

SIRLENE VIANA DE FARIA

**CARACTERIZAÇÃO FENOTÍPICA E MOLECULAR E ANÁLISE DE GWAS
E GWS PARA CARACTERES DE RAIZ EM UM CONJUNTO DE LINHAGENS
ENDOGÂMICAS DE MILHO TROPICAL**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

Orientador: Rodrigo Oliveira de Lima

VIÇOSA – MINAS GERAIS

2021

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

F224
2021 Faria, Sirlene Viana, 1987-
Caracterização fenotípica e molecular e análise de GWAS e
GWS para caracteres de raiz em um conjunto de linhagens
endogâmicas de milho tropical / Sirlene Viana Faria. – Viçosa,
MG, 2021.

115 f. : il. (algumas color.) ; 29 cm.

Orientador: Rodrigo Oliveira de Lima.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Milho - Melhoramento genético. 2. Diversidade genética.
3. Desequilíbrio de ligação. 4. Polimorfismo de Nucleotídeo
Único. I. Universidade Federal de Viçosa. Departamento de
Agronomia. Programa de Pós-Graduação em Genética e
Melhoramento. II. Título.

CDD 22. ed. 633.1592

SIRLENE VIANA DE FARIA


**CARACTERIZAÇÃO FENOTÍPICA E MOLECULAR E ANÁLISE DE GWAS
E GWS PARA CARACTERES DE RAIZ EM UM CONJUNTO DE
LINHAGENS ENDOGÂMICAS DE MILHO TROPICAL**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 29 de janeiro de 2021.

Assentimento:


Sirlene Viana de Faria
Autora


Rodrigo Oliveira de Lima
Orientador

A Deus, por me guiar em todo meu caminho

OFEREÇO

Ao meu marido Bruno e ao meu filho amado Luís Gustavo pelo amor e carinho, por serem meu porto seguro e razão da minha felicidade

DEDICO

AGRADECIMENTOS

A Deus, pois ele é responsável por tudo que sou e tenho e por estar presente em todos os momentos.

Ao meu marido Bruno pelo companheirismo, carinho e apoio em todos esses anos juntos, por compreender todos os meus momentos e dificuldades.

Ao meu filho Luís Gustavo, agradeço pelo amor incondicional, pelos sorrisos diários, por me ensinar a cada dia e pelo carinho que sempre me estimularam nos momentos difíceis.

Aos meus pais, Maria Aparecida e João Dimas, aos meus sogros Rita Maria e Marcelo, as minhas sobrinhas do coração Emmily e Amanda pelo apoio, incentivo, pelas orações e por estarem presentes nessa caminhada.

Ao orientador Prof. Rodrigo Oliveira de Lima, pelos ensinamentos, amizade e por ser um modelo de profissional a ser seguido.

Aos professores, Maicon Nardino, Felipe Lopes da Silva, a Dra. Lisandra Magna Moura e ao Dr. Kaio Olimpio das Graças Dias pela disponibilidade e sugestões para o enriquecimento do trabalho.

À equipe do Programa Milho-UFV, estagiários, bolsistas de iniciação científica e pós-graduandos, pelo apoio, amizade e por todo conhecimento compartilhado, em especial ao Dr. Leandro T. Zuffo por toda ajuda para a concretização deste trabalho.

À Universidade Federal de Viçosa, em especial ao programa de pós-graduação em Genética e Melhoramento, pela oportunidade de realização do doutorado.

Aos secretários do Programa de Pós-graduação em Genética e Melhoramento da Universidade Federal de Viçosa Marco Túlio e Odilon por todo apoio durante o doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) e à Fundação Arthur Bernardes (Funarbe), pelo financiamento dos projetos e concessão da bolsa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

BIOGRAFIA

SIRLENE VIANA DE FARIA, filha de João Dimas de Faria e Maria Aparecida Viana de Faria, nasceu em Viçosa/MG, no dia 18 de setembro de 1987. Ingressou no curso de Agronomia da Universidade Federal de Viçosa/UFV em março de 2008 e concluiu em setembro de 2013. Em março de 2014 ingressou no Mestrado em Genética e Melhoramento da Universidade Federal de Viçosa, concluiu em fevereiro de 2016. Em agosto do mesmo ano ingressou no Doutorado em Genética e Melhoramento da Universidade Federal de Viçosa, concluindo-o em janeiro de 2021.

RESUMO

FARIA, Sirlene Viana, M.Sc., Universidade Federal de Viçosa, janeiro de 2021. **Caracterização fenotípica e molecular e análise de GWAS e GWS para caracteres de raiz em um conjunto de linhagens endogâmicas de milho tropical.** Orientador: Rodrigo Oliveira de Lima.

A arquitetura de raiz do milho é importante para a ancoragem das plantas e para a absorção de nutrientes e água. Ferramentas como a associação genômica ampla (GWAS) e a seleção genômica ampla (GWS) podem ajudar os melhoristas a tomar decisões sobre quais genótipos devem avançar em seus programas de melhoramento. Foi realizada a caracterização fenotípica e molecular de um conjunto de 187 linhagens endogâmicas de milho tropical do programa de melhoramento público da Universidade Federal de Viçosa e posteriormente a análise de GWAS para identificar polimorfismo de nucleotídeo único (SNP) associados a caracteres de morfologia de raiz relacionados a eficiência no uso de nitrogênio e possíveis genes candidatos que afetam o desenvolvimento das raízes e avaliar a acurácia da seleção genômica (GWS). Os resultados mostraram que existe uma grande variação fenotípica e genotípica no conjunto de linhagens. Também foi encontrado alta diversidade genética ($GD = 0,34$) e baixos coeficientes de parentesco entre as linhagens de milho. O declínio do LD com o aumento da distância física em todos os dez cromossomos em todo o conjunto de linhagens de milho com $r^2 = 0,2$ foi de 86.899 pb. Em relação à estrutura populacional, os resultados do STRUCTURE e da análise de componentes principais distinguiram as linhagens endogâmicas em três subpopulações, com muita consistência entre os dois resultados. Além disso, os resultados da análise de agrupamento com base em dados fenotípicos e moleculares agruparam as linhagens em 14 e 22 agrupamentos de divergência genética, respectivamente. Na GWAS, um total de 10 SNPs acima do limite de significância foi detectado. Esses SNPs levaram a 16 genes candidatos. Na GWS, as acurácias para o modelo que incorpora todos os 3.083 SNPs variaram de -0,22 a 0,16. Quando foi considerado apenas os SNPs associados significativamente as acurácias variaram de -0,17 a 0,46. A análise de GWAS revelou marcadores associadas aos caracteres de morfologia de raiz, localizados em regiões que contém genes ou modelos de genes com funções celulares relacionados de proteção das plantas a estresses ambientais, incluindo estresse de nitrogênio. Na GWS, as acurácias do modelo que utilizaram apenas os SNPs associados aos caracteres foram significativamente maiores do que as acurácias obtidas no modelo que incorporaram

todos os SNPs para todos os 14 caracteres avaliados nas condições de BN e AN. Os genes encontrados podem ser mais estudados para ajudar a entender a base genética do desenvolvimento radicular e auxiliar no desenvolvimento de cultivares de milho mais eficientes no uso de nitrogênio.

Palavras-chave: Diversidade genética. Estrutura de população. Desequilíbrio de ligação. Associação genômica ampla. Seleção genômica.

ABSTRACT

FARIA, Sirlene Viana, M.Sc., Universidade Federal de Viçosa, January, 2021. **Phenotypic and molecular characterization and analysis of GWAS and GWS for root traits in a set of tropical maize inbred lines.** Adviser: Rodrigo Oliveira de Lima.

The root architecture of maize is important for plant anchorage and for the absorption of nutrients and water. Tools such as Genome Wide Association Studies (GWAS) and Genomic Wide Selection (GWS) can help breeders to make decisions about which genotypes should advance in their breeding programs. We carried out the phenotypic and molecular characterization of a set of 187 inbred lines of tropical maize from the public breeding program of the Universidade Federal of Viçosa (UFV) and later the GWAS analysis to identify single nucleotide polymorphism (SNP) associated with root morphology characters related to the efficiency in the use of nitrogen and possible candidate genes that affect root development and evaluate the accuracy of genomic selection. The results showed that there is a great phenotypic and genotypic variation in the set of inbred lines of tropical maize from the UFV corn breeding program. We also found a high genetic diversity ($GD = 0.34$) and low coefficients of kinship between maize line (only about 4.00% of the kinship relative to the pairs was above 0.50). The LD decay distance on all ten chromosomes across the set of maize lines with $r^2 = 0.2$ was 86,899 kb. Regarding the population structure, the results of STRUCTURE and the analysis of main components distinguished inbred lines in three subpopulations, with much consistency between the two results. In addition, the results of the UPGMA cluster analysis based on phenotypic and molecular data grouped the lines into 14 and 22 genetic divergence clusters, respectively. At GWAS, a total of 10 SNPs above the significance limit were detected. These SNPs led to 16 candidate genes. At GWS, the accuracy for the model that incorporates all 3,083 SNPs ranged from -0.22 to 0.16. When only the significantly associated SNPs were considered, the accuracy varied from -0.17 to 0.46. The GWAS analysis revealed markers associated with the root morphology characters, located in regions that contain genes or models of genes with related cellular functions to protect plants from environmental stresses, including nitrogen stress. At GWS, the accuracy of the model that used only the SNPs associated with the characters was significantly higher than the accuracy obtained in the model that incorporated all SNPs for all 14 characters evaluated under the conditions of BN and AN. The genes found can

be further studied to help understand the genetic basis of root development and assist in the development of corn cultivars that are more efficient in using nitrogen.

Keywords: Genetic diversity. Population structure. Linkage Disequilibrium. Genome wide association studies. Genomic wide selection.

SUMÁRIO

1.	INTRODUÇÃO GERAL	12
2.	OBJETIVOS	15
2.1.	Objetivo Geral.....	15
2.2.	Objetivos Específicos.....	15
3.	REVISÃO DE LITERATURA	16
3.1.	Cultura do Milho	16
3.2.	Melhoramento com foco em eficiência no uso de nitrogênio	17
3.3.	Diversidade Genética	19
3.4.	Estrutura Populacional	23
3.5.	Desequilíbrio de Ligação	26
3.6.	Estudos de Associação genômica ampla (GWAS)	30
3.7.	Seleção genômica ampla (GWS)	31
4.	REFERÊNCIAS BIBLIOGRAFICAS	33
5.	CAPÍTULO 1: Caracterização molecular e fenotípica de um conjunto de linhagens de milho tropical	43
	RESUMO	44
5.1.	INTRODUÇÃO	45
5.2.	MATERIAL E MÉTODOS	47
5.2.1.	Material vegetal.....	47
5.2.2.	Dados Fenotípicos	48
5.2.3.	Dados Genotípicos	49
5.2.4.	Análises estatísticas dos dados fenotípicos	50
5.2.5.	Análises estatísticas dos dados genotípicos	51
5.3.	RESULTADOS	53
5.3.1.	Variação fenotípica e parâmetros genéticos com base em caracteres fenotípicos 53	
5.3.2.	Caracterização molecular	55
5.3.3.	Parentesco relativo	57
5.3.4.	Desequilíbrio de ligação (LD).....	58
5.3.5.	Estrutura populacional	59
5.3.6.	Análise de componentes principais	61

5.3.7. Diversidade genética com base em dados fenotípicos e dados moleculares.....	62
5.4. DISCUSSÃO	65
5.5. CONCLUSÃO.....	71
5.6. REFERÊNCIAS BIBLIOGRÁFICAS.....	72
6. CAPÍTULO 2: Estudo de associação genômica e análise da predição genômica para caracteres de raiz relacionados a eficiência no uso de nitrogênio em um painel de linhagens de milho tropical.....	82
RESUMO	83
6.1. INTRODUÇÃO.....	84
6.2. MATERIAL E MÉTODOS.....	86
6.2.1. Material Genético.....	86
6.2.2. Dados fenotípicos.....	87
6.2.3. Genotipagem	87
6.2.4. Mapeamento da Associação Genômica Ampla.....	88
6.2.5. Seleção genômica.....	90
6.2.6. Análise de Genes Candidatos.....	90
6.3. RESULTADOS	91
6.3.1. Caracteres fenotípicos em AN e BN	91
6.3.2. Estudos de Associação em todo o genoma e genes candidatos	91
6.3.3. Seleção genômica.....	99
6.4. DISCUSSÃO	101
6.4.1. Associação genômica e genes candidatos	101
6.4.2. Seleção genômica.....	106
6.5. CONCLUSÃO.....	107
6.6. REFERÊNCIA BIBLIOGRÁFICA.....	107
7. CONCLUSÃO GERAL	115

1. INTRODUÇÃO GERAL

O milho (*Zea mays ssp. mays*) é um dos cereais mais importantes cultivados para alimentos, rações e biocombustíveis em todo o mundo. O Brasil é o terceiro maior produtor mundial de milho, depois dos Estados Unidos e da China (FAO, 2020). Atualmente, o milho híbrido é responsável por mais de 95% da área de produção de milho no Brasil. A expansão dos locais e da época de cultivo do milho, faz com que uma fração significativa da produção do cereal seja conduzido em condições estressantes por fatores abióticos (Monneveux et al., 2005; Giacomini et al., 2015). Entre os diversos tipos de estresses abióticos, a deficiência de nitrogênio (N) é comum nos solos brasileiros, somada a baixa eficiência na adubação nitrogenada, geralmente menor que 50% devido a perda por volatilização, desnitrificação e lixiviação (Gallais e Hirel, 2004), afeta negativamente o desenvolvimento das plantas e reduz a produtividade (Fritsche-Neto e Borém, 2012). Por meio do melhoramento genético, é possível desenvolver cultivares de milho mais eficientes no uso de nitrogênio (EUN), ou seja, que produzam mais ou a mesma quantidade de grãos, com menor demanda por N.

Uma estratégia para aumentar a eficiência do uso de nitrogênio (EUN) no milho é o melhoramento genético da aquisição de N por meio da morfologia radicular (Torres et al., 2019). Caracteres relacionados à morfologia do sistema radicular podem fornecer informações relevantes sobre a EUN em genótipos de milho (Majerowicz et al., 2002). A morfologia da raiz pode ter um impacto considerável na EUN em milho, principalmente em solos secos e empobrecidos, uma vez que as características da raiz desempenham um papel essencial na aquisição de N (Saengwilai et al., 2014; Morosini et al., 2017; Sanchez et al., 2018).

A arquitetura do sistema radicular pode ser de grande importância para melhorar a EUN em cereais (Garnett et al., 2009). Isto porque o sistema radicular exerce papel essencial na captação do N via água (Barber e Mackay, 1986; Horst et al., 2003). Assim, sob estresse nutricional ou hídrico, por exemplo, um maior desenvolvimento radicular e, conseqüentemente, da área superficial das raízes pode auxiliar a obter o nutriente em camadas de solo mais profundas e evitar a deficiência (Horst et al., 2001; Chun et al., 2005; Pace et al., 2015a). Foi descoberto que a demanda por nutrientes está correlacionada com o crescimento da raiz (Peng et al., 2010) e intimamente associado ao rendimento de grãos (Cai et al., 2012). Sendo assim, regiões genômicas que afetam o

desenvolvimento e o crescimento das raízes podem afetar a EUN (Pace et al., 2015a). Os caracteres da raiz são quantitativos, controlados por vários genes e muito difíceis de serem mensurados (Hochholdinger et al., 2018). A caracterização de genótipos com base em dados moleculares pode contribuir para otimizar o uso de germoplasma em relação à seleção para EUN (Morosini et al., 2017).

Ferramentas como a associação genômica ampla (GWAS) e a seleção genômica ampla (GWS) podem auxiliar os melhoristas a tomar decisões sobre quais genótipos devem avançar em seus programas de melhoramento. O mapeamento de associação bem-sucedido requer a escolha ou criação de uma coleção de germoplasma e uma compreensão profunda da diversidade genética e o parentesco, estrutura populacional e a extensão do decaimento do desequilíbrio de ligação (LD) desse germoplasma (Ersoz et al., 2009; Zhu et al., 2008; Flint-Garcia et al., 2003; 2005). O milho tem uma grande diversidade genética e rápido declínio do LD, o que o torna uma cultura modelo para análise de associação (Flint-Garcia et al. 2005; Yan et al., 2011).

A GWAS é uma metodologia baseada na associação entre marcadores de polimorfismo de nucleotídeo único (SNP) de alta densidade e o fenótipo de interesse (Korte e Farlow, 2013) em nível populacional por meio de testes de hipóteses visando detectar efeitos de SNPs significativos. Essa metodologia vem sendo extensivamente usada em estudos genéticos de plantas para condições de estresses abióticos (Xue et al., 2013; Zhang et al., 2015; Laperche et al., 2017). Wang et al. (2019) avaliaram 210 linhagens de milho para tolerância à seca no estágio de muda e identificaram um total de 413 SNPs associados aos caracteres nas condições de alta disponibilidade de água e 696 SNPs nas condições de estresse hídrico. Eles encontraram informações valiosas para a compreensão da variação genética da tolerância ao estresse hídrico no milho. Morosini et al. (2017) avaliaram 74 linhagens de milho tropicais em campo contrastando para EUN. A análise de GWAS revelou regiões genômicas associadas com EUN em condições de campo.

A GWS ou predição genômica, proposta por Meuwissen et al. (2001) permite a predição simultânea dos efeitos dos marcadores, é um método de seleção que pode aumentar o ganho genético por unidade de tempo e reduzir custos (Wang et al., 2019). Segundo Pace et al. (2015b), a predição genômica permite identificar genótipos para caracteres de difícil mensuração em um grande número de indivíduos. Para eles a GWS não é apenas uma abordagem promissora para fins de melhoramento, mas também para

pesquisa básica. Pace et al. (2015b) aplicaram GWS em caracteres de raiz de uma população de treinamento composta por 384 linhagens consanguíneas de milho e encontraram acurácias médias entre 0,38 e 0,55.

O programa de melhoramento de milho da UFV tem se concentrado na introdução e desenvolvimento de reservatórios de germoplasma, mapeamento de populações de QTL, variedades de polinização aberta e linhagens endogâmicas para condições tropicais, principalmente para ambientes com baixo teor de fósforo e nitrogênio, seca e calor. O programa possui uma coleção de germoplasma, de quase 200 linhagens endogâmicas adaptadas aos ambientes tropicais do Brasil. A maioria dessas linhagens desenvolvidas pelo programa foi caracterizada fenotipicamente quanto à eficiência no uso de nitrogênio e morfologia de raiz em casa de vegetação e condições tropicais de campo (Rodrigues et al., 2017; Torres et al., 2018; 2019), e algumas delas também foram avaliadas em combinações híbridas em condições de estresse, especialmente baixo nitrogênio, calor e alta população de plantas. No entanto, nenhum estudo molecular foi realizado para fazer qualquer diferenciação quantitativa entre a diversidade genética nas linhagens e, portanto, não temos informações sobre a diversidade genética, grupos heteróticos, LD para posterior estudo de associação e predição genômica neste conjunto de linhagens do atual programa de melhoramento da UFV.

2. OBJETIVOS

2.1. Objetivo Geral

Realizar uma caracterização fenotípica e molecular de um conjunto de linhagens de milho tropical e realizar as análises de seleção genômica e de associação genômica ampla para identificar SNPs associados a caracteres de morfologia de raiz e genes candidatos relacionados a esses marcadores.

2.2. Objetivos Específicos

Quantificar a variação genética para caracteres agrônomico e avaliar a diversidade fenotípica e genética em um conjunto de 187 linhagens de milho tropical do banco de germoplasma de milho da UFV;

Estimar as correlações genética entre os caracteres agrônomicos avaliados para as 187 linhagens endogâmicas.

Investigar a estrutura de população e padrões heteróticos e o grau de parentesco entre as linhagens de milho tropical.

Avaliar se este conjunto de linhagens endogâmicas pode compreender um painel de linhagens endogâmicas de milho tropical para análise de mapeamento de associação.

Determinar a extensão e distribuição genômica do desequilíbrio de ligação nesse conjunto de linhagens para posterior mapeamento de associação;

Identificar SNPs associados a caracteres de morfologia de raiz relacionados a eficiência no uso de nitrogênio;

Investigar a localização dos marcadores SNPs associados para possíveis genes candidatos que afetam o desenvolvimento das raízes e fazer a anotação funcional desses SNPs;

Avaliar a acurácia da seleção genômica em diferentes conjuntos de populações de treinamento.

3. REVISÃO DE LITERATURA

3.1. Cultura do Milho

O milho (*Zea mays* L.) é um dos cereais de maior importância econômica, caracterizada pelas diversas formas de utilização, que vão desde a alimentação humana e animal, até a indústria de alta tecnologia (Môro e Fristsche-Neto, 2015).

O germoplasma de milho pode ser classificado principalmente em dois grupos: milho tropical e milho temperado, com base na latitude e nas características ambientais onde é cultivado, principalmente a duração do dia (Paliwal, 2000). No germoplasma temperado, o desenvolvimento de grupos heteróticos começou há cerca de 100 anos, e o padrão heterótico dos Estados Unidos Stiff Stalk x Non-Stalk Stalk tem sido amplamente utilizado na maior parte do mundo temperado (Beckett et al., 2017). Em contraste, no milho tropical, o trabalho em grupo heterótico começou há apenas três décadas pelo Centro Internacional de Melhoramento do Milho e do Trigo (CIMMYT), pois até o início dos anos 1990, o programa de melhoramento do CIMMYT enfatizou o desenvolvimento de germoplasma de milho com uma base genética ampla para criação de variedades de polinização aberta sem considerar o padrão heterótico (Ertiro et al., 2017, Wu et al., 2016; Semagn et al., 2012; Lu et al., 2009). Então, com o aumento da adoção de híbridos de milho nas áreas tropicais, o CIMMYT mudou seu foco para o desenvolvimento de linhagens nos últimos 30 anos, mas a maioria das linhagens de milho tropical desenvolvidas por eles foram inicialmente derivadas de amplos pools de germoplasma, independentemente da origem racial ou padrão heterótico (Wu et al., 2016; Semagn et al., 2012). Além disso, a autofecundação de híbridos comerciais para produzir novas linhagens tem sido uma prática muito comum na maioria dos programas de melhoramento ao redor do mundo, principalmente no Brasil (Guimarães et al., 2018). Esses procedimentos complicaram a atribuição limpa de linhagens endogâmicas de milho tropical em grupos heteróticos, resultando na mistura desses, principalmente em programas de melhoramento de milho brasileiro.

Atualmente, a alocação de linhagens endogâmicas de reprodução tropical, com origens complexas e diversificadas, em grupos heteróticos tem sido baseada em seu desempenho de testcross com vários testadores de elite representativos. No entanto, a confiabilidade da atribuição de linhagens em conjuntos heteróticos com base na capacidade combinada depende de vários fatores: i) o histórico genético das linhagens

endogâmicas, ii) o tipo e número de testadores, iii) número de ambientes para ensaios de produção, e iv) falta de um híbrido de verificação comum em diferentes ensaios de rendimento que podem ser usados para comparar resultados entre empresas, instituições e melhoristas (Leng et al., 2019; Ertiro et al., 2017; Wu et al., 2016). Assim, para complementar a avaliação baseada na capacidade de combinação, vários autores sugeriram o emprego de marcadores moleculares em agrupamento heterótico, principalmente para germoplasma com fundo genético complexo como linhagens endogâmicas de milho tropical (Adu et al., 2019; Leng et al., 2019; Wu et al., 2016; Mundim et al., 2015; Olmos et al., 2013; Lu et al., 2009). Em um estudo de simulação, Viana et al. (2013) concluíram que, com um número adequado de marcadores moleculares, a análise de estrutura populacional é muito eficiente para designar indivíduos em sua população de origem. Por outro lado, a análise de agrupamento é muito útil em um programa de melhoramento de milho, uma vez que um dendrograma preciso com base na distância genética e nas relações entre as linhagens pode ajudar os melhoristas a entender o histórico genético de linhagens derivadas de híbridos comerciais (Breckett et al., 2017), auxiliá-los na redução do número de cruzamentos a serem feitos e avaliados (Ajala et al., 2020; Silva et al., 2020).

3.2. Melhoramento com foco em eficiência no uso de nitrogênio

Extensivas áreas de plantio de milho no mundo são cultivadas em condições de solos tropicais, nos quais predominam a ocorrência de seca e de baixa fertilidade natural, principalmente de nitrogênio (N) e fósforo. O N é o nutriente absorvido em maior quantidade pela cultura do milho. Ele é constituinte de aminoácidos, proteínas, enzimas e ácidos nucléicos e, além disso, está associado a processos fisiológicos de fotossíntese, respiração e absorção iônica (Taiz e Zeiger, 2010). Por participar de processos biológicos essenciais das plantas, a deficiência de N geralmente provoca redução da área foliar, da fotossíntese e da biomassa final e, conseqüentemente, na produtividade de grãos (Boomsma et al., 2009; Ciampitti; 2011). Na cultura do milho, ele é necessário desde os estádios fenológicos iniciais de desenvolvimento da planta até a fase de enchimento de grãos (Basso e Ceretta, 2000). Assim, há uma enorme demanda por alternativas sustentáveis, como a utilização de plantas mais eficientes no uso de N (EUN), ou seja, que apresentem produtividades satisfatórias em baixa disponibilidade desse nutriente, ou ainda, que produzam a mesma quantidade, mas com menor demanda de adubo.

Os cultivares modernos de milho foram desenvolvidos em condições de alta fertilidade do solo, com foco em altas produtividades. O germoplasma de milho selecionado em condições ótimas de cultivo geralmente não é adequado para utilização em condições de estresses abióticos como a deficiência nutricional de nitrogênio (Cecarelli, 1996; Fidelis et al., 2007). Alguns genes importantes que controlam a produtividade de grãos em condições de estresses abióticos são diferentes daqueles em condições ótimas (Atlin e Frey, 2000; Fritsche-Neto et al., 2010; Souza et al., 2008), o que faz com que um material selecionado em condições favoráveis não necessariamente seja o melhor em condições de estresse (Souza et al., 2009). Portanto, para que se obtenha sucesso na seleção para EUN, a mesma deve ser feita submetendo as plantas não apenas às condições ideais, mas também, à condição estressante. Nesse contexto, o uso de métodos de seleção precoce e seleção indireta para EUN é de grande interesse, pois aceleram o processo de seleção, e permite o descarte de imediato dos genótipos inferiores e possibilita concentrar os recursos nos potencialmente superiores (Fritsche-Neto et al., 2012).

Alguns autores têm constatado ampla diversidade genética entre plantas quanto a mecanismos fisiológicos e morfológicos da raiz e de parte aérea, quando elas são submetidas às condições adversas de fertilidade do solo (DoVale et al., 2012; Medici et al., 2004). Dentre os caracteres que tem potencial para serem usados na seleção para EUN, pode-se destacar o maior desenvolvimento do sistema radicular. Esse caractere é altamente responsivo à disponibilidade de nutrientes e sua distribuição no solo (Linkohr et al., 2002). Foi descoberto que a demanda por nutrientes está correlacionada com o crescimento da raiz (Peng et al. 2010) e intimamente associados ao rendimento de grãos (Cai et al., 2012). Segundo Toaldo et al. (2013) a extensão e natureza da variação genética do sistema radicular possuem grande implicação na obtenção e absorção de água e nutrientes para o aumento da produtividade sob estresse hídrico ou baixa fertilidade do solo. Assim, o crescimento do sistema radicular determina a habilidade da planta em explorar recursos que estão mal distribuídos. Com isso, o comprimento radicular pode ser um indicador do potencial de absorção de água e nutrientes, sendo proporcionalmente maior o volume ocupado e explorado do solo quanto maior for o comprimento deste (Zonta et al., 2006). As características da raiz são quantitativas e controladas por vários genes e muito difíceis de medir (Hochholdinger et al., 2018).

A caracterização detalhada dos genótipos com base em dados moleculares pode contribuir estrategicamente para otimizar a utilização do germoplasma disponível em programas de melhoramento de milho visando alta EUN (Romay et al., 2013). Assim, para o aperfeiçoamento da relação custo-benefício, a otimização do tempo e a realização de uma seleção com acurácia, deve-se lançar mão de metodologias precisas que amostram o genoma e o correlacionem com as características fenotípicas. Ferramentas como a associação genômica ampla (GWAS) e a seleção genômica ampla (GWS) podem ajudar os melhoristas a tomar decisões sobre quais genótipos devem avançar em seus programas de melhoramento. Pace et al. (2015a) e Sanchez et al. (2018) encontraram 268 e 39 SNPs, respectivamente, associados a caracteres de raízes de plântulas de milho localizadas ou ligadas a modelos de genes relacionados ao desenvolvimento da raiz.

3.3. Diversidade Genética

A diversidade genética pode ser definida como a amplitude de variação genética existente em uma determinada espécie. Assim, torna-se necessário o conhecimento da variabilidade e da sua interação com o ambiente para a obtenção do material genético a ser utilizado em programas de melhoramento. A quantificação da variação genética presente nos bancos de germoplasma é indispensável para o sucesso de qualquer programa de melhoramento. A caracterização fenotípica e a avaliação da diversidade genética do germoplasma é o primeiro passo para o conhecimento da magnitude dos parâmetros genéticos, essenciais em um programa de melhoramento.

O conhecimento da diversidade genética entre as linhagens de milho está associado à busca do vigor híbrido no F1 proveniente do cruzamento de parentais que exibem alta divergência entre si (Paterniani et al., 2008). Os grupos heteróticos são grupo de plantas que quando cruzadas entre si os híbridos não apresentam heterose, mas quando cruzadas com plantas de outro grupo há heterose significativa (Melchinger e Gumber, 1998). A avaliação da diversidade é essencial para definir estratégias de melhoramento genético e identificar pais superiores para o desenvolvimento de novas cultivares. Para o sucesso de qualquer programa de melhoramento a avaliação precisa da diversidade genética pode auxiliar o melhorista na alocação de linhagens de milho em grupos heteróticos, seleção de testadores eficientes para testes de linhagens em combinações híbridas, inferências sobre a introgressão de genes desejáveis de diversas fontes de germoplasma na base genética disponível, identificação de diversas combinações

parentais para criar progênies segregantes com máxima variabilidade genética para posterior seleção e a estimativa da perda de diversidade genética durante a conservação ou seleção (Adu et al., 2019; Semang et al., 2012).

