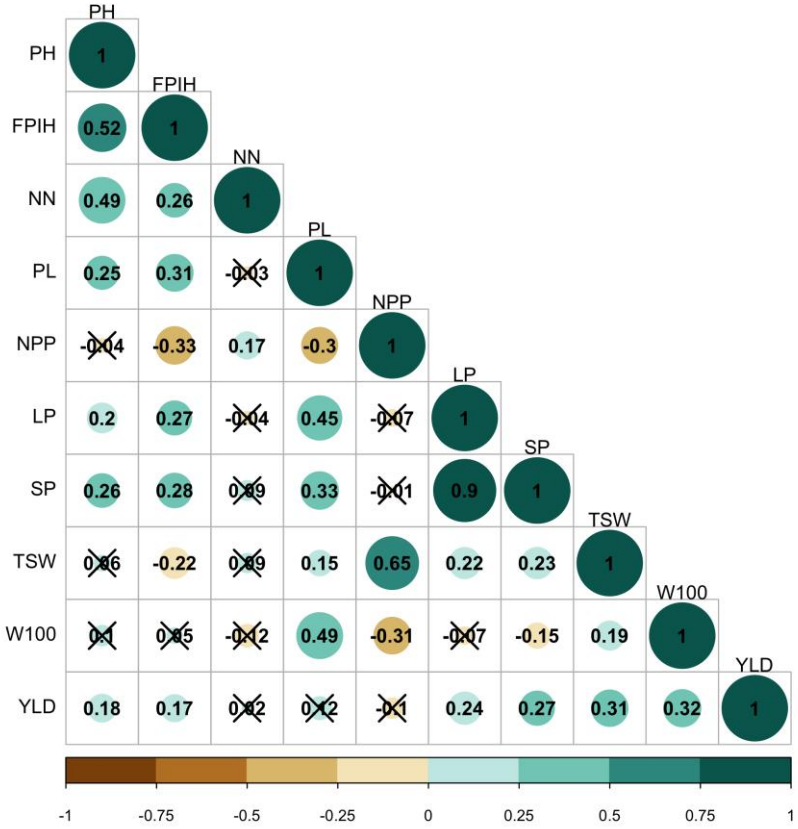
[1]PH = plant height (cm), FPIH = first pod insertion height (cm), NN = number of nodules, PL = pod length (cm), NPP = total number of pods per plant, LP = number of locules per pod, SP = number of seeds per pod, TSW = total seed weight per plant (gm), W100 = 100-seed weight (gm) and YLD = grain yield (kg.ha^{-1}). [2]F_{amb}, F_{set}, $F_{amb*set}$, $F_{rep(amb*set)}$, $F_{trat(set)}$, $F_{amb*trat(set)}$ represent 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; [3]CV (%) = coefficient of variation. *P < 0.01, **P < 0.001, ***P < 0.0001, and ns not significant.

Figure 5.1 Frequency distribution of the ten morpho-agronomic traits evaluated in different locations and seasons in common bean accessions belonging to the Brazilian Diversity Panel (BDP).



PH = plant height (cm), FPIH = first pod insertion height (cm), NN = number of nodules, PL = pod length (cm), NPP = total number of pods per plant, LP = number of locules per pod, SP = number of seeds per pod, TSW = total seed weight per plant (gm), W100 = 100-seed weight (gm) and YLD = grain yield ($kg.ha^{-1}$). Environments, 2018 rainy season crops: LDA_A18 = Londrina, PG_A18 = Ponta Grossa and GUA_18 = Guarapuava; 2018/2019 dry season crop: PG_S19 = Ponta Grossa.

Figure 5.2 Pearson's correlation analysis for morpho-agronomic traits of a common bean accession belonging to the Brazilian Diversity Panel (BDP).



(X) = not significant at 5% probability. PH = plant height (cm), FPIH = first pod insertion height (cm), NN = number of nodules, PL = pod length (cm), NPP = total number of pods per plant, LP = number of locules per pod, SP = number of seeds per pod, TSW = total seed weight per plant (gm), W100 = 100-seed weight (gm) and YLD = grain yield (kg.ha^{-1}).

The 64 QTNs identified explained a low percentage of phenotypic variation (PVE): PH (*n* = 11; PVE = 1.4−11.3%), FPIH (*n* = 4; PVE = 3.03−10.05%), NN (*n* = 11; PVE = 3.33.10^{-8}−11.33%), PL (*n* = 8; PVE = 1.08−13.21%), NPP (*n* = 6; PVE = 2.55−13.04%), LP (*n* = 8; PVE = 7.03.10^{-8}−9.48%), SP (*n* = *2;* PVE = 2.5−11.2%), TSW (*n* = 6; PVE = 2.89−15.42%), W100 (*n* = 2; PVE = 4.04−12.3%), and YLD (*n* = 6; PVE = 1.47−13.31%). Two QTNs were considered pleiotropics, since they were identified in more than one trait, being them PL-LP (*S08_9375624*) and NPP-TSW (*S11_1617681*) localized in the Pv08 and Pv11 chromosomes, respectively.

Table 5.2 QTNs associated with morpho-agronomic traits detected at least three times via different methods and in different environments in common bean accessions belonging to the Brazilian Diversity Panel (BDP).

Trait.	SNP	Chr	Position (bp)	QTN Effect[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]	T-test[7]
PH	S01_5305887	1	5305887	-8.13 ~ -6.83	4.78 ~ 6.56	6.55 ~ 9.33	0.063	CC	1	1,2,3	1,5
	S04_2493297	4	2493297	2.59 ~ 3.9	3.49 ~ 5.09	2.2 ~ 4.18	0.152	GG	1,5	2,4	1,2,3,5
	S04_373114	4	373114	-3.34 ~ -1.73	3.84 ~ 5.66	1.4 ~ 4.6	0.433	GG	3,5	1,4	1,3,5
	S05_39604389	5	39604389	1.65 ~ 3.76	3.04 ~ 6.12	2.9 ~ 8.19	0.455	GG	1,2,3,5	1,2,3,4	1,2,3,4,5
	S05_39680093	5	39680093	3.16 ~ 4.28	4.86 ~ 7.85	5.51 ~ 10.08	0.352	GG	1	1,2,3	1,2,3,4,5
	S06_25397668	6	25397668	2.43 ~ 2.81	3.43 ~ 5.09	2.57 ~ 8.4	0.354	CC	5	1,3,4	1,2,4,5
	S07_17942068	7	17942068	5.59 ~ 7.5	5.24 ~ 5.92	2.85 ~ 6.82	0.051	CC	3,4,5	1,4	1,2,3,4,5
	S08_62021856	8	62021856	4.08 ~ 8.43	4.52 ~ 6.2	1.84 ~ 9.62	0.062	GG	3,5	3,4	1,2,3,4,5
	S10_43645690	10	43645690	2.16 ~ 4.45	4.17 ~ 6.92	3.97 ~ 8.46	0.242	CC	3,5	1,2,3,4	2,3,5
	S10_44036828	10	44036828	-2.83 ~ -1.69	3.04 ~ 4.97	3.89 ~ 10.61	0.393	TT	2	2,3,4	1,2,3,5
	S11_1465100	11	1465100	2.69 ~ 3.06	4.78 ~ 5.57	9.01 ~ 11.3	0.318	AA	2	1,2,3	2,3,4,5
FPIH	S01_1292307	1	1292307	0.55 ~ 0.59	3.35 ~ 3.43	4.77 ~ 5.28	0.1761	AA	5	1,3,4	4,5
	S01_20991636	1	20991636	-0.46 ~ -0.37	3.33 ~ 3.8	3.62 ~ 5.53	0.4375	CC	5	2,3,4	5
	S06_20814429	6	20814429	-0.61 ~ -0.49	3.84 ~ 6.86	6.52 ~ 10.05	0.483	AA	5	1,2,4	5
	S09_27171634	9	27171634	0.51 ~ 0.78	3.09 ~ 3.63	3.03 ~ 6.93	0.118	CC	2	2,3,4	2
NN	S01_5448199	1	5448199	0.14 ~ 0.17	3.33 ~ 4.96	2.67 ~ 4.24	0.2584	CC	5	1,2,3	2,4,5
	S02_25464609	2	25464609	0.26 ~ 0.36	4 ~ 4.71	4.59 ~ 8.52	0.4148	GG	1	2,3,4	1
	S02_48537121	2	48537121	0.17 ~ 0.26	3.03 ~ 3.67	2.4 ~ 5.55	0.2898	CC	2	1,3,4	2,5
	S04_2503984	4	2503984	0.13 ~ 0.29	3.67 ~ 5.43	3.03 ~ 7.81	0.3807	CC	2,5	1,2,3	2,3,4,5
	S07_474203	7	474203	0.12 ~ 0.19	3.12 ~ 4.81	2.07 ~ 5	0.2841	TT	5	1,3,4	1,2,4,5
	S07_495888	7	495888	-0.35 ~ 0	3.61 ~ 6.22	0 ~ 10.76	0.3523	GG	2	1,2,3,4	2,5
	S08_44008378	8	44008378	-0.47 ~ -0.32	3.84 ~ 5.58	3.42 ~ 7.1	0.0909	CC	2	1,2,3,4	2,3,4,5
	S08_61614494	8	61614494	0.15 ~ 0.25	3.93 ~ 7.72	4.14 ~ 11.33	0.4045	TT	5	1,2,3	3,4,5
	S09_5349386	9	5349386	0.25 ~ 0.3	3.07 ~ 5.28	2.73 ~ 7.29	0.125	GG	5	1,2,3,4	1,2,3,4,5
	S10_44010107	10	44010107	-0.26 ~ -0.23	3.48 ~ 4.48	4.79 ~ 5.96	0.3708	CC	2	1,2,3	2,5
	S11_8369504	11	8369504	-0.19 ~ -0.14	3.3 ~ 4.87	2.9 ~ 5.43	0.264	AA	5	1,2,3	
PL	S02_1040748	2	1040748	-0.26 ~ -0.19	3.43 ~ 4.19	1.02 ~ 4.9	0.1023	TT	2	1,3,4	1,2,3,5
	S02_47586597	2	47586597	-0.33 ~ -0.22	4.92 ~ 8.61	1.63 ~ 10.93	0.1685	AA	2,3,5	1,2,3,4	1,2,3,5
	S02_49538733	2	49538733	0.13 ~ 0.23	3.1 ~ 5.31	2.9 ~ 8.78	0.5	GG	1,2,3	1,2,3	1,2,3
	S07_30515591	7	30515591	-0.31 ~ -0.23	3 ~ 3.65	3.91 ~ 7.18	0.1292	GG	3	1,2,3	1,2,3,5
	S08_1159693	8	1159693	-0.23 ~ -0.19	3.48 ~ 4.51	1.52 ~ 5.23	0.1591	CC	2	1,3,4	1,2
	S08_62432046	8	62432046	-0.23 ~ -0.16	4.01 ~ 4.24	1.08 ~ 5.14	0.1534	TT	2	1,2,3,4	2
	S08_9375624	8	9375624	-0.54 ~ -0.41	5.89 ~ 6.76	2.25 ~ 13.21	0.0571	CC	5	2,3,4	1,2,3,4,5
	S11_2437959	11	2437959	0.17 ~ 0.18	3.27 ~ 3.39	3.5 ~ 3.89	0.2429	AA	4	1,2,4	2,3,4,5

(Continue)

(Continuation)