Existem duas estratégias para inferir sobre a divergência genética: de forma quantitativa e de forma preditiva. Entre as de natureza quantitativa, citam-se as análises dialélicas (Griffing, 1956), baseiam-se nos conceitos de capacidade geral e específica de combinação. As quais são necessários cruzamentos entre genitores e sua posterior avaliação. A avaliação das linhagens quanto à capacidade de combinação é a etapa mais trabalhosa e onerosa dos programas de melhoramento de milho com foco no desenvolvimento de híbridos. As de natureza preditiva têm por base as diferenças morfológicas, fisiológicas ou moleculares quantificadas com emprego de alguma medida de dissimilaridade que possa expressar o grau de diversidade genética entre os genitores (Cruz et al., 2012). A diversidade genética pode ser estimada por meio de informações fenotípicas, genealógicas e genotípicas.

A diversidade entre genótipos, com base em dados fenotípicos, é quantificada a partir de medidas de dissimilaridade e as mais utilizadas são a distância euclidiana e de Mahalanobis. Após análise da dissimilaridade, faz-se necessário a alocação dos genótipos em grupos. A forma mais conhecida e utilizada para representar os agrupamentos com base na distância entre pares de genótipos é através da média das distâncias entre todos os pares de genótipos analisados, denominado de método UPGMA (*Unweighted Pair Group Method using Arithmetic averages*) (Bertan et al., 2006). A determinação do número de grupos pode ser definida a priori, por meio do conhecimento que se tenha dos dados, com base na inspeção visual das ramificações do dendrograma, ou recorrer a algum procedimento estatístico, como o método de Mojena (1977).

A avaliação da diversidade genética a partir de marcadores moleculares pode ser estimada por diferentes tipos de marcadores moleculares (Nyombayire et al., 2016). Atualmente, os avanços na tecnologia molecular têm mostrado uma mudança em direção a utilização dos marcadores SNPs. Estes marcadores são bialélicos e com boa cobertura no genoma do milho (Van Inghelandt et al., 2010). Eles apresentam especificidade de loco, baixo custo, codominância, abundância genômica elevada, potencial para análise de alto rendimento e menores taxas de erro de genotipagem (Chagné et al., 2007). Assim, os SNPs vêm sendo empregado com sucesso em estudos de diversidade genética e estrutura populacional, construção de mapas genéticos (Gore et al., 2009), detecção de genes/QTLs

e análise de arquitetura genética de caracteres complexos (Cook et al., 2012; Peiffer et al., 2014). Dada a relevância de unir análises fenotípicas e fisiológicas a moleculares, diferentes abordagens têm sido utilizadas. De forma geral, elas implicam em genotipagem com alta densidade de marcadores no genoma e o uso de métodos baseados em desequilíbrio de ligação (LD). Semagn et al. (2012) relataram que os marcadores SNP são uma ferramenta poderosa em estudos de diversidade genética. Estes autores genotiparam 450 linhagens de milho com emprego de 1.065 marcadores SNPs com objetivo de avaliar a utilidade dos SNPs e identificar um subconjunto de marcadores SNP altamente informativos para a genotipagem de rotina e de baixo custo do germoplasma do CIMMYT.

Dos estudos mais antigos, que se basearam apenas em caracteres morfológicos até os mais recentes, que são realizados por meio de marcadores moleculares, todos têm como objetivo o entendimento da diversidade genética do milho. Os primeiros estudos usaram a morfologia como método para agrupar as várias raças de milho conhecidas, e contribuíram para o início da organização da variabilidade dessa espécie. Goodman e Bird (1977) realizaram um importante trabalho separando mais de 200 raças latino-americanas de milho em 14 grupos, partindo de 20 caracteres morfológicos de espiga e das localidades originais das plantas. Eles também analisaram os problemas que a classificação por meio da morfologia traz e mostraram como isso pode inserir erros nos agrupamentos. Os principais problemas, destacados por eles, em estudos morfológicos são: a possibilidade de uma mesma planta ser classificada com nomes diferentes, a necessidade de seleção de características que não receberam grande influência do ambiente, e a posição intermediária de algumas raças com relação a dois grupos próximos. Os autores concluíram que a principal fonte de variabilidade do milho está concentrada nos grupos “dent” e “flint” e que o grande número de introduções, vindas na América do Norte, Europa e México, torna difícil a caracterização do germoplasma brasileiro.

O primeiro trabalho que avaliou a diversidade genética de linhagens tropicais de milho foi o de Warburton et al. (2002). Estes autores caracterizaram 57 linhagens com marcadores SSR e encontraram alto nível de diversidade genética, o que indica que foi incorporado diversidade genética considerável no germoplasma de milho. Eles também observaram que apenas linhagens intimamente relacionadas por pedigree foram agrupadas e que a população de origem e o grupo heterótico não foram associados aos agrupamentos formados com base nos marcadores SSR, resultado consistente com o alto

nível de diversidade nas populações de origem das linhagens. Os autores concluíram que embora isto dificulte a atribuição das linhagens endogâmicas a grupos heteróticos existentes, os marcadores podem ser úteis para refinar os grupos heteróticos em grupos adicionais e mais uniformes.

Van Inghelandt et al. (2010) examinaram a diversidade genética e a estrutura de população em 1.537 linhagem de milho genotipadas com 359 SSR e 8.244 marcadores SNP. Os autores concluíram que a organização do germoplasma em quatro grupos heteróticos foi consistente com resultados de análises de estrutura populacional baseadas em marcadores SSR e SNP. Yang et al. (2011) avaliaram 527 linhagens endogâmicas de milho com origens tropicais, subtropicais e temperadas, representando a diversidade global do milho. As linhagens foram genotipadas com 1.536 marcadores SNP, e três subpopulações foram consistentemente identificadas por diferentes métodos estatísticos. Semagn et al. (2012) relataram em seu trabalho que os marcadores SNP são uma ferramenta poderosa em estudos de diversidade genética. Wu et al. (2016) utilizaram um painel de 538 linhagens endogâmicas e 362.008 SNPs para compreender a diversidade genética, a estrutura populacional e os padrões heteróticos dessa coleção de milho. Os resultados indicaram que a ampla distância genética reflete a singularidade da maioria das linhagens, a estrutura populacional e a divergência genética entre o subgrupo temperado e os subgrupos tropicais eram claras, a diversidade genética dos três subgrupos tropicais foi semelhante e maior que a do subgrupo temperado e os padrões heteróticos em cada grupo de adaptação ambiental estimado usando os SNPs foram apenas parcialmente consistentes com os padrões estimados com base nos testes de capacidade de combinação e na informação de pedigree. Os autores concluíram que a combinação de informações heteróticas com base em testes de habilidade de combinação e as relações genéticas inferidas de análises de marcadores moleculares podem ser a melhor estratégia para definir grupos heteróticos para a futura melhora do milho tropical. E que as informações resultantes desta pesquisa ajudarão os melhoristas a entender melhor como utilizar as linhagens para selecionar parentais, substituir testadores, designar grupos heteróticos e criar um conjunto básico de germoplasma de reprodução. Nyombayire et al. (2016) quantificaram a diversidade genética predominante entre 71 linhagens endogâmicas de diferentes fontes que foram genotipadas com 92 marcadores SNP. Os autores concluíram que as informações de agrupamento genético adquiridas são adequadas também para

outros programas de melhoramento de milho tropical. E que os resultados podem orientar para padrões e grupos heteróticos adequados.

Vários estudos de caracterização molecular do milho foram realizados utilizando diferentes germoplasmas e vários tipos de marcadores moleculares (Lu et al., 2009, Mir et al., 2013, Warburton et al., 2005, 2008, Wen et al., 2011, Wu, et al., 2014, Xia et al., 2004, Yan et al., 2009). Esses estudos revelaram a estrutura populacional entre as regiões temperada e tropical/subtropical, a maior diversidade genética do milho tropical e os padrões heteróticos no germoplasma de milho temperado. No entanto, falta informação clara sobre o germoplasma de milho tropical para os padrões de agrupamento baseados no fenótipo, na adaptação ambiental e no grupo heterótico. As análises de marcadores moleculares fornecem uma abordagem alternativa importante para caracterizar a diversidade genética, a estrutura da população e as relações genéticas entre os materiais de reprodução de elite dentro de uma determinada coleção de germoplasma de milho. No entanto, a maioria dos estudos anteriores se concentrou principalmente em milho temperado que não incluía um número suficiente de linhagens para representar plenamente o milho tropical melhorado. Assim, torna-se de extrema importância a caracterização molecular e fenotípica do germoplasma de milho tropical.

3.4. Estrutura Populacional

A estrutura de uma população pode ser entendida a partir das frequências dos alelos que compõem os diferentes genótipos dos diferentes indivíduos integrantes da população. De acordo com Brown (1978), a estrutura genética de uma população é definida como a forma pela qual a variabilidade genética é distribuída entre e dentro dos níveis hierárquicos de subdivisão de uma espécie. Como os indivíduos de uma espécie raramente se distribuem de maneira homogênea no espaço, eles quase sempre estão associados, agregados e, assim, a estrutura genética de populações refere-se à heterogeneidade na distribuição dos genótipos e do grau de endogamia dentro de populações e entre estas (Robinson, 1998).

A estrutura populacional de qualquer espécie domesticada é influenciada pela história de domesticação de seus ancestrais, bem como pelo sistema de melhoramento utilizado (Xie et al., 2008). No melhoramento genético, a informação da estrutura populacional permite ao melhorista conhecer a estrutura genética de uma população,

detectar a estrutura familiar, estimar níveis de migração e dispersão, além de auxiliar na manutenção do banco de germoplasma e no diagnóstico evolutivo (Cruz et al., 2011).

A complexa história do melhoramento de muitas culturas importantes e o fluxo gênico limitado na maioria das plantas silvestres criaram estratificação complexa dentro do germoplasma, o que complica os estudos de associação (Sharbel et al., 2000). Historicamente, as análises de associação têm sido amplamente utilizadas na genética médica, mas praticamente não têm aplicação na genética de plantas. Um obstáculo à sua aplicação são as populações estruturadas frequentemente encontradas em plantas, o que pode levar a associações espúrias não-funcionais (Thornberry et al., 2001). Segundo Gupta et al. (2005) e Nordborg e Weigel (2008) é necessário considerar a estrutura populacional para inferências sobre o desequilíbrio de ligação (LD), pois a estrutura da população interfere grandemente na sua mensuração. A existência de associações falso-positivas entre locos não ligados é um dos problemas que ocorrem quando não é realizado o controle efetivo da influência deste fator sobre o LD (Rafalski, 2002), exatamente como pode acontecer nas análises tradicionais de desenvolvimento de modelos de GWS. A mistura de indivíduos oriundos de diferentes fragmentos homogêneos (subpopulações) de uma população pode levar a associações preferenciais. A presença de LD altamente significativo entre polimorfismos em diferentes cromossomos pode produzir associações entre um marcador e um fenótipo, embora o marcador não esteja fisicamente ligado aos locos responsáveis pela variação fenotípica (Pritchard e Rosenberg, 1999). Para o mapeamento associativo o controle eficiente da estrutura de população é imprescindível, de modo que apenas associações entre locos ligados sejam considerados para a seleção assistida por marcadores.

De acordo com Viana et al. (2013), a análise da estrutura populacional foi proposta como procedimento a ser utilizado antes do mapeamento de associação com o objetivo de descartar associações significativas entre marcadores e genes, quando esses são realmente independentes (Pritchard e Rosenberg, 1999; Pritchard et al., 2000). Pritchard et al. (2000) publicaram uma das ferramentas mais difundidas e importantes para abordar a estrutura populacional: o método de agrupamento baseado em modelo conhecido como STRUCTURE. Eles usaram uma abordagem de agrupamento Bayesiana para resolver o problema de associações espúrias. Esse software detecta a estrutura populacional, atribui indivíduos a populações e identifica migrantes e indivíduos que possuem genoma de duas ou mais populações. Além disso, permite a inclusão de informações a priori, o que tende

a tornar as análises mais precisas (Viana et al., 2013). O método pode ser aplicado a vários tipos de marcadores moleculares, mas assume que os locos dos marcadores estejam desvinculados e em equilíbrio de ligação entre si dentro das populações. Também assume o equilíbrio de Hardy-Weinberg dentro das populações (Pritchard et al., 2000). Na primeira aplicação desse método, modificado para caracteres quantitativos, Thornberry et al. (2001) avaliaram o florescimento em milho. Eles usaram abordagens de associação para avaliar polimorfismos da sequência *dwarf8* em 92 linhagens endogâmicas de milho. A estrutura populacional foi estimada usando uma análise Bayesiana e 141 marcadores SSR. Os resultados indicaram que um conjunto de polimorfismos no gene do milho *dwarf8* se associa à variação no florescimento das plantas, que incluem uma deleção que pode alterar um domínio chave na região codificadora. A distribuição de polimorfismos não-sinônimos sugere que *dwarf8* tenha sido alvo de seleção. A incidência de falsos positivos criados pela estrutura populacional foi reduzida em até 80% como resultado do método de Pritchard et al. (2000). Além desse, outros métodos baseados em marcador molecular também têm sido desenvolvidos para o controle da estrutura de população, com o objetivo de estudar o mapeamento associativo. Por exemplo, o método baseado em modelos mistos, permite estimar a estrutura populacional e a matriz de parentesco (Yu et al., 2006). Segundo Viana et al. (2013) a maioria das análises de estrutura populacional, no melhoramento de plantas, envolveu linhagens endogâmicas e revelaram subpopulações de acordo com a adaptação geográfica e climática, informações de pedigree, origem do germoplasma e grupos heteróticos. E eles afirmam que em alguns estudos, a análise descartou com sucesso as associações espúrias. Flint-Garcia et al. (2005) descreveram a estrutura populacional de 302 linhagens endogâmicas de milho e investigaram a relação entre a estrutura populacional e várias medidas de valor fenotípico e genético. Eles concluíram que a inclusão da estrutura populacional em modelos de associação é fundamental para análises significativas e no planejamento de projetos futuros para explorar a diversidade presente no milho. Com o objetivo de identificar grupos heteróticos em germoplasma de milho, a partir da análise da estrutura da população, Yang et al. (2011) avaliaram 527 linhas endogâmicas de origem tropical, subtropical e temperada, representando a diversidade global do milho. As linhagens endogâmicas foram genotipadas com 1.536 SNPs, e três subpopulações foram consistentemente identificadas por diferentes métodos estatísticos. Van Inghelandt et al. (2010) genotiparam 1.537 linhagens com 359 marcadores SSR e 8.244 SNPs e

concluíram que a organização do germoplasma em quatro grupos heteróticos foi consistente com resultados de análises de estrutura populacional baseadas em marcadores moleculares. Semagn et al. (2012) investigaram a estrutura populacional e os padrões de relacionamento do germoplasma em 450 linhagens de milho desenvolvidas pelo CIMMYT usando 1.065 marcadores SNP. E concluíram que os resultados serão úteis na seleção de melhores combinações parentais para novos cruzamentos, mapeando o desenvolvimento populacional e o melhoramento assistido por marcadores. Wu et al. (2016) avaliaram a diversidade genética e a estrutura populacional de um painel com 538 linhagens de milho tropical do CIMMYT e seis linhagens endogâmicas temperadas. Os resultados indicaram que a estrutura populacional e a divergência genética entre o subgrupo temperado e os subgrupos tropicais foram claras. A diversidade genética dos três subgrupos tropicais foi semelhante e superior à do subgrupo temperado.

3.5. Desequilíbrio de Ligação

Desequilíbrio de ligação (linkage disequilibrium, LD), também é conhecido como desequilíbrio da fase gamética, desequilíbrio gamético e associação alélica. O termo desequilíbrio de ligação foi definido pela primeira vez por Lewontin e Kojima (1960) e é uma das definições mais utilizadas. Esses autores observaram que a ligação entre locos era fator determinante para uma população atingir o equilíbrio (Hedrick, 2011). Eles a definiram com base na análise das frequências de pares de alelos durante a formação de gametas, por comparação entre as frequências observadas com as esperadas sob independência, obtidas com base nas frequências de cada um dos alelos que compõem esses pares. Quando as frequências gaméticas observadas em determinada população são idênticas àquelas esperadas, obtidas pelo produto das frequências alélicas envolvidas, diz-se que os locos estão em equilíbrio de ligação. Quando isto não ocorre têm-se a falta de equilíbrio de ligação, ou seja, o LD. Assim, quando há desvio nas frequências observadas dos haplótipos quando comparadas às esperadas tem-se o desequilíbrio de ligação (Lewontin e Kojima, 1960; Lewontin, 1964; Weir, 1979). Flint-Garcia et al. (2003) definiram LD como a associação não aleatória entre alelos de diferentes locos em uma população. LD e a ligação física são termos frequentemente confundidos, pois são, muitas vezes, encarados como sinônimos. Isso normalmente ocorre porque locos intimamente ligados estão em elevado LD (Flint-Garcia et al., 2003). O fato de dois locos quaisquer estarem em LD não significa que estão fisicamente ligados. Do mesmo modo, o fato de

dois locos estarem fisicamente ligados em determinado cromossomo ou haplótipo não significa que estão em LD (Pritchard e Przeworski, 2001; Slatkin, 2008; Hedrick, 2011).

Para o melhoramento genético, de modo geral, o desequilíbrio de ligação é fator determinante para a seleção com base em marcadores genéticos (Hartl e Clark, 2010; Resende et al., 2010). Ele auxilia no desenvolvimento de estratégias, com vistas à redução de custos, o que contribui para o aumento da eficiência dos programas de melhoramento genético (Rosa, 2011). Vários trabalhos que visam maior eficiência dos métodos de seleção em programas de melhoramento são embasados no LD (Carneiro e Vieira, 2002; Hedrick, 2011).

Por interferir diretamente na dinâmica genética das populações o LD deve ser entendido como um fenômeno genético. Seu efeito é visualizado na segregação não independente dos alelos dos diferentes locos (Hartl e Clark, 2010), o que resulta em uma correlação entre eles durante a formação dos haplótipos (Flint-Garcia et al., 2003). Assim, sua existência acaba interferindo na formação dos genótipos dos indivíduos nas populações. É importante ressaltar que qualquer fator que altere as frequências alélicas pode interferir na dinâmica do LD. Consequentemente, todos os fatores evolutivos considerados nos estudos de genética de população, bem como todas as ferramentas integrantes do processo de melhoramento genético que modifiquem estas frequências, atuam sobre esta dinâmica.

Antigamente, quando a disponibilidade de dados moleculares ainda era limitada, o LD detectado em uma população era normalmente atribuído à seleção simultânea de locos. Nos dias de hoje, em especial para locos intimamente ligados, é reconhecido que mutação e deriva genética são fatores importantes no surgimento de LD, assim como recombinação na sua queda ao longo das gerações (Hedrick, 2011). No entanto, esses fatores não são os únicos. Tamanho populacional, endogamia, sistema reprodutivo, inversão e conversão de genes também podem afetar o LD de uma população (Pritchard e Przeworski, 2001; Gallais, 2003; Flint-Garcia et al., 2003; Hartl e Clark, 2007; Slatkin, 2008; Hedrick, 2011). Outro fator que afeta substancialmente o LD é a estrutura populacional. Possivelmente, esta é a principal causa da ocorrência de falsos positivos em estudos de mapeamento associativo (Abdurakhmonov e Abdugarimov, 2008). A mistura de indivíduos oriundos de diferentes fragmentos homogêneos (subpopulações) de uma população pode culminar em associações preferenciais.

Existem diversas maneiras de aferir a magnitude do desequilíbrio de ligação. O coeficiente de correlação r é uma medida relativa do LD. Esta medida proposta por Hill e Robertson (1968) é comumente expressa em termos quadráticos r^2 ou Δ^2 . Ele mede o grau de associação entre locos (covariância) de acordo com a variação dos seus alelos. Essa medida depende do conceito básico do LD, o qual é baseado nas frequências dos haplótipos. Obtido por: $r^2_{XY} = \text{Cov}(X,Y)/V(X).V(Y)$, em que: r^2_{XY} (coeficiente de determinação, ou quadrado do coeficiente de correlação) é a medida relativa do LD entre dois locos bialélicos quaisquer X e Y, $\text{Cov}(X, Y)$ é a covariância entre esses locos, $V(X)$ e $V(Y)$ são suas respectivas variâncias. O r^2 pode variar entre 0 e 1 ($0 \leq r^2 \leq 1$), $r^2=1$, indica associação preferencial máxima entre alelos de pares de locos, pode-se dizer que dois dos quatro haplótipos estão ausentes na amostra, o que caracteriza uma situação de perfeito LD (Ardlie et al., 2002).

Vários estudos têm sido realizados para investigar o LD em milho com emprego de diferentes tipos de populações e de marcadores. Os padrões de LD variam substancialmente com a população escolhida (Flint-Garcia et al., 2003). Os primeiros estudos de associação em milho foram publicados em 2001 (Tenaillon et al., 2001; Remington et al., 2001). Tenaillon et al. (2001) avaliaram nove linhagens endogâmicas dos EUA e 16 raças exóticas para inferir sobre o impacto da seleção, recombinação e LD na modelagem da diversidade do milho. Eles mediram a diversidade de sequências em 21 locos distribuídos ao longo do cromossomo 1 do milho. Os resultados mostraram que a diversidade de sequências foi heterogênea entre os locos, mas não houve padrão de diversidade ao longo do mapa genético do cromossomo 1. A recombinação foi suficiente para quebrar o desequilíbrio de ligação entre os SNPs. Os autores mostraram que LD tipicamente quebra dentro de algumas centenas de pares de bases e que o desequilíbrio de ligação declina dentro de 100-200 pb em média, sugerindo que os levantamentos genômicos para análises de associação requerem SNPs a cada 100-200 pb. Segundo Remington et al. (2001), estudos de associação com base em LD podem fornecer alta resolução para a identificação de genes que podem contribuir para a variação fenotípica. Em um estudo realizado com 102 linhagens de milho eles observaram que o LD caiu rapidamente com a distância ($r^2 < 0,1$ dentro de 1.500 pb), porém, as taxas de declínio foram muito variáveis entre os genes. Portanto, estudos de associação com base em desequilíbrio de ligação podem permitir a identificação dos genes representados por QTLs. Em geral, o LD ocorre de forma aleatória no genoma e depende da espécie e da

população em estudo (Oraguzie et al., 2007). Stich et al. (2005) estudaram o desequilíbrio de ligação (LD) em 147 linhagens elites de milho da Europa (dent e flint) e dos Estados Unidos (stiff stalk e non stiff stalk). Foram utilizados 100 marcadores SSR. Os autores encontraram LD significativo para 49% dos pares de marcas nas 80 linhagens flint e para 56% dos pares de marcas nas 57 linhagens dent. O número total de blocos de desequilíbrio foi maior no grupo flint, em comparação ao grupo dent. Contudo, o grupo dent apresentou blocos de LD de maior tamanho. Os maiores blocos de LD foram encontrados nos cromossomos 2 (105 cM) e 8 (103 cM), em relação aos grupos flint e dent, respectivamente. O tamanho médio dos grupos de LD foi de 33 cM. Wang et al. (2008) utilizaram 145 marcadores SSR para avaliar a diversidade genética, estrutura populacional e o LD de um conjunto de 95 linhagens de milho que representavam as linhagens de milho chinês. Os autores encontram LD significativo entre 63,89% dos pares de SSR e concluíram que entre os fatores que influenciam o LD, a ligação gênica foi a principal causa de LD dos loci SSR. Stich et al. (2006), investigaram a extensão e a distribuição do LD entre pares de marcadores. Eles utilizaram 72 linhagens europeias genotipadas com 452 marcadores AFLP e 93 SSR. O LD foi significativo ($P < 0,05$) para 15% dos pares de marcadores AFLP e para 49% dos pares de SSR em cada um dos dois grupos de germoplasma utilizados, flint e dent. Em ambos os grupos, a proporção de pares de locos ligados e não ligados em LD foi maior para AFLPs que para SSRs. A observação de LD devido à ligação para ambos os tipos de marcadores, sugeriu que o uso de mapeamento associativo é possível utilizando tanto marcadores AFLPs quanto SSRs. Lu et al. (2011) analisaram um total de 287 linhagens tropicais e 160 linhagens temperadas que foram genotipadas com 1.943 marcadores SNPs e comparadas quanto a queda de LD. Eles encontraram significativa variação na queda da distância de LD (2-100kb) em todo o genoma. Em média a queda da distância de LD (10-100kb) no germoplasma temperado foi de duas a dez vezes maior do que no germoplasma tropical (5-10kb). Os autores concluíram que o milho tropical possui um rápido decaimento do LD que fornece maiores possibilidades para a seleção. Morosini et al. (2017) estimaram o padrão de LD entre os cromossomos e o genoma inteiro pelo coeficiente de correlação ao quadrado (r^2). Esses autores avaliaram 64 linhagens de milho tropical para identificar marcadores SNP associados a caracteres de eficiência no uso de nitrogênio (EUN) e para estudos de GWAS. Os autores observaram um rápido declínio no LD com o aumento da distância física (Mb) em todos os cromossomos. E que o comprimento médio do decaimento de

LD em todos eles foi de 80–100 kb ($r^2=0,13$). Eles afirmaram que a extensão do LD limitada exige um número substancialmente maior de marcadores, de modo que o poder da análise para detectar associações entre marcadores e locais causais possa ser assegurado. Eles concluíram que a pequena média geral de LD contrastada pelo alto número de SNPs utilizados juntamente com as estimativas de herdabilidade dos caracteres avaliados, indicou uma boa resolução para a identificação de SNPs relacionados à EUN.

O LD é uma base para a seleção genômica, imputação de marcador genômico, seleção assistida por marcador, mapeamento de locos de caracteres quantitativos (QTL), testes de parentesco e estudos de associação do genoma completo. Assim, o estudo da extensão do LD é muito importante, e possibilita determinar a resolução de mapeamento e a densidade de marcadores necessários na identificação de QTL's (Nordborg et al., 2002; Wei et al., 2006; Zhu et al., 2008). Se o LD permanecer em curtas distâncias no genoma, uma alta resolução de mapeamento será esperada, mas uma elevada quantidade de marcadores será necessária. Em contrapartida, se o LD se estender a maiores distâncias no genoma, a resolução de mapeamento tenderá a ser baixa, mas uma quantidade menor de marcadores será requerida (Yu e Buckler, 2006; Oraguzie et al., 2007; Zhu et al., 2008).

3.6. Estudos de Associação genômica ampla (GWAS)

Com o rápido desenvolvimento das tecnologias de sequenciamento e métodos computacionais, emerge a GWAS (Genome Wide Association Studies), utilizada como uma poderosa ferramenta para detectar a variação natural relacionada aos caracteres complexos em diversas culturas (Rafalski, 2010). Em populações genéticas, LD é a associação não aleatória de alelos em loci diferentes. Sua extensão determina não apenas a resolução de mapeamento que pode ser alcançada, mas também o número de marcadores necessários para uma cobertura adequada do genoma em um estudo de associação genômica (Whitt e Buckler, 2003).

A abordagem GWAS permite a identificação de regiões cromossômicas associadas a um determinado fenótipo sem conhecimento da mutação causal da variação genética, de forma que os SNPs identificados podem não ter efeito direto sobre a característica em estudo (Korte e Farlow, 2013). Como não são feitas suposições sobre a localização genômica das variantes causais, essa abordagem estima a associação entre locos e o caractere fenotípico em nível populacional por meio de testes de hipóteses,

visando detectar efeitos com significância estatística (Resende et al., 2013). Nesse contexto, os estudos GWAS para identificação de QTLs podem auxiliar o desenvolvimento mais rápido de cultivares com maior EUN, pois permitem o entendimento do controle genético e fisiológico dessa característica. Isso porque, após a identificação dos SNPs em LD com caracteres relacionados a EUN, é possível o monitoramento de regiões cromossômicas específicas, genes candidatos, que afetam a expressão do caráter, fornecendo informações sobre o número e a localização dos locos que o controlam (Fritsche-Neto e Borém, 2011). Ainda, por meio do uso de bancos de dados, é possível realizar a caracterização dos marcadores significativos resultantes através de anotação funcional. A GWAS vem sendo usada para encontrar genes associados a caracteres para EUN como uma ferramenta poderosa que conecta informações de genótipo-fenótipo (Zhang et al., 2015; Laperche et al., 2017). Morosini et al. (2017) avaliaram 74 linhagens de milho tropicais em campo contrastando para EUN. A análise de GWAS revelou regiões genômicas associadas com EUN em condições de campo.

3.7. Seleção genômica ampla (GWS)

A seleção genômica ampla (GWS) ou seleção genômica (GS) foi proposta por Meuwissen et al. (2001) como uma forma de aumentar a eficiência e acelerar o melhoramento genético. Com os avanços nas técnicas de genética molecular, o GWS se tornou uma ferramenta promissora para encurtar o intervalo de geração, reduzir o custo por ciclo e economizar custos de mão de obra (Wang et al. 2018). A Seleção Genômica Ampla (Genome Wide Selection - GWS) consiste na predição simultânea (sem o uso de testes de significância para marcas individuais) dos efeitos genéticos de grande número de marcadores genéticos dispersos em todo o genoma de um organismo, de forma a capturar os efeitos de todos os locos, de pequenos e grandes efeitos, e explicar grande parte da variação genética de um caráter quantitativo (Meuwissen et al., 2001). Para que isso seja possível é necessário que haja desequilíbrio de ligação, em nível populacional, entre alelos dos marcadores e alelos dos genes que controlam o caráter.

A predição dos efeitos genéticos é realizada com base em dados genotípicos e fenotípicos de indivíduos pertencentes a uma amostra da população de seleção. Esses efeitos genéticos dos marcadores sobre fenótipos de caracteres quantitativos são somados e usados na predição de valores genéticos de indivíduos apenas genotipados, candidatos

à seleção nos programas de melhoramento. A predição e a seleção podem ser realizadas em fases juvenis, pois atua precocemente sobre os genes expressos na idade adulta (seleção precoce direta), acelerando assim o processo de melhoramento genético. Ao contrário da seleção precoce tradicional indireta, via avaliação fenotípica, que atua sobre genes ativados na idade precoce, esperando que esses informem parcialmente sobre genes expressos na idade adulta (Resende et al., 2012). A GWS é superior e apresenta maior acurácia seletiva que a seleção fenotípica, por usar a matriz de parentesco real de cada caráter, propiciar repetição experimental sem uso de testes clonais e de progênies, por usar maior número de informações, combinando informação fenotípica, genotípica e genealógica para corrigir e desregressar os dados e por usar a seleção precoce direta, que aumenta o ganho genético por unidade de tempo. Após a obtenção dos marcadores, seus efeitos são estimados, baseados em dados fenotípicos de uma população conhecida como de estimação. Uma vez estimados os efeitos, eles são testados em uma população de validação e, então, selecionados os marcadores que explicam grande parte da variância genética do caráter em estudo para que sua informação seja efetivamente incorporada à etapa de seleção do programa de melhoramento (Resende et al., 2010).

Existem numerosas abordagens de predição genômica, procedimentos de modelo bayesiano e misto. O modelo RR-BLUP (Random Regression Best Linear Unbiased Predictor) (Fernando et al., 2007), tem se mostrado um modelo confiável no fornecimento de boa acurácia de predição para caracteres quantitativos. O modelo misto usado pelo RR-BLUP reduz todos os efeitos do marcador igualmente com a mesma variância para um caractere. Isso explica o bom desempenho do RR-BLUP para predição de caracteres complexos (Clark et al., 2011). Computacionalmente, RR-BLUP é menos exigente o que o torna um dos métodos mais acessíveis e sugeridos para a predição genômica (Heslot et al., 2012; Pace et al., 2015). Pace et al. (2015b) aplicaram GWS em características de raiz em um subconjunto do painel de linhagens de milho e encontraram precisões médias entre 0,38 e 0,55. Cantelmo et al. (2017) avaliaram 838 e 797 híbridos simples nas safras de inverno e verão, respectivamente, em diferentes locais para estimar a precisão preditiva usando diferentes populações de treinamento e validação. Eles encontraram intervalos de correlações de 0,82 a 0,89 na safra de inverno, 0,56 a 0,76 na safra de verão, e entre diferentes safras e locais foram de 0,53.

4. REFERÊNCIAS BIBLIOGRÁFICAS

Abdurakhmonov, I.Y.; Abdulkarimov, A. Application of association mapping to understanding the genetic diversity of plant germoplasm resources. *International Journal of plant Genomics*, Cairo, p. 1-18, 2008. Adu, G.B.; Badu-Apraku, B.; Akromah, R.; Garcia-Oliveira, A. L.; Awuku, F. J.; Gedil, M. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *Plos One*. p.1-12, 2019. Allard, R.W.; Bradshaw, A.D. Implications of genotype-environmental interactions in applied plant breeding. *Crop Science*, v.4, p. 503-508, 1964. Arunachalam, V. Genetic distance in plant breeding. *The Indian journal of genetics and plant breeding*, v.41, p.226-236, 1981. Ardlie, K.; Kruglyak, L.; Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*, New York, v.112, p.876-884, 2002. Atlin, G.N.; Frey, KJ. Breeding crop varieties for low-input agriculture traits in maize. *Crop Science*, v.47, p.1082-1090, 2007. Barber, S., Mackay, A. Root growth and phosphorus and potassium uptake by two corn genotypes in the field. *Fertilizer Research* 10, 217–230 (1986). Balding, D. J.; Nichols, R. A. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, v.96, p. 3–12, 1995. Basso, CJ; Ceretta, CA. Nitrogen fertilization management for corn grown after winter cover crops under no-tillage. *Rev. Bras. Ciênc. Solo* vol.24 no.4, 200022 Beadle, G.W. Teosinte and the origin of maize. In: Walden, D.B (ed.) *Maize breeding and genetics*. New York, John Wiley and Sons, 8:113-141, 1978. Beckett TJ, Morales AJ, Koehler KL, Rocheford TR. Genetic relatedness of previously PlantVariety-Protected commercial maize inbreds. *PLoS ONE*, 2017, 12(12):0189277.

Bertan, I.; Carvalho, F. I. F.; Oliveira, A.C.; Vieira, E.A.; Hartwig, I.; Silva, J.A.G.; Shimidt, D.A.M.; Valério, I.P.; Busato, C.C.; Ribeiro, G. Comparação de métodos de agrupamento na representação da distância genética morfológica entre genótipos de trigo. *Revista Brasileira de Agrociência*, Pelotas, v.12, n.3, p.279-286, 2006. Bered, F.; Barbosa Neto, J.F.; Carvalho, F.I.F. Marcadores moleculares e sua utilização no melhoramento genético de plantas. *Ciência Rural*, Santa Maria, v.27, n.3, p.513-520, 1997. Bortoletto, A.M., Alcarde, A.R. Dominante nos EUA, etanol de milho é opção, no Brasil, para safra excedente. *Rev. Visão Agrícola* p.135–137, 2015. Boomsma, C.R. Maize per plant and canopy-level morpho-physiological responses simultaneous stresses of intense crowding and low nitrogen availability. *Agronomy Journal*, v.101,p.1426-1452, 2009. Bradbury, P. J.; Zhang, Z.; Kroon, D. E.; Casstevens, T. M.; Ramdoss, Y.; Buckler, E. S. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatic*, v.1, p.2633-2635, 2007.

Brown, A. H. D. Isozymes, plant populations genetics structure and genetic conservation. *Theoretical and Applied Genetics*, New York, v. 52, n. 4, p. 145- 157, 1978. Brown, W.L. Maize variability of potential interest to plant molecular geneticists. *Maydica*, v.30, p.225-233, 1985. Brown, A. H. D., Hodgkin, T. Indicators of Genetic Diversity, Genetic Erosion, and Genetic Vulnerability for Plant Genetic Resources. p. 25–53, 2015. Carneiro, M. S.; Vieira, M. L. C. Mapas genéticos em plantas. *Bragantia*, Campinas-SP, v.61, n.2, p.89-100, 2002. Cecarelli, S. Adaptation to low/high input cultivation. *Euphytica*, v.92, p.203-214, 1996. Conab (Companhia Nacional de Abastecimento). Acompanhamento da safra brasileira de grãos, v6-safra 2018/2019, oitavo levantamento, maio 2019. Brasília, DF. Disponível: 23 <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>. Acesso em 05 de junho 2019. Corder, E.H.; Saunders, A.M.; Risch, N.J.; Strittmatter, W.J.; Schmechel, D.E. Protective effect of apolipoprotein-E type-2 allele for late-onset Alzheimer-disease. *Nat. Genet.* V.7, p.180–84, 1994. Ciampitti, L. et al. Genetic analysis of maize root characteristics in response to low nitrogen stress. *Plant and Soil*, v.276, p.369-382, 2005. Chagné, D.; Batley, J.; Edwards, D.; Forster, J. W.; Single nucleotide polymorphism genotyping in plants, p.77-94. In: *Association mapping in plants*. Springer, 2007. Chun, L.; Mi, G.; Li, J.; Chen, F.; Zhang, F. Genetic analysis of maize root characteristics in response to low nitrogen stress. *Plant and Soil*, v.276, p.369-382, 2005. Cruz, C.D. Programa genes. *Biometria*. Viçosa, UFV, 382p. 2006. Cruz, C. D.; Ferreira, F. M.; Pessoni, L. A.; *Biometria aplicada ao estudo da diversidade genética*. Visconde do Rio Branco, MG: Suprema, 620p, 2011. Cruz, C. D.; Regazzi, A. J.; Carneiro, P. C. S. Modelos biométricoa aplicados ao melhoramento genético: Volume 1, 4th ed. UFV, Viçosa, 2012. Doebley, J.; Bacigalupo, A.; Stec, A. Inheritance of kernel weight in two maize – teosinte hybrid population: implications for crop Evolution. *Journal of Heredity*, Oxford, v.85, p.191-195, 1994. DoVale, J.C. Et al. Melhoramento para eficiência no uso de nitrogênio. In: *Fritsche-Neto, R.; Borém, A. melhoramento de plantas para condições de estresses abióticos*. Suprema, p.81- 99, 2011. Duvick, D.N. Biotechnology in the 1930s: the development of hybrid maize. *Nature Genetics*, New York, v.2, n.1, p.69-74, 2001. Duvick, D.N. Genetic contributions to advances in yield in U.S. maize. *Maydica*, Bergamo, v.37, n.1, p.69–79, 1992.24 Earl, D. A.; VonHoldt, B. M. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour*, v.4, p.359–361, 2012. Embrapa. Cultivares de Milho Disponíveis no Mercado Brasileiro na safra 2016/2017. Disponível: <https://www.embrapa.br/busca-de-publicacoes/-/publicação/1060346/mercado-de-sementes-de-milho-no-brasil-safra-20162017>. Acesso

em: 10 de março 2017. Ertiro BT, Semagn K, Das B, Olsen M, Labuschagne M, Worku M, Wegary D, Azmach G, Ogugo V, Keno T, Abebe B, Chibsa T, Menkir A. Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. *BMC Genomics*, 2017; 18:777. Evanno, G.; Regnaut, S.; GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. 2005. Ersoz ES, Yu J, Buckler ES, Applications of Linkage Disequilibrium and Association Mapping in Maize, In: Kriz AL, Larkins BA. (eds.) *Molecular Genetic Approaches to Maize Improvement. Biotechnology in Agriculture and Forestry*, 63:173-195, 2009. FAO. *Agricultural production: primary crops*. Disponível em: . Acesso em: 12 janeiro de 2020. Fan, J. B. et al. Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*, v. 68, p. 69-78, 2003. ISSN 0091-7451. Disponível em: <://WOS:000222969300010 >. Fidelis, RR et al. Fontes de germoplasma de milho para estresses abióticos. *Suprema*, 250p. 2011. Ferreira, M.E.; Grattapaglia, D. *Introdução ao uso de marcadores moleculares em análise genética*. 3 ed. Brasília: Embrapa-Cenargen, 1998. 220p. Flint-Garcia S. A.; Thornsberry J.M.; Buckler IV E. S. Structure of Linkage Disequilibrium in Plants. *Plant Biol*. V.54, p. 357– 374, 2003. Flint-Garcia, S. A.; Thuillet, A. C.; Yu, J.; Pressoir, G.; Romero, S. M.; Mitchell, S. E.; Doebley, J.; Kresovich, S.; Goodman, M. M.; Buckler, E. S. Maize association population: 25 a high-resolution platform for quantitative trait locus dissection. *The Plant Journal*, v.44, p.1054-1064, 2005. Fritsche-Neto, R. et al. Seleção genômica ampla e novos métodos de melhoramento do milho. *Ceres*, v. 59, p. 794-802, 2012. Fritsche-Neto, R.; Borém, A. *Plant Breeding for Abiotic Stress Tolerance*. 1th ed. Heidelberg: Springer, 172 p., 2012. Garnett, T.; Conn, V.; Kaiser, B.N. Root based approaches to improving nitrogen use efficiency in plants. *Plant, Cell and Environment*, v.32, p.1272-83, 2009. Garcia, D. E.; Medina, M. G.; Clavero, T.; Cova, L. J.; Dominguez, C.; Baldizan, A. Nutritional characterization of six fodder species foliage with emphasis in their polyphenolic profiles. *Revista Científica, FCV-LUZ*, v.18 (2), p.188-196, 2008. Gallais, A. *Quantitative genetics and breeding methods in autopolyploid plants*. Paris: Inra Editions. 2003. 515p. Ganal, M. W.; Durstewitz, G.; Polley, A.; Bérard, A.; Buckler, E. S.; Charcosset, A.; Clarke, J. D.; Graner, E. M.; Hansen, M.; Joets, J; Paslier, M. C. L.; McMullen, M. D.; Montalent, P.; Rose, M.; Schon, C. C.; Sun, Q.; Walter, H.; Martin, O. C.; Falque, M. A Large Maize (*Zea mays* L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. *Plos one*, v.6, e28334, 2011. Giacomo, R.; Pereira, M.G.; Guareschi, R.F.; Machado, D.L. Atributos químicos e físicos do solo, estoques de carbono e nitrogênio e

frações húmicas em diferentes formações vegetais. *Ciencia Florestal*, v.25(3), p.617-631, 2015. Goodman, M.M.; Bird, R.M. The Races of Maize IV: Tentative Grouping of 219 Latin American Races. *Economic Botany*, v.31, p.204-221, 1977. Griffing, B, A Generalized Treatment of the Use of Diallel Crosses in Quantitative Inheritance., *Heredity*, 1956. Gore, M. A.; Chia, J. M.; Elshire, R. J.; Sun, Q.; Ersoz, E. S.; Hurwitz, B. L.; Peiffer, J. A.; McMullen, M. D.; Grills, G. S.; Ross-Ibarra, J.; Ware, D. H.; Buckler, E. S. A First-Generation Haplotype Map of Maize. *Science*, v. 326, 2009.26 Guimarães LJM, Trindade RS, Parentoni, SN, Guimarães, PEO. Desenvolvimento de linhagens. In: DeLima RO, Borém A. (Eds.). *Melhoramento de Milho*. Viçosa, MG: Editora UFV, 2018:396p. Gupta, P. K.; Rustgi, S.; Kulwal, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology*, v.57, p.461- 485, 2005. Hallauer, A. R., & Miranda, J. B. *Quantitative Genetics in Maize Breeding* (2nd ed.). Iowa, Ames. USA: Iowa State University Press. 1988. Hartl, D. L.; Clark, A. G. *Princípios de genética de populações*. Porto Alegre: Editora Artmed, 4ª edição, 659pp, 2010. Hochholdinger F, Yu P, Marcon C. Genetic Control of Root System Development in Maize. *Trends Plant Sci* 23:79–88 2018. Hedrick, P. W. *Genetics of Population*. 4a ed. USA: Jones and Bartlett Publishers, Inc., 2011. 675 p. Hedrick, P.W. Gametic Disequilibrium Measures: Proceed With Caution. *Genetics*, Pittsburg, n.117. p.331-341, 1987. Hill, W.G.; Robertson, A. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, New York, v.38, p.226-231, 1968. Hochholdinger F, Yu P, Marcon C. Genetic Control of Root System Development in Maize. *Trends Plant Sci* 23:79–88, 2018. Horst, W. J.; Kamh, M.; Jibrin, J. M.; Chude, V. O. Agronomic measures for increasing P availability to crops. *Plant Soil*, v.237, p.211-233, 2001. Horst, W.J. et al. Genotypic differences in nitrogen use-efficiency in crop plants. In: Lynch, J.M. Schepers, J.S.; Unver, I (eds). *Innovative soil-plant systems for sustainable agricultural production*. Paris: OECD publishing, p.25-92. 2003. Johnson, S. A.; Cubberley, G.; Bentley, D. L. Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. *Molecular Cell*, v.33, p.215-26, 2009.27 Jombart, T.; Kamvar, Z.N.; Collins, C.; Lustrik, R.; Beugin, M.P.; Knaus, B.J.; Soly-mos, P.; Mikryukov, V.; Schliep, K.; Maié, T.; Morkovsky, L.; Ahmed, I.; Cori, A.; Calboli, F.; Ewing, R.J. *adegenet: a R package for the multivariate analysis of genetic markers*, 2018. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, v.9, p.29, 2013. Kopelman, N. M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N. A.; Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, v.15, p. 1179-1191, 2015.

Laurie, C. C.; Doheny, K. F.; Mirel, D. B.; Pugh, E. W.; Bierut, L.J.; Bhangale, T.; Boehm, F.; Caporaso, N. E.; Cornelis, M. C.; Edenberg, H. J.; Gabriel, S. B.; Harris, E. L.; Hu, F. B.; Jacobs, K.; Kraft, P.; Landi, M. T.; Lumley, T.; Manolio, T. A.; McHugh, C.; Painter, I.; Paschall, J.; Rice, J.P.; Rice, K. M.; Zheng, X.; Weir, B. S. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol.* 34(6): pag.591-602. 2010.

Langley, C.H.; Lazzaro, B.P.; Phillips, W.; Heikkinen, E.; Braverman, J.M. Linkage disequilibria and the site frequency spectra in the *su(s)* and *su(wa)* regions of the *Drosophila melanogaster* X chromosome. *Genetics*, v.156, p.1837–52, 2000.

Laperche, A.; Aigu, Y.; Jubault, M. Clubroot resistance QTL are modulated by nitrogen input in *Brassica Napus*. *Theor Appl Genet*, v.130, p.669–684, 2017.

Leal, A.; Mangolin, C.; do Amaral Júnior, A.; Gonçalves, L.; Scapim, C.; Mott, A.; Eloi, I.; Cordovés, V.; Da Silva, M. Efficiency of RAPD versus SSR markers for determining genetic diversity among popcorn lines. *Genet Mol Res*, v. 9, p.9-18, 2010.

Lewontin, R.C.; Kojima, K. The Evolutionary Dynamics of Complex Polymorphisms. *Evolution, Massachusetts*, v. 14, n. 4, p. 458-472, 1960.

Lewontin, R.C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*. v.49, p.49–67, 1964.

Li, Y., Shi, Y.S., Cao, Y.S. and Wang, T.Y. (2002) A phenotypic diversity analysis of maize germplasm preserved in China. *Maydica* 47: 107-114.

Liu, K.; Goodman, M.; Muse, S.; Smith, S.; Buckler, E.; Doebley, J. Genetic Structure and Diversity Among Maize Inbred lines as Inferred from DNA Microsatellites. *Genetics*, v.165, p.2117-2128, 2003.

Linkohr, B.I.; Dunn, TN. An introduction to plant sphingolipids and review of recent advances in understanding their metabolism and function. *New phytologist*, v.161, p.677-702, 2004.

Long, A.D.; Lyman, R.F.; Langley, C.H.; Mackay, T.F. Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics*, v.149, p.999–1017, 1998.

Lu, Y.; Shah, T.; Hao, Z.; Taba, S.; Zhang, S.; Gao, S; Liu, J.; Cao, M.; Wang, J.; Prakash, A. B.; Rong, T.; Xu, Y. Comparative SNP and Haplotype Analysis Reveals a Higher Genetic Diversity and Rapider LD Decay in Tropical than Temperate Germplasm in Maize. *Plosone*, 6(9): e24861. 2011.

Mangelsdorf, P.C. *Corn, its origin, evolution and improvement*. Cambridge, Massachusetts: Harvard University Press, 1974. 262p.

Melchinger, A.E.; Gumber, R.K. Overview of heterosis and heterotic group in agronomic crops. In LAMKEY, K.R.; STAUB, J.E. (Ed.). *Concepts and breeding of heterosis in crop plants*. Madison: ACSESS, 1998. p. 29-44.

Melo, W.M.C.; Pinho, R.G.V.; Santos, J.B.; Ferreira, D.F. Utilização de caracteres morfoagronômicos e marcadores moleculares para a avaliação da divergência genética entre híbridos de milho. *Revista Ceres, Viçosa*, v.48, n.276, p.195-207, 2001.

Miranda Filho, J. B.; Viégas, G. P. Milho híbrido. In: PArteniani, E.;

Viègas, G. P. (Eds.). *Melhoramento e produção de milho*. Campinas: Fundação Cargill, 1987. v. 1, p. 275-340. Meuwissen, T.H.E.; Karlsen, A.; Lien, S.; Olsaker, I; Goddard, M. E. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, v161, p.373-379, 2001. Medici, L.O. et al. The influence of nitrogen supply on antioxidant enzymes in plant roots. *Functional plant biology*, v.31, p.1-9, 2004.29 Mojena, R. Hierarchical grouping methods and stopping rules: an evaluation. *Computer Journal*, v.20, p.359–363, 1977. Monneveux, P.; Zaidi, P.H.; Sanchez, C. Population density and low nitrogen affects yield-associated traits in tropical maize. *CropScience*, v.45, p.535-545, 2005. Montgomery, D.C.; Peck, E.A. *Introduction to linear regression analysis*. New York: J. Wiley, 1981. 504p. Morosini, J.S.; Mendonça, L.F; Lyra, D.H.; Galli, G.; Vidotti, M.S.; Fritsche-Neto, R. Association mapping for traits related to nitrogen use efficiency in tropical maize lines under field conditions. *Plant Soil*, v.421, p.453–463, 2017. Mõro. G. V.; Fritsche-Neto, R. Importância e usos do milho no Brasil. In: Galvão. J. C. C.; Borém, A.; Pimentel, M. A. *Milho: Do plantio à colheita*. Viçosa, MG: Ed. UFV, p. 9-25, 2015. Mundim, G. B.; Viana, M. S.; Maia, C.; Paes, G. P.; Delima, R. O.; Valente, M. S. F. Inferring tropical popcorn gene pools based on molecular and phenotypic data. *Euphytica*, v.202, p. 55-68, 2015. Nei, M. Genetic Distance between Populations. *The American Naturalist*, v.106, p. 283-292, 1972. Nordborg, M.; Borevitz, J.O.; Bergelson, J.; Berry, C.C.; Chory, J. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 30:190–93, 2002. Nordborg, M.; Weigel, D. Next-generation genetics in plants. *Nature*, New York, v. 456, n. 7223, p. 720-723, 2008. Nunes, J.L.S. (2016). *Tecnologia de sementes -Produção de sementes de milho*. Disponível em: https://www.agrolink.com.br/sementes/tecnologia-sementes/producao-de-sementes-de-milho_361338.html. Acessado em: 20 de Abril de 2019. Nyombayire, A.; Derera, J.; Sibiya, J.; Gasura, E.; Ngaboyisonga, C. Genetic diversity among maize inbred lines selected for the midaltitudes and highlands of Rwanda. *Maydica*, v.61, 2016.30 Olmos, S. E.; Delucchi, C.; Ravera, M.; Negri, M. E.; Mandolino, C.; Eyhéabide, G. H. Genetic relatedness and pop structure within the Argentinean maize inbred lines. *Maydica*, v. 59, p.16-31, 2014. Oraguzie, N.C.; Rikkerink, E.H.A.; Gardiner, S.E.; Silva, H.N. *Association Mapping in Plants*. Springer, Tokio and New York, 277 pp., 2007. Pace, J.; Gardner, C.; Romay, C.; Ganapathysubramanian, B.; Lübberstedt, T. Genome-wide association analysis of seedling root development in maize (*Zea mays* L.). *BMC Genomics*, v.16 p.47, 2015a. Pace, J.; Yu, X.; Lübberstedt, T. Genomic prediction of seedling root length in maize (*Zea mays* L.). *Plant J* 83:903–912, 2015b. Paliwal RL, Granados G, Lafitte HR, Violic AD, Marathée JP. Tropical

maize: improvement and production, pp.374, 2000. Peng Y, Niu J, Peng Z, et al. Shoot growth potential drives N uptake in maize plants and correlates with root growth in the soil. *F Crop Res* 115:85–93, 2010. Paterniani, E.; Campos, M.S. Melhoramento de Milho. In: Borém, A. Melhoramento de espécies cultivadas. Viçosa: Ufv, 1999, 817p. Paterniani, M.E.A.G.Z. Use of heterosis in maize breeding: history, methods and perspectives – A review. *Crop Breeding and Applied Biotechnology*, v.1, n.2, p.159-178, 2001. Paterniani, M.E.A.G. Z; Guimarães, P.S.; Lüders, R.R.; Gallo, P.B.; Souza, A.P.; Laborda, P.R.; Oliveira, K.M. Capacidade combinatória, divergência genética entrelinhagens de milho e correlação com heteroses. *Bragantia*, Campinas, v.67, n.3, p.639-648, 2008. Pearson, K. On lines and planes of closet fit to systems of points in space. *Philosophical Magazine*, v.2, p.559-572, 1901. Pereira Filho, A.I.; Borghi, E. Sementes de Milho no Brasil – A dominância dos transgênicos. Empresa Brasileira de Pesquisa Agropecuária. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2018. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/186601/1/doc223.pdf> Acesso em 05 de junho de 2019. 31 Pioneer Sementes. Como escolher um híbrido. Disponível em: <http://www.pioneersementes.com.br/milho/como-escolher-um-hibrido>. Acessado em: 19 de maio de 2017, 2014. Prazeres, C. S.; Coelho, C. M. M. Divergência genética e heterose relacionada à qualidade fisiológica em sementes de milho. *Bragantia*, v. 75, p.411–417, 2016. Pritchard, J.K.; Przeworski, M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* 69:1–14, 2001. Pritchard, J.K.; Rosenberg, N.A. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet.* v.65, n.1, p.220-8, 1999. Pritchard, J. K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, v.155, p.945-959, 2000. R Core Team. The R project for statistical computing- R version 3.1.1. Software. [Internet]. 2015. Available: <http://www.r-project.org> Rafalski, A. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, v.5, p.94-100, 2002. Reif, J. C.; Hallauer, A. R.; Melchinger, A. E. Heterosis and heterotic patterns in maize. *Maydica* v.50, p.215–223, 2005. Remington, D.V.; Thornsberry, J. M.; Matsuoka, Y.; Wilson, L.M.; Whitt, S.R.; Doebley, J.; Kresovich, S.; Goodman, M.M.; Bucler IV, E. S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the Natural Academy Sciences*, v.20, p.11479-1148, 2001. Resende, M.D.V.; Resende Júnior, M.F.R.; Aguiar, A.M.; Abad, J.I.M.; Missiaggia, A.A.; Sansaloni, C.; Petroli, C.; Grattapaglia, D. Computação da seleção genômica ampla (GWS). Colombo: Embrapa Florestas 2010. 79p. Resende, M. D. V. de, Silva, F. F., Lopes, P. S.; Azevedo, C. F. Seleção Genômica ampla (GWS)

via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial. Viçosa, Universidade Federal de Viçosa/Departamento de Estatística. 2012, 291 p.32

Resende, M. D. V. de, Silva, F. F., Azevedo, C. F. Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTLGWAS, estatística espacial e temporal, competição, sobrevivência. Suprema, 2014.

Robinson, I.P. Aloenzimas na genética de populações de plantas. In: Alfenas, A.C. (Ed.). Eletroforese de isoenzimas e proteínas afins: fundamentos e aplicações em plantas e microrganismos. Viçosa, MG: UFV, p. 329-380, 1998.

Rodrigues, M.C.; Rezende, W.M.; Silva, M.E.J.; Faria, S.V.; Zuffo, L.T.; Galvão, J.C.C.; DeLima R.O. Genotypic variation and relationships among nitrogen-use efficiency and agronomic traits in tropical maize inbred lines. GRM, 2017.

Rosenberg, N. A. distruct: a program for the graphical display of population structure. Molecular Ecology Resources, v.4, p.137-138, 2004.

Sanchez, D.L., Liu, S.; Ibrahim, R.; Blanco, M.; Lübberstedt, T. Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (*Zea mays* L.). Plant Sci, v.268, p.30–38, 2018.

Slatkin, M. Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics, New York, v.9, p.477-485, 2008.

Sharbel, T.F.; Haubold, B.; Mitchell-Olds, T. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Mol. Ecol. v.9, p.2109–18, 2000.

Santos, F.M.C. Capacidade de combinação de híbridos comerciais de milho visando à obtenção de híbridos de F2. Tese de Mestrado. Instituto Agrônomo de Campinas. 80p. 2009.

Semagn, K.; Magorokosho, C.; Vivek, B. S.; Makumbi, D.; Beyene, Y.; Mugo, S.; Prasanna, B. M.; Warburton, M. L. Molecular characterization of diverse CIMMY maize lines using SNP. BMC Genomics, v.13, 2012.

Silva, M. J., Pastina, M. M.; Souza, V. F.; Schaffert, R. E.; Carneiro, P. C. S.; Noda, R. W.; Carneiro, J. E. S.; Damasceno, C. M. B.; Parrella, R A. C. Phenotypic and molecular characterization sorghun accessions for bionergy. Plos One, 2017.33

Singh, Y. P.; Kumar, A.; Chauhan, B. P. S. Genetic divergence in pearl millet. The Indian J. of Genet and Plant Breed, v.41, p.186-190, 1981.

Souza L.V. et al. Genetic control of grain yield and nitrogen use efficiency in tropical maize. Pesquisa Agropecuaria Brasileira, v.43, p.1517-1523, 2008

Souza, LV. Et al. Combining ability of maize grain yield under different levels of environmental stress. Pesquisa Agropecuaria Brasileira, v.44, p.1297-1303, 2009.

Smith, J.S.C.; Smith, O.S. The use of morphological, biochemical and genetic characteristics to measure distance and to test for minimum distance between inbred lines of maize (*Zea mays* L.). Maydica, Bergamo, v.34, n.1, p.141–150, 1989.

Smith, S.