Trait.	SNP	Chr	Position (bp)	QTN Effect[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]	T-test[7]
NPP	S03_12044967	3	12044967	-2.28 ~ -1.62	3.28 ~ 4.46	4.9 ~ 9.69	0.0966	AA	3	1,2,3,4	3,5
	S05_40466290	5	40466290	1.04 ~ 1.41	3.12 ~ 5.36	3.96 ~ 7.21	0.2102	TT	3	1,2,3,4	3
	S05_445417	5	445417	-1.43 ~ -0.92	3.11 ~ 4.96	2.55 ~ 6.02	0.0562	AA	5	1,2,3	3,5
	S07_38456082	7	38456082	-1.56 ~ -0.48	3.11 ~ 3.85	3.31 ~ 11.71	0.4432	GG	3,4,5	1,2,3,4	3
	S08_2493035	8	2493035	-2.45 ~ -2.07	5.78 ~ 6.16	4.1 ~ 11.68	0.1067	CC	4	1,2,4	1,4,5
	S11_1617681	11	1617681	-0.97 ~ -0.73	4.94 ~ 7.45	7.5 ~ 13.04	0.4602	AA	5	1,2,3,4	1,2,3,4,5
LP	S01_221817	1	221817	0.07 ~ 0.14	3.3 ~ 5.73	1.92 ~ 7.72	0.2921	GG	1	1,2,4	1,5
	S02_41632778	2	41632778	0 ~ 0.18	3.03 ~ 3.78	0 ~ 6.56	0.0629	TT	5	1,3,4	2,4,5
	S07_33862545	7	33862545	-0.17 ~ -0.1	3.41 ~ 5.07	3.24 ~ 9.48	0.1854	AA	3	1,2,3	2,3,4,5
	S08_9375624	8	9375624	-0.24 ~ -0.15	3.02 ~ 4.86	4.08 ~ 9.32	0.0514	CC	5	1,2,3,4	2,3,5
	S10_4876917	10	4876917	0.12 ~ 0.14	3.41 ~ 3.84	5.38 ~ 6.67	0.2045	AA	3	2,3,4	2,3,5
	S10_4911729	10	4911729	0.09 ~ 0.12	3.55 ~ 4.21	4.86 ~ 8.64	0.24	GG	5	1,2,3,4	2,3,5
	S11_29062	11	29062	-0.13 ~ -0.07	3.42 ~ 9.02	3.78 ~ 7.81	0.3371	TT	1,5	2,3,4	1,5
	S11_52195944	11	52195944	-0.1 ~ -0.07	3.68 ~ 4.5	4.56 ~ 8.66	0.3616	GG	5	1,2,3	1,5
SP	S01_112055	1	112055	0.12 ~ 0.2	3.97 ~ 4.81	3.56 ~ 11.2	0.2727	CC	1	2,3,4	1,2,4,5
	S08_10350174	8	10350174	-0.37 ~ -0.13	3.27 ~ 4.79	2.5 ~ 10.42	0.0568	GG	2,5	1,2,3,4	1,2,3,5
TSW	S01_44752890	1	44752890	1.98 ~ 2.51	3.53 ~ 4.97	4.17 ~ 8.25	0.1486	CC	4	1,2,3,4	2,4,5
	S02_2242481	2	2242481	-1.92 ~ -1.31	3.4 ~ 6.12	2.89 ~ 6.49	0.2147	TT	4	1,2,4	1,4,5
	S04_41451220	4	41451220	-2.03 ~ -1.53	3.18 ~ 5.58	5.77 ~ 10.15	0.1136	CC	1	2,3,4	1,5
	S10_18589262	10	18589262	-3.58 ~ -1.56	3.48 ~ 4.62	4.49 ~ 8.69	0.0506	GG	3,5	1,2,3,4	3,5
	S11_1617681	11	1617681	-1.21 ~ -0.73	3.75 ~ 7.98	5.09 ~ 15.42	0.4571	AA	5	1,2,3,4	2,4,5
	S11_52088416	11	52088416	-1.75 ~ -1.48	4.14 ~ 4.14	5.38 ~ 7.53	0.2247	AA	3	1,2,3	3,5
W100	S03_11484802	3	11484802	-0.76 ~ -0.69	4.42 ~ 5.35	5.42 ~ 6.48	0.4034	CC	3	1,2,3,4	1,2,3,4,5
	S03_4682037	3	4682037	0.96 ~ 1.68	3.33 ~ 7.64	4.04 ~ 12.3	0.0686	CC	5	2,3,4	1,4,5
YLD	S01_44911599	1	44911599	360.06 ~ 529	3.93 ~ 6.9	6.17 ~ 13.31	0.0966	AA	4	1,2,3,4	1,2,3,4,5
	S01_51067135	1	51067135	77.81 ~ 123.27	3.53 ~ 4.87	2.65 ~ 6.45	0.2898	GG	5	1,2,3,4	4,5
	S02_34513049	2	34513049	-186.17 ~ -114.43	3.33 ~ 4.19	1.47 ~ 5.84	0.1875	TT	3	2,3,4	1,2,3
	S03_2802438	3	2802438	113.41 ~ 141.54	3.74 ~ 5.64	6.09 ~ 9.49	0.3466	CC	1	2,3,4	1,3,4,5
	S03_49731981	3	49731981	-176.06 ~ -133.48	4.83 ~ 5.89	5.24 ~ 11.65	0.2443	CC	5	1,2,3,4	5
	S07_34450891	7	34450891	166.38 ~ 260.05	3.13 ~ 8.89	3.77 ~ 12.15	0.1307	AA	2,4,5	1,2,3,4	1,2,4,5

PH = plant height (cm), FPIH = first pod insertion height (cm), NN = number of nodules, PL = pod length (cm), NPP = total number of pods per plant, LP = number of locules per pod, SP = number of seeds per pod, TSW = total seed weight per plant (gm), W100 = 100-seed weight (gm) and YLD = grain yield (kg.ha^{-1}). [1]*Quantitative trait nucleotide effect;* [2]LOD value, the significant threshold for *P-value* transformed; [3]PVE (%): *Phenotypic variation explained*; [4]*Minor allele frequency*; [5]Environments: 1-LDA, 2-PG, 3-GUA, 4-LSmeans; [6]Methods: 1-FASTmrMLM, 2-ISIS EM-BLASSE, 3-mrMLM, 4-pLARmEB; [7]QTNs with significant effect on the t-test. Pleiotropic QTNs, related to more than one mineral, are in bold.

Besides the identification of pleiotropic QTNs, 27 QTNs presented an overlap of the genomic region. In the Pv05, Pv07, and Pv10 chromosomes, QTNs that overlap for the same trait were observed for PH, NN, and LP, respectively. The PH and NN traits shared the genomic region around the QTNs in three chromosomes: Pv01, Pv04, and Pv10. Other overlaps were observed for SP-LP (Pv01), TSW-YLD (Pv01), W100-NPP (Pv03), LP-YLD (Pv07), PH-LP-PL (Pv08), TSW-LP (Pv11). In addition, the pleiotropic QTN identified for NPP-TSW (Pv11) shared the genomic region with a QTN identified for PH.

The highest number of QTNs was identified in the LSmeans dataset, followed by PG_18, GUA, 18, LDA_18, and PG_19. In the GWAS multi-locus methods, the ISIS-EM-BLASSO method detected the highest number of SNPs, followed by pLARmEB, mrMLM, and FASTmrMLM. Considering only the 64 reliable QTNs, the environment ranking remained the same; that is, LSmeans was the environment that detected the highest number of QTNs. For the methods, the number of stable QTNs detected was similar among the different methodologies, varying between 53 and 61. Looking at the efficiency of these methods, in other words, the number of QTNs considered reliable in relation to the initial number, the FASTmrMLM method stood out from the others (56%), followed by mrMLM (47%), pLARmEB (40%), and ISIS-EM-BLASO (34%).

5.4.4 IDENTIFICATION OF FAVORABLE ALLELIC VARIATIONS AND CANDIDATE GENES

In the 64 QTNs considered reliable, 39 presented significant results for the t-test in at least three environments and were considered stable (Table 5.2). These stable QTNs were used to investigate alleles favorable for the traits PH (*n* = 10), NN (*n* = 6), PL (n = 6), NPP (*n* = 2), LP (*n* = 4), SP (*n* = 2), TSW (n = 3), W100 (*n = 2),* and YLD (*n* = 4). For FPIH, stable QTNs following the established criteria were not observed. Thus, only for this trait, three QTNs that presented significant results through the t-test for the overall mean (LSmeans) were used, resulting in a total of 42 stable QTNs (Figure 5.3).

All QTNs that had a positive effect (increased values) were considered as favorable alleles. The only exception was PH, for which common bean breeding programs search for plants with smaller size. The accumulation of these alleles in the same accession resulted in a gradual increase, or decrease in the case of PH, in all traits (Figure 5.3). Comparing the mean values between genotypes with zero and those with the maximum number of favorable alleles, PH was the most influenced trait, revealing a difference of 51.7% between the two genotype groups, reducing the values from 92.8 to 48 cm. For those traits where higher values are

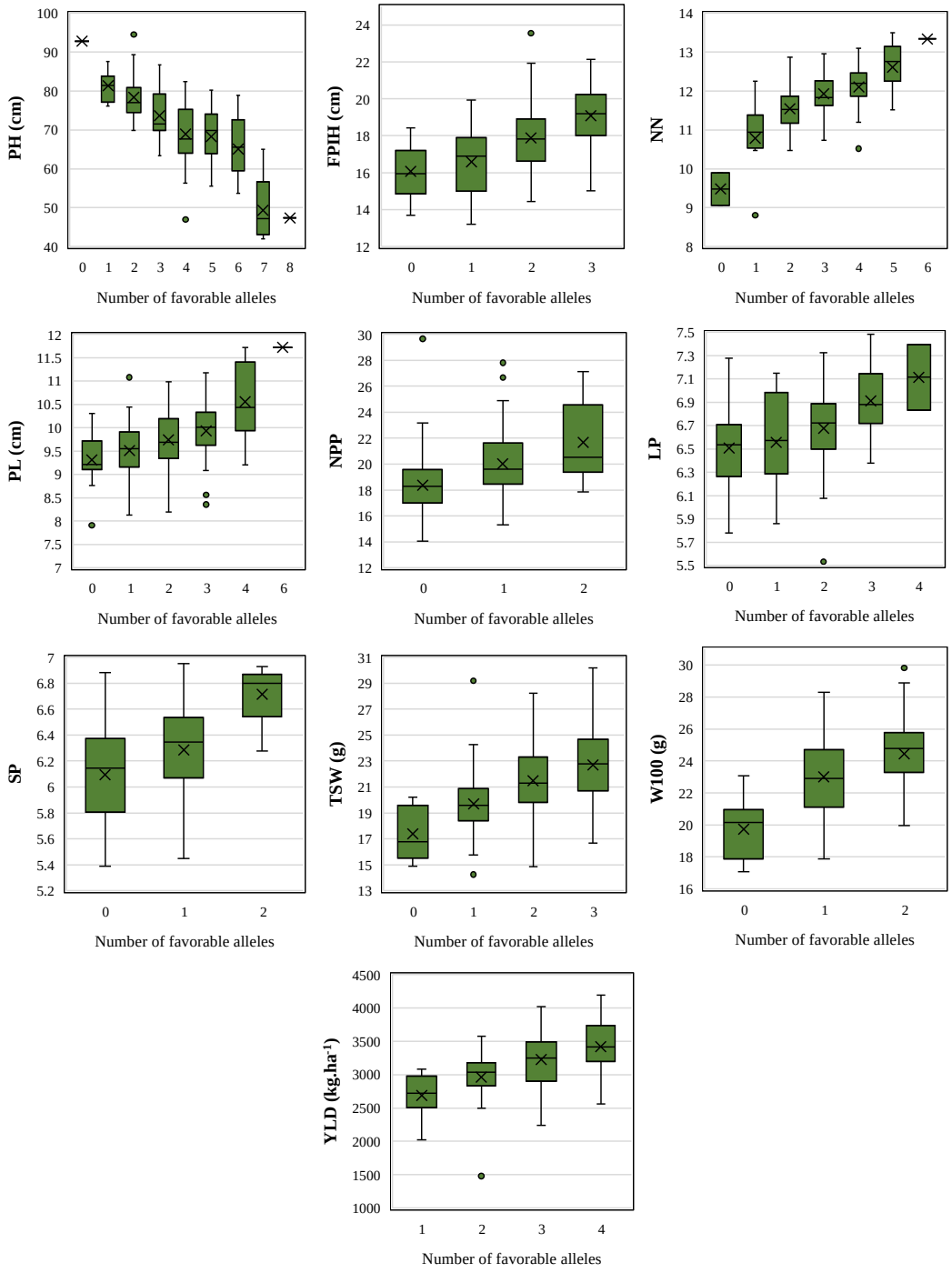
favorable, the greatest increase was observed for NN (34%, from 9.47 to 12.7), followed by TSW (30%, from 17.4 to 22.7 gm), YLD (27%, from 2,688 to 3,419 kg.ha^{-1}), W100 (24%, from 19.7 to 24.5 g), FPIH (19%, from 16 to 19 cm), NPP (18%, from 18.4 to 21.7), PL (14%, from 9.3 to 10.6 cm), LP (10%, from 6.5 to 7.14), and SP (10%, from 6.1 to 6.7).

The distance determined using LD decay (296 kb) was used to select potential candidate genes at a specific QTN distance. Since the search was conducted in a large genomic region around the QTNs, many genes were identified for the ten traits evaluated in this study, resulting in 1,528 genes with known putative functions, and of these, 74% were identified more than once in regions of overlap between QTNs. In the functions of the identified genes, 55% had a molecular function, 32% had functions related to biological processes, and 13% were cellular components. In the molecular function category, the main functions detected were protein binding, ATP binding, and protein kinase activity; for biological processes, the functions were related to protein phosphorylation, oxidation-reduction processes, and transcription regulation, and for cellular components, the functions were related to the membrane and integral components of the membrane.

5.5 DISCUSSION

Although several studies already identified QTNs associated with morpho-agronomic traits in common beans using GWAS (KAMFWA; CICHY; KELLY, 2015a; LEI et al., 2020; MOGHADDAM et al., 2016; NASCIMENTO et al., 2018; NEMLI et al., 2014; RESENDE et al., 2018a; SOLTANI et al., 2016; WU et al., 2020), panels composed exclusively of common beans of Mesoamerican origin and adapted to environmental conditions in Brazil have not been sufficiently explored (VALDISSER et al., 2020). Moreover, few studies using the GWAS multi-locus approach have been conducted in common bean. The use of the GWAS multi-locus methods has grown in recent years, becoming one of the main tools to identify molecular markers associated with traits of interest, especially for traits considered complex, i.e., controlled by several genes of small effect and highly influenced by the environment (WEN et al., 2018; YANG et al., 2020; ZHANG; JIA; DUNWELL, 2019).

Figure 5.3 Accumulation of favorable alleles in relation to adjusted means (LSmeans) for morpho-agronomic traits of common beans detected in accessions belonging to the Brazilian Diversity Panel (BDP).



PH = plant height (cm), FPIH = first pod insertion height (cm), NN = number of nodules, PL = pod length (cm), NPP = total number of pods per plant, LP = number of locules per pod, SP = number of seeds per pod, TSW = total seed weight per plant (gm), W100 = 100-seed weight (gm) and YLD = grain yield (kg.ha^{-1}).

A GE interaction was observed for most of the morpho-agronomic traits evaluated in the present study, indicating that the accessions' differential behavior depends on the evaluated environments. The presence of the GE interaction is frequently observed in GWAS studies, consequently interfering with the occurrence of the QTN x Environment interaction (FATTAHI; FAKHERI, 2019; PAN et al., 2018). The estimates of h^2 obtained in the present study were similar to those observed in literature (ASFAW et al., 2017; RANA et al., 2015). The h^2 is the central parameter of any breeding program, used to estimate the response to selection and explain the proportion of phenotypic variation due to genetic variations (FALCONER; MACKAY, 1996).

Correlations were observed among the traits related to the pods (PL, LP, and SP), the plant architecture traits (PH, FPIH, and NN), and the production components (TSW, W100, and YLD). Similar results were reported in several studies on common beans (ASFAW et al., 2017; NADEEM et al., 2020; RANA et al., 2015). The positive correlations observed between YLD and LP, SP, TSW, and W100 corroborate the possibility of indirect selection of YLD through these traits. However, Asfaw et al. (2017), when studying the relationship between morpho-agronomic traits in 202 accessions of Andean and Mesoamerican origin, reported that the correlation between YLD and W100 occurs only in Andean common beans. Furthermore, the same authors recommend the indirect selection of YLD through NPP, independent of the gene pool. Although no positive correlation between YLD and NPP was found in this study, moderate correlations with TSW and SP were observed.

Quantitative genetics assumes that the genetic correlations between traits can be attributed to gene binding and/or pleiotropy (SALTZ; HESSEL; KELLY, 2017). If pleiotropy is the main reason for genetic correlations between two traits, the same QTN can be identified in both traits. However, if gene binding is the main reason, an overlap of the location between QTNs is expected. Thus, the pleiotropic QTNs identified between the PL-LP and NPP-TSW traits can be considered one of the causes of the correlations observed in these traits. On the other hand, the high number of QTNs identified in overlapping genomic regions indicates that the genetic binding is the leading cause of the observed genetic correlations in the other assessed traits. Pleiotropy or binding effects have been reported for many traits such as productivity, biomass, and plant height (SOLTANI et al., 2016).

In the ML-GWAS methods used, the ISIS-EM-BLASSO detected the highest number of SNPs. However, it was the least efficient. On the other hand, the FASTmrMLM method, even with the lowest number of QTNs detected, was considered the most efficient of the methods evaluated. Several studies comparing the ISIS-EM-BLASSO, pLARmEB, mrMLM,

and FASTmrMLM methods have already been performed in many crops and the results are similar to those observed in the present study (FANG et al., 2020; MA et al., 2018; MISRA et al., 2018; ZHANG et al., 2018b). Although the ML-GWAS methods have similar approaches, the differential identification of QTNs is related to different screening and estimation models of each method (ZHANG et al., 2018b). From the present study results, FASTmrMLM can be considered the most reliable method, as it presented a low rate of false-positive associations. The FASTmrMLM method results from an improvement of the mrMLM method, which is a faster, more reliable, with high statistical power, high estimation accuracy, and low false-positive rate (TAMBA; ZHANG, 2018).

Most of the QTNs identified in this study were observed in only one environment, indicating the presence of QTN x Environment interaction. Several studies also reported the presence of this interaction in morpho-agronomic traits in common beans, suggesting that the gene expression of these QTNs is influenced by the evaluation environment (MACQUEEN et al., 2020; WU et al., 2020). The presence of the QTN x Environment interaction is considered one of the main aggravators in selecting QTNs in breeding programs, as these QTNs are considered more unstable. On the other hand, considering only the stable QTNs, the accumulation of the favorable alleles provided a gradual improvement in all the evaluated traits.

Significant increases in productivity and its components were observed, and alleles that cause a significant PH reduction were identified. PH is an essential factor in the formation of production components, and at the same time, it promotes or inhibits other components, affecting mainly the resistance to lodging and NPP (FANG et al., 2020). Small-sized plants are associated with a determined growth habit, less susceptibility to lodging, and shorter cycles (CHANG et al., 2018). Over the years, with the technification of agriculture, common bean breeding programs have sought to develop plants with these traits, as they facilitate management and mechanized harvesting, reducing harvest losses and susceptibility to some diseases, allowing an increase in the number of crops per year due to a reduction in the cycle (TEIXEIRA; RAMALHO; ABREU, 1999).

Most QTNs identified in this study are considered as small effect, confirming the complex and quantitative nature of the main morpho-agronomic traits in common beans (GUPTA et al., 2020; MACQUEEN et al., 2020). As most of these traits are controlled by polygenes, the effect of each locus individually is relatively small. Nevertheless, it is vital to identify small-effect loci that cumulatively can explain the variation in the trait (NAKANO; KOBAYASHI, 2020). The selection of higher-effect QTNs is preferable for the selection assisted by molecular markers (SAM) (NADEEM et al., 2018; OLADOSU et al., 2019).

However, the use of small-effect QTNs associated with the traits of interest is considered an important strategy in approaches to genomic selection (GS) since only these QTNs can replace the need for high-density genotyping by random SNPs and thus reduce genotyping costs. Moreover, models of GS using only SNPs known to be associated with the traits of interest showed greater accuracy of prediction since they showed lower background noise in constructing these models (ALI et al., 2020; HE et al., 2019b).

In the candidate genes identified in this study, nine (*Phvul.001G189200*, *Phvul.001G192200*, *Phvul.003G039900*, *Phvul.006G098300*, *Phvul.007G246700*, *Phvul.008G013300*, *Phvul.008G268700*, *Phvul.008G277352*, and *Phvul.011G020500*) were previously identified in other studies of GWAS for morpho-agronomic traits in common beans (CICHY et al., 2015; MACQUEEN et al., 2020; MOGHADDAM et al., 2016; SOLTANI et al., 2017; TOCK et al., 2017). The candidate gene *Phvul.003G039900*, associated with W100 in the present study, was also related to the weight of seeds in MacQueen et al. (2020). The same authors also associated the gene *Phvul.006G098300* with the PH, whereas this gene was associated with FPIH in the present study. The gene *Phvul.003G039900* has as function methyltransferase activity and the gene *Phvul.006G098300* is related to transferase activity and transferring acyl groups other than amino-acyl groups.

The candidate gene *Phvul.008G013300*, related to PL in this study, was also associated with the weight of seeds as reported by Moghaddam et al. (2016) and has serine-type endopeptidase activity and proteolysis functions. The candidate gene *Phvul.011G020500* was associated with PH, NPP, and TSW, while this same gene was associated with the aerial part's biomass trait in the observations of Soltani et al. (2017). Several functions were reported for this gene, such as DNA-binding transcription factor activity, transcription regulator complex, regulation of transcription, DNA-templated and cell cycle.

Considering that the genomic regions around the significant QTNs for the different traits assessed in this study overlap one another, 74% of the identified genes were also detected for more than one trait. Due to the strong LD observed in common beans, it is difficult to assign a gene precisely to a trait, especially when it comes to polygenic traits (IKRAM et al., 2020). The genes located in the genomic regions around the identified QTNs may serve as promising targets for studying molecular mechanisms responsible for morpho-agronomic traits in common beans.

In this study, 64 QTNs were identified for ten morpho-agronomic traits in common beans. Thirty-nine of them were identified as favorable alleles that can significantly increase productivity and its components in the cultivation of common beans through allele pyramiding.

The results reinforce the importance of conducting phenotyping of individuals in multiple environments, using multiple detection methods to increase the reliability of QTNs obtained in GWAS studies. The QTNs identified proved adequate for implementation in common bean breeding programs, mainly for improving Mesoamerican common beans from the Black and Carioca commercial classes, which are the primary targets in Brazil.

Table S5.1 QTNs associated with morpho-agronomic traits detected at least twice via different methods and in different environments in common bean accessions belonging to the Brazilian Diversity Panel (BDP).