D.; Murray, S. C.; Heffner, E. Molecular analysis of genetic diversity in a Texas maize breeding program. *Maydica*, v.60, 2015. Stich, B.; Melchinger, A. E.; Frisch, M.; Maurer, H. P.; Heckenberger, M.; Reif, J. C. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theoretical and Applied Genetics*, v.111, p. 723-730, 2005. Stich, B.; Maurer, H. P.; Melchinger, A. E.; Frisch, M.; Heckenberger, M.; Van Der Voort, J. R.; Peleman, J.; Sorensen, A. P.; Reif, J. C. Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Molecular Breeding*, v.7, p.217-226, 2006. Torres, L.G.; Rodrigues, M.C.; Lima, N. L.; Trindade, T.F.H.; Silva, F. F.; Azevedo, C. F.; DeLima, R. O. Multi-trait multi-environment Bayesian model reveals G x E interaction for nitrogen use efficiency components in tropical maize. *Plos One*, June 27, 2018. Torres, L.G.; Caixeta, D.G.; Rezende, W. M.; Schuster, A.; Azevedo, C. F.; Silva, F. F.; DeLima, R. O. Genotypic variation and relationships among traits for root morphology in a panel of tropical maize inbred lines under contrasting nitrogen levels. *Euphytica*, 215:51, 2019. Toaldo, D. et al. Selection in early Generation and the occurrence of heterosis for the character root distribution. *Euphytica*, v190, p.335-344, 2013.34 Thornsberry, J.M.; Goodman, M.M.; Doebley, J.; Kresovich, S.; Nielsen, D. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* v.28, p.286–89, 2001. USDA (United States Department of Agriculture). Grain: World Markets and Trade, 10.03.2018. Disponível: <https://apps.fas.usda.gov/psdonline/circulars/grain-corn-coarsegrains.pdf>. Acesso em 10 de março 2018. Van Inghelandt, D.; Melchinger, A. E.; Lebreton, C.; Stich, B. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet*, v.120, p.1289-1299, 2010. Viana, J.M.S.; Valente, M.S.F.; Silva, F.F.; Mundim, G.B.; Paes, G.P. Efficacy of population structure analysis with breeding populations and inbred lines. *Genetica*. v.141, p.389–399, 2013. Wang, R.; Yu, Y.; Zhao, J.; Shi, Y.; Song, Y.; Wang, T., Li, Y. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor Appl Genet*, v.117, p.1141–1153, 2008. Warburton, M.L., Xianchun, X., Crossa, J., Franco, J., Melchinger, A.E., Frisch, M., Bohn, M. and Hoisington D. Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci.* v.42, p.1832-1840, 2002. Warburton, M.L.; Ribaut, J.M.; Franco, J.; Crossa, J.; Dubreuil, P.; Betrán, F.J. Genetic characterization of 218 elite CIMMYT inbred maize lines using RFLP markers. *Euphytica* 142:97–106, 2005. Wei, X.; Jackson, P.A.; McIntyre, C.L. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theoretical and Applied Genetics*, v.114, p.155-164, 2006. Wen, W.; Franco,

J.; Chavez-Tovar, V. H.; Yan, J.; Taba, S. Genetic Characterization of a Core Set of a Tropical Maize Race Tuxpenõ for Further Use in Maize Improvement. *Plos one*, v.7, 2012. Weir, B.S. *Genetic Data Analysis II*. Sunderland, MA: Sinauer. 376 pp., 1996.35 White, S.E.; Doebley, J.F. Of genes and genomes and the origin of maize. *Trends in Genetics*, Oxford, v. 14, n.8, 1998. Whitt, S.R.; Buckler, E.S. Using natural allelic diversity to evaluate gene function. In: Grotewold E, editor. *Plant Functional Genomics*. New Jersey: Humana Press; 2003. pp. 123±140. <https://doi.org/10.1385/1-59259-413-1:123> Wu, Y.; Vicente, F. S.; Huang, K.; Dhliwayo, T.; Costich, D. E.; Semagn, K.; Sudha, N.; Olsen, M.; Prasanna, B. M.; Zhang, X.; Babu, R. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theor Appl Genet*, v.129, p.753- 765, 2016. Xie, C.; Warburton, M.; Li, M.; Li, X.; Xiao, M.; Hao, Z.; Zhao, Q.; Zhang, S. An analysis of Population structure and LD in maize inbred lines. *Mol Breeding*, v.21, p. 407-418, 2008. Xue, Y. et al. Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. *Theoretical and Applied genetics*, v.126, p.2587-2596,2013. Yang, X.; Gao, S.; Xu, S.; Zhang, Z.; Prasanna, B. M.; Li, L.; Li, J.; Yan, J. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol Breeding*, v.28, p.511-526, 2011. Yu, J.M.; Buckler, E.S. Genetic association mapping and genome organization of maize. *Current Opinion in biotechnology*, v.17, p.155-160, 2006. Zapata, C. The D´ measure of overall gametic disequilibrium between pairs of multiallelic loci. *Evolution*, Lawrence, v.54, p.1809-1812, 2000. Zheng, D; Van, K.; Lee, S. - Molecular diversity and relationships among elite maize inbreds from US and CIMMYT populations and current heterotic groups in China. *Hereditas*, v.145, p.182-193, 2008. Zonta, E. et al. O sistema radicular e suas interações com o ambiente edáfico. In: Fernandes, MS, ed. *Nutrição Mineral de plantas*, Sociedade Brasileira de ciencias do solo, 2006, p7- 52. Zhu, C.; Gore, M.; Buckler, I.V.; Yu, J. Status and prospects of association mapping in plants, *The Plant Genome*, v.1, p.5-20, 2008.

5. CAPÍTULO 1

Caracterização molecular e fenotípica de um conjunto de linhagens de milho tropical

RESUMO

A caracterização da diversidade genética e a estrutura populacional de linhagens de milho são de grande importância para o sucesso de qualquer programa de melhoramento pois auxiliam os melhoristas a manter e potencialmente aumentar a taxa de ganho genético. Assim, o objetivo do presente estudo foi realizar a caracterização fenotípica e molecular de um conjunto de linhagens endogâmicas de milho tropical do banco de germoplasma de milho do programa de melhoramento público da Universidade Federal de Viçosa (UFV), com base em 18 caracteres agronômicos e 3.083 marcadores moleculares SNPs (*Single Nucleotide Polymorphism*) para: (i) quantificar a variação genética para caracteres agronômicos ; ii) estimar a relação genética entre as linhagens; iii) determinar o decaimento de LD; iv) investigar a estrutura populacional; v) avaliar a diversidade fenotípica e genética e vi) avaliar se esse conjunto de linhagens endogâmicas pode compreender um painel de linhagens endogâmicas de milho tropical para estudos de mapeamento de associação. Os resultados mostraram que existe uma grande variação fenotípica e genotípica no conjunto de linhagens endogâmicas de milho tropical do programa de melhoramento de milho da UFV. Também encontramos uma alta diversidade genética ($GD = 0,34$) e baixos coeficientes de parentesco entre as linhagens de milho (apenas cerca de 4% do parentesco relativo aos pares estava acima de 0,50). A distância de decaimento de LD em todos os dez cromossomos em todo o conjunto de linhagens de milho com $r^2 = 0,2$ foi 86.899 kb. Em relação à estrutura populacional, os resultados do STRUCTURE e da análise de componentes principais distinguiram as linhagens endogâmicas em três subpopulações, com muita consistência entre os dois métodos. Além disso, os resultados da análise de agrupamento UPGMA com base nos dados moleculares e fenotípicos agruparam as linhagens em 14 e 22 agrupamentos de divergência genética, respectivamente. Esses resultados indicam que o conjunto de linhagens endogâmicas do programa de melhoramento de milho da UFV pode compreender um painel de linhagens de milho tropical e que este é adequado para o estudo de associação de todo o genoma, principalmente em ambientes de clima tropical. Além disso, as informações resultantes deste estudo serão úteis para auxiliar na seleção das melhores combinações parentais para novos cruzamentos, mapeamento de populações, cruzamentos orientados para obtenção de híbridos altamente heteróticos e produtivos e na predição de híbridos não testados no programa de melhoramento de milho da UFV.

5.1. INTRODUÇÃO

A avaliação da diversidade genética, a estrutura populacional e as relações dentro e entre as linhagens de milho são de grande importância para o sucesso de qualquer programa de melhoramento e para estudos de associação genômica ampla em milho. Informações precisas da diversidade genética podem auxiliar o melhorista na alocação de linhagens de milho em grupos heteróticos, na seleção de testadores eficientes para testes de linhagens em combinações híbridas e é útil para ampliar a base genética dos programas de melhoramento (Adu et al., 2019; Semagn et al., 2012; Wu et al., 2016). Diversas estratégias têm sido usadas, ao longo dos anos, para avaliar a diversidade genética entre e dentro de populações, com base em diferenças morfológicas, agrônômicas, geográficas, bioquímicas e moleculares.

A caracterização fenotípica fornece uma gama de informações sobre a variabilidade genética. Entretanto, os efeitos do ambiente, a interação genótipos x ambientes e os erros de medição também contribuem para as diferenças observadas entre os genótipos (Adu et al., 2019; Nyombayire et al., 2016; Schlotterer, 2004; Silva et al., 2017). Por outro lado, a caracterização molecular é livre de efeitos ambientais e têm sido uma ferramenta indispensável para caracterizar recursos genéticos e fornecer aos melhoristas informações mais detalhadas que podem auxiliar na seleção de linhagens mais divergentes (Leal et al., 2010). Vários autores relataram que o uso combinado das informações de marcadores moleculares e de caracteres fenotípicos pode ser mais eficiente para quantificar as diferenças genéticas entre as linhagens de milho (Barata e Carena, 2006; Franco et al., 2001; Hartings et al., 2008; Mundim et al., 2015; Wu et al., 2014).

Diferentes marcadores moleculares têm sido amplamente utilizados para avaliar a diversidade genética em milho. No entanto, os marcadores SNP (*Single Nucleotide Polymorphism*) apresentam algumas vantagens, como por exemplo, o baixo custo por *data point*, alta abundância genômica, especificidade de *loci*, codominância, potencial para análise de alto rendimento e menores taxas de erro de genotipagem (Chagné et al., 2007; Van Inghelandt et al., 2010; Wu et al., 2014). Semagn et al. (2012) relataram que os marcadores SNPs são uma ferramenta poderosa em estudos de diversidade genética. Esses autores utilizaram 1.065 SNPs para genotipar 450 linhagens de milho com objetivo de avaliar a utilidade desses marcadores e identificar um subconjunto de SNPs altamente informativos para a genotipagem de rotina e de baixo custo do germoplasma do milho.

No melhoramento genético, a informação da estrutura de população permite ao melhorista conhecer a estrutura genética de uma população, detectar estrutura familiar, estimar níveis de migração e dispersão, além de auxiliar na manutenção do banco de germoplasma e no diagnóstico evolutivo. Também é utilizada para identificar grupos heteróticos. Além disso, a estrutura populacional estimada a partir de marcadores moleculares, fornece informações úteis para controlar associações espúrias que limitam estudos de mapeamento associativo. A existência de associações falso-positivas entre locos não ligados é um dos problemas que ocorrem quando não é realizado o controle efetivo da influência deste fator sobre o desequilíbrio de ligação (LD) (Rafalski, 2002), exatamente como pode acontecer nas análises tradicionais de seleção genômica (GWS). LD é a associação não aleatória de alelos em diferentes loci, ele desempenha um papel integral no mapeamento de associação e determina a resolução de um estudo de associação (Flint-Garcia et al., 2003). Vários estudos mostraram a eficácia da estrutura de população na classificação dos agrupamentos (Van Inghelandt et al., 2010; Viana et al., 2013; Xie et al., 2008). Wang et al. (2008) utilizaram 145 marcadores SSR para avaliar a diversidade genética, estrutura populacional e LD de um conjunto de 95 linhagens endogâmicas de milho. Eles identificaram quatro subgrupos que correspondem às quatro principais origens do germoplasma chinês.

Diversos estudos de caracterização molecular e estrutura populacional em milho foram realizados com diferentes germoplasmas e vários tipos de marcadores moleculares (Liu, et al., 2003; Olmos et al., 2014; Smith et al., 2015; Warburton et al., 2002; Wen et al., 2012; Yan et al., 2009; Yang et al., 2011). No entanto, a maioria dos estudos anteriores se concentrou principalmente em milho temperado e, portanto, há uma carência de informação sobre o germoplasma de milho tropical, principalmente no Brasil. O germoplasma de milho tropical tem maior diversidade genética, menor distância de decaimento de LD e, conseqüentemente, contém mais alelos raros do que o milho temperado (Zhang et al., 2016; Wu et al., 2014; Lu et al., 2011; Yan et al., 2009). Assim, torna-se de extrema importância a caracterização molecular e fenotípica do germoplasma de milho tropical (Wu et al., 2016).

O Programa Milho®, Programa de Melhoramento de Milho da Universidade Federal de Viçosa (UFV), possui um banco de germoplasma de milho que é constituído por linhagens obtidas de diferentes fontes de germoplasma de milho tropical. Há algum tempo, esse banco vem sendo mantido e renovado. Esse conjunto de linhagens foram

fenotipadas para caracteres de raiz e algumas delas foram fenotipadas para caracteres de eficiência no uso de nitrogênio (Rodrigues et al., 2017; Torres et al., 2018; Torres et al., 2019). Entretanto, estudo detalhado sobre a diversidade genética existente entre essas linhagens ainda não foi realizado. Além disso, análises moleculares que pudessem certificar a existência de grupos divergentes de linhagens, a partir dos quais poderiam ser escolhidos pares de linhagens que pudessem atuar como genitores, ainda não foram conduzidas. Contudo, pouca informação sobre a diversidade genética, estrutura populacional e LD é conhecida nesse conjunto de linhagens endogâmicas para posterior estudo de mapeamento de associação.

Assim, os objetivos deste estudo foram: (i) quantificar a variação genética para caracteres agronômicos; ii) estimar a relação genética entre as linhagens; iii) avaliar a diversidade fenotípica e genética; iv) investigar a estrutura populacional e padrões heteróticos; v) determinar a extensão e o decaimento de LD das linhagens e vi) avaliar se este conjunto de linhagens endogâmicas pode compreender um painel de linhagens endogâmicas de milho tropical para posterior análise de mapeamento de associação.

5.2. MATERIAL E MÉTODOS

5.2.1. Material vegetal

Neste estudo foi utilizado um conjunto de 187 linhagens endogâmicas de milho tropical do Banco Ativo de Germoplasma do Programa Milho[®], Programa de melhoramento de milho da Universidade Federal de Viçosa (UFV), Departamento de Fitotecnia, Viçosa, Minas Gerais, Brasil. As linhagens foram obtidas de diferentes fontes de germoplasma de milho tropical: híbridos comerciais de milho (convencionais), populações melhoradas e variedades de polinização aberta: 166 linhagens foram derivadas de 49 híbridos comerciais plantados no Brasil (áreas tropicais), nove linhagens foram derivadas de duas variedades de polinização aberta (BR106 e Nitroflint) e sete linhagens foram derivadas de duas populações sintéticas (CMS28 e CMS50). Todas as 187 linhagens são resultantes de pelo menos oito gerações de autofecundação e, em teoria, aproximadamente todos os locos encontram-se em homozigose. Elas foram desenvolvidas usando um método de pedigree modificado na mesma região onde foram avaliadas. As linhagens representam uma amostra do germoplasma usado em nosso programa de melhoramento de milho para desenvolver híbridos, novas linhagens

consanguíneas, populações melhoradas e populações para mapeamento de QTL e estudos de herança para ambientes de estresse, principalmente nitrogênio.

5.2.2. Dados Fenotípicos

Para a caracterização fenotípica foram utilizadas 187 linhagens endogâmicas de milho tropical avaliadas em três locais do sudoeste do Estado de Minas Gerais e em duas safras, o que totalizou quatro ambientes. Foram avaliadas na safra de 2016 e na safra 2016/2017 na Estação Experimental de Coimbra, Coimbra-MG (20°49'46,5" S, 42°45'51,1" O, altitude 715 m). Além disso, na safra de 2016/2017, as linhagens foram avaliadas na Estação Experimental da Horta Nova (20°45'47,6" S, 42°49'25,1" O, altitude 665 m), localizado no distrito de São José do Triunfo, Viçosa-MG; e na Estação Experimental do Aeroporto (20°45'47,6" S, 42°49'25,1" O, altitude 665 m), Viçosa-MG. Todas as estações experimentais pertencem ao Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV). Os experimentos foram conduzidos em condições de sequeiro, e nenhuma irrigação foi aplicada. O gerenciamento da trilha era o mesmo para todos os experimentos que empregavam práticas agrícolas padrão. Todas as sementes de milho foram tratadas com CropStar® (Imidacloprid, Tiodicarbe) e Maxim Advanced® (Metalaxyl-M, Tiabendazol, Fludioxonil). Em todos os ambientes, o ensaio foi realizado em delineamento experimental Alfa-látice (17x11) com duas repetições. Cada parcela era uma única linha de 4 m com linhas espaçadas de 0,80 m. O plantio das parcelas foi manual e desbastadas no estágio V3 para uma população de plantas de 62.500 plantas ha⁻¹, que é a população de plantas ideal para o cultivo de milho nesses ambientes.

Foram avaliados os seguintes caracteres, medidos em cinco plantas do centro de cada parcela: Um caractere fisiológico, conteúdo de clorofila na folha (SPAD), que foi mensurado na folha abaixo da espiga superior, 15 dias após o florescimento feminino, usando o medidor portátil de clorofila (SPAD-502); 10 caracteres de arquitetura de planta, dias até o florescimento masculino (FM) e feminino (FF), dias desde a semeadura até o florescimento masculino para 50% das plantas em uma parcela e da semeadura até a emergência dos cabelos (estilo-estigma visíveis) para 50% das plantas dentro de uma parcela, respectivamente; comprimento foliar (CF) e largura foliar (LF) foram medidos em centímetros do início da lígula até a ponta da folha da espiga superior e na seção mais larga da folha da espiga, respectivamente e a área foliar (AF), foi estimado em centímetros quadrados usando a fórmula de Montgomery (1909): [(comprimento x largura) x 0,75];

altura de plantas (AP), altura de espigas (AE), foram medidos em centímetros como a distância do nível do solo até o ponto de inserção da folha bandeira e até o ponto de inserção da espiga superior, respectivamente; diâmetro de colmo (DC) foi medido em milímetros no primeiro entrenó acima do solo usando um paquímetro digital e foram registrados o número de nós acima (NNAE) e abaixo (NNBE) da folha da espiga, respectivamente. Para os componentes da produção, seis caracteres foram mensurados: Prolificidade (PRL) ou número de espigas por planta foi estimado dividindo o número total de espigas por parcela pelo número de plantas colhidas; comprimento da espiga (CE) medido em centímetros por meio de uma régua da base à ponta da espiga; número de fileiras (NF), diâmetro de espiga (DE) e diâmetro de sabugo (DS) foram medidos em milímetros usando um paquímetro digital; peso de 1000 grãos (P1000) medido em gramas com o peso de 1000 grãos ajustado para 14,5% de umidade e por fim produtividade de grãos (PG), foram descascadas e registrado o peso e a porcentagem de umidade dos grãos para todas as espigas nas parcelas na maturidade fisiológica e o PG foi corrigido para 14,5% de umidade.

5.2.3. Dados Genotípicos

Para a caracterização molecular, foram genotipadas 182 linhagens endogâmicas de 187 avaliadas quanto aos caracteres fenotípicos. Amostras de tecido foliar foram obtidas a partir de cinco plantas para cada linhagem endogâmica plantadas na casa de vegetação no campo experimental Diogo Alves de Melo, UFV e as amostras foram coletadas no estágio de três folhas completamente expandidas (V3) e enviadas para a empresa DuPont Pioneer®, onde a extração de DNA e a genotipagem foram realizadas. Todas as 182 linhagens foram genotipadas usando a plataforma GoldenGate (Illumina, San Diego, CA, EUA) contendo 3.713 SNPs distribuídos em 10 cromossomos (Fan et al. 2003). A informação dos SNP utilizada neste estudo foi encontrada nos trabalhos de Jones et al. (2009); Gore et al. (2009) e Ganal et al. (2011). As informações sobre os SNPs estão disponibilizadas publicamente no banco de dados MaizeGDB (www.maizegdb.org) e NCBI (National Center for Biotechnology Information) (www.ncbi.nlm.nih.gov). Dos 3.713 SNPs genotipados, um conjunto de 3.083 SNPs com frequência de alelo menor (MAF) maior que 0,01 e Call rate de 90% foi selecionado para análise de caracterização genética de todas as linhagens endogâmicas neste estudo.

5.2.4. Análises estatísticas dos dados fenotípicos

Para estimar os componentes de variância e prever os valores genotípicos de cada linhagem endogâmica em todos os ambientes foi utilizado um modelo misto implementado no pacote R “lme4” (Bates et al., 2015). O ambiente foi incluído no modelo como efeito fixo, enquanto bloco, repetição e linhagem foram considerados como efeitos aleatórios. As interações entre os efeitos fixos e aleatórios foram incluídas no modelo como efeitos aleatórios. Os valores fenotípicos foram obtidos de acordo com a seguinte equação:

$$y_{irmk} = \mu + \alpha_i + \gamma_k + \alpha\gamma_{ik} + \varphi_{r(k)} + \delta_{m(rk)} + \varepsilon_{ikrm}$$

em que: μ é a média geral, α_i é o efeito aleatório da linhagem endogâmica, γ_k é o efeito fixo do ambiente, $\alpha\gamma_{ik}$ é o efeito aleatório da interação linhagem x ano, $\varphi_{r(k)}$ é o efeito fixo da repetição dentro do ano, $\delta_{m(rk)}$ é o efeito aleatório do bloco dentro da repetição dentro do ambiente, e ε_{ikrm} é o efeito aleatório do erro.

Uma análise de deviance e teste de razão de verossimilhança foi usada para testar os efeitos aleatórios por meio da estatística qui-quadrado (Resende, 2007). Intervalos e valores médios foram baseados em BLUPs. A herdabilidade no sentido amplo (\hat{h}^2) com base na média das linhagens foi estimada para cada caractere pela seguinte expressão (Hallauer et al., 2010):

$$\hat{h}_x^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \frac{\hat{\sigma}_{GxE}^2}{n} + \frac{\hat{\sigma}^2}{nr}}$$

em que: $\hat{\sigma}_G^2$, $\hat{\sigma}_{GxE}^2$ e $\hat{\sigma}^2$ são as estimativas de variância genotípica, estimativas de variância devido à interação linhagem x ano e estimativas de variância do erro, respectivamente; n e r são o número de ambientes e repetições, respectivamente.

Os coeficientes de correlação de Pearson entre pares de caracteres foram estimados usando as melhores estimativas lineares não enviesadas de predição (BLUP) de cada caractere. Os coeficientes de correlação foram estimados usando o pacote R “agricolae” (Mendiburu, 2019). Após a obtenção dos coeficientes de correlação, foi feito um *Heatmap* no software R (R Core Team, 2015). Além disso, para evitar os efeitos adversos das relações lineares entre linhas ou colunas da matriz e as correlações residuais na análise da diversidade genética, a multicolinearidade foi diagnosticada com base em uma matriz de correlação genotípica (Montgomery e Peck, 2001).

A avaliação da diversidade genética entre as 187 linhagens com base em dados fenotípicos foi realizada usando os pacotes R “ade4” (Dray et al., 2018) e “adegenet”

(Jombart et al., 2018). Geramos a matriz de distâncias por meio da distância generalizada de Mahalanobis, responsável pelas correlações residuais entre os caracteres (Mahalanobis, 1936). Em seguida, uma matriz de distância de Mahalanobis foi usada como dados de entrada para análise de agrupamento com base no método de agrupamento da ligação média entre grupos (UPGMA). O dendrograma UPGMA foi gerado com base na distância generalizada de Mahalanobis para estimar o nível de parentesco entre as linhagens endogâmicas usando o pacote R “ape” (Paradis et al. 2018). O método de Mojena (1977) foi usado para alocar as linhagens em grupos. De acordo com este método, o dendrograma deve ser cortado em função do valor médio da distância genética dos níveis de fusão e do desvio padrão dos valores de distância. A importância relativa das características na quantificação da diversidade genética foi estimada de acordo com o método de Singh (Singh et al., 1981) usando o pacote R " biotools" (Silva et al. 2017).

5.2.5. Análises estatísticas dos dados genotípicos

Usamos o conjunto de dados de 3.083 SNPs filtrados para estimar a frequência de alelos, o conteúdo de informação polimórfica (PIC), a diversidade de genes e a taxa de heterozigosidade. A análise da frequência alélica dos 3.083 SNPs foi realizada usando o software TASSEL 5.0 (Bradbury et al. 2007). A diversidade de genes e a taxa de heterozigosidade observada foi calculada utilizando o pacote “Poppr” (Kamvar et al., 2019) implementado no software R. Os valores de conteúdo de informação polimórfica (PIC) foram obtidos por:

$$PIC=1-\sum_{j=1}^{n-1} P_{ij}^2 - \sum_{j=1}^{n-1} \sum_{k=j+1}^n 2P_{ij}^2 P_{ik}^2$$

em que: P_{ij} e P_{ik} são as frequências dos j -ésimo e k -ésimo alelos do marcador i , respectivamente; a soma estende-se sobre n alelos.

Para acessar as relações genéticas entre todas as linhagens endogâmicas de milho, o coeficiente de parentesco relativo aos pares entre elas foi estimado com todo o conjunto de dados de 3.083 SNP usando o software TASSEL 5.0 (Bradbury et al., 2007). Um parentesco próximo de 0 indica nenhuma relação entre elas, e valores mais próximos de 1 indicam uma relação entre as linhagens.

O desequilíbrio de ligação (LD) entre cada par de SNPs em cada cromossomo foi estimado pelo coeficiente de correlação de Pearson ao quadrado (r^2) com 3.083 SNPs usando o software TASSEL 5.0 (Bradbury et al., 2007). A distância média de decaimento LD dentro e ao longo de dez cromossomos com $r^2 = 0,2$ foi usada para medir a diferença

da distância de decaimento LD entre todas as linhagens endogâmicas. Uma janela deslizante de 50 kb foi usada para determinar a largura da janela em um lado do local de início, o espaçamento entre dois loci no mesmo cromossomo foi segmentado a uma distância de 50 kb e o LD médio foi avaliado para cada janela. Os valores de LD entre cada par de SNPs foram plotados usando os pacotes R “ggplot2”, “Mass” e “scales” (Wickham, 2018; Wickham, 2019; Ripley et al., 2019;).

A estrutura populacional das 182 linhagens foi investigada usando o software STRUCTURE (Pritchard et al., 2000) e análise de componentes principais (PCA) com base em 3.083 SNPs. Na primeira abordagem, um método de agrupamento baseado em um modelo de mistura implementado no STRUCTURE foi executado para K número de grupos variando de 1 a 10, e cada K foi executado 20 vezes com um período de burn-in de 10.000 e 100.000 MCMC (cadeias de Markov e Monte Carlo). A estatística ad hoc delta K (ΔK) foi usada para determinar o número mais provável de grupos (Evanno et al. 2005) usando o software Structure Harvester (Earl e VonHoldt, 2011). O software Clumpak (Kopelman et al, 2015) foi usado para alinhar os rótulos de agrupamento nas 20 execuções antes de plotar os dados. As frequências alélicas foram consideradas correlacionadas e os loci considerados não vinculados. Foi estimado as atribuições de cada linhagem a cada grupo de acordo com seus valores médios de probabilidade de adesão (Pritchard et al., 2000). Linhagens endogâmicas com probabilidades de adesão maiores ou iguais a 0,60 foram atribuídas ao mesmo grupo, enquanto aquelas com probabilidades de adesão menores que 0,60 foram designadas a um grupo "misto" (Segman et al., 2012; Yang et al., 2011; Wu et al., 2014; Wu et al., 2016). Na segunda abordagem para inferir a estrutura populacional, PCA com base em 3.083 SNPs foi realizada usando o pacote R "pcaMethods" (Price et al., 2006), e os dois primeiros componentes principais foram ilustrados para o exame visual do padrão de agrupamento das linhagens.

A diversidade genética a partir dos dados moleculares foi obtida com base na distância genética de Nei (Nei, 1972) entre as 182 linhagens para os 3.083 SNPs, utilizando o pacote “adegenet” (Jombard et al., 2018) no software R. Em seguida, o dendrograma UPGMA para todas as 182 linhagens de milho foi gerado com base nas distâncias genéticas de Nei usando o pacote "ape" (Paradis et al., 2019). O método de Mojena (1977) foi usado para alocar as linhagens em grupos.

5.3. RESULTADOS

5.3.1. Variação fenotípica e parâmetros genéticos com base em caracteres fenotípicos

Houve uma enorme variação para todos os caracteres mensurados neste conjunto de linhagens endogâmicas de milho, conforme mostrado pelos valores genotípicos em quatro ambientes (Tabela 1). Florescimento masculino e feminino (FM e FF) variaram em torno de 66 a 73 dias. Altura de planta e de espiga (AP e AE) variaram de 102,1 a 170,2 e 40,5 a 100,6 cm, respectivamente. O comprimento e largura da folha (CF e LF) e a área foliar (AF) variaram de 60,2 a 96,6 cm, 7,3 a 12,7 cm e 401,7 a 710,3 cm², respectivamente.

Em relação aos componentes da produção, número de fileiras (NF) e peso de 1000 grãos (P1000) variaram de 9,7 a 21,9 linhas e de 158,7 a 350,1 g, respectivamente. A produtividade de grãos variou de 1.208,8 a 3.605,9 kg ha⁻¹. O coeficiente de variação (CV) variou de 2,5% (FM) a 28,6% (PG), e a maioria dos caracteres apresentou CVs baixos, esperado para SPAD (11,9%), AF (13,4%), AE (11,9%) e PRL (18,3%), que exibiu valores de CV intermediários, e PG que apresentou um valor de CV alto.

Os componentes de variância devido às linhagens foram altamente significativos ($P < 0,01$) pelo teste da razão de verossimilhança para todas os caracteres (Tabela 1). Os componentes de variância devido à interação linhagens x ambientes foram significativos ($P < 0,05$) para sete dos 18 caracteres avaliados: LF, AP, AE, DE, DS, P1000 e PG, e, portanto, as linhagens endogâmicas tiveram desempenho relativo diferente devido a variação ambiental para esses caracteres. Em geral, as estimativas de herdabilidade no sentido amplo ($\hat{h}_{\bar{x}}^2$) foram intermediárias a altas e variaram de 0,44 (PRL) a 0,79 (FF). Os valores de $\hat{h}_{\bar{x}}^2$ foram consistentemente maiores que 0,50 para quase todos os caracteres, exceto para PRL (0,44) e PG (0,47).

Tabela 1 – Estimativas de componentes de variância, herdabilidade, coeficiente de variação (CV), média, mínimo e máximo para os BLUPs dos 18 caracteres avaliados em quatro ambientes em 187 linhagens endogâmicas de milho

Caracteres ¹	Mínimo	Médio	Máximo	$\hat{\sigma}_G^2$	$\hat{\sigma}_{GE}^2$	\hat{h}^2	CV%
FM	66,2	72,8	80,8	9,1***	0,5 ^{ns}	0,79	2,5
FF	65,8	73,0	80,1	9,3***	0,6 ^{ns}	0,76	2,7
CF	60,2	76,3	96,6	39,1***	6,8 ^{ns}	0,58	9,2
LF	7,3	9,3	12,7	0,8***	0,1*	0,67	8,7
AF	401,7	529,9	710,3	3.229,0***	712,2 ^{ns}	0,53	13,4
PRL	0,8	1,1	1,4	0,0***	0,0 ^{ns}	0,44	18,3
AP	102,1	136,8	170,2	249,3***	21,1*	0,75	8,8
AE	40,5	71,7	100,6	133,5***	13,7***	0,76	11,9
DC	17,4	21,3	26,1	3,4***	0,3 ^{ns}	0,59	9,8
SPAD	30,3	44,1	55,0	25,8***	1,5 ^{ns}	0,64	11,9
NNBE	5,5	7,1	9,4	0,5***	0,1 ^{ns}	0,66	9,8
NNAE	4,4	5,9	7,4	0,3***	0,0 ^{ns}	0,68	8,6
CE	10,1	13,1	16,9	1,8***	0,3*	0,67	8,0
NF	9,7	13,9	21,9	3,4***	0,2 ^{ns}	0,76	8,8
DE	30,5	37,5	45,2	7,5***	1,8***	0,69	5,1
DS	16,0	23,4	28,8	5,3***	0,3 ^{ns}	0,76	6,4
P1000	158,7	239,0	350,1	1.314,6***	335,0***	0,71	9,8
PG	1028,8	2263,3	3605,9	289.626,0***	94102***	0,47	28,6

*, ** e ***: significativo à 10, 5 e 1 % de probabilidade pelo teste da razão de verossimilhança (TRV), respectivamente. ¹FM-florescimento masculino, FF-florescimento feminino, CF-comprimento foliar, LF-largura foliar, AF-área folia, PRL-prolificidade, AP-altura de planta, AE-altura de espiga, DC-diâmetro de colmo, SPAD-teor de clorofila na folha, NNBE-número de nós abaixo da espiga, NNAE-número de nós acima da espiga, CE-comprimento da espiga, NF-número de fileiras, DE-diâmetro de espiga, DS-diâmetro de sabugo, PG-produtividade de grãos e P1000-peso de 1.000 grãos.