Trait.	SNP	Chr	Position (bp)	QTN Effect[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]
PH	S01_5305887	1	5305887	-8.13 ~ -6.83	4.78 ~ 6.56	6.55 ~ 9.33	0.063	CC	1	1,2,3
	S02_33714137	2	33714137	1.91 ~ 3.34	3.41 ~ 4.98	1.32 ~ 5.36	0.253	AA	3,4	1,4
	S03_13423968	3	13423968	-3.18 ~ -2.99	3.98 ~ 4.45	6.17 ~ 6.97	0.180	GG	5	1,2
	S03_40723746	3	40723746	2.49 ~ 3.09	3.61 ~ 4.74	3.43 ~ 5.29	0.348	CC	1	1,2
	S04_10370006	4	10370006	2.38 ~ 4.53	3 ~ 4.18	2.35 ~ 11.84	0.449	CC	3	3,4
	S04_2493278	4	2493278	2.93 ~ 5.28	3.71 ~ 3.87	4.66 ~ 7.89	0.136	CC	1,5	1,3
	S04_2493297	4	2493297	2.59 ~ 3.9	3.49 ~ 5.09	2.2 ~ 4.18	0.152	GG	1,5	2,4
	S04_373114	4	373114	-3.34 ~ -1.73	3.84 ~ 5.66	1.4 ~ 4.6	0.433	GG	3,5	1,4
	S05_39604389	5	39604389	1.65 ~ 3.76	3.04 ~ 6.12	2.9 ~ 8.19	0.455	GG	1,2,3,5	1,2,3,4
	S05_39624676	5	39624676	3.19 ~ 3.44	7.87 ~ 10.98	11.72 ~ 13.67	0.449	GG	5	1,2
	S05_39680093	5	39680093	3.16 ~ 4.28	4.86 ~ 7.85	5.51 ~ 10.08	0.352	GG	1	1,2,3
	S06_25397668	6	25397668	2.43 ~ 2.81	3.43 ~ 5.09	2.57 ~ 8.4	0.354	CC	5	1,3,4
	S06_25791997	6	25791997	-7.54 ~ -5.18	4.73 ~ 8.19	6.8 ~ 12.64	0.118	TT	1	2,4
	S07_17942068	7	17942068	5.59 ~ 7.5	5.24 ~ 5.92	2.85 ~ 6.82	0.051	CC	3,4,5	1,4
	S08_3952225	8	3952225	-2.74 ~ -2.7	3.15 ~ 3.87	4.6 ~ 4.72	0.455	TT	3	1,2
	S08_58030052	8	58030052	2.67 ~ 5.09	3.06 ~ 4.19	2.25 ~ 7.55	0.148	AA	3	1,3
	S08_62021856	8	62021856	4.08 ~ 8.43	4.52 ~ 6.2	1.84 ~ 9.62	0.062	GG	3,5	3,4
	S10_42297372	10	42297372	3.35 ~ 3.7	3.83 ~ 4.73	5 ~ 6.08	0.244	AA	1	1,3
	S10_43645690	10	43645690	2.16 ~ 4.45	4.17 ~ 6.92	3.97 ~ 8.46	0.242	CC	3,5	1,2,3,4
	S10_44036828	10	44036828	-2.83 ~ -1.69	3.04 ~ 4.97	3.89 ~ 10.61	0.393	TT	2	2,3,4
	S11_1465100	11	1465100	2.69 ~ 3.06	4.78 ~ 5.57	9.01 ~ 11.3	0.318	AA	2	1,2,3
FPIH	S01_1292307	1	1292307	0.55 ~ 0.59	3.35 ~ 3.43	4.77 ~ 5.28	0.1761	AA	5	1,3,4
	S01_20991636	1	20991636	-0.46 ~ -0.37	3.33 ~ 3.8	3.62 ~ 5.53	0.4375	CC	5	2,3,4
	S01_2896512	1	2896512	-0.7 ~ -0.67	3.16 ~ 3.66	4.21 ~ 4.66	0.092	AA	5	1,3
	S01_48972163	1	48972163	0.55 ~ 0.63	4.04 ~ 4.69	5.3 ~ 6.89	0.2069	AA	5	1,3
	S03_4648326	3	4648326	0.4 ~ 0.6	3.39 ~ 3.43	3.1 ~ 6.69	0.2159	TT	2	2,3
	S05_39601221	5	39601221	0.4 ~ 0.54	3.55 ~ 3.95	3 ~ 5.5	0.2241	GG	5	2,3
	S06_20814429	6	20814429	-0.61 ~ -0.49	3.84 ~ 6.86	6.52 ~ 10.05	0.483	AA	5	1,2,4
	S06_29875948	6	29875948	-0.79 ~ -0.55	3.72 ~ 3.76	3.72 ~ 7.43	0.125	CC	2	2,3
	S09_27171634	9	27171634	0.51 ~ 0.78	3.09 ~ 3.63	3.03 ~ 6.93	0.118	CC	2	2,3,4
	S10_15424493	10	15424493	0.45 ~ 0.72	3.05 ~ 3.35	2.71 ~ 6.78	0.1307	AA	2	2,3
NN	S01_5448199	1	5448199	0.14 ~ 0.17	3.33 ~ 4.96	2.67 ~ 4.24	0.2584	CC	5	1,2,3
	S02_25464609	2	25464609	0.26 ~ 0.36	4 ~ 4.71	4.59 ~ 8.52	0.4148	GG	1	2,3,4
	S02_25870688	2	25870688	0.35 ~ 0.36	3.41 ~ 3.97	3.64 ~ 3.84	0.1517	GG	4	2,4
	S02_38214788	2	38214788	-0.27 ~ -0.25	4.55 ~ 5.7	4.9 ~ 5.52	0.264	GG	2	2,4

(Continue)

(Continuation)

Trait.	SNP	Chr	Position (bp)	QTN Effect[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]
	S02_48537121	2	48537121	0.17 ~ 0.26	3.03 ~ 3.67	2.4 ~ 5.55	0.2898	CC	2	1,3,4
	S04_2503984	4	2503984	0.13 ~ 0.29	3.67 ~ 5.43	3.03 ~ 7.81	0.3807	CC	2,5	1,2,3
	S07_3943786	7	3943786	-0.14 ~ -0.14	3.53 ~ 4.08	3.46 ~ 3.65	0.4663	AA	5	1,2
	S07_474203	7	474203	0.12 ~ 0.19	3.12 ~ 4.81	2.07 ~ 5	0.2841	TT	5	1,3,4
	S07_495888	7	495888	-0.35 ~ 0	3.61 ~ 6.22	0 ~ 10.76	0.3523	GG	2	1,2,3,4
	S08_44008378	8	44008378	-0.47 ~ -0.32	3.84 ~ 5.58	3.42 ~ 7.1	0.0909	CC	2	1,2,3,4
	S08_61614494	8	61614494	0.15 ~ 0.25	3.93 ~ 7.72	4.14 ~ 11.33	0.4045	TT	5	1,2,3
	S09_22575298	9	22575298	-0.4 ~ -0.23	3.53 ~ 4.47	2.38 ~ 7.16	0.1348	CC	2	1,3
	S09_5349386	9	5349386	0.25 ~ 0.3	3.07 ~ 5.28	2.73 ~ 7.29	0.125	GG	5	1,2,3,4
	S10_415221	10	415221	0.3 ~ 0.37	4.18 ~ 5.1	7.6 ~ 11.3	0.427	GG	3	1,3
	S10_44010107	10	44010107	-0.26 ~ -0.23	3.48 ~ 4.48	4.79 ~ 5.96	0.3708	CC	2	1,2,3
	S11_8369504	11	8369504	-0.19 ~ -0.14	3.3 ~ 4.87	2.9 ~ 5.43	0.264	AA	5	1,2,3
PL	S02_1040748	2	1040748	-0.26 ~ -0.19	3.43 ~ 4.19	1.02 ~ 4.9	0.1023	TT	2	1,3,4
	S02_47586597	2	47586597	-0.33 ~ -0.22	4.92 ~ 8.61	1.63 ~ 10.93	0.1685	AA	2,3,5	1,2,3,4
	S02_49444827	2	49444827	0.16 ~ 0.19	3.29 ~ 4.1	1.98 ~ 6.66	0.4943	AA	2	3,4
	S02_49538733	2	49538733	0.13 ~ 0.23	3.1 ~ 5.31	2.9 ~ 8.78	0.5	GG	1,2,3	1,2,3
	S03_1420507	3	1420507	-0.14 ~ -0.13	4.5 ~ 5.06	1.05 ~ 4.11	0.4124	TT	5	2,4
	S04_1926044	4	1926044	-0.44 ~ -0.34	3.83 ~ 3.86	2.45 ~ 6.72	0.0568	CC	3	3,4
	S07_30515591	7	30515591	-0.31 ~ -0.23	3 ~ 3.65	3.91 ~ 7.18	0.1292	GG	3	1,2,3
	S07_30758337	7	30758337	-0.31 ~ -0.3	4.22 ~ 4.84	5.71 ~ 6.25	0.0955	CC	2	1,3
	S08_1159693	8	1159693	-0.23 ~ -0.19	3.48 ~ 4.51	1.52 ~ 5.23	0.1591	CC	2	1,3,4
	S08_26923941	8	26923941	0 ~ 0.17	3.08 ~ 3.24	0 ~ 0.6	0.0674	CC	2	1,4
	S08_61609638	8	61609638	-0.28 ~ -0.27	3.09 ~ 4.39	1.45 ~ 3.66	0.0674	CC	2	1,4
	S08_62432046	8	62432046	-0.23 ~ -0.16	4.01 ~ 4.24	1.08 ~ 5.14	0.1534	TT	2	1,2,3,4
	S08_9375624	8	9375624	-0.54 ~ -0.41	5.89 ~ 6.76	2.25 ~ 13.21	0.0571	CC	5	2,3,4
	S11_2437959	11	2437959	0.17 ~ 0.18	3.27 ~ 3.39	3.5 ~ 3.89	0.2429	AA	4	1,2,4
	S11_47294839	11	47294839	-0.38 ~ -0.24	3.32 ~ 4.14	2.87 ~ 7.12	0.0621	CC	5	2,3
NPP	S01_25820006	1	25820006	0.73 ~ 0.78	3.16 ~ 3.62	2.84 ~ 3.3	0.4719	CC	3	2,4
	S03_12044967	3	12044967	-2.28 ~ -1.62	3.28 ~ 4.46	4.9 ~ 9.69	0.0966	AA	3	1,2,3,4
	S05_40466290	5	40466290	1.04 ~ 1.41	3.12 ~ 5.36	3.96 ~ 7.21	0.2102	TT	3	1,2,3,4
	S05_445417	5	445417	-1.43 ~ -0.92	3.11 ~ 4.96	2.55 ~ 6.02	0.0562	AA	5	1,2,3
	S05_4994208	5	4994208	-1.39 ~ -1.22	3.79 ~ 4.92	2.05 ~ 6.01	0.1854	CC	4	2,4
	S07_38456082	7	38456082	-1.56 ~ -0.48	3.11 ~ 3.85	3.31 ~ 11.71	0.4432	GG	3,4,5	1,2,3,4
	S08_2493035	8	2493035	-2.45 ~ -2.07	5.78 ~ 6.16	4.1 ~ 11.68	0.1067	CC	4	1,2,4
	S08_60292408	8	60292408	1.25 ~ 1.73	3.45 ~ 4.05	3.1 ~ 4.26	0.0506	CC	3,5	1
	S11_1617681	11	1617681	-0.97 ~ -0.73	4.94 ~ 7.45	7.5 ~ 13.04	0.4602	AA	5	1,2,3,4
LP	S01_221817	1	221817	0.07 ~ 0.14	3.3 ~ 5.73	1.92 ~ 7.72	0.2921	GG	1	1,2,4
	S01_45434768	1	45434768	0.18 ~ 0.22	3.47 ~ 4.97	4.39 ~ 7	0.0674	TT	3	2,4
	S02_41632778	2	41632778	0 ~ 0.18	3.03 ~ 3.78	0 ~ 6.56	0.0629	TT	5	1,3,4
	S02_42213511	2	42213511	-0.13 ~ -0.1	3.27 ~ 5.53	2.96 ~ 5.33	0.1854	AA	1	1,4

(Continue)

(Continuation)

Trait.	SNP	Chr	Position (bp)	QTN Effect[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]
	S04_3837889	4	3837889	-0.14 ~ -0.11	3.8 ~ 4	5.09 ~ 7.31	0.3427	GG	2	2,3
	S05_36623179	5	36623179	-0.13 ~ -0.09	3.18 ~ 3.46	3.58 ~ 6.07	0.2472	GG	3	1,3
	S07_136162	7	136162	0.12 ~ 0.13	3.51 ~ 5.46	5.14 ~ 5.56	0.2022	AA	3	2,4
	S07_33862545	7	33862545	-0.17 ~ -0.1	3.41 ~ 5.07	3.24 ~ 9.48	0.1854	AA	3	1,2,3
	S07_3955474	7	3955474	0.12 ~ 0.14	3.42 ~ 4.09	4.84 ~ 6.07	0.1685	CC	3	2,3
	S07_5382811	7	5382811	0.11 ~ 0.13	3.67 ~ 4.48	5.66 ~ 7.92	0.4607	CC	2	1,3
	S08_9375624	8	9375624	-0.24 ~ -0.15	3.02 ~ 4.86	4.08 ~ 9.32	0.0514	CC	5	1,2,3,4
	S10_41807590	10	41807590	0.12 ~ 0.21	3.08 ~ 3.61	2.34 ~ 6.55	0.0899	GG	2	2,3
	S10_4876917	10	4876917	0.12 ~ 0.14	3.41 ~ 3.84	5.38 ~ 6.67	0.2045	AA	3	2,3,4
	S10_4911729	10	4911729	0.09 ~ 0.12	3.55 ~ 4.21	4.86 ~ 8.64	0.24	GG	5	1,2,3,4
	S11_26987117	11	26987117	-0.24 ~ 0	3.13 ~ 5.86	0 ~ 5.91	0.0625	GG	2	3,4
	S11_29062	11	29062	-0.13 ~ -0.07	3.42 ~ 9.02	3.78 ~ 7.81	0.3371	TT	1,5	2,3,4
	S11_30320686	11	30320686	0.04 ~ 0.14	3.09 ~ 3.15	1.07 ~ 9.14	0.3933	TT	3	1,3
	S11_52195944	11	52195944	-0.1 ~ -0.07	3.68 ~ 4.5	4.56 ~ 8.66	0.3616	GG	5	1,2,3
SP	S01_112055	1	112055	0.12 ~ 0.2	3.97 ~ 4.81	3.56 ~ 11.2	0.2727	CC	1	2,3,4
	S01_14589174	1	14589174	-0.16 ~ -0.09	3.28 ~ 4.73	3.05 ~ 8.67	0.4382	GG	2	1,3
	S01_465127	1	465127	0.2 ~ 0.28	3.58 ~ 4.4	5.07 ~ 9.01	0.0787	CC	3	2,3
	S05_19626048	5	19626048	0.14 ~ 0.19	3.93 ~ 4.26	5.16 ~ 9	0.191	CC	3	2,3
	S05_36420518	5	36420518	0.1 ~ 0.26	3.59 ~ 5.91	2.55 ~ 7.7	0.0843	GG	4,5	2
	S07_17126884	7	17126884	0.11 ~ 0.19	3.28 ~ 3.43	3.3 ~ 3.66	0.1124	AA	4,5	4
	S08_10350174	8	10350174	-0.37 ~ -0.13	3.27 ~ 4.79	2.5 ~ 10.42	0.0568	GG	2,5	1,2,3,4
	S10_42038210	10	42038210	0.08 ~ 0.11	3.99 ~ 4.5	5.14 ~ 8.44	0.3933	GG	5	2,3
TSW	S01_44752890	1	44752890	1.98 ~ 2.51	3.53 ~ 4.97	4.17 ~ 8.25	0.1486	CC	4	1,2,3,4
	S02_2242481	2	2242481	-1.92 ~ -1.31	3.4 ~ 6.12	2.89 ~ 6.49	0.2147	TT	4	1,2,4
	S03_51482697	3	51482697	-0.98 ~ -0.65	3.04 ~ 4.15	3.53 ~ 7.5	0.2429	TT	5	2,3
	S04_41451220	4	41451220	-2.03 ~ -1.53	3.18 ~ 5.58	5.77 ~ 10.15	0.1136	CC	1	2,3,4
	S06_640352	6	640352	0 ~ 2.81	3.75 ~ 4.84	0 ~ 6.42	0.0847	AA	4	1,2
	S10_18589262	10	18589262	-3.58 ~ -1.56	3.48 ~ 4.62	4.49 ~ 8.69	0.0506	GG	3,5	1,2,3,4
	S11_1617681	11	1617681	-1.21 ~ -0.73	3.75 ~ 7.98	5.09 ~ 15.42	0.4571	AA	5	1,2,3,4
	S11_36831666	11	36831666	-1.09 ~ -1	3.3 ~ 3.3	4.07 ~ 4.82	0.2079	AA	1	1,2
	S11_52088416	11	52088416	-1.75 ~ -1.48	4.14 ~ 4.14	5.38 ~ 7.53	0.2247	AA	3	1,2,3
W100	S02_2696078	2	2696078	1.15 ~ 1.54	4.37 ~ 6.73	5.78 ~ 12.4	0.0625	GG	1	3,4
	S02_3942810	2	3942810	0.92 ~ 1.15	3.71 ~ 4.13	3.75 ~ 5.76	0.0899	CC	2	2,3
	S03_11484802	3	11484802	-0.76 ~ -0.69	4.42 ~ 5.35	5.42 ~ 6.48	0.4034	CC	3	1,2,3,4
	S03_4682037	3	4682037	0.96 ~ 1.68	3.33 ~ 7.64	4.04 ~ 12.3	0.0686	CC	5	2,3,4
	S04_41872476	4	41872476	-0.85 ~ -0.76	3.14 ~ 5.75	5.43 ~ 6.43	0.2921	AA	2,4	2,4
	S05_262416	5	262416	0.61 ~ 0.87	3.23 ~ 4.03	3.35 ~ 6.97	0.4802	CC	4	2,3
	S06_1330875	6	1330875	0.57 ~ 0.74	3.85 ~ 4.04	3.42 ~ 5.91	0.1943	TT	5	3,4
	S06_23116097	6	23116097	0.46 ~ 0.87	3.73 ~ 4.45	2.77 ~ 9.67	0.3989	CC	2	2,3
	S08_62432046	8	62432046	-0.85 ~ -0.82	3.38 ~ 5.44	4.59 ~ 5.05	0.1517	TT	2	2,3

(Continue)

(Continuation)

Trait.	SNP	Chr	Position (bp)	QTN Effect[1]	LOD score[2]	PVE (%)[3]	MAF[4]	Gen	Env[5]	Methods[6]
YLD	S01_1973183	1	1973183	175.24 ~ 289.53	3.04 ~ 3.23	1.41 ~ 5.75	0.0682	GG	3	3,4
	S01_44911599	1	44911599	360.06 ~ 529	3.93 ~ 6.9	6.17 ~ 13.31	0.0966	AA	4	1,2,3,4
	S01_51067135	1	51067135	77.81 ~ 123.27	3.53 ~ 4.87	2.65 ~ 6.45	0.2898	GG	5	1,2,3,4
	S01_6568871	1	6568871	-180.61 ~ -173.95	3.36 ~ 4.29	6.04 ~ 6.52	0.2416	CC	3	2,3
	S02_2153254	2	2153254	136.73 ~ 150.71	3.53 ~ 3.58	1.89 ~ 3.02	0.0674	CC	5	2,4
	S02_34513049	2	34513049	-186.17 ~ -114.43	3.33 ~ 4.19	1.47 ~ 5.84	0.1875	TT	3	2,3,4
	S02_8413809	2	8413809	167.94 ~ 222.58	3.87 ~ 3.87	5.84 ~ 9.95	0.2191	TT	2	1,3
	S03_2802438	3	2802438	113.41 ~ 141.54	3.74 ~ 5.64	6.09 ~ 9.49	0.3466	CC	1	
	S03_3510073	3	3510073	81.7 ~ 126.8	3.13 ~ 3.92	3.52 ~ 8.21	0.4663	CC	5	2,3
	S03_3530662	3	3530662	97.86 ~ 123.61	3.33 ~ 4.79	3.82 ~ 4.58	0.4551	TT	2,5	2,4
	S03_49731981	3	49731981	-176.06 ~ -133.48	4.83 ~ 5.89	5.24 ~ 11.65	0.2443	CC	5	1,2,3,4
	S04_47436704	4	47436704	157.01 ~ 180.08	3.28 ~ 3.89	4.15 ~ 5.46	0.191	AA	3	2,3
	S07_34450891	7	34450891	166.38 ~ 260.05	3.13 ~ 8.89	3.77 ~ 12.15	0.1307	AA	2,4,5	1,2,3,4
	S10_2336773	10	2336773	-182.4 ~ -146.16	3.73 ~ 3.73	5.52 ~ 8.34	0.309	GG	2	1,3
	S11_43243770	11	43243770	136.79 ~ 170.39	3.06 ~ 4.47	3.73 ~ 5.78	0.2079	CC	2	2,4
	S11_53567276	11	53567276	-256.31 ~ -192.42	3.25 ~ 3.33	3.29 ~ 5.84	0.2022	AA	4	2,3

PH = plant height (cm), FPIH = first pod insertion height (cm), NN = number of nodules, PL = pod length (cm), NPP = total number of pods per plant, LP = number of locules per pod, SP = number of seeds per pod, TSW = total seed weight per plant (gm), W100 = 100-seed weight (gm) and YLD = grain yield (kg.ha^{-1}). [1]*Quantitative trait nucleotide effect;* [2]LOD value, the significant threshold for *P-value* transformed; [3]PVE (%): *Phenotypic variation explained*; [4]*Minor allele frequency*; [5]Environments: 1-LDA, 2-PG, 3-GUA, 4-LSmeans; [6]Methods: 1-FASTmrMLM, 2-ISIS EM-BLASSE, 3-mrMLM, 4-pLARmEB; Pleiotropic QTNs, related to more than one mineral, are in bold.

6 CONSIDERAÇÕES FINAIS

O painel de diversidade desenvolvido nesse estudo, denominado *Brazilian Diversity Panel* (BDP), o qual representa grande parte da diversidade brasileira de feijão, apresentou grande diversidade genética e se mostrou adequado para estudos que visem a identificação de regiões genômicas relacionadas a características de interesse.

O *BDP* se mostrou eficiente para GWAS relacionado a características nutricionais e agronômicas, apresentando alta variabilidade genética entre os acessos para as diferentes características avaliadas, permitindo a detecção de diversos QTNs.

O uso de diferentes métodos multi-locus de GWAS e a exploração de diferentes ambientes demonstrou a confiabilidade dos marcadores associados as características avaliadas.

Os QTNs identificados se mostraram promissores para aplicação em programas de melhoramento genético visando a piramidação dos alelos superiores, que poderá ser monitorada por meio da seleção assistida por marcadores, sendo úteis principalmente para o melhoramento de feijões mesoamericanos das classes comerciais preto e carioca, que são os principais alvos no Brasil. Além disso, esses locos também podem ser incorporados em painéis de SNPs relacionados a características de interesse na seleção genômica. Estudos adicionais seriam úteis para validar a importância e funcionalidade dos genes candidatos no acúmulo de nutrientes nos grãos e na expressão das características agronômicas.

7 REFERÊNCIAS

ABDURAKHMONOV, I. Y.; ABDUKARIMOV, A. Application of association mapping to understanding the genetic diversity of plant germplasm resources. **International Journal of Plant Genomics**, v. 2008, n. May, 2008.

AIDAR, H. **Cultivo do feijoeiro comum: características da cultura**, 2007.

ALI, M. et al. Genomic prediction for grain yield and yield-related traits in chinese winter wheat. **International Journal of Molecular Sciences**, v. 21, n. 4, p. 1–18, 2020.

ALIPOUR, H. et al. Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. **Frontiers in Plant Science**, v. 8, n. August, p. 1–14, 2017.

ARIANI, A.; BERNY MIER Y TERAN, J. C.; GEPTS, P. Genome-wide identification of SNPs and copy number variation in common bean (*Phaseolus vulgaris* L.) using genotyping-by-sequencing (GBS). **Molecular Breeding**, v. 36, n. 7, 2016.