Embora a maioria dos coeficientes de correlação de Pearson entre as características não tenham sido significativos ($P > 0,05$) ou tenham apresentado baixa magnitude, encontramos correlações positivas moderadas a fortes ($r > 0,45$) para alguns pares de caracteres (Figura 1). Fortes correlações foram observadas entre FM e FF ($r = 0,91$), DS e DE ($r = 0,81$), AE e AP ($r = 0,77$) e AE e NNBE ($r = 0,76$), enquanto correlações moderadas foram apresentadas entre NNBE e AP (0,45), AF e CF (0,57), AF e LF (0,68), NF e DS (0,45), NF e DE (0,55), e entre PG e DE (0,47) e PG e PRL (0,48). A produtividade de grãos apresentou baixa ou nenhuma correlação com outros caracteres testados.

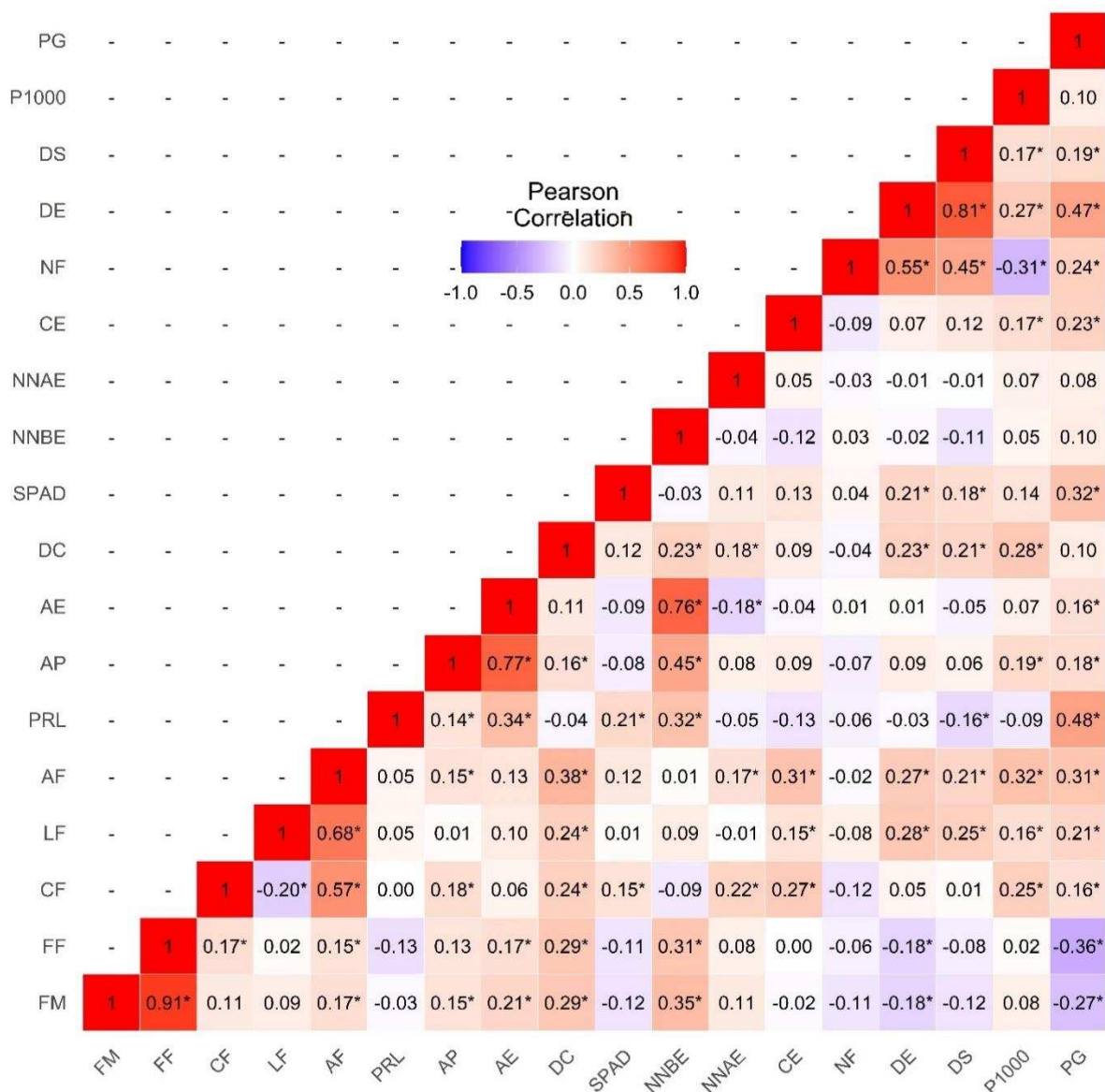


Figura 1- Heatmap de correlações fenotípicas entre 18 caracteres morfoagronômicos. As cores atribuídas a um ponto na grade do Heatmap indica a intensidade de uma correlação particular entre duas características. O nível de correlação é indicado por vermelho para correlações positivas e azul para correlações negativas, conforme representado na chave colorida. FM-florescimento masculino, FF-florescimento feminino, CF-comprimento foliar, LF-largura foliar, AF-área foliar, PRL-prolificidade, AP-altura de planta, AE-altura de espiga, DC-diâmetro de colmo, SPAD-teor de clorofila na folha, NNBE-número de nós abaixo da espiga, NNAE-número de nós acima da espiga, CE-comprimento da espiga, NF-número de fileiras, DE-diâmetro de espiga, DS-diâmetro de sabugo, PG-produtividade de grãos e P1000-peso de 1.000 grãos.

5.3.2. Caracterização molecular

Após o controle de qualidade dos dados, 3.083 SNPs de alta qualidade com frequência de alelo menor (MAF) maior que 0,01 e call rate de 90% foram selecionados para posterior análise de caracterização genética. O número de SNPs apresentou variação

moderada entre os 10 cromossomos do milho e variou de 207 no cromossomo 10 a 494 no cromossomo 1, com uma média de 308 SNPs por cromossomo (Tabela 2). *Missing rate* variou de 0,84% nos cromossomos 3 e 7 a 1,04% no cromossomo 10, com um valor médio de 0,94%, enquanto a taxa de heterozigosidade observada variou de 1,61% no cromossomo 3 a 2,61% no cromossomo 2, com um valor médio de 1,97%. Os valores de PIC variaram de 0,26 nos cromossomos 1 e 6 a 0,35 no cromossomo 8, com média de 0,28. A MAF e a GD apresentaram pouca variação entre os cromossomos, com média de 0,25% e 0,34, respectivamente.

Tabela 2 – Heterozigosidade observada, menor frequência alélica (MAF), diversidade de genes, Missing rate e PIC por cromossomo

Cromossomo	Número de SNPs	Missing rate (%)	MAF (%)	Diversidade de Genes	Heterozigosidade observada (%)	PIC
1	494	0,93	0,24	0.32	2,10	0,26
2	340	1,03	0,25	0.34	2,61	0,27
3	369	0,84	0,25	0.33	1,61	0,27
4	355	0,98	0,26	0.35	1,82	0,28
5	304	0,87	0,25	0.33	1,16	0,27
6	247	0,87	0,24	0.32	2,31	0,26
7	254	0,84	0,25	0.34	2,15	0,35
8	277	0,96	0,27	0.35	2,19	0,28
9	236	1,00	0,24	0.33	1,62	0,27
10	207	1,04	0,26	0.34	2,15	0,27
Média	308,30	0,94	0,25	0.34	1,97	0,28

5.3.3. Parentesco relativo

Os coeficientes de parentesco relativo aos pares entre as 182 linhagens de milho estimadas com 3.083 SNPs variaram de 0 e 2,00, mas sua distribuição entre 0,00 e 0,50 representa 96,34% dos dados (Figura 2). Os resultados mostraram que um total de 10,80% do parentesco relativo aos pares foram iguais a 0; 66,88% deles variaram de 0,00 a 0,10; 48,16% variaram de 0,10 a 0,50, e apenas cerca de 4,00% estavam acima de 0,50. Essas informações de parentesco genético indicam a maioria das linhagens endogâmicas são fracamente ou modestamente relacionadas entre si, e apenas algumas linhas mostraram fortes semelhanças entre elas neste conjunto de linhagens endogâmicas de milho tropical.

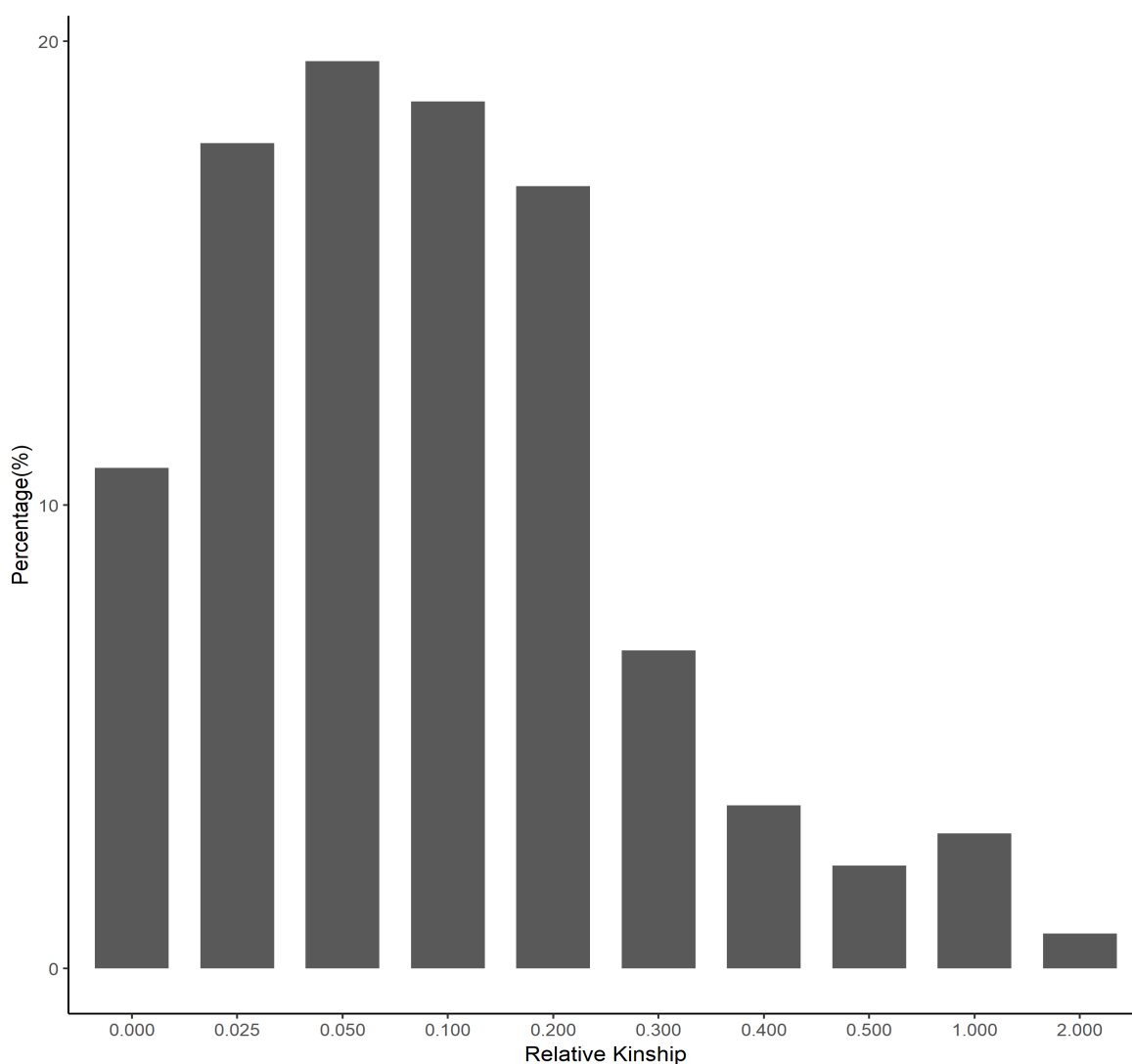


Figura 2 – Distribuição de parentesco relativo para 182 linhagens endogâmica de milho tropical usando 3.083 SNPs.

5.3.4. Desequilíbrio de ligação (LD)

A distância de decaimento LD em todos os dez cromossomos em todo o conjunto de linhagens de milho com $r^2 = 0,2$ foi 86.899 bp e variou de 163 bp (no cromossomo 7) a 102.750 bp (no cromossomo 8) (Fig. 2). A distância de decaimento do LD foi muito semelhante em oito cromossomos, mas um decaimento LD relativamente rápido com uma distância curta foi observado nos cromossomos 7 (163 bp) e 2 (6.244 bp).

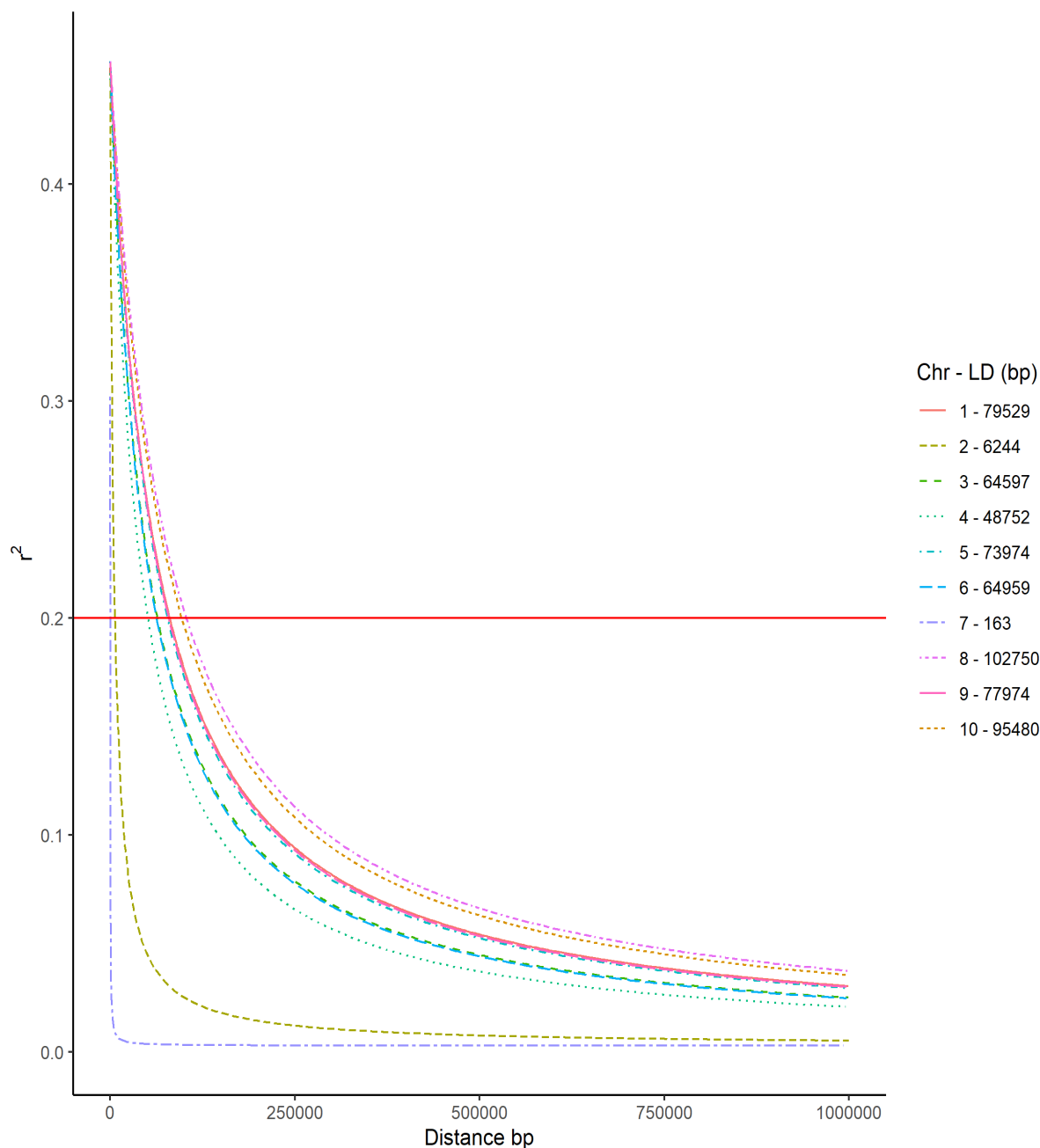
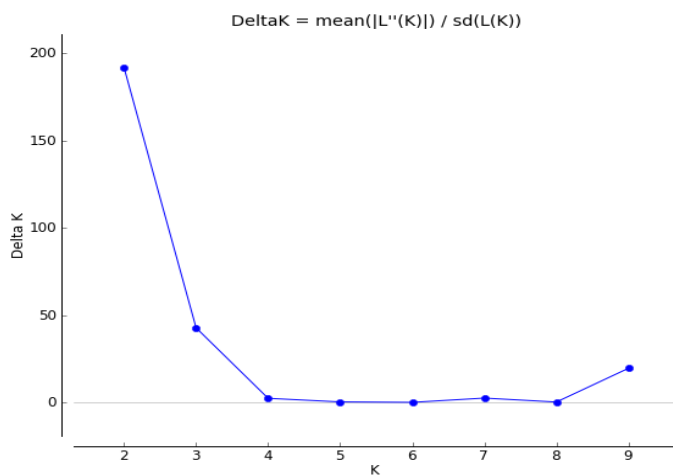


Figura 3 - Decaimento do LD para todos os 10 cromossomos e a distância LD em bp quando $r^2 = 0,2$.

5.3.5. Estrutura populacional

Os resultados da análise de estrutura de população de todas as 182 linhagens de K (número de grupos ou populações) variando de 1 a 10 são apresentados na Figura 4a. A estatística *ad hoc* ΔK apresentou valores de maior verossimilhança quando k foi igual a 2, com decréscimo acentuado quando k aumentou de 2 para 4. No entanto, em $k = 3$, Δk foi significativamente maior do que em $k = 4$ e, portanto, $k = 2$ e $k = 3$ podem ser considerados os melhores números possíveis de subpopulações. Assim, o conjunto de linhagens endogâmicas de milho pode ser dividido em duas ou três subpopulações (Figura 4a). O primeiro nível de agrupamento ($k = 2$) separou as linhagens nas subpopulações 1 e 2, com 141 (77,5%) e 25 (13,7%) linhagens de milho em cada subpopulação, respectivamente, e uma subpopulação mista (probabilidades de adesão $<0,60$) com 16 (8,8%) linhagens (Figura 4b). Enquanto em $k = 3$, as subpopulações 1, 2 e 3 consistiam em 84 (46,2%), 47 (25,8%) e 21 (11,5%) linhagens endogâmicas, respectivamente, e as 30 (16,5%) linhagens restantes tinham probabilidades de adesão $<0,60$ e foram classificadas em uma subpopulação mista (Figura 4c). As linhagens agrupadas nas subpopulações 1, 2 e 3, do agrupamento ($k=3$) são derivadas de diferentes origens, elas vieram de 35, 24 e 13 híbridos comerciais, respectivamente. Embora alguns híbridos de milho tenham contribuído com linhagens para mais de uma subpopulação, em geral, linhagens endogâmicas derivadas do mesmo híbrido e de híbridos diferentes, mas desenvolvidas pela mesma empresa de sementes, tenderam a se agrupar na mesma subpopulação. Assim, é provável que as linhagens pertencentes a mesma subpopulação derivadas dos híbridos da mesma empresa compartilham uma ancestralidade comum (por exemplo, o mesmo pai) e, a partir daí, um parentesco genético. Com relação ao germoplasma de origem, todas as sete linhagens derivadas de Nitroflint, três de quatro de CMS50 e duas de quatro de CMS28 foram agrupadas na subpopulação 1, enquanto as outras linhagens derivadas de CMS50 e CMS28 foram agrupadas na subpopulação 2. As duas linhagens derivadas de BR106 foram agrupadas na subpopulação mista. Os híbridos que mais tiveram linhagens extraídas foram AG4051, DKB435, GARRA, Nitroflint, POINTER, Z8480, 30F87, BALU551, BALU184, CMS28 e CMS50 e todas foram para a subpopulação 1.

a)



(b)



c)

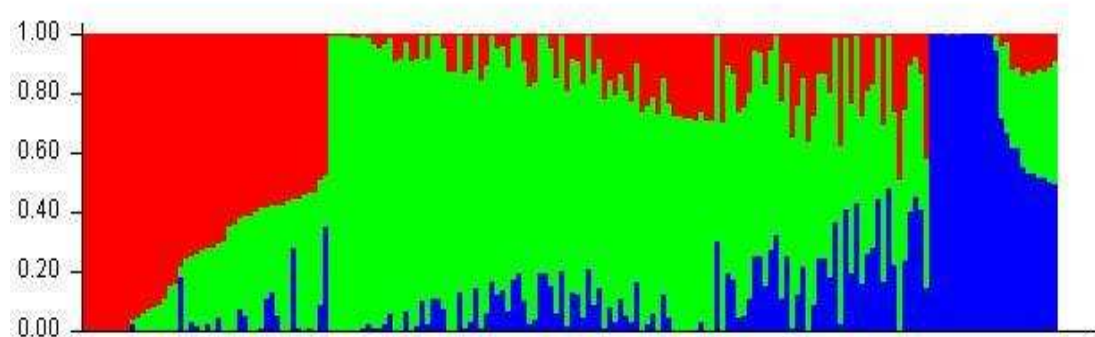


Figura 4 - Estrutura populacional de 182 linhagens endogâmicas de milho estimadas com 3.083 SNPs. (a) valores de Δk (delta K) para K variando de 1 a 10. (b) A estrutura populacional das 182 linhagens para $k=2$ representada por uma única linha vertical que é particionada em segmentos coloridos no eixo x, com comprimentos proporcionais à associação de probabilidade estimada em cada uma das populações inferidas de K (eixo y). As cores vermelha e verde representam as populações 1 e 2, respectivamente.

5.3.6. Análise de componentes principais

A estrutura populacional revelada pela análise de componentes principais (PCA) com base nos SNPs distinguiram claramente as 182 linhagens endogâmicas de milho tropical em três subpopulações (Figura 5). Os primeiros dois componentes principais (PCs) explicaram 7,21% da variação total do SNP em todo o conjunto de linhagens de milho. O agrupamento com base no PCA foi muito consistente com os resultados obtidos pelo STRUCTURE, e as subpopulações 1 (98 linhagens), 2 (54 linhagens) e 3 (30 linhagens) compreenderam todas as linhagens de milho da subpopulação 1, 2 e 3 dos resultados do STRUCTURE, respectivamente. Além disso, as linhagens que foram classificadas como subpopulação mista pelo STRUCTURE foram separadas entre as três subpopulações com base nos resultados de PCA, e 14, sete e nove linhagens alocadas nas subpopulações 1, 2 e 3, respectivamente.

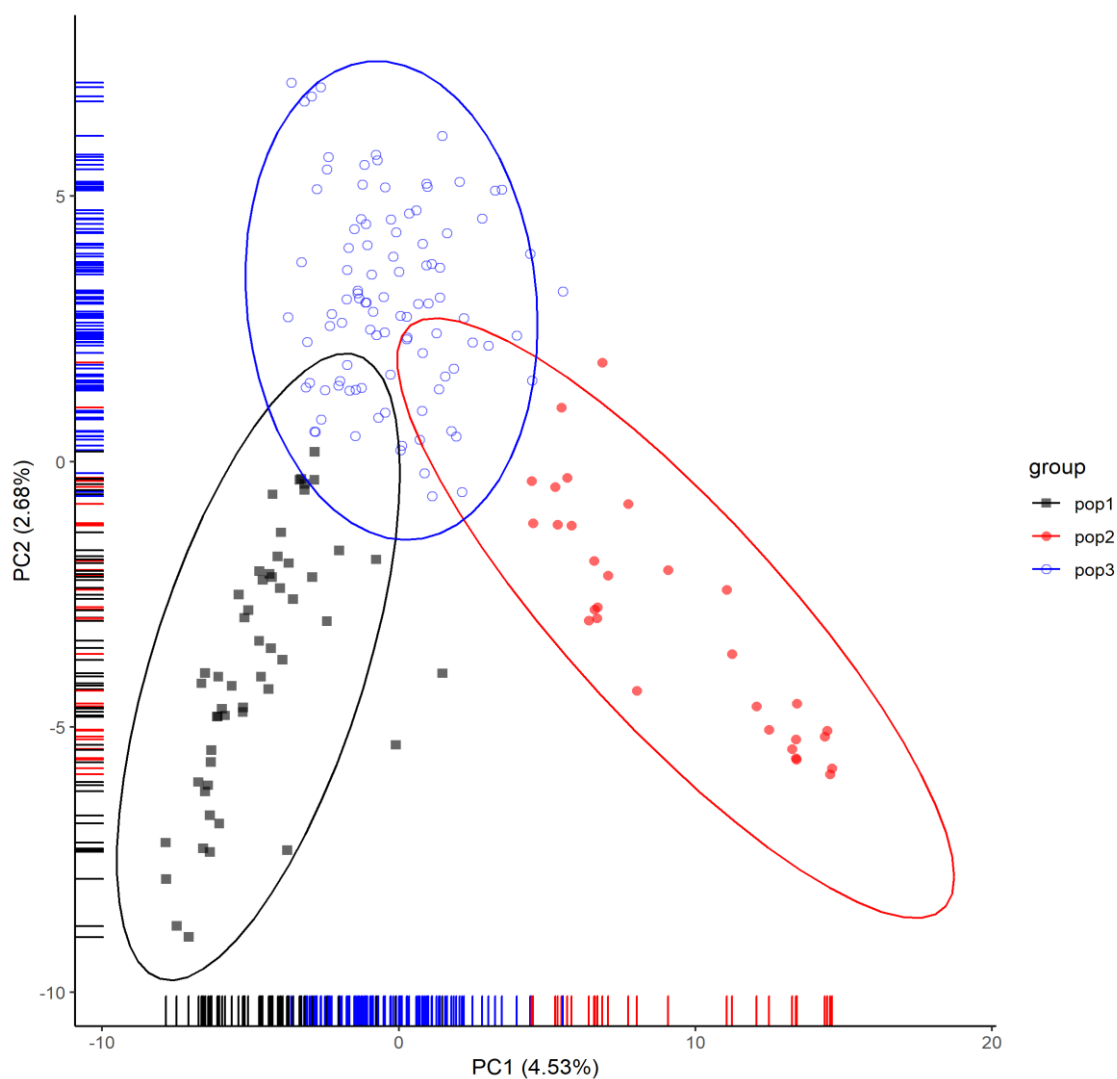


Figura 5 - Associações entre as linhagens reveladas pela análise de componentes principais com base nas semelhanças genéticas. PC1 e PC2 foram o primeiro e o segundo

componente principal que juntos representaram cerca de 7,21% da variação total entre as linhagens.

5.3.7. Diversidade genética com base em dados fenotípicos e dados moleculares

As distâncias genéticas de Nei entre as comparações de pares de todas as 182 linhagens endogâmicas de milho foram calculadas a partir de 3.083 marcadores SNPs (Figura 6). Distância genética média geral foi de 0,40 e variou de 0,003 a 0,48. Com base no método de Mojena (1977), o dendrograma UPGMA gerado a partir da matriz de distância genética de Nei agrupou as 182 linhagens endogâmicas em 14 grupos. A maior distância genética (0,480) foi obtida entre as linhagens endogâmicas VML006 e VML079, as quais foram alocadas nos clusters I e II. Estas linhagens apresentam origens diferentes, elas foram derivadas dos híbridos AG4051 e BALU184, respectivamente. Enquanto as linhagens mais similares foram VML001 e VML036 que apresentaram a menor distância (0,003) e foram alocados no agrupamento I. Curiosamente, essas linhagens foram derivadas de diferentes híbridos de milho (30F87 e AG1051, respectivamente). Os grupos III, I e V, os maiores grupos formados, consistiram de 69, 35 e 28 linhagens, respectivamente. Os grupos VI (VML019 e VML084), VIII (VML030 e VML177), XI (VML072 e VML102), XII (VML086 e VML095), XIII (VML115 e VML139), com duas linhagens cada, e XIV (VML182), com uma linhagem, eram os menores agrupamentos de diversidade genética. Quase todos esses agrupamentos compreendem linhagens endogâmicas derivadas de diferentes híbridos comerciais, exceto que ambas as linhagens pertencem ao agrupamento VIII, que foram derivadas do híbrido P3041. Os resultados do agrupamento com base nos dados de marcadores moleculares foram parcialmente consistentes com os resultados da estrutura populacional, e descobrimos que todas as linhagens endogâmicas agrupadas no agrupamento I foram alocadas na subpopulação 2 com base nos resultados da estrutura de população (STRUCTURE e PCA), e quase todas as linhagens endogâmicas nos agrupamentos IV, V e VII, e todas as linhagens nos agrupamentos VI, X, XI, XII, XIII e XIV foram alocadas na subpopulação 1. No geral, as linhagens endogâmicas derivadas do mesmo híbrido comercial mostraram uma leve tendência a se agrupar nos mesmos grupos. Assim, todas as linhagens derivadas do híbrido BRS1010 (2 linhagens), DKB199 (3), C333 (7) e Z8420 (4) foram alocadas nos grupos III, I, III e III, respectivamente. Além disso, quase todas as linhagens derivadas dos híbridos Balu184 (6 de 9 linhagens), AG8080 (5 de 6), Balu551 (4 de 7), DKB435 (5 de 6) e P30F90 (6 de 7), e de população Nitroflint (5 de 7)

foram agrupados nos mesmos grupos. Em contraste, as 16 linhagens derivadas do híbrido P3041, sete derivadas de P30F87 e seis de Garra foram agrupadas em sete, seis e quatro grupos diferentes, respectivamente.