ARIANI, A.; MIER TERAN, J. C. B.; GEPTS, P. Spatial and temporal scales of range expansion in wild Phaseolus vulgaris. **Molecular Biology and Evolution**, v. 35, n. 1, p. 119–131, 2017.

ASFAW, A. et al. Trait associations in diversity panels of the two common bean (*Phaseolus vulgaris* l.) gene pools grown under well-watered and water-stress conditions. **Frontiers in Plant Science**, v. 8, n. May, p. 1–15, 2017.

ASSEFA, T. et al. A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, v. 39, n. 2, 2019.

BARDOU, P. et al. SOFTWARE Open Access jvenn: an interactive Venn diagram viewer. **BMC Bioinformatics**, v. 15, n. 293, p. 1–7, 2014.

BEEBE, S. E. et al. Phenotyping common beans for adaptation to drought. **Frontiers in Physiology**, v. 4 MAR, n. March, p. 1–21, 2013.

BEEBE, S.; GONZALEZ, A. V.; RENGIFO, J. Research on trace minerals in the common bean. **Food and Nutrition Bulletin**, v. 21, n. 4, p. 387–391, 2000.

BEER, H. et al. Consuming iron-biofortified pearl millet increased hemoglobin concentrations and prevented a decline in energy efficiency in Indian girls. **FASEB Journal**, v. 28, n. 1, p. Supplement 646.7, 2014.

BERNY MIER Y TERAN, J. C. et al. Root and shoot variation in relation to potential intermittent drought adaptation of Mesoamerican wild common bean (*Phaseolus vulgaris* L.). **Annals of Botany**, p. 1–16, 2018.

BHATTA, M. et al. Unlocking the novel genetic diversity and population structure of synthetic Hexaploid wheat. **BMC Genomics**, v. 19, n. 1, p. 1–12, 2018.

BHATTARAI, U.; SUBUDHI, P. K. Identification of drought responsive QTLs during vegetative growth stage of rice using a saturated GBS-based SNP linkage map. **Euphytica**, v. 214, n. 38, p. 1–17, 2018.

BITOCCHI, E. et al. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. **New Phytologist**, v. 197, n. 1, p. 300–313, 2013.

BLAIR, M. W. et al. Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, v. 23, n. 2, p. 197–207, 2009.

BLAIR, M. W. et al. QTL for seed iron and zinc concentration and content in a Mesoamerican common bean (*Phaseolus vulgaris* L.) population. **Theoretical and Applied Genetics**, v. 121, n. 6, p. 1059–1070, 2010.

BLAIR, M. W. Mineral biofortification strategies for food staples: The example of common bean. **Journal of Agricultural and Food Chemistry**, v. 61, n. 35, p. 8287–8294, 2013.

BLAIR, M. W. et al. A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, v. 126, n. 2, p. 535–548, 2013.

BLAIR, M. W. et al. Uneven recombination rate and linkage disequilibrium across a reference SNP map for common bean (*Phaseolus vulgaris* L.). **PLoS ONE**, v. 13, n. 3, p. 1–21, 2018.

BOUIS, H. E.; SALTZMAN, A. Improving nutrition through biofortification: A review of evidence from HarvestPlus, 2003 through 2016. **Global Food Security**, v. 12, n. January, p. 49–58, 2017.

BRADBURY, P. J. et al. TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633–2635, 2007.

BROUGHTON, W. J. et al. Beans (*Phaseolus* spp.) - model food legumes. **Plant and Soil**, v. 252, n. 1, p. 55–128, 2003.

BROWNING, B. L.; BROWNING, S. R. Genotype imputation with millions of reference samples. **American Journal of Human Genetics**, v. 98, n. 1, p. 116–126, 2016.

BROWNING, B. L.; ZHOU, Y.; BROWNING, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. **American Journal of Human Genetics**, v. 103, n. 3, p. 338–348, 2018.

BURGHARDT, L. T.; YOUNG, N. D.; TIFFIN, P. A guide to genome-wide association mapping in plants. **Current Protocols in Plant Biology**, v. 2, n. March, p. 22–38, 2017.

BURLE, M. L. et al. Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. **Theoretical and Applied Genetics**, v. 121, n. 5, p. 801–813, 2010.

CAKMAK, I. Enrichment of cereal grains with zinc: Agronomic or genetic biofortification? **Plant and Soil**, v. 302, n. 1–2, p. 1–17, 2008.

CAMPA, A.; MURUBE, E.; FERREIRA, J. J. Genetic diversity, population structure, and linkage disequilibrium in a Spanish common bean diversity panel revealed through genotyping-by-sequencing. **Genes**, v. 9, n. 11, 2018.

CHANG, F. et al. Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. **Frontiers in Plant Science**, v. 9, n.

August, p. 1–13, 2018.

CHENG, J. et al. Development of a SNP array and its application to genetic mapping and diversity assessment in pepper (*Capsicum* spp.). **Scientific Reports**, v. 6, n. August, p. 1–11, 2016.

CHHATRE, V. E.; EMERSON, K. J. StrAuto: Automation and parallelization of STRUCTURE analysis. **BMC Bioinformatics**, v. 18, n. 1, p. 1–5, 2017.

CICHY, K. A. et al. QTL analysis of seed iron, zinc, and phosphorus levels in an andean bean population. **Crop Science**, v. 49, n. 5, p. 1742–1750, 2009.

CICHY, K. A. et al. A *Phaseolus vulgaris* diversity panel for Andean bean Improvement. **Crop Science**, v. 55, 2015.

CICHY, K. A.; WIESINGER, J. A.; MENDOZA, F. A. Genetic diversity and genome - wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L .). **Theoretical and Applied Genetics**, v. 128, n. 8, p. 1555–1567, 2015.

COLLARD, B. C. Y. et al. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. **Euphytica**, v. 142, n. 1–2, p. 169–196, 2005.

CONAB, COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira - Safra 2018/2019**. [s.l: s.n.]. v. 6

CONTRERAS-SOTO, R. I. et al. Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (*Glycine max*). **Euphytica**, v. 213, n. 8, 2017a.

CONTRERAS-SOTO, R. I. et al. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-Based haplotype analysis. **PLoS ONE**, v. 12, n. 2, p. 1–22, 2017b.

CORTÉS, A. J.; CHAVARRO, M. C.; BLAIR, M. W. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, v. 123, n. 5, p. 827–845, 2011.

CU, S. T. et al. Genetic dissection of zinc, iron, copper, manganese and phosphorus in wheat (*Triticum aestivum* L.) grain and rachis at two developmental stages. **Plant Science**, v. 291, n. November 2019, p. 110338, 2020.

CUI, Y.; ZHANG, F.; ZHOU, Y. The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. **Frontiers in Plant Science**, v. 9, n. October, p. 1–9, 2018.

DANECEK, P. et al. The variant call format and VCFtools. **Bioinformatics**, v. 27, n. 15, p. 2156–2158, 2011.

DE CAMPOS, T. et al. Inheritance of growth habit detected by genetic linkage analysis using microsatellites in the common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, v. 27, n. 4, p. 549–560, 2011.

DELFINI, J. et al. Distinctness of Brazilian common bean cultivars with carioca and black grain by means of morphoagronomic and molecular descriptors. **PLoS ONE**, v. 12, n. 11, 2017.

DELFINI, J. et al. Estimation of genetic parameters and prediction of genotypic values in

common beans using mixed models. **Emirates Journal of Food and Agriculture**, v. 30, n. 12, p. 1026–1035, 2018.

DELFINI, J. et al. Diversity of nutritional content in seeds of Brazilian common bean germplasm. **PLoS ONE**, n. I, p. 1–13, 2020.

DELFINI, J. et al. Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. **Scientific Reports**, v. 11, 2021.

DESROUSSEAUX, D. et al. **LDcorSV: Linkage disequilibrium corrected by the structure and the relatedness. R package version 1.3.2**, 2017. Disponível em: <https://cran.r-project.org/package=LDcorSV>

DINIZ, A. L. et al. Evidence for strong kinship influence on the extent of linkage disequilibrium in cultivated common beans. **Genes**, v. 10, n. 1, p. 1–16, 2019.

DOERGE, R. W. Multifactorial Geneticsmapping and Analysis of Quantitative Trait Loci in Experimental Populations. **Nature Reviews Genetics**, v. 3, n. 1, p. 43–52, 2002.

DWIVEDI, S. L. et al. Nutritionally Enhanced Staple Food Crops. **Plant Breeding Reviews**, v. 36, p. 169–291, 2012.

EDWARDS, D.; BATLEY, J. Plant genome sequencing: Applications for crop improvement. **Plant Biotechnology Journal**, v. 8, n. 1, p. 2–9, 2010.

ELSHIRE, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS ONE**, v. 6, n. 5, p. 1–10, 2011.

ELTAHER, S. et al. Genetic diversity and population structure of F3:6 Nebraska Winter wheat genotypes using genotyping-by-sequencing. **Frontiers in Genetics**, v. 9, n. 76, p. 1–9, 2018.

ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **The Plant Genome Journal**, v. 4, n. 3, p. 250, 2011.

ERDOGMUS, S. et al. Genome-wide association studies of Ca and Mn in the seeds of the common bean (*Phaseolus vulgaris* L.). **Genomics**, v. 112, n. 6, p. 4536–4546, 2020.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. **Molecular Ecology**, v. 14, n. 8, p. 2611–2620, 2005.

FALCONER, D. S. .; MACKAY, T. F. C. **Introduction to quantitative genetics**. Essex, England: Longmam, 1996.

FANG, Y. et al. Linkage Analysis and Multi-Locus Genome-Wide Association Studies Identify QTNs Controlling Soybean Plant Height. **Frontiers in Plant Science**, v. 11, n. February, p. 1–14, 2020.

FAO, FOOD AND AGRICULTURE ORGANIZATION. **FAOSTAT: FAO Statistical Databases**. Disponível em: <http://www.fao.org/faostat/>. Acesso em: 25 jul. 2020.

FATTAHI, F.; FAKHERI, B. A. Evolutionary dynamics models in biometrical genetics supports QTL × environment interactions. **Journal of Genetics**, v. 98, n. 2, p. 1–9, 2019.

FINKELSTEIN, J. L. et al. A randomized trial of iron-biofortified pearl millet in school

children in India. **Journal of Nutrition**, v. 145, p. 1576–1581, 2015.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. TH. Structure of linkage disequilibrium in plants. **Annu Rev Plant Biol**, v. 54, p. 357–374, 2003.

GANNON, B. et al. Biofortified orange maize is as efficacious as a vitamin A supplement in Zambian children even in the presence of high liver reserves of vitamin A: A community-based, randomized placebo-controlled trial. **American Journal of Clinical Nutrition**, v. 100, n. 6, p. 1541–1550, 2014.

GAUR, R. et al. High density linkage mapping of genomic and transcriptomic SNPs for synteny analysis and anchoring the genome sequence of chickpea. **Scientific Reports**, v. 5, n. 10531, p. 1–11, 2015.

GEPTS, P. et al. Genomics of *Phaseolus* Beans, a Major Source of Dietary Protein and Micronutrients in the Tropics. **Genomics of Tropical Crop Plants**, n. January, p. 113–143, 2008.

GIOIA, T. et al. 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**, v. 14, n. 1, p. 1–21, 2019.

GLAHN, R. et al. Iron bioavailability studies of the first generation of iron-biofortified beans released in Rwanda. **Nutrients**, v. 9, 2017.

GLAUBITZ, J. C. et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. **PLoS ONE**, v. 9, n. 2, 2014.

GUJARIA-VERMA, N. et al. Gene-based SNP discovery in tepary bean (*Phaseolus acutifolius*) and common bean (*P. vulgaris*) for diversity analysis and comparative mapping. **BMC Genomics**, v. 17, p. 239, 2016.

GUO, C. et al. Transversions have larger regulatory effects than transitions. **BMC Genomics**, v. 18, n. 1, p. 1, 2017.

GUPTA, N. et al. Marker association study of yield attributing traits in common bean (*Phaseolus vulgaris* L.). **Molecular Biology Reports**, 2020.