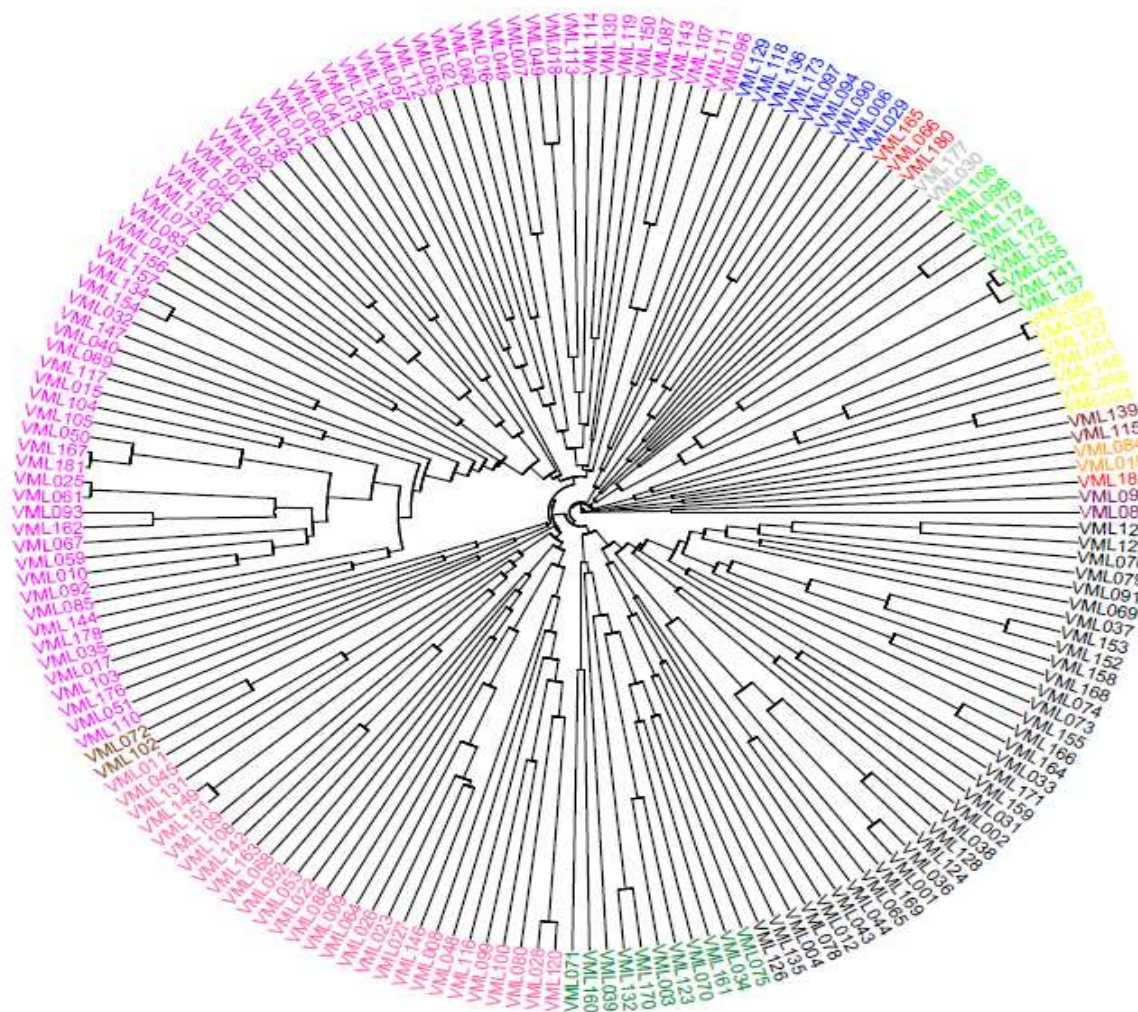


Figura 6- Dendrograma UPGMA usando dados genotípicos. Dendrograma do método de grupos de pares não ponderados de agrupamento aritmético para 182 linhagens endogâmicas de milho a partir das distâncias genéticas de Nei calculadas a partir de 3.083 marcadores SNPs.

Os coeficientes de distância de Mahalanobis foram estimados entre a comparação de pares de todas as 187 linhagens de milho com base em 18 caracteres agrônômicos medidos em quatro ambientes (Figura 7). A distância genética média foi de 61,81, variou de 4,66 (VML001 e VML036) a 219,46 (VML089 e VML1170). Com base no método de Mojena (1977), a análise de agrupamento alocou as 187 linhagens endogâmicas de milho em 22 agrupamentos (ponto de corte = 46,91). As linhagens endogâmicas VML089 e VML170, as linhagens mais distantes, foram alocadas nos grupos I e II, respectivamente, enquanto que as linhagens mais similares VML001 e VML036 foram

alocadas no grupo I. Essas linhagens também apresentaram menor distância e foram alocadas no mesmo grupo com base na análise de agrupamento com base em marcadores moleculares. Os grupos I e II foram os maiores grupos e consistiram em 68 e 28 linhagens, respectivamente, enquanto os grupos XIV (VML043), XVIII (VML097), XIX (VML139), XXI (VML161) e XXII (VML182) consistiram de apenas uma linhagem endogâmica cada. A VML182 também foi alocada sozinha em um grupo com base no agrupamento a partir de marcador molecular. Além disso, os grupos IV (9), V (10), XII (9) e XVI (11) consistiam em aproximadamente dez linhagens de milho, enquanto o número de linhagens em outros grupos variou de duas a sete. Curiosamente, cerca de 50% (35 linhagens) das linhagens no agrupamento I, e todas as linhagens nos agrupamentos VIII (5 linhagens) e XI (7 linhagens) foram alocadas na subpopulação 1 com base nos resultados de PCA e estrutura populacional. Embora quase 50% (32 linhagens) das linhagens no agrupamento I com base nos dados fenotípicos tenham sido alocadas no mesmo agrupamento (III) com base no agrupamento de marcadores moleculares, houve baixa concordância entre o agrupamento de linhagens com base em ambos os conjuntos de dados. Além disso, não observamos correspondência entre germoplasma de origem (híbridos comerciais) e o agrupamento de linhagens de milho com base em dados fenotípicos.

milho da UFV em um painel de associação pode ajudar os melhoristas a aumentar a taxa de ganho genético (Brescaglio e Sorrells, 2006; Inghelandt et al., 2010; Beckett et al., 2017).

Neste estudo, realizamos uma caracterização fenotípica e genotípica de um conjunto de linhagens endogâmicas de milho do banco de germoplasma de milho da UFV que apresentam boa adaptação para a maioria das condições de cultivo em ambientes tropicais brasileiros e são atualmente utilizadas no programa de melhoramento. Encontramos variação genotípica substancial para todos os 18 caracteres medidos entre ambientes, e esses mostraram valores intermediários a altos de herdabilidade, mesmo aqueles que apresentaram uma interação linhagens x ambientes significativa. Esses valores de $(\hat{h}_{\bar{x}}^2)$ refletem que podemos ter estimativas precisas de efeitos genéticos em experimentos de QTL com este conjunto de linhagens endogâmicas, e diferentes associações genômicas podem ser detectadas em ambientes para esses caracteres. A enorme variação fenotípica apresentada para os caracteres agrônômicos neste conjunto de linhagens e os valores de $(\hat{h}_{\bar{x}}^2)$ para alguns caracteres foram semelhantes aos relatados em outros painéis de associação de milho ao redor do mundo (Peiffer et al., 2014; Hansey et al., 2011; Yang et al., 2011; 2010; Flint-Garcia et al., 2005).

O GD médio (0,34) do conjunto de 182 linhagens endogâmicas de milho tropical foi semelhante aos valores obtidos por Zhang et al. (2016) e Wu et al. (2014) GD em torno de 0,35 para dois conjuntos de linhagens de milho e superior ao GD (variando de 0,22 a 0,32) observado em quatro conjuntos de linhagens endogâmicas (Wu et al., 2016; Lu et al., 2009; Hansey et al., 2011, Inghelandt et al., 2010; Hamblin et al., 2007) consistindo de germoplasma de milho tropical e temperado, mas abaixo dos valores de GD de 0,39 encontrados por Lu et al. (2011) em um conjunto de 527 linhagens de milho. No entanto, as linhagens endogâmicas tiveram uma diversidade de genes mais alta do que a diversidade genética demonstrada por um conjunto de 94 linhagens de milho brasileiras adaptadas às condições tropicais (GD = 0,27; Lu et al., 2009) e 94 linhagens de milho tropical de maturação precoce do Oeste e África Central (GD = 0,22; Adu et al., 2019). Além disso, embora a maioria das linhagens avaliadas fossem derivadas de híbridos comerciais, elas mostraram uma diversidade genética muito maior do que um grupo heterótico chinês (GD = 0,21), que consiste em linhagens puras essencialmente derivadas de híbridos modernos dos EUA (Wu et al., 2011). Assim, pode-se dizer que a abundante diversidade genética apresentada pelas 182 linhagens endogâmicas se deve

principalmente à ampla base genética entre os híbridos comerciais usados como germoplasma fonte dessas linhagens. Esses híbridos foram desenvolvidos por várias empresas de sementes, e há uma grande variabilidade genética entre o germoplasma de milho usado em seu programa de melhoramento comercial (Andrade et al., 2016).

O valor PIC fornece uma estimativa do poder discriminatório de um marcador e é sinônimo de diversidade genética. O PIC pode ser calculado para indivíduos resultantes de um grande número de autofecundações, em que há poucos heterozigotos e em que a variação se revela na existência de homozigotos para diferentes alelos. No presente estudo, os valores de PIC variaram de 0,02 a 0,38 a média do PIC foi de 0,27, pode-se inferir que as 182 linhagens analisadas com os marcadores SNPs apresentam elevada variabilidade genética. Os valores encontrados são semelhantes aos descrito por outros autores. Wu et al. (2016) encontraram valores de PIC variando de 0,09 a 0,38, com média de 0,25. Para Nyombaryire et al. (2016), o alto valor de PIC revelado em seu estudo pode ser uma indicação relevante do potencial dos marcadores SNPs para discriminar as linhagens endogâmicas de origens diversas.

Os baixos coeficientes de parentesco entre as 182 linhagens indicam claramente a falta de linhagens redundantes no germoplasma estudado e a singularidade genética da maioria desses acessos no banco de germoplasma da UFV. Isso sugere que cada linhagem neste conjunto de linhagens endogâmicas de milho tropical tem potencial para contribuir com novos alelos para nosso programa de melhoramento. Nossos coeficientes de parentesco foram semelhantes aos de Wu et al. (2014) entre 367 linhagens amplamente utilizadas no melhoramento do milho na China, Wu et al. (2019) entre 226 acessos derivados de recursos tropicais e temperados, e Ertiro et al. (2017) entre 265 linhagens amplamente utilizadas no programa de melhoramento de milho da Etiópia, e menores do que os coeficientes de parentesco encontrados por Semagn et al (2012), que relataram que 79% dos coeficientes de parentesco variaram de 0,05 a 0,50 entre 450 linhagens do CIMMYT. No entanto, os coeficientes de parentesco entre as 182 linhagens avaliadas nesse estudo foram um pouco maiores do que os coeficientes de parentesco relatados por Wu et al. (2016) entre as 538 linhagens de milho que representam a coleção CIMMYT em todo o mundo, Yan et al. (2009) entre 632 linhagens de programas de melhoramento público temperado, tropical e subtropical, e Zhang et al. (2016) entre 362 linhagens do programa de melhoramento da China. Nesses três estudos anteriores, eles descobriram que cerca de 60% dos coeficientes de parentesco entre pares de linhagens eram iguais a

zero, enquanto neste estudo apenas 10,80% deles eram iguais a zero. Em contraste com esses autores, avaliamos um conjunto menor de linhagens de milho tropical desenvolvidas e liberadas exclusivamente pelo programa de melhoramento de milho da UFV, além disso, algumas linhagens foram derivadas do mesmo híbrido comercial. E apresentam assim, uma divergência genética mais estreita do que as coleções de milho usadas nos estudos anteriores.

O declínio de LD médio foi de 86.899 pb em todo o conjunto de 182 linhagens de milho tropical com $r^2 = 0,2$, valor menor do que 391 kb obtidos em um painel com 367 linhagens endogâmicas (Wu et al., 2014). Lu et al. (2011) analisaram um total de 287 linhagens tropicais e 160 linhagens temperadas que foram genotipadas com 1.943 SNPs e comparadas quanto ao declínio de LD. Eles encontraram significativa variação no decréscimo do LD com o aumento da distância física (2-100kb) em todo o genoma. Em média o decréscimo de LD no germoplasma temperado (10-100kb) foi de duas a dez vezes maior do que no germoplasma tropical (5-10 kb). Os autores concluíram que o milho tropical possui um rápido declínio do LD que fornece maiores possibilidades para a seleção. No entanto, a distância média de declínio de LD neste estudo foi maior do que os 41,15 kb em 226 acessos de milho de regiões tropicais e temperadas (Wu et al., 2019), e muito maiores do que 5–10 kb estimados em uma coleção global de milho altamente diversa de 632 linhagens endogâmicas (Yan et al., 2009), e 3,76 kb na coleção global CIMMYT de 538 linhagens de milho (Wang et al., 2016). Como pode ser observado, os estudos mencionados acima avaliaram as extensões de LD em grandes coleções de linhagens de milho que são muito mais diversas e provavelmente contêm mais alelos raros do que as 182 linhagens deste estudo. No geral, o declínio do LD é muito afetada pela composição do germoplasma de origem e pela intensidade da seleção durante o processo de melhoramento (Beckett et al., 2017; Gupta et al., 2005; Flint-Garcia et al., 2003), e o decréscimo do LD tende a ser consideravelmente maior no germoplasma de melhoramento do que em coleções de germoplasma mais diversas, uma vez que em programas de melhoramento as populações sofrem uma alta pressão de seleção (Leng et al., 2019; Inghelandt et al., 2011). Leng et al. [31] (2019) encontraram um declínio de LD médio de 1.050 kb em um painel de 157 linhagens de criação avançadas do programa da China. As 182 linhagens avaliadas foram derivadas de um conjunto de híbridos comerciais brasileiros que foram previamente formados por linhagens puras elite com vários anos de seleção para adaptação aos ambientes tropicais brasileiros. Além disso, a

maioria delas foram selecionadas para eficiência no uso de nitrogênio, caracteres de raiz com estresse de nitrogênio e fósforo e caracteres agronômicos em condições tropicais de campo (Torres et al., 2019; 2018; Rodrigues et al., 2017) e, portanto, alguns alelos raros podem ser perdidos durante o processo de seleção. No nível cromossômico, descobrimos que o declínio do LD com o aumento da distância física para os cromossomos 7 (533 bp) e 2 (18.450 bp) foi muito mais rápido do que o declínio do LD para os outros cromossomos (cujo decréscimo variou de 48.752 a 102.750 bp). Esta variação também foi relatada em outros estudos com germoplasma de milho, e pode ser causada por diferentes padrões de LD entre os cromossomos e indica que a faixa de pressão de seleção em diferentes cromossomos foi devido aos diferentes objetivos no melhoramento do milho (Zhang et al., 2016; Wu et al., 2014; Lu et al., 2011; Yan et al., 2010).

A maioria das linhagens endogâmicas de milho tropical usadas nesse estudo foram derivadas de híbridos comerciais e ainda não há nenhuma informação heterótica sobre elas. Nesse cenário, os marcadores moleculares obtidos por meio da análise da estrutura populacional podem ser usados como uma abordagem alternativa para alocar as linhagens em grupos heteróticos (Leng et al., 2019; Beckett et al., 2017; Wu et al., 2016; Olmos et al., 2013; Viana et al., 2013a; Inglelandt et al., 2010; Lu et al., 2009; Reif et al., 2003ab). A estrutura populacional definida pelos marcadores moleculares parece ser mais plausível para o agrupamento heterótico do que os métodos tradicionais baseados em estudos de linhagem e capacidade de combinação (Adu et al., 2019; Ertiro et al., 2017; Wu et al., 2016). Assim, a estrutura populacional de um conjunto de 182 linhagens endogâmicas de milho tropical, a maioria delas derivadas de híbridos comerciais, foi realizada usando os métodos baseados em modelo STRUCTURE e PCA baseados em 3.083 SNPs, e ambos os métodos distinguiram as 182 linhagens em três subpopulações com alta consistência entre seus resultados. O agrupamento de linhagens endogâmicas de milho tropical em três grandes grupos heteróticos também foi proposto em outros estudos, uma vez que devido aos altos níveis de diversidade no germoplasma de milho tropical e a origem mista das populações usadas para desenvolver linhagens, os padrões heteróticos em milho ainda não está claro em comparação com os padrões heteróticos bem definidos apresentados em milho temperado (Ertiro et al., 2017; Dinesh et al., 2016; Wu et al., 2016; Segman et al., 2012). Adu et al. (2019) atribuíram 94 linhagens endogâmicas de milho tropical derivadas de 12 populações com origens mistas em três grupos heteróticos. A estrutura populacional presente no conjunto de linhagens endogâmicas avaliado e sua atribuição

em três grupos ajudará a entender como utilizar essas linhagens endogâmicas para escolher testadores, atribuir grupos heteróticos e, conseqüentemente, desenvolver híbridos de alto rendimento. Assim, linhagens alocadas no mesmo grupo e com bom desempenho fenotípico podem ser recombinadas para desenvolver populações melhoradas: uma população sintética e/ou várias populações biparentais. Estas podem ser usadas como fonte de germoplasma para o desenvolvimento de novas linhagens, e as populações sintéticas podem ser usadas como populações base para iniciar um programa de seleção recorrente recíproca entre sintéticos de diferentes subpopulações (Kolawole et al., 2017; Viana et al., 2013; Dona et al., 2012; Romay et al., 2011; Hinze et al. 2005). Por outro lado, as linhagens alocadas em um grupo podem ser cruzadas diretamente com linhagens de outro grupo para desenvolver híbridos altamente heteróticos e produtivos para as condições tropicais brasileiras.

Análise de agrupamento com base em dados fenotípicos e moleculares, deste estudo, mostraram níveis altamente variáveis de distância genética entre pares de linhagens de milho com um grande número de agrupamentos de divergência genética (14 e 22 agrupamentos de dados moleculares e fenotípicos, respectivamente). Este resultado concorda com a baixa relação entre as linhagens endogâmicas obtidos pelo coeficiente de parentesco neste estudo e sugere a presença de uma baixa porcentagem de linhagens endogâmicas redundantes que pouco contribuem para a variação genotípica e divergência genética no germoplasma de milho do programa de melhoramento da UFV. A baixa correlação entre o agrupamento com base em dados fenotípicos e moleculares observada concorda com os resultados de estudos anteriores em milho (Hartlings et al., 2008; Reboug et al., 2003), e não deve ser considerada como uma limitação para avaliar a diversidade genética entre acessos de reprodução em um banco de germoplasma. De fato, a avaliação da diversidade genética entre os acessos e seu agrupamento com base em dados fenotípicos fornecem informações adicionais que geralmente são independentes das informações genotípicas, e é uma ferramenta complementar que deve ser combinada com dados moleculares para auxiliar nos esforços do programa de melhoramento (Ajala et al., 2020; Mundim et al., 2015; Wen et al., 2012; Silva et al., 2017; Wen et al., 2012; Geleta et al., 2006). Além disso, de acordo com Hansey et al. (2011), um painel de linhagens de milho destinado ao mapeamento de associação deve utilizar o máximo de diversidade fenotípica possível, e a análise de agrupamento mostrou que as linhagens de milho do programa de melhoramento da UFV são amplamente diversas. Para Singh et al.

(1991), a melhor maneira de identificar divergências entre genótipos é o uso combinado de dados fenotípicos e moleculares.

Em contraste com o agrupamento baseado em dados fenotípicos, encontramos uma concordância moderada entre a análise de agrupamento de marcadores moleculares e os resultados da estrutura populacional em termos de membros de cada grupo. Todas as linhagens endogâmicas alocadas no grupo 1 pertencem à subpopulação 2, e o maior número de grupos de divergência obtido pela análise de agrupamento (14) em comparação com as três subpopulações foi principalmente devido à subdivisão da subpopulação 1 em quase uma dúzia de grupos de divergência. Embora o padrão de agrupamento da análise da estrutura populacional pareça ser mais confiável do que a análise de agrupamento molecular para o estabelecimento de grupos heteróticos (Leng et al., 2019; Segman et al., 2012), a avaliação da diversidade genética é uma abordagem importante que pode ajudar os melhoristas a selecionar os cruzamentos mais promissores em um programa de melhoramento de milho, uma vez que o desempenho dos híbridos é maior quando os pais são geneticamente divergentes (Laude e Carena, 2015; George et al., 2011; Makumbi et al., 2011; Reif et al., 2005). De acordo com Silva et al. (2020), distâncias genéticas baseadas em SNP fornecem informações importantes para selecionar as melhores combinações parentais, e evita cruzamentos entre linhagens endogâmicas de milho semelhantes dentro ou entre grupos heteróticos. Nesse estudo, linhagens endogâmicas de milho tropical que pertencem a um grupo heterótico definido pela análise da estrutura populacional devem ser cruzadas com linhagens endogâmicas de outros grupos, e o melhorista de milho deve priorizar o cruzamento entre linhagens endogâmicas mais divergentes que apresentam características agrônômicas desejáveis. Por outro lado, as linhagens menos divergentes dentro de cada subpopulação devem ser cruzadas em pares para produzir populações biparentais que serão usadas para o desenvolvimento de novas linhagens de milho tropical (Hallauer et al., 2010).

5.5. CONCLUSÃO

Neste estudo, encontramos grande variação fenotípica e genotípica no conjunto de linhagens endogâmicas de milho tropical do programa de melhoramento de milho da UFV que pode ser explorada para fins de melhoramento de milho. Foram identificados três grupos heteróticos que, juntamente com a análise de agrupamento com base nos

marcadores moleculares, serão úteis para guiar cruzamentos em nosso programa de melhoramento de milho para selecionar testadores, desenvolver populações de melhoramento e de mapeamento e desenvolver híbridos altamente heteróticos e produtivos para ambientes tropicais brasileiros. Os resultados também mostraram que há baixos coeficientes de parentesco e alta distância genética (com base em dados fenotípicos e genotípicos) entre a maioria dos pares de linhagens endogâmicas, juntamente com curta distância de decaimento de LD em todo o conjunto dessas linhagens. Isso indica claramente a singularidade da maioria das linhagens endogâmicas que podem contribuir com novos alelos para o programa de melhoramento da UFV e, além disso, podem compreender um painel de linhagens endogâmicas de milho tropical que é adequado para o estudo de associação genômica ampla de caracteres quantitativos complexos em milho, principalmente em ambientes tropicais.

5.6. REFERÊNCIAS BIBLIOGRÁFICAS

- Adu, G.B.; Badu-Apraku, B.; Akromah, R; Garcia-Oliveira, A. L.; Awuku, F. J.; Gedil, M. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *Plos One*. p.1-12, 2019.
- Andrade LRB, Fritsche Neto R, Granato ISC, Sant’Ana GC, Morais PPP, Borém A. Genetic Vulnerability and the Relationship of Commercial Germplasms of Maize in Brazil with the Nested Association Mapping Parents. *PLoS ONE*, 2016,11(10): e0163739.
- Balding, D. J.; Nichols, R. A. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, v.96, p. 3–12, 1995.
- Barata, C.; Carena, M.J. Classification of North Dakota maize inbred lines into heterotic groups based on molecular and testcross data *Euphytica*, v.151, p.339–349, 2006.
- Bradbury, P. J.; Zhang, Z.; Kroon, D. E.; Casstevens, T. M.; Ramdoss, Y.; Buckler, E. S. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatic*, v.1, p.2633-2635, 2007.

- Bracco, M.; Lia, V. V.; Gottlieb, A. M.; Hernández, J. C.; Poggio, L. Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina. *Genetica*, v.135(1), 39-49, 2009.
- Beckett TJ, Morales AJ, Koehler KL, Rocheford TR. Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds. *PLoS ONE*, 2017, 12(12):0189277.
- Bertan I, Carvalho FIF, Oliveira AC. Parental selection strategies in plant breeding programs. *J Crop Sci Biotechnol*. 2007, 10: 211-222.
- Breseghele F, Sorrells ME. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics*, 2006, 172(2):1165-1177.
- Cantelmo NF, Von Pinho RG, Balestre M. Genome-wide prediction for maize single-cross hybrids using the GBLUP model and validation in different crop seasons *Molecular Breeding*, 2017, 37 (51)
- Chagné, D.; Batley, J.; Edwards, D.; Forster, J. W.; Single nucleotide polymorphism genotyping in plants, p.77-94. In: *Association mapping in plants*. Springer, 2007.
- Crossa, J.; Vasal, S.K.; Beck, D.L. Combining ability estimates of CIMMYT'S tropical late yellow maize germplasm. *Maydica*, v.35:273-278, 1990.
- Cruz, C. D. GENES: software para análise de dados em estatística experimental e em genética quantitativa. *Acta Sci., Agron.* [online]. 2013, vol.35, n.3, pp.271-276.
- DeLima RO, Borem, A. *Melhoramento de Milho*. Editora UFV, Viçosa, 2018
- Dray, S.; Béatrice, A.; Thioulouse, J.; Jombart, T.; Pavoine, S.; Lobry, J. R.; Ollier, S.; Borcard, D.; Legendre, P.; Bougeard, S.; Siberchicot. *Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences*. v.1.7-16, 2020.
- Dias KOG, Piepho HP, Guimarães LJM, Guimarães PEO, Parentoni, SN, Pinto MO, Noda RW, Magalhães JV, Guimarães CT, Garcia, AAF, Patina MM. Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. *Theoretical and Applied Genetics*, 2020,133:443–455.

- Dinesh A, Patil A, Zaidi P, Kuchanur PH, Vinayan MT, Seetharam K. Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study. *Maydica*, 2016;61.3
- Earl, D. A.; VonHoldt, B. M. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour*, v.4, p.359–361, 2012.
- Ertiro BT, Semagn K, Das B, Olsen M, Labuschagne M, Worku M, Wegary D, Azmach G, Ogugo V, Keno T, Abebe B, Chibsa T, Menkir A. Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. *BMC Genomics*, 2017; 18:777.
- Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. 2005.
- Fan, J. B. et al. Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*, v. 68, p. 69-78, 2003. ISSN 0091-7451. Disponível em: <<Go to ISI>://WOS:000222969300010 >.
- Franco, J.; Crossa, J.; Ribaut, J.M.; Bertran, J.; Warburton, M.L.; Khairallah, M. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. *Theor Appl Genet.*; v.103, p.944±952, 2001.
- Flint-Garcia, S. A.; Thornsberry J.M.; Buckler, I.V.E.S. Structure of Linkage Disequilibrium in Plants. *Plant biol.* V.54, p. 357– 374, 2003.
- Flint-Garcia, S. A.; Thuillet, A. C.; Yu, J.; Pressoir, G.; Romero, S. M.; Mitchell, S. E.; Doebley, J.; Kresovich, S.; Goodman, M. M.; Buckler, E. S. Maize association population: a high-resolution platform for quantitative trait locus dissection. *The Plant Journal*, v.44, p.1054-1064, 2005.
- Ganal, M. W.; Durstewitz, G.; Polley, A.; Bérard, A.; Buckler, E. S.; Charcosset, A.; Clarke, J. D.; Graner, E. M.; Hansen, M.; Joets, J.; Paslier, M. C. L.; McMullen, M. D.; Montalent, P.; Rose, M.; Schon, C. C.; Sun, Q.; Walter, H.; Martin, O. C.;

- Falque, M. A Large Maize (*Zea mays* L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. *Plos one*, v.6, e28334, 2011.
- Gore, M. A.; Chia, J. M.; Elshire, R. J.; Sun, Q.; Ersoz, E. S.; Hurwitz, B. L.; Peiffer, J. A.; McMullen, M. D.; Grills, G. S.; Ross-Ibarra, J.; Ware, D. H.; Buckler, E. S. A First-Generation Haplotype Map of Maize. *Science*, v. 326, 2009.
- Gupta, P. K.; Rustgi, S.; Kulwal, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology*, v.57, p.461-485, 2005.
- Guimarães LJM, Trindade RS, Parentoni, SN, Guimarães, PEO. Desenvolvimento de linhagens. In: DeLima RO, Borém A. (Eds.). *Melhoramento de Milho*. Viçosa, MG: Editora UFV, 2018:396p.
- Hallauer AR, Carena MJ, Miranda JB. *Quantitative genetics in maize breeding*. 3rd ed. Iowa: Iowa State University Press/Springer Science, 2010.
- Hallauer AR, Carena MJ. Maize. In: Carena JM, editor. *Cereals*. New York, NY: Springer US; 2009. p. 3–98.
- Hartings, H.; Berardo, N.; Mazzinelli, G.F.; Valoti, P.; Verderio, A.; Motto, M. Assessment of genetic diversity and relationships among maize (*Zea mays* L.) Italian landraces by morphological traits and AFLP. *Theor Appl Genet*, v.117, p.831–842, 2008.
- Henderson, C.R. *Applications of linear models in animal breeding*. Guelph: University of Guelph, 1984. 462p.
- Jones, E.; Chu, W.C.; Ayele, M.; Ho, J.; Bruggeman, E.; Yourstone, K.; Rafalski, A.; Smith, O. S.; McMullen, M.D.; Bezawada, C.; Warren, J.; Babayev, J.; Basu, S.; Smith, S. Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.) germplasm. *Mol Breeding*, v.24, p.165–176, 2009.

- Kopelman, N. M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N. A.; Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, v.15, p. 1179-1191, 2015.
- Leng Y, Lu C, Li, Xiang Y, Xia, Wei R, Rong T, Lan H. Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). *Mol Breeding*, 2019, 39, 38.
- Lu Y, Yan J, Guimarães CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek, BS, Magorokosho C, Mugo S, Parentoni SN, Shah T, Rong T, Crouch JH, Xu Y. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theoretical and Applied Genetics*, 2009, 120:93–115.
- Lu, Y.; Shah, T.; Hao, Z.; Taba, S.; Zhang, S.; Gao, S.; Liu, J.; Cao, M.; Wang, J.; Prakash, A. B.; Rong, T.; Xu, Y. Comparative SNP and Haplotype Analysis Reveals a Higher Genetic Diversity and Rapider LD Decay in Tropical than Temperate Germplasm in Maize. *Plos one*, 6(9): e24861, 2011.
- Liu, K.; Goodman, M.; Muse, S.; Smith, S.; Buckler, E.; Doebley, J. Genetic Structure and Diversity Among Maize Inbred lines as Inferred from DNA Microsatellites. *Genetics*, v.165, p.2117-2128, 2003.
- Mantel, N. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res.* v.27, p.209±220, 1967.
- Mahalanobis, P. C. On the generalized distance in statistics. *Proceedings of the National Institute of Science of India.* v.12, p.49–55, 1936.
- Masuka B, Magorokosho C, Olsen M, Atlin GN, Bänziger M, Pixley KV, Vivek BS, Labuschagne M, Matemba-Mutasa R, Burgueño J, Macrobert J, Prasanna BM, Das B, Makumbi D, Tarekegne A, Crossa J, Zaman-Allah M, van Biljon A, Cairns JA. Gains in Maize Genetic Improvement in Eastern and Southern Africa: II. CIMMYT Open-Pollinated Variety Breeding Pipeline. *Crop Sci.*, 2017, 57:180–191.

- Mendiburu, F. *Statistical Procedures for Agricultural Research*, v.1.3-3, 2020.
- Mojena, R. Hierarchical grouping methods and stopping rules: an evaluation. *Computer Journal*, v.20, p.359–363, 1977.
- Montgomery, D.C.; Peck, E.A. *Introduction to linear regression analysis*. New York: J. Wiley, 1981. 504p.
- Morosini, J. S.; Mendonça, L. F.; Lyra, D. H.; Galli, G.; Vidotti, M. S.; Fritsche-Neto, R. Association mapping for traits related to nitrogen use efficiency in tropical maize lines under field conditions. *Plant Soil*, v.421:453–463, 2017.
- Mundim, G. B.; Viana, M. S.; Maia, C.; Paes, G. P.; Delima, R. O.; Valente, M. S. F. Inferring tropical popcorn gene pools based on molecular and phenotypic data. *Euphytica*, v.202, p. 55-68, 2015.
- Nei, M. Genetic Distance between Populations. *The American Naturalist*, v.106, p. 283-292, 1972.
- Nyombayire, A.; Derera, J.; Sibiya, J.; Gasura, E.; Ngaboyisonga, C. Genetic diversity among maize inbred lines selected for the midaltitudes and highlands of Rwanda. *Maydica*, v.61, 2016.
- Olmos, S. E.; Delucchi, C.; Ravera, M.; Negri, M. E.; Mandolino, C.; Eyhéribide, G. H. Genetic relatedness and pop structure within the Argentinean maize inbred lines. *Maydica*, v. 59, p.16-31, 2014.
- Paliwal RL, Granados G, Lafitte HR, Violic AD, Marathée JP. *Tropical maize: improvement and production*, 2000 pp.374 pp.
- Pearson, K. On lines and planes of closet fit to systems of points in space. *Philosophical Magazine*, v.2, p.559-572, 1901.
- Pritchard, J. K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, v.155, p.945-959, 2000.

- R Core Team. The R project for statistical computing- R version 3.1.1. Software. [Internet]. 2019. Available: <http://www.r-project.org>
- Rodrigues, M.C.; Rezende, W.M.; Silva, M.E.J.; Faria, S.V.; Zuffo, L.T.; Galvão, J.C.C.; DeLima R.O. Genotypic variation and relationships among nitrogen-use efficiency and agronomic traits in tropical maize inbred lines. GRM, 2017.
- Rafalski, A. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, v.5, p.94-100, 2002.
- Rana, M.K.; Singh, V.P.; Bhat, K.V. Assessment of genetic diversity in upland cotton (*Gossypium hirsutum* L.) breeding lines by using amplified fragment length polymorphism (AFLP) markers and morphological characteristics. *Genet Resour Crop Evol.* v.52:989±997, 2005.
- Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK: Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Sci.* 2003, 43:1275-1282.
- Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK: Use of SSRs for establishing heterotic groups in subtropical maize. *Theor Appl Genet.* 2003, 107: 947-957.
- Rosenberg, N. A. *distruct*: a program for the graphical display of population structure. *Molecular Ecology Resources*, v.4, p.137-138, 2004.
- Segman, K.; Magorokosho, C.; Vivek, B. S.; Makumbi, D.; Beyene, Y.; Mugo, S.; Prasanna, B. M.; Warburton, M. L. Molecular characterization of diverse CIMMY maize lines using SNP. *BMC Genomics*, v.13, 2012.
- Silva, M. J., Pastina, M. M.; Souza, V. F.; Schaffert, R. E.; Carneiro, P. C. S.; Noda, R. W.; Carneiro, J. E. S.; Damasceno, C. M. B.; Parrella, R A. C. Phenotypic and molecular characterization sorghun accessions for bionergy. *Plos One*, 2017.

- Silva LM, Oliveira LR, Rodrigues F, Benett KSS, Benett CGS. Relative chlorophyll index of maize and soybean according to weather conditions at different times of day. *Revista de Agricultura Neotropical (Journal of Neotropical Agriculture)*. 2020,7(2).
- Singh, Y. P.; Kumar, A.; Chauhan, B. P. S. Genetic divergence in pearl millet. *The Indian J. of Genet and Plant Breed*, v.41, p.186-190, 1981.
- Singh, S.P.; Gutierrez, J. A.; Molina, A.; Urrea, C.; Gepts, P. Genetic Diversity in Cultivated Common Bean: II. Marker-Based Analysis of Morphological and Agronomic Traits. *Crop Science*, v.31, 1991.
- Schlotterer, C. The evolution of molecular markers - just a matter of fashion? *Nat Rev Genet*. v.5, p.63-69, 2004.
- Smith, S. D.; Murray, S. C.; Heffner, E. Molecular analysis of genetic diversity in a Texas maize breeding program. *Maydica*, v.60, 2015.
- Torres, L.G.; Rodrigues, M.C.; Lima, N. L.; Trindade, T.F.H.; Silva, F. F.; Azevedo, C. F.; DeLima, R. O. Multi-trait multi-environment Bayesian model reveals G x E interaction for nitrogen use efficiency components in tropical maize. *Plos One*, June 27, 2018.
- Torres, L.G.; Caixeta, D.G.; Rezende, W. M.; Schuster, A.; Azevedo, C. F.; Silva, F. F.; DeLima, R. O. Genotypic variation and relationships among traits for root morphology in a panel of tropical maize inbred lines under contrasting nitrogen levels. *Euphytica*, 215:51, 2019.
- Thornsberry, J.M.; Goodman, M.M.; Doebley, J.; Kresovich, S.; Nielsen, D. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.*,2001,28:286–89.
- Van Inghelandt, D.; Melchinger, A. E.; Lebreton, C.; Stich, B. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet*, v.120, p.1289-1299, 2010.

- Viana, J.M.S.; Valente, M.S.F.; Silva, F.F.; Mundim, G.B.; Paes, G.P. Efficacy of population structure analysis with breeding populations and inbred lines. *Genetica*. v.141, p.389–399, 2013.
- Viana JMS, Pereira HD, Mundim GB. Efficiency of genomic prediction of non-assessed single crosses. *Heredity (Edinb)*, 2018;120:283-295.
- Viana JMS, Pereira HD, Piepho HP, Silva FF. Efficiency of Genomic Prediction of Nonassessed Testcrosses. *Crop Breeding & Genetics*, 2019, 59:2020–2027.
- Viana JMS, Risso LA, Delima RO, Silva FF. Factors affecting heterotic grouping with cross-pollinating crops. *Agronomy Journal*, 2020.
- Wang, R.; Yu, Y.; Zhao, J.; Shi, Y.; Song, Y.; Wang, T., Li, Y. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor Appl Genet*, v.117, p.1141–1153, 2008.
- Warburton, M.L., Xianchun, X., Crossa, J., Franco, J., Melchinger, A.E., Frisch, M., Bohn, M. and Hoisington D. Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci*. v.42, p.1832-1840, 2002.
- Wen, W.; Franco, J.; Chavez-Tovar, V. H.; Yan, J.; Taba, S. Genetic Characterization of a Core Set of a Tropical Maize Race Tuxpenõ for Further Use in Maize Improvement. *Plos one*, v.7, 2012.
- Whitt, S.R.; Buckler, E.S. Using natural allelic diversity to evaluate gene function. In: Grotewold E, editor. *Plant Functional Genomics*. New Jersey: Humana Press; 2003.
- Wu, X.; Li, Y.; Shi, Y.; Song, Y.; Wang, T.; Huang, Y.; Li, Y. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theor Appl Genet* 127:621–631, 2014.
- Yan, J.; Shah, T.; Warburton, M.; Buckler, E.S.; McMullen, M.D.; Crouch, J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One*, v.4, e8451, 2009.

- Yang, X.; Gao, S.; Xu, S.; Zhang, Z.; Prasanna, B. M.; Li, L.; Li, J.; Yan, J.
Characterization of a global germplasm collection and its potential utilization for
analysis of complex quantitative traits in maize. *Mol Breeding*, v.28, p.511-526,
2011.
- Zhu C, Gore M, Buckler IV, Yu J. Status and prospects of association mapping in
plants, *The Plant Genome*, 2008,1:5-20.

6. CAPÍTULO 2: Estudo de associação genômica e análise da predição genômica para caracteres de raiz relacionados a eficiência no uso de nitrogênio em um painel de linhagens de milho tropical

RESUMO

O sistema radicular é importante para a ancoragem das plantas ao solo e na aquisição e absorção de água e nutrientes, como nitrogênio. A deficiência de nitrogênio é um dos estresses abióticos mais críticos para a cultura do milho. O desenvolvimento de genótipos mais eficientes no uso do nitrogênio é essencial para se obter uma maior produtividade e a sustentabilidade. Análise de associação genômica ampla do genoma (GWAS) e a seleção genômica ampla (GWS) permite analisar a diversidade alélica de caracteres complexos e identificar alelos superiores e auxiliar na tomada de decisão quanto aos genótipos que devem avançar nos programas de melhoramento. Este estudo teve como objetivo identificar marcadores de polimorfismo de nucleotídeo único (SNP) associados a caracteres de morfologia de raiz em condições de baixa e alta disponibilidade de N e genes candidatos relacionados a esses marcadores por estudo de associação do genoma (GWAS). Um painel com 150 linhagens de milho tropical genotipadas com 3.083 marcadores SNPs e 28 caracteres fenotipados para morfologia de raiz e do caule em estágio de plântula em alta (AN) e baixa (BN) disponibilidade de nitrogênio foram utilizados para executar análise de GWAS e GWS. Foi utilizado o modelo linear misto (MLM) para realizar um estudo GWAS para identificar marcadores SNPs associados aos caracteres de morfologia de raiz em condições de baixa e alta disponibilidade de N. Um total de 13 SNPs acima do limite significativo ($p \leq 2,8 \times 10^{-4}$) foi detectado. Esses SNPs levaram a 15 genes candidatos. Na GWS, os modelos de predição que incorporam associações de marcadores de traço apresentaram maior precisão de predição do que os modelos de predição usando um número equivalente de SNPs selecionados. As acurácias do modelo que utilizaram apenas os SNPs associados aos caracteres foram significativamente maiores do que as acurácias obtidas no modelo que incorporaram todos os SNPs para todos os 14 caracteres avaliados nas condições de BN e AN. A análise de GWAS revelou marcadores associadas aos caracteres de morfologia de raiz, localizados em regiões que contém genes ou modelos de genes com funções celulares relacionados de proteção das plantas a estresses ambientais, incluindo estresse de nitrogênio. Os genes encontrados podem ser mais estudados para ajudar a entender a base genética do desenvolvimento radicular e auxiliar no desenvolvimento de cultivares de milho mais eficientes no uso de nitrogênio.

6.1. INTRODUÇÃO

O milho é uma das safras de cereais mais importantes cultivadas para alimentos, rações e biocombustíveis em todo o mundo, e o Brasil é o terceiro maior produtor mundial de milho, depois dos Estados Unidos e da China (FAO, 2020). Devido ao seu alto potencial produtivo, ele apresenta alta exigência nutricional. Entre os nutrientes considerados como essenciais, o nitrogênio (N) é o requerido em maior quantidade pela cultura do milho. O N é constituinte de diversos compostos como: aminoácidos, clorofila e proteínas e envolvido em diversas rotas metabólicas importantes para o desenvolvimento da planta (Taiz e Zeiger, 2015). Entre os estresses abiótico, a deficiência de N, comum nos solos brasileiros, aliado ao cultivo de milho de segunda safra, o qual ocorre em regiões e épocas com maior frequência de estresses abióticos, podem limitar significativamente o desenvolvimento das plantas e reduzir a produtividade (Fritsch-Neto e Borém, 2012). Além disso, a baixa eficiência na adubação nitrogenada, geralmente menor que 50% devido aos processos de perda por volatilização, desnitrificação e lixiviação (Gallais e Hirel, 2004). Assim, o desenvolvimento de genótipos mais eficientes no uso de nitrogênio (EUN) é essencial para se obter um cenário mais produtivo e sustentável. Por meio do melhoramento genético, é possível desenvolver cultivares de milho mais EUN, ou seja, que produzam mais ou a mesma quantidade com menor demanda por N. A identificação de linhagens EUN e permite a obtenção de híbridos eficientes no uso desse nutriente, e dessa forma, reduzir o custo com adubação nitrogenada e os impactos ambientais gerados com o uso excessivo de adubos nitrogenados (Mundin et al., 2018).

A EUN foi definida por Moll et al. (1982) como a produtividade de grãos por unidade de nitrogênio do solo. Os dois componentes principais da EUN são a eficiência de absorção de N, que é a razão entre o N total na planta e o suprimento de N, e a eficiência de utilização de N que é expressa como a produtividade de grãos por N total na planta. O controle genético da EUN envolve muitos genes e eles são expressos de acordo com a disponibilidade de N (Gallais e Hirel, 2004). Portanto, requer a avaliação de genótipos sob condições de baixo e ótimo disponibilidade de N para fazer inferências.

Uma estratégia para aumentar a eficiência do uso de nitrogênio (EUN) no milho é o melhoramento genético da aquisição de N por meio da morfologia radicular (Torres et al., 2019). A morfologia da raiz pode ter um impacto considerável na EUN em milho, principalmente em solos secos e empobrecidos, uma vez que as características da raiz

desempenham um papel essencial na aquisição de N (Saengwilai et al., 2014; Morosini et al., 2017, Sanchez et al., 2018). A arquitetura do sistema radicular pode ser de grande importância para melhorar a EUN em cereais (Garnett et al., 2009; Ma et al., 2020). Segundo Pace et al. (2015a), a arquitetura e o desenvolvimento da raiz são componentes-chave para a eficiência no uso de N. O aumento do tamanho da raiz e, conseqüentemente, a área superficial das raízes pode ser uma estratégia usada pelas plantas para aumentar a eficiência de absorção, quando os nutrientes como o N são limitantes (Horst et al., 2001; Pace et al., 2015a). Sendo assim, regiões genômicas que afetam o desenvolvimento e o crescimento das raízes podem afetar a EUN (Pace et al., 2015a). A caracterização de genótipos baseada em dados moleculares pode contribuir para otimizar o uso de germoplasma em relação à seleção para EUN (Morosini et al., 2017).

Ferramentas como estudos de associação do genoma (GWAS) e a seleção genômica ampla (GWS) podem ajudar na tomada de decisão sobre quais genótipos devem avançar nos programas de melhoramento (Ma et al., 2020; Wang et al., 2019). Esta metodologia é baseada na associação entre marcadores de polimorfismo de nucleotídeo único (SNP) de alta densidade e o fenótipo de interesse (Korte e Farlow, 2013). GWAS se tornou uma abordagem popular e eficiente para estudar as variações genéticas de vários caracteres complexos em várias plantas, como *Arabidopsis*, milho e arroz (Yang et al., 2020). No milho, a GWAS é adequada devido à alta diversidade e à rápida decomposição do desequilíbrio de ligação (LD) nesta espécie, o que a torna uma cultura modelo para análise de associação (Flint-Garcia et al., 2005; Yan et al., 2011). Pace et al. (2015a) e Sanchez et al. (2018) encontraram SNPs associados a caracteres de raízes de plântulas de milho localizadas ou ligadas a modelos de genes que codificam proteínas associadas ao desenvolvimento das raízes. Wang et al. (2019) avaliaram 210 linhagens de milho para tolerância à seca no estágio de muda e identificaram um total de 413 SNPs associados aos caracteres nas condições de alta disponibilidade de água e 696 SNPs nas condições de estresse hídrico. Eles encontraram informações valiosas para a compreensão da variação genética da tolerância ao estresse hídrico no milho. A GWAS vem sendo usada para encontrar genes associados a caracteres para EUN como uma ferramenta poderosa que conecta informações de genótipo-fenótipo (Zhang et al., 2015; Laperche et al., 2017). Morosini et al. (2017) avaliaram 74 linhagens de milho tropicais em campo contrastando para EUN. A análise de GWAS revelou regiões genômicas associadas com EUN em condições de campo.

A GWS ou predição genômica, proposta por Meuwissen et al. (2001) permite a predição simultânea dos efeitos dos marcadores é um método de seleção que pode aumentar o ganho genético por unidade de tempo e reduzir custos (Wang et al., 2019). Segundo Pace et al. (2015b), a predição genômica permite identificar genótipos para caracteres de difícil mensuração em um grande número de indivíduos. Para eles a GWS não é apenas uma abordagem promissora para fins de melhoramento, mas também para pesquisa básica. Existem numerosas abordagens de predição genômica, procedimentos de modelo bayesiano e misto. O modelo RR-BLUP (Random Regression Best Linear Unbiased Predictor; Fernando et al., 2007), tem se mostrado um modelo confiável no fornecimento de boa acurácia de predição para caracteres quantitativos. O modelo misto usado pelo RR-BLUP reduz todos os efeitos do marcador igualmente com a mesma variância para um caractere. Isso explica o bom desempenho do RR-BLUP para predição de caracteres complexos (Clark et al., 2011). Computacionalmente, RR-BLUP é menos exigente o que o torna um dos métodos mais acessíveis e sugeridos para a predição genômica (Heslot et al., 2012; Pace et al., 2015). Pace et al. (2015b) aplicaram GP em caracteres de raiz de uma população de treinamento composta por 384 linhagens endogâmicas de milho e encontraram acurácias médias entre 0,38 e 0,55.

A fim de avaliar a tolerância ao estresse de nitrogênio e identificar genes associados a caracteres de interesse que sejam úteis para a melhoria do germoplasma de milho do programa de melhoramento de milho da Universidade Federal de Viçosa, um painel com 150 linhagens endogâmicas de milho foi usado neste estudo para análises de GWAS e GWS. Esse painel de linhagens foi fenotipado para 14 caracteres de morfologia de raiz e de caule em plântulas em dois ambientes, em alta e baixa disponibilidade de N por Torres et al. (2019). Esse painel de linhagens também foi genotipado com 3.713 SNPs e realizado a caracterização molecular e fenotípica por Faria et al. (2020). Assim, os objetivos do presente estudo foram: i) identificar SNPs associados a caracteres de morfologia de raiz relacionados a tolerância ao estresse de nitrogênio, ii) investigar a localização dos marcadores SNPs associados para possíveis genes candidatos que afetam o desenvolvimento das raízes e iii) avaliar a acurácia da seleção genômica em diferentes conjuntos de populações de treinamento.

6.2. MATERIAL E MÉTODOS

6.2.1. Material Genético

Um painel de 150 linhagens endogâmicas de milho tropical oriundas do Banco Ativo de Germoplasma do Programa Milho®, Programa de Melhoramento de Milho da Universidade Federal de Viçosa, Departamento de Fitotecnia foram utilizadas. Essas linhagens foram selecionadas por suas origens, desempenho '*per se*' para caracteres agrônômicos de arquitetura de planta e produção de grãos.

Foram conduzidos dois experimentos em casa de vegetação localizada no Campo Experimental Diogo Alves de Melo, pertencente ao Departamento de Fitotecnia da Universidade Federal de Viçosa (latitude 20°46'S; longitude 42°52'W; 640 metros de altitude). Resumidamente, foi conduzido um experimento em condição de alto (AN) e o outro em condição de baixo (BN) disponibilidade de N. Em cada experimento foi utilizado o delineamento de blocos ao acaso com três repetições. As plantas foram coletadas no estágio vegetativo de cinco folhas completamente expandidas (V5). Cada unidade experimental foi constituída por um vaso de polietileno. Dez dias após a semeadura, as plântulas passaram a ser adubadas com 0,3 litros de solução nutritiva em intervalos de dois dias. A solução nutritiva utilizada foi adaptada e descrita por Chun et al. (2005). As plântulas foram cortadas e as raízes separadas da parte aérea. As raízes foram avaliadas utilizando-se o software WinRHIZO Pro 2009a (Basic, Reg, Pro & Arabidopsis for Root Measurement). Os detalhes do experimento foram descritos por Torres et al. (2019).

6.2.2. Dados fenotípicos

Os seguintes caracteres foram avaliados em AN e BN por Torres et al. (2019): altura de planta (AP, cm), diâmetro do colmo (DC, mm), densidade do tecido radicular (DTR, mg cm⁻³), massa de raiz seca (MRS, mg), massa da parte aérea seca (MPAS, mg), massa total seca (MTS, mg), relação raiz:parte aérea (RSR, mg.mg⁻¹), crescimento diário (CD, cm), comprimento radicular axial (CRA, cm), comprimento radicular lateral (CRL, cm), comprimento radicular total (CRT, cm), diâmetro radicular médio (DRM, mm), área superficial radicular (ASR, cm²) e volume radicular (VR, cm³).

6.2.3. Genotipagem

Um painel com 182 linhagens de milho foram genotipadas por meio da plataforma Illumina GoldenGate® (Illumina, San Diego, CA, USA), com um array de 3.713 SNPs distribuídos pelos 10 cromossomos de milho (Fan et al., 2003). As fontes dos marcadores

são do SNP chip de 56K. As informações dos marcadores foram publicadas por Jones et al. (2009); Gore et al. (2009) e por Ganal et al. (2011) que escolheram os marcadores que funcionam bem com diversos germoplasmas para colocar na matriz SNP. As informações sobre os SNPs estão disponibilizadas publicamente no banco de dados MaizeGDB (www.maizegdb.org) e NCBI (National Center for Biotechnology Information) (www.ncbi.nlm.nih.gov). Após o controle de qualidade com 1% de frequência do alelo raro (MAF, Minor Allele Frequency), e Call rate de 90%. Finalmente, restaram 3.083 marcadores SNPs distribuídos pelos 10 cromossomos do genoma do milho para calcular a estrutura populacional, o parentesco, o desequilíbrio de ligação, resultados podem ser vistos no Capítulo 1 dessa tese e por fim os dados foram usados para realizar a análise de associação genômica para encontrar e genes associados aos caracteres de interesse e para a análise de predição genômica.

6.2.4. Mapeamento da Associação Genômica Ampla

Estudos de associação genômica ampla foram realizados nas 150 linhagens endogâmicas usando o modelo linear misto proposto por Yu et al. (2006). Neste modelo, cada marcador (efeito fixo) é avaliado individualmente, um de cada vez (isto é, regressão de marcador único). A equação deste modelo pode ser expressa como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{m} + \mathbf{Q}\mathbf{v} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

em que: \mathbf{y} é um vetor de observações fenotípicas; $\boldsymbol{\beta}$ é um vetor de efeitos fixos que não SNP sob teste ou efeitos de grupo populacional; \mathbf{m} é um vetor de efeito SNP em avaliação (efeito fixo); \mathbf{v} é um vetor de efeitos de subpopulação (efeito fixo); \mathbf{u} é um vetor de efeitos de fundo de polígono (efeitos aleatórios - proporção dos valores de reprodução não contabilizados pelo marcador SNP); \mathbf{e} é um vetor de efeitos residuais; \mathbf{Q} é uma matriz de incidência de escores de componentes principais (vetores próprios) que relacionam os níveis de subpopulação com os efeitos de subpopulação; e \mathbf{X} , \mathbf{W} e \mathbf{Z} são matrizes de incidência de um e zeros relacionados \mathbf{y} a $\boldsymbol{\beta}$, \mathbf{m} e \mathbf{u} , respectivamente. Supõe-se que a variação dos efeitos aleatórios seja $Var(u) = 2K\sigma_A^2$, onde \mathbf{K} é uma matriz de coeficientes de parentesco que define o grau de covariância genética entre um par de indivíduos, para determinar se são independentes. As matrizes \mathbf{Q} e \mathbf{K} foram ajustadas no modelo linear misto para controlar associações espúrias devido a estrutura de população e parentesco, respectivamente.

O modelo linear misto (MLM) e a estimativa da matriz de parentesco foram realizados usando o pacote ‘GAPIT’ - Genomic Association e Prediction Integral Tool do software R, desenvolvido por Lipka et al. (2012). Para explicar a relação entre indivíduos, avaliamos três abordagens diferentes para estimar a matriz de parentesco (K) como matriz de variância-covariância para efeitos aleatórios de genótipos: VanRaden (VanRaden, 2008), Loiselle (Loiselle et al., 1995) e a associação eficiente de modelos mistos Abordagens de parentesco (EMMA) (Kang et al., 2008). Como a estrutura populacional pode resultar em associações espúrias, foi considerado o uso dos três primeiros componentes principais que juntos contribuíram com 7,21% da variação total. Assim, vários modelos de GWAS, “Q + K”, “K”, “Q” e sem estrutura populacional e parentesco, foram avaliados para determinar o modelo mais adequado com base na expectativa de que os valores de p sigam uma distribuição uniforme entre 0 e 1 (isto é, valores $\sim U [0,1]$) sob a hipótese nula de não associação. Gráficos quantil-quantil esperados versus observados foram desenhados para avaliar a adequação do modelo.

Para determinar um limiar apropriado em termos de genoma que controla a taxa de erro do tipo experimental, implementamos o método simpleM (Gao et al., 2008). O método SimpleM demonstrou ser uma ferramenta eficaz de controlar a taxa de erro experimental em estudos de associação em todo o genoma (Gao et al., 2010; Johnson et al., 2010; Gao, 2011). Esse método aplica uma correção de Bonferroni ao número real de teste independente (ou seja, o número efetivo de teste) estimando o desequilíbrio de ligação entre cada par de marcadores e aplicando a análise de componentes principais para obter os valores próprios.

Neste estudo, o número efetivo cumulativo de testes em todos os cromossomos foi 177; portanto, o limiar em termos de genoma é igual a $0,05 / 177$, ou seja, $2,8 \times 10^{-4}$ ($\alpha = 0,05$). O número efetivo de teste para nossos SNPs foi igual ao número de autovalores necessários para explicar 99% da variância. Além disso, os testes de permutação não são apropriados na presença da subestrutura da população, porque o embaralhamento dos dados do genótipo em relação aos dados fenotípicos interrompe não apenas a associação SNP-característica, mas também a relação traço-pedigree (por exemplo, parentesco) (Aulchenko et al., 2007). Marcadores acima do limiar sugerido para MLM foram considerados marcadores SNP significativamente associados a características e candidatos a polimorfismos causais.

6.2.5. Seleção genômica

A análise de seleção genômica para os 14 caracteres avaliados em AN e BN foi realizada usando o pacote rrBLUP (Ridge regression-best linear unbiased prediction) no software R (Endelman, 2011). O método rrBLUP tem uma distribuição t multivariada marginal com média zero, o modelo misto usado pelo RR-BLUP reduz todos os efeitos do marcador igualmente com a mesma variância para um caractere (Wang et al., 2018). De acordo com Clark et al. (2011), isso explica o bom desempenho do RR-BLUP para predição de caracteres complexos. Computacionalmente, RR-BLUP é menos exigente, o que o torna um dos métodos mais acessíveis e sugeridos de predição genômica (Heslot et al., 2012). Dois conjuntos de SNP foram usados para a análise de seleção genômica, foram utilizados os 3.083 SNPs distribuídos uniformemente pelo genoma, sem valores ausentes, e menor frequência alélica (0,05) e os SNPs significativamente associados aos caracteres detectados pelo GWAS a partir do modelo MLM + Q + K. A predição genômica foi estimada como o coeficiente de correlação entre os valores genéticos genômicos GEBV e os fenótipos observados por divisão aleatória em 60-40% como conjuntos de treinamento e validação, respectivamente, e repetido 100 vezes. O modelo usado para estimar foi: $y = X\beta + Z\alpha + \varepsilon$, em que: y é um vetor de fenótipos; β é um vetor de efeitos fixos não genéticos; X é uma matriz de incidência para os efeitos fixos; α é um vetor de coeficientes de regressão aleatória de todos os efeitos de marcadores; Z é uma matriz genotípica n x k de marcadores; e ε é um vetor de resíduos.

6.2.6. Análise de Genes Candidatos

Além do mapeamento de associação ampla do genoma, onde todos os marcadores foram testados para associação com os 14 caracteres avaliados em AN e BN, também realizamos o mapeamento de associação de genes candidatos para os caracteres significativos (Thornsberry et al., 2001). O banco de dados GDB do milho (<http://www.maizegdb.org>) foi usado para encontrar genes candidatos vinculados a cada SNP com base no genoma de referência B73 RefGen_V3 do milho, também para encontrar correspondências com B73 RefGen_V4. Após a delimitação da região a ser considerada, foram apontados os possíveis genes candidatos que flanqueiam o marcador em questão. Em seguida, com o ID do gene, a ferramenta Ensembl Biomart (Kinsella et al., 2011) foi usada para obter informações funcionais dos genes candidatos.

6.3. RESULTADOS

6.3.1. Caracteres fenotípicos em AN e BN

Torres et al. (2019) observaram em seus resultados que os componentes de variância obtidos para todas as 150 linhagens endogâmicas de milho tropical avaliadas foram significativos ($P = 0,01$) pelo teste da razão de verossimilhança para todos os caracteres de raiz e parte aérea em AN e BN. Também observaram amplas faixas de valores genotípicos para todos os caracteres em ambas as disponibilidades de N (Tabela 1). A maioria dos caracteres de raiz e parte aérea foram significativamente afetados ($P = 0,05$) pela disponibilidade de N.