HAAS, J. D. et al. Consuming iron biofortified beans increases iron status in Rwandan women after 128 days in a randomized controlled feeding trial. **The Journal of nutrition**, n. 146, p. 1586–92, 2016.

HALLAUER, A. R.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. Ames: Iowa State University Press, 1988.

HART, J. P.; GRIFFITHS, P. D. Genotyping-by-Sequencing Enabled Mapping and Marker Development for the Potyvirus Resistance Allele in Common Bean. **The Plant Genome**, v. 8, n. 1, p. 1–14, 2015.

HE, J. et al. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Frontiers in Plant Science**, v. 5, n. September, 2014.

HE, L. et al. Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.). **Frontiers in Plant Science**, v. 9, n. January, p. 1–15, 2019a.

HE, L. et al. Evaluation of genomic prediction for Pasmo resistance in flax. **International Journal of Molecular Sciences**, v. 20, n. 2, 2019b.

HILL, W. G.; WEIR, B. S. Variances and covariances of squared linkage disequilibria in finite populations. **Theoretical Population Biology**, v. 33, n. 1, p. 54–78, 1988.

HOYOS-VILLEGAS, V.; SONG, Q.; KELLY, J. D. Genome-wide Association Analysis for Drought Tolerance and Associated Traits in Common Bean. **The Plant Genome**, v. 10, n. 1, p. 1–17, 2017.

HU, Z. et al. An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. **Plant Genome**, v. 12, n. 1, 2019.

HYTEN, D. L. et al. Highly variable patterns of linkage disequilibrium in multiple soybean populations. **Genetics**, v. 175, n. 4, p. 1937–1944, 2007.

IKRAM, M. et al. Identification of QTNs and Their Candidate Genes for 100-Seed Weight in Soybean (*Glycine max* L.) Using Multi-Locus Genome-Wide Association Studies. **Genes**, v. 11, n. 714, 2020.

ISLAM, F. M. A. et al. Seed compositional and disease resistance differences among gene pools in cultivated common bean. **Genetic Resources and Crop Evolution**, v. 49, n. 3, p. 285–293, 2002.

JAISWAL, V. et al. Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (*Setaria italica* L.). **Journal of Cereal Science**, v. 85, p. 48–55, 2019.

JENSEN, E. S. et al. **Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review**. [s.l: s.n.]. v. 32

KAMFWA, K.; CICHY, K. A.; KELLY, J. D. Genome-wide association study of agronomic traits in common bean. **The plant genome**, v. 8, n. 2, p. 1–12, 2015a.

KAMFWA, K.; CICHY, K. A.; KELLY, J. D. Genome-wide association analysis of symbiotic nitrogen fixation in common bean. **Theoretical and Applied Genetics**, v. 128, n. 10, p. 1999–2017, 2015b.

KATUURAMU, D. N. et al. Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (*Phaseolus vulgaris* L.). **Molecular Breeding**, v. 38, n. 4, 2018.

KELLER, B. et al. Genomic Prediction of Agronomic Traits in Common Bean (*Phaseolus vulgaris* L.) Under Environmental Stress. **Frontiers in Plant Science**, v. 11, n. July, p. 1–15, 2020.

KOLDE, R. **pheatmap: Pretty Heatmaps. R package version 1.0.12.**, 2019.

KORTE, A.; ASHLEY, F. The advantages and limitations of trait analysis with GWAS : a review Self-fertilisation makes Arabidopsis particularly well suited to GWAS. **Plant methods**, v. 9, n. 1, p. 29, 2013.

KWAK, M.; GEPTS, P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). **Theoretical and Applied Genetics**, v. 118, n. 5, p.

979–992, 2009.

LAN, S. et al. Genomic prediction accuracy of seven breeding selection traits improved by QTL identification in flax. **International Journal of Molecular Sciences**, v. 21, n. 5, 2020.

LEI, L. et al. Marker-Trait Association Analysis of Seed Traits in Accessions of Common Bean (*Phaseolus vulgaris* L.) in China. **Frontiers in Genetics**, v. 11, n. June, p. 1–12, 2020.

LI, C. et al. Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton (*Gossypium hirsutum* l.). **Frontiers in Plant Science**, v. 9, n. August, p. 1–16, 2018a.

LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589–595, 2010.

LI, J. et al. Genome-wide association studies for five forage quality-related traits in sorghum (*Sorghum bicolor* l.). **Frontiers in Plant Science**, v. 9, n. August, p. 1–8, 2018b.

LI, L. et al. Loci and candidate gene identification for resistance to Phytophthora sojae via association analysis in soybean [*Glycine max* (L.) Merr.]. **Molecular Genetics and Genomics**, v. 291, n. 3, p. 1–9, 2016.

LIOI, L. et al. Genotyping-by-sequencing reveals molecular genetic diversity in Italian common bean landraces. **Diversity**, v. 11, n. 9, p. 1–14, 2019.

LOW, J. W. et al. A food-based approach introducing orange-fleshed sweet potatoes increased vitamin A intake and serum retinol concentrations in young children in rural Mozambique. **The Journal of nutrition**, v. 137, p. 1320–1327, 2007.

LÜ, H. et al. Genome-wide association studies of photosynthetic traits related to phosphorus efficiency in soybean. **Frontiers in Plant Science**, v. 9, n. August, 2018.

LUO, Z. et al. Genetic diversity and population structure of a Camelina sativa spring panel. **Frontiers in Plant Science**, v. 10, n. February, p. 1–12, 2019.

LYNCH, M.; WALSH, B. **Genetics and Analysis of Quantitative Traits.** Sunderland: Sinauer Associates, 1998.

MA, L. et al. Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. **Frontiers in Plant Science**, v. 9, n. April, p. 1–15, 2018.

MACQUEEN, A. H. et al. Genetic associations in four decades of multienvironment trials reveal agronomic trait evolution in common bean. **Genetics**, v. 215, n. 1, p. 267–284, 2020.

MAPA, MINISTÉRIO DA AGRICULTURA PECUÁRUA E ABASTECIMENTO. **Plano nacional de desenvolvimento da cadeia do feijão e pulses.** Brasília, DF, 2018.

MCCLEAN, P. E. et al. Phenotypic diversity for seed mineral concentration in North American dry bean germplasm of middle american ancestry. **Crop Science**, v. 57, n. 6, p. 3129–3144, 2017.

MIKLAS, P. N. et al. Common bean breeding for resistance against biotic and abiotic stresses: From classical to MAS breeding. **Euphytica**, v. 147, n. 1–2, p. 105–131, 2006.

MISRA, G. et al. Deciphering the genetic architecture of cooked rice texture. **Frontiers in Plant Science**, v. 9, n. October, 2018.

MIYAZAWA, M. et al. Análises químicas de tecido vegetal. In: SILVA, F. C. (Ed.). . **Manual de análises químicas de solos, plantas e fertilizantes**. Brasília: Embrapa, 1999. p. 171–223.

MOGHADDAM, S. M. et al. Genome-wide association study identifies candidate loci underlying agronomic traits in a middle American diversity panel of common bean. **Plant Genome**, v. 9, n. 3, p. 1–21, 2016.

MOGHADDAM, S. M. et al. Genetic architecture of dietary fiber and oligosaccharide content in a middle american panel of edible dry bean. **Plant Genome**, v. 11, n. 1, p. 1–11, 2018.

MUKAMUHIRWA, F.; TUSIIME, G.; MUKANKUSI, M. C. Inheritance of high iron and zinc concentration in selected bean varieties. **Euphytica**, v. 205, n. 2, p. 349–360, 2015.

NADEEM, M. A. et al. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. **Biotechnology and Biotechnological Equipment**, v. 32, n. 2, p. 261–285, 2018.

NADEEM, M. A. et al. Phenotypic characterization of 183 Turkish common bean accessions for agronomic, trading, and consumer-preferred plant characteristics for breeding purposes. **Agronomy**, v. 10, n. 2, 2020.

NAKANO, Y.; KOBAYASHI, Y. Genome-wide association studies of agronomic traits consisting of field-and molecular-based phenotypes. **Reviews in Agricultural Science**, v. 8, n. 1, p. 28–45, 2020.

NASCIMENTO, M. et al. Quantile regression for genome-wide association study of flowering time-related traits in common bean. **PLoS ONE**, v. 13, n. 1, p. 1–14, 2018.

NEMLI, S. et al. Association mapping for five agronomic traits in the common bean (*Phaseolus vulgaris* L.). **Journal of the Science of Food and Agriculture**, v. 94, n. 15, p. 3141–3151, 2014.

NICOLAS, S. D. et al. Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L) diversity panel newly designed for association studies. **BMC Plant Biology**, v. 16, n. 1, p. 1–19, 2016.

NORDBORG, M. et al. The extent of linkage disequilibrium in Arabidopsis thaliana. **Nature Genetics**, v. 30, n. 2, p. 190–3, 2002.

OLADOSU, Y. et al. Drought Resistance in Rice from Conventional to Molecular Breeding: A Review. **International Journal of Molecular Sciences**, v. 20, n. 14, 2019.

OLADZAD, A. et al. Single and multi-trait GWAS identify genetic factors associated with production traits in common bean under abiotic stress environments. **G3: Genes, Genomes, Genetics**, v. 9, n. 6, p. 1881–1892, 2019.

ORAGUZIE, N. et al. **Association mapping in plants**. New York: Springer, 2007.

PAN, L. et al. Efficient QTL detection of flowering date in a soybean RIL population using the novel restricted two-stage multi-locus GWAS procedure. **Theoretical and Applied Genetics**, v. 131, n. 12, p. 2581–2599, 2018.

PARKER, T.; PALKOVIC, A.; GEPTS, P. Determining the Genetic Control of Common Bean Early-Growth Rate using Unmanned Aerial Vehicles. **Remote Sensing**, v. 12, n. 11, p. 1748, 2020.

PAVAN, S. et al. A distinct genetic cluster in cultivated chickpea as revealed by genome-wide marker discovery and genotyping. **Plant Genome**, v. 10, n. 2, p. 1–9, 2017.

PEREIRA-DIAS, L. et al. Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). **Horticulture Research**, v. 6, n. 1, 2019.

PÉREZ-VEGA, E. et al. Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, v. 120, n. 7, p. 1367–1380, 2010.

PERSEGUINI, J. M. K. C. et al. Developing a common bean core collection suitable for association mapping studies. **Genetics and Molecular Biology**, v. 38, n. 1, p. 67–78, 2015.

PERSEGUINI, J. M. K. C. et al. Genome-Wide association studies of anthracnose and angular leaf spot resistance in common bean (*Phaseolus vulgaris* L.). **PLoS ONE**, v. 11, n. 3, p. 1–19, 2016.

PETRY, N. et al. Review: The potential of the common bean (*Phaseolus vulgaris*) as a vehicle for iron biofortification. **Nutrients**, v. 7, n. 2, p. 1144–1173, 2015.

PFEIFFER, W. H.; MCCLAFFERTY, B. HarvestPlus: Breeding crops for better nutrition. **Crop Science**, v. 47, n. SUPPL. DEC., 2007.

PINHEIRO, C. et al. Diversity of seed mineral composition of Phaseolus vulgaris L. germplasm. **Journal of Food Composition and Analysis**, v. 23, n. 4, p. 319–325, 2010.

PIPAN, B.; MEGLIČ, V. Diversification and genetic structure of the western-to-eastern progression of European Phaseolus vulgaris L. germplasm. **BMC Plant Biology**, v. 19, n. 1, p. 1–16, 2019.

POLAND, J. A. et al. Genotyping-by-Sequencing for Plant Breeding and Genetics. **The Plant Genome Journal**, v. 5, n. 3, p. 92, 2012.

PRICE, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, v. 38, n. 8, p. 904–909, 2006.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, n. 2, p. 945–959, 2000.