Tabela 1. Estatísticas dos caracteres de raiz e parte aérea medidas em alta (AN) e baixa (BN) disponibilidade de N para 150 linhagens endogâmicas de milho tropical

Caracteres ¹	AN				BN			
	Min.	Média	Max.	h ²	Min.	Média	Max.	h ²
AP	11,63	18,00	24,78	0,82	11,03	19,69	28,95	0,85
DC	6,17	8,04	10,23	0,74	6,3	9,31	11,9	0,84
MRS	275,71	468,47	913,74	0,68	236,95	443,53	733,24	0,65
MPAS	822,17	1.511,73	2.668,67	0,71	1.065,3	2.098,93	3.716,34	0,78
MTS	1.102,5	1.980,2	3.607,84	0,71	1.319,3	2.542,47	4365,2	0,78
RSR	0,25	0,32	0,49	0,34	44,67	0,14	0,22	0,59
CD	0,4	0,55	0,72	0,78	0,43	0,65	0,88	0,82
CRL	1.815,42	2.346,78	3.282,55	0,54	2154,78	2.804,87	3.644,15	0,56
CRA	627,81	938,63	1.495,12	0,71	792,34	1.006,6	1.434,41	0,65
CRT	2.499,68	3.286,26	4.600,94	0,58	2.926,78	3.812,88	5.094,02	0,59
DRM	0,43	0,52	0,59	0,60	0,43	0,5	0,55	0,55
ASR	344,07	530,17	842,69	0,71	431,36	597,61	865,99	0,67
VR	3,61	6,91	12,04	0,75	4,94	7,56	11,5	0,68
DTR	56,74	66,29	77,76	0,54	15,01	44,58	57,23	0,54

¹AP - altura da planta (cm), DC - diâmetro do caule (mm), MRS - massa seca de raiz (mg), MPAS - massa seca da parte aérea (mg), MTS - massa seca total (mg), RSR - relação massa seca da raiz/ parte aérea, CD - crescimento diário (cm), CRL - comprimento da raiz lateral (cm), CRA - comprimento da raiz axilar (cm), CRT - comprimento total da raiz (cm), DRM - diâmetro médio da raiz (mm), ASR - área superficial radicular (cm²), VR - volume da raiz (cm³) e DTR - densidade do tecido radicular (mg cm⁻³)

Em geral, as estimativas de herdabilidade no sentido amplo (Tabela 1) obtidas por Torres et al. (2019) para esses caracteres variaram de 0,34 a 0,82, 0,54 a 0,85 e de 0,61 a 0,91, para BN, AN e entre os níveis de N, respectivamente. O caractere RSR foi o mais

afetado pelo estresse de N, foi o único que apresentou valor de herdabilidade menor que 0,50 em BN (0,34).

6.3.2. Estudos de Associação em todo o genoma e genes candidatos

A fim de identificar SNPs associados a caracteres relacionados a EUN, a GWAS foi realizada com oito abordagens diferentes a partir do modelo linear misto (MLM). Os gráficos quantil-quantil dos valores de P comparando a distribuição uniforme do valor esperado $-\log_{10}(\text{P-value})$ com o valor observado $-\log_{10}(\text{P-value})$ para os caracteres que apresentaram SNPs significativos, AP, ASR, CD, CRA, CRL, CRT, MPAS, MRS, MTS e RSR são apresentados na Figura 1, para BN e Figura 2, para os caracteres DC, DRM e DTR em AN. O modelo misto que utiliza a matriz Q e os modelos Q + K ajustaram bem os dados para os caracteres avaliados. Esses modelos mostraram uma melhoria significativa no ajuste do MLM em relação ao modelo no qual a estrutura da população foi ignorada (Q[0]). No geral, as diferenças na adequação do modelo entre as três abordagens de parentesco, VanRaden, Loiselle e EMMA, foram muito pequenas e exibiram um bom controle da taxa de erro do tipo I. Os modelos Q3 e Loiselle + K se ajustaram melhor do que os outros modelos avaliados.

Figura 1. Comparação de gráficos QQ plots para diferentes modelos de GWAS de três caracteres (DC, DRM e DTR) em baixa disponibilidade de nitrogênio (BN)

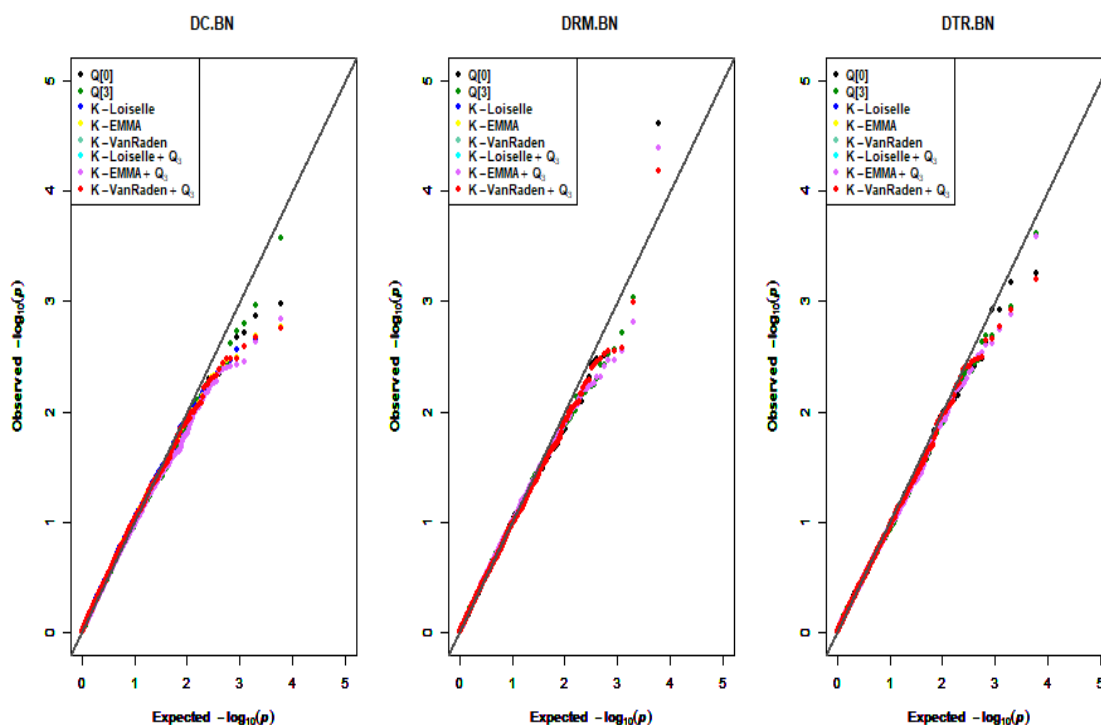
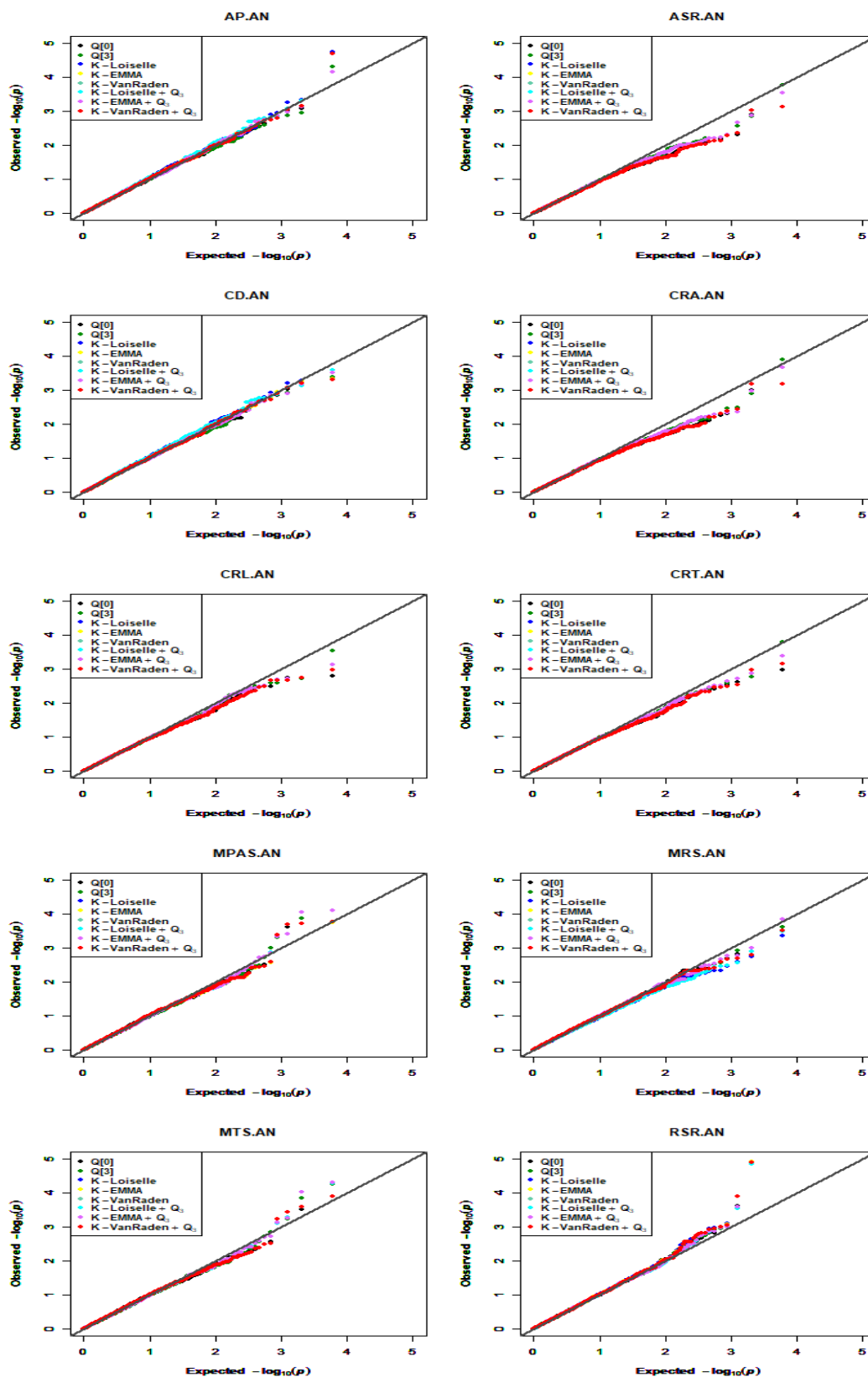


Figura 2. Comparação de gráficos QQ plots para diferentes modelos de GWAS de 10 caracteres (AP, ASR, CD, CRA, CRL, CRT, MPAS, MRS, MTS e RSR) em alta disponibilidade de nitrogênio (AN)



O modelo MLM com as abordagens Q3 e Loiselle + K identificaram 10 SNPs acima do limite de significância entre os caracteres avaliados. A maioria dos caracteres que detectaram SNPs significativos foram avaliados em alta disponibilidade de nitrogênio (Figura 3), foram esses AP, ASR, CD, CRA, CRL, CRT, MPAS, MRS, MTS e RSR. Para o caractere RSR (relação raiz/parte aérea) foram detectados três SNPs, PZE 105120851 no cromossomo 5 ($P = 2,64 \times 10^{-4}$), PZE-107054297 no cromossomo 7 ($P = 3,69 \times 10^{-8}$) e PZA03035.5 ($P = 1,21 \times 10^{-5}$) no cromossomo 8. O marcador PZA02151.3, cromossomo 4, foi significativamente associado a cinco caracteres, todos para alta disponibilidade de nitrogênio, a saber, ASR ($P = 1,66 \times 10^{-4}$), CRL ($P = 2,88 \times 10^{-4}$), CRA ($P = 1,24 \times 10^{-4}$), CRT ($P = 1,64 \times 10^{-4}$) e MRS ($P = 2,50 \times 10^{-4}$). Os SNPs PZA02151.3 e PZE-105021600 em alto nitrogênio, foram significativos para os caracteres MPAS ($P = 7,7 \times 10^{-5}$ e $P = 1,33 \times 10^{-4}$, respectivamente) e MTS ($P = 7,7 \times 10^{-5}$ e $P = 1,33 \times 10^{-4}$, respectivamente).

Para baixa disponibilidade de nitrogênio os SNPs PZE-109009258, cromossomo 9 ($P = 2,68 \times 10^{-4}$) para o caractere DC, PHM12904.7, cromossomo 8 ($P = 4,15 \times 10^{-5}$) para DRM e PZA00466.1 no cromossomo 9 ($P = 2,47 \times 10^{-4}$) para DTR (Figura 4). O caractere VR foi o único que não detectou SNPs significativos em nenhuma das condições de nitrogênio. Para os demais caracteres avaliados não foram encontrados SNPs significativos utilizando o SimpleM ($P = 2,8 \times 10^{-4}$) como limiar.

Figura 3. Gráficos Manhattan mostrando os SNPs significativos em todo o genoma associado a 10 caracteres (AP, ASR, CD, CRA, CRL, CRT, MPAS, MRS, MTS e RSR) para AN

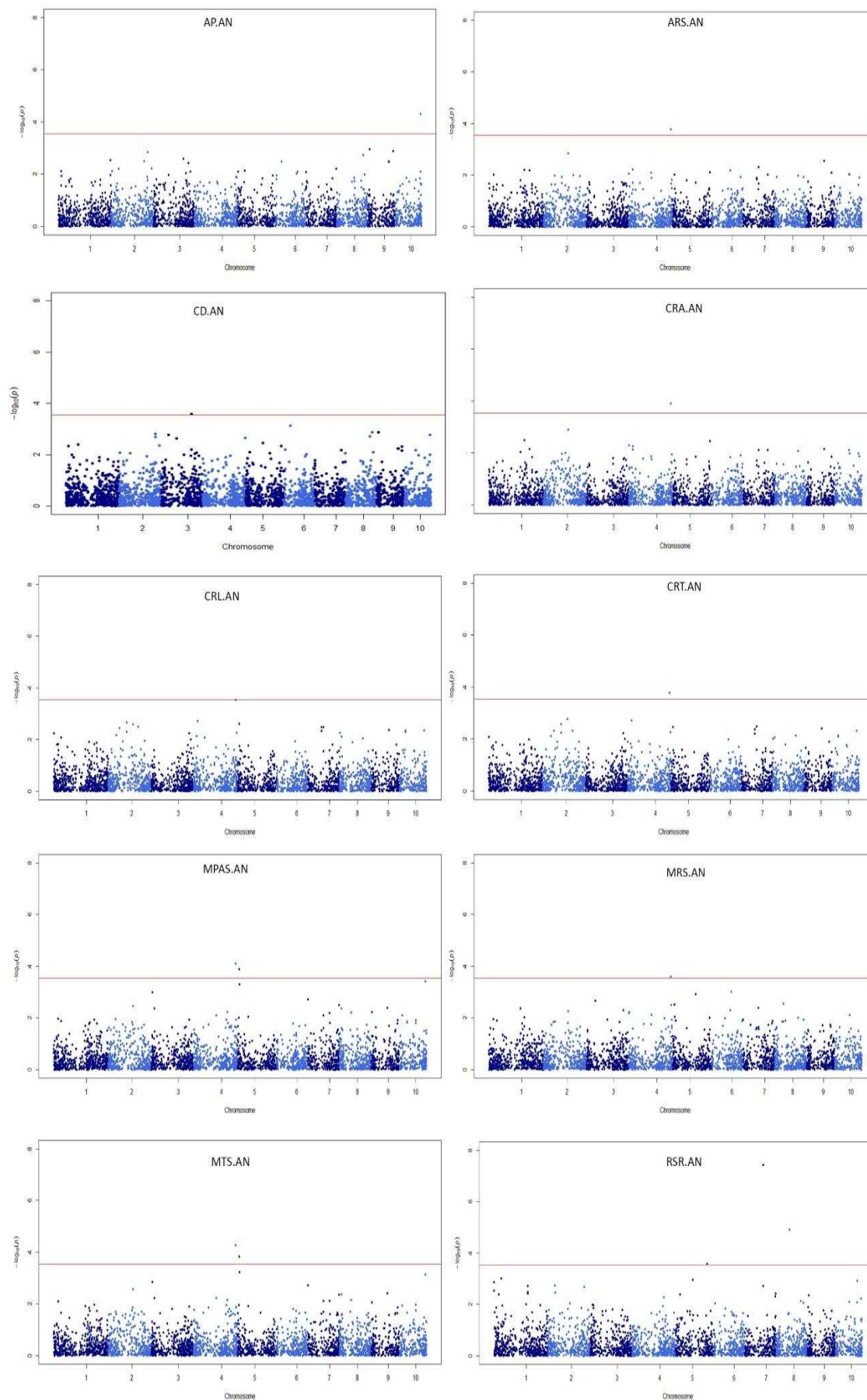
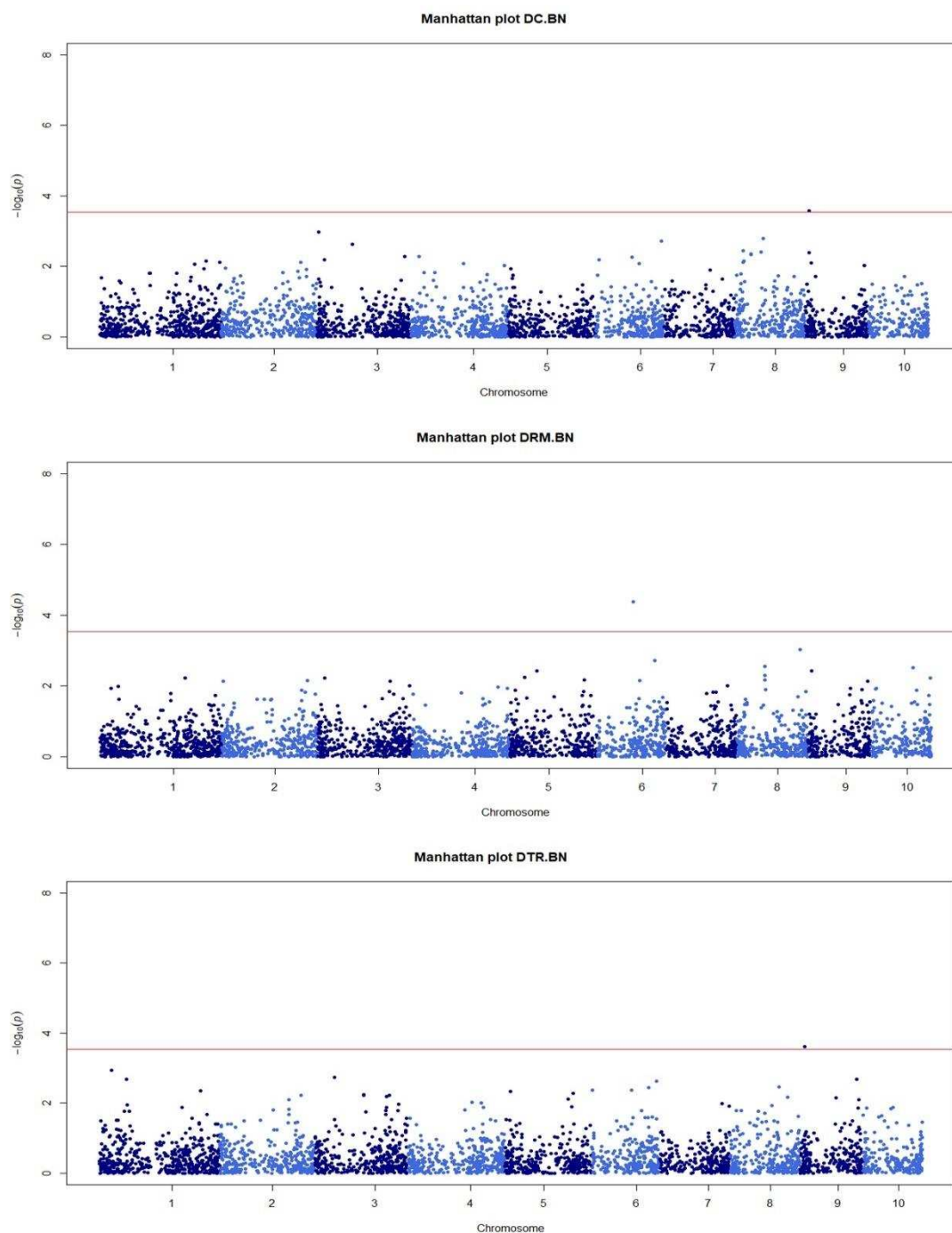


Figura 4. Gráficos de Manhattan mostrando SNPs significativos em todo o genoma associado a três caracteres (DC, DRM e DTR) em BN



A GWAS revelou um conjunto de genes candidatos identificados nos cromossomos 3, 4, 5, 8, 9 e 10 (Tabela 2). Dos 10 SNPs acima do limite de significância, encontramos 16 genes candidatos (genoma de referência B73 v3). Foram considerados genes candidatos apenas os SNPs identificados nas regiões gênicas.

Para o CD o marcador significativo PZE-103108877 foi encontrado nos modelos de gene AC210163.3_FG005 (Zm00001d042574), localizado entre 169777887 a 169778231 bp e codifica a proteína induzidas por auxina e GRMZM2G038309 (Zm00001d042575), localizado entre 169786827 a 169791501 bp codifica a proteína Serina / treonina-proteína quinase RIO2. Ambos no cromossomo 3.

O marcador PZA02151.3 que foi associado aos caracteres MPAS e MRS, foi encontrado nos modelos de gene GRMZM2G078832 (Zm00001d053620) localizado entre 177663621 a 177668212 bp identificado com uma proteína Peptidil-prolil cis-trans isomerase CYP95 e GRMZM2G153871 (Zm00001d053619) localizado entre 232233178 a 232235607 bp e codifica a proteína Fenilalanina amônia-liase 1. Ambos no cromossomo 4.

Para os caracteres MPAS e MTS o marcador PZE-105021600 associado significativamente foi encontrado nos modelos de gene GRMZM5G801651 (Zm00001d016972) que codifica a proteína da família de transportadores de colina na membrana plasmática, GRMZM2G459166 (Zm00001d016973) que codifica a proteína F-box GID2, GRMZM2G086246 (Zm00001d016975) que codifica a proteína DNA - (sítio apurínico ou apirimidínico) liase cloroplástico e GRMZM5G861269 (Zm00001d016977) que codifica a proteína da família de cadeia leve de dineína tipo 1. Todos os genes estão no cromossomo 5 localizados entre 177633413 a 177781361 bp.

Para o caractere RSR o marcador associado significativamente, PZA03035.5, foi encontrado dentro do modelo de gene GRMZM2G017821 (Zm00001d009743), localizado entre 77106424 a 77108655 bp no cromossomo 8 e codifica a proteína contendo repetição de pentatricopeptídeo.

Para o caractere DC o marcador significativo PZE-109009258 foi encontrado no modelo de gene GRMZM2G146358 (Zm00001d045069), localizado entre 12027561 a 12030163 bp cromossomo 9 e codifica a proteína Ribosomal L19.

Para o caractere AP o marcador significativo PZA02049.1 foi encontrado em diferentes modelos de gene GRMZM2G445575, GRMZM2G143102, GRMZM2G443819, GRMZM2G177393, GRMZM2G143071 e GRMZM2G037226, todos no cromossomo 10, localizados entre 143801191 a 143992151 bp, cada um codificando uma proteína diferente, fator de transcrição TGA4, Provável galacturonosiltransferase 10, proteína semelhante a poliproteína de histona ligante micronuclear, Proteína fator de padrão epidérmico, formação da proteína 1 das junções

cristas e a proteína de ligação a RNA AKIP 1% 3B, proteína de ligação a RNA AKIP1 isoforma 2.

Tabela 2. Genes candidatos para marcadores significativos encontrados em GWAS

Ch r	SNP	B73 Gene ID ¹	Zm Gene ID ²	Anotação ³
3	PZE-103108877	AC210163.3_FG005	Zm00001d042574	Proteína induzível por auxina
3	PZE-103108877	GRMZM2G038309	Zm00001d042575	Serina / treonina-proteína quinase RIO2
4	PZA02151.3	GRMZM2G078832	Zm00001d053620	Peptidyl-prolyl cis-trans isomerase CYP95
4	PZA02151.3	GRMZM2G153871	Zm00001d053619	Fenilalanina amônia-liase 1
5	PZE-105021600	GRMZM5G801651	Zm00001d016972	Proteína da família de transportadores de colina de membrana plasmática
5	PZE-105021600	GRMZM2G459166	Zm00001d016973	Proteína F-box GID2
5	PZE-105021600	GRMZM2G086246	Zm00001d016975	DNA- (sítio apurínico ou apirimidínico) liase cloroplástico
5	PZE-105021600	GRMZM5G861269	Zm00001d016977	Proteína da família de cadeia leve de dineína tipo 1
8	PZA03035.5	GRMZM2G017821	Zm00001d009743	Proteína contendo repetição de pentatricopeptídeo
9	PZE-109009258	GRMZM2G146358	Zm00001d045069	Proteína ribossomal L19
10	PZA02049.1	GRMZM2G445575	Zm00001d026398	Fator de transcrição TGA4
10	PZA02049.1	GRMZM2G143102	Zm00001d026394	Provável galacturonosiltransferase 10
10	PZA02049.1	GRMZM2G443819	Zm00001d026395	Proteína semelhante a poliproteína de histona ligante micronuclear
10	PZA02049.1	GRMZM2G177393	Zm00001d026401	Proteína FATOR DE PADRÃO EPIDÉRMICO 2
10	PZA02049.1	GRMZM2G143071	Zm00001d026396	Formação da proteína 1 das junções cristas
10	PZA02049.1	GRMZM2G037226	Zm00001d026397	Proteína de ligação a RNA AKIP1% 3B proteína de ligação a RNA AKIP1 isoforma 2

¹Baseado em B73 RefGen_v2; ² Baseado em B73 RefGen_v4; ³Obtido em MaizeGDB and Gramene.

6.3.3. Seleção genômica

Para o presente estudo, dois bancos de dados genotípicos foram utilizados para as análises de seleção genômica dos 28 caracteres a partir do método rrBLUP. As acurácias de predição genômica para o modelo que incorpora todos os 3.083 SNPs foram maiores para os caracteres CD, RSR e MPAS, todos em BN, 0,19, 0,17 e 0,16, respectivamente. A menor acurácia (-0,22) foi encontrada para VR (AN). As acurácias para as condições de AN foram menores que para BN, com média de -0,01 e 0,04, respectivamente. Quando foi considerado apenas os SNPs associados significativamente as acurácias para as condições de AN foram maiores que para BN, com média de 0,27 e 0,08, respectivamente. E os caracteres que apresentaram maiores acurácias foram AP, MPAS, MTS e MRS/CRL, todos em AN, 0,46, 0,42, 0,40 e 0,33, respectivamente. A menor acurácia (-0,17) foi encontrada para ASR (BN). As acurácias do modelo que incorporam os SNPs selecionados que foram associados aos caracteres avaliados, obtidos pela GWAS, foram significativamente maiores do que as acurácias do modelo que incorporaram todos os SNPs, exceto para CRT e CRL, avaliados em BN, que apresentaram uma redução de 0,14 e 0,11, respectivamente.

Figura 5. Acurácia da seleção genômica (eixo y) para os caracteres AP, DC, MRS, MPAS, MTS, RSR, CD e CRT em AN e BN usando todos os SNPs (All) e os SNPs significativamente associados (Sig)

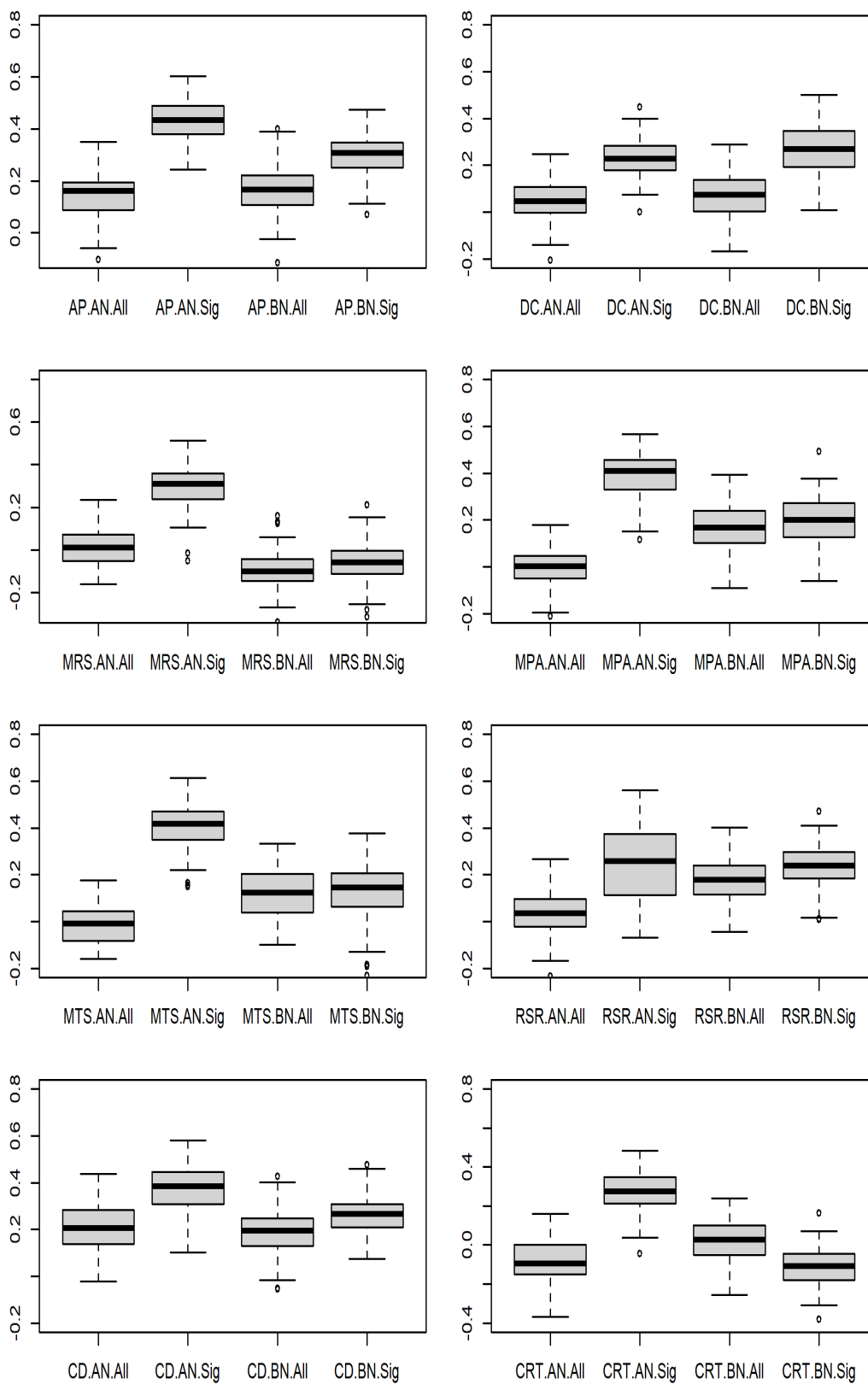