PURCELL, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

R CORE TEAM. **R: A language and environment for statistical computing.**Vienna, AustriaR Foundation for Statistical Computing, , 2020. Disponível em: <https://www.r-project.org/>

RAATZ, B. et al. Analyses of African common bean (*Phaseolus vulgaris* L.) germplasm using a SNP fingerprinting platform: diversity, quality control and molecular breeding. **Genetic Resources and Crop Evolution**, v. 66, n. 3, p. 707–722, 2019.

RAGGI, L. et al. Genome-wide association study reveals candidate genes for flowering time variation in common bean (*Phaseolus vulgaris* L.). **Frontiers in Plant Science**, v. 10, n. July, p. 1–14, 2019.

RAMALHO, M. A. P.; DIAS, L. A. DOS S.; CARVALHO, B. L. Contributions of plant breeding in Brazil – progress and perspectives. **Crop Breeding and Applied Biotechnology**, v. S2, p. 111–120, 2012.

RAMBAUT, A. **FigTree v1.4.4.** EdinburghInstitute of Evolutionary Biology, University of Edinburgh, , 2018.

RANA, J. C. et al. Characterisation of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. **Euphytica**, v. 205, n. 2, p. 441–457, 2015.

RASHEED, A. et al. Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. **Molecular Plant**, v. 10, n. 8, p. 1047–1064, 2017.

RAWAL, V.; NAVARRO, D. K. **The Global Economy of Pulses**. [s.l.] FAO, 2019.

RESENDE, R. T. et al. Genome-wide association and Regional Heritability Mapping of plant architecture, lodging and productivity in phaseolus vulgaris. **G3: Genes, Genomes, Genetics**, v. 8, n. 8, p. 2841–2854, 2018a.

RESENDE, R. T. et al. Genome-wide association and Regional Heritability Mapping of plant architecture, lodging and productivity in phaseolus vulgaris. **G3: Genes, Genomes, Genetics**, v. 8, n. 8, p. 2841–2854, 2018b.

RIBEIRO, N. D. et al. Combined selection for grain yield, cooking quality and minerals in the common bean. **Revista Ciencia Agronomica**, v. 44, n. 4, p. 869–877, 2013.

RIBEIRO, N. D. et al. Evaluation of special grains bean lines for grain yield, cooking time and mineral concentrations. **Crop Breeding and Applied Biotechnology**, v. 14, n. 1, p. 15–22, 2014.

RISCH, N.; MERIKANGAS, K. The future of genetic studies of complex human diseases. **Science (New York, N.Y.)**, v. 273, n. 5281, p. 1516–1517, 1996.

SALTZ, J. B.; HESSEL, F. C.; KELLY, M. W. Trait Correlations in the Genomics Era. **Trends in Ecology and Evolution**, v. 32, n. 4, p. 279–290, 2017.

SAS INSTITUTE. **JMP IN 4.0.3.** Cary, North Carolina, 2000.

SCHMUTZ, J. et al. A reference genome for common bean and genome-wide analysis of dual domestications. **Nature genetics**, v. 46, n. 7, p. 707–13, 2014.

SCHRÖDER, S. et al. Optimization of genotyping by sequencing (GBS) data in common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, v. 36, n. 1, p. 1–9, 2016.

SHI, C. et al. Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. **BMC Plant Biology**, v. 11, n. 1, p. 52, 2011.

SIADJEU, C.; MAYLAND-QUELLHORST, E.; ALBACH, D. C. Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). **BMC Plant Biology**, v. 18, p. 1–14, 2018.

SILVA, C. A. et al. Chemical composition as related to seed color of common bean. **Crop Breeding and Applied Biotechnology**, v. 12, n. 2, p. 132–137, 2012.

SILVA, C. A.; ABREU, Â. DE F. B.; RAMALHO, M. A. P. Genetic control of zinc and iron concentration in common bean seeds. **African Journal of Agricultural Research**, v. 8, n. 11, p. 1001–1008, 2013.

SINGH, S. P.; NODARI, R.; GEPTS, P. Genetic Diversity in Cultivated Common Bean: I. Allozymes. **Crop Science**, v. 31, n. 1, p. 19, 1991.

SOLTANI, A. et al. Targeted analysis of dry bean growth habit: Interrelationship among architectural, phenological, and yield components. **Crop Science**, v. 56, n. 6, p. 3005–3015, 2016.

SOLTANI, A. et al. Genetic architecture of flooding tolerance in the dry bean middle-American diversity panel. **Frontiers in Plant Science**, v. 8, n. July, p. 1–15, 2017.

SOLTANI, A. et al. Genetic analysis of flooding tolerance in an andean diversity panel of dry bean (*Phaseolus vulgaris* L.). **Frontiers in Plant Science**, v. 9, n. June, 2018.

SONAH, H. et al. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. **Plant Biotechnology Journal**, v. 13, n. 2, p. 211–221, 2015.

STANSELL, Z. et al. Genotyping-by-sequencing of Brassica oleracea vegetables reveals unique phylogenetic patterns, population structure and domestication footprints. **Horticulture Research**, v. 5, n. 1, 2018.

STAPLEY, J. et al. Adaptation genomics: The next generation. **Trends in Ecology and Evolution**, v. 25, n. 12, p. 705–712, 2010.

TAJIMA, F. Statistical method for testing the neutral mutation hy- pothesis by DNA polymorphism. **Genetics**, v. 123, n. 3, p. 585–595, 1989.

TALSMA, E. F. et al. Biofortified yellow cassava and vitamin A status of Kenyan children : a randomized controlled trial. **American Journal of Clinical Nutrition**, v. 103, n. January, p. 258–267, 2016.

TAMBA, C. L.; NI, Y. L.; ZHANG, Y. M. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. **PLoS Computational Biology**, v. 13, n. 1, p. 1–20, 2017.

TAMBA, C. L.; ZHANG, Y.-M. A fast mrMLM algorithm for multi-locus genome-wide association studies. p. 341784, 2018.

TEIXEIRA, F. F.; RAMALHO, M. A. P.; ABREU, Â. D. F. B. Genetic control of plant architecture in the common bean (*Phaseolus vulgaris* L.). **Genetics and Molecular Biology**, v. 22, n. 4, p. 577–582, 1999.

TOCK, A. J. et al. Genome-wide linkage and association mapping of halo blight resistance in common bean to race 6 of the globally important bacterial pathogen. **Frontiers in Plant Science**, v. 8, n. July, p. 1–17, 2017.

TSUTSUMI, C. Y.; BULEGON, L. G.; PIANO, J. T. Melhoramento genético do feijoeiro:

avanços, perspectivas e novos estudos, no âmbito nacional. **Nativa**, v. 3, n. 3, p. 217–223, 2015.

VALDISSER, P. A. et al. Genome-wide association studies detect multiple QTLs for productivity in Mesoamerican diversity panel of common bean under drought stress. **Frontiers in Plant Science**, 2020.

VALDISSER, P. A. M. R. et al. SNP discovery in common bean by restriction-associated DNA (RAD) sequencing for genetic diversity and population structure analysis. **Molecular Genetics and Genomics**, v. 291, n. 3, p. 1277–1291, 2016.

VALDISSER, P. A. M. R. et al. In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. **BMC Genomics**, v. 18, n. 1, p. 1–19, 2017.

VALENÇA, A. W. et al. Agronomic biofortification of crops to fight hidden hunger in sub-Saharan Africa. **Global Food Security**, v. 12, p. 8–14, 2017.

VALENTINI, G. et al. Population structure and genetic diversity of common bean accessions from Brazil. **Plant Molecular Biology Reporter**, v. 36, n. 5–6, p. 897–906, 2018.

VAN JAARSVELD, P. J. et al. Beta-carotene-rich orange-fleshed sweet potato improves the vitamin A status of primary school children assessed with the modified-relative-dose-response test. **The American journal of clinical nutrition**, v. 81, n. 5, p. 1080–1087, 2005.

VIEIRA, C. **Cultura do feijão**. 2. ed. Viçosa: UFV, 1983.

VILHORDO, B. W. .; BURIN, M. E. .; GANDOLFI, V. H. Morfologia. In: ZIMMERMANN, M. J. O. .; ROCHA, M.; YAMADA, T. (Eds.). . **Cultura do feijoeiro: fatores que afetam a produtividade**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1988. p. 87–123.

VILLORDO-PINEDA, E. et al. Identification of novel drought-tolerant-associated SNPs in common bean (Phaseolus vulgaris). **Frontiers in Plant Science**, v. 6, n. JULY, p. 1–9, 2015.

VLASOVA, A. et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. **Genome Biology**, v. 17, n. 1, p. 1–18, 2016.

VOYSEST, O. V. **Mejoramiento genetico del frijol (Phaseolus vulgaris) legado de variedades de América Latina (1930-1999)**. Cali: CIAT, 2000.

WANG, M. et al. Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. **Theoretical and Applied Genetics**, v. 124, n. 2, p. 233–246, 2012.

WANG, S. et al. Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. **Genome Biology**, v. 16, n. 1, 2015.

WANG, S. B. et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. **Scientific Reports**, v. 6, n. January, p. 1–10, 2016.

WARD, B. P. et al. Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. **PLoS ONE**, v. 14, n. 2, p. 1–28, 2019.

WEI, T.; SIMKO, V. **R package "corrplot": Visualization of a Correlation Matrix (Version 0.84)**, 2017.

WEIR, B. S.; COCKERHAM, C. C. Estimating F-statistics for the analysis of population structure. **Evolution**, v. 38, n. 6, p. 1358–1370, 1984.

WEN, Y. J. et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. **Briefings in bioinformatics**, v. 19, n. 4, p. 700–712, 2018.

WU, J. et al. Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. **Nature Genetics**, v. 52, n. 1, p. 118–125, 2020.

XU, Q. et al. The genetic diversity and structure of indica rice in China as detected by single nucleotide polymorphism analysis. **BMC Genetics**, v. 17, n. 1, p. 1–8, 2016.

YA-WEN, Z.; PEI, L.; YUAN-MING, Z. **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.**, 2019.

YANG, Y. et al. Multi-Locus GWAS of Quality Traits in Bread Wheat: Mining More Candidate Genes and Possible Regulatory Network. **Frontiers in Plant Science**, v. 11, n. July, p. 1–19, 2020.

YU, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, v. 38, n. 2, p. 203–208, 2006.

ZHANG, D. et al. Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. **The Plant Genome Journal**, v. 3, n. 2, p. 117, 2010.

ZHANG, J. et al. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. **BMC Genomics**, v. 16, n. 1, p. 1–11, 2015.

ZHANG, J. et al. PLARmEB: Integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. **Heredity**, v. 118, n. 6, p. 517–524, 2017a.

ZHANG, K. et al. Identification of QTLs for 14 agronomically important traits in Setaria italica based on SNPs generated from high-throughput sequencing. **G3: Genes, Genomes, Genetics**, v. 7, n. 5, p. 1587–1594, 2017b.

ZHANG, K. et al. Identification of QTNs controlling seed protein content in soybean using multi-locus genome-wide association studies. **Frontiers in Plant Science**, v. 871, n. November, p. 1–14, 2018a.

ZHANG, Y. et al. Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. **Frontiers in Plant Science**, v. 9, n. May, 2018b.

ZHANG, Y. M. et al. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). **Genetics**, v. 169, n. 4, p. 2267–2275, 2005.

ZHANG, Y. M.; JIA, Z.; DUNWELL, J. M. Editorial: The applications of new multi-locus gwas methodologies in the genetic dissection of complex traits. **Frontiers in Plant Science**, v.

10, n. February, p. 1–6, 2019.

ZHENG, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. **Bioinformatics**, v. 28, n. 24, p. 3326–3328, 2012.

ZHU, C. et al. Status and Prospects of Association Mapping in Plants. **The Plant Genome**, v. 1, n. 1, p. 5, 2008.

ZUIDERVEEN, G. H. et al. Genome-wide association study of anthracnose resistance in Andean beans (*Phaseolus vulgaris*). **PLoS ONE**, p. 1–17, 2016.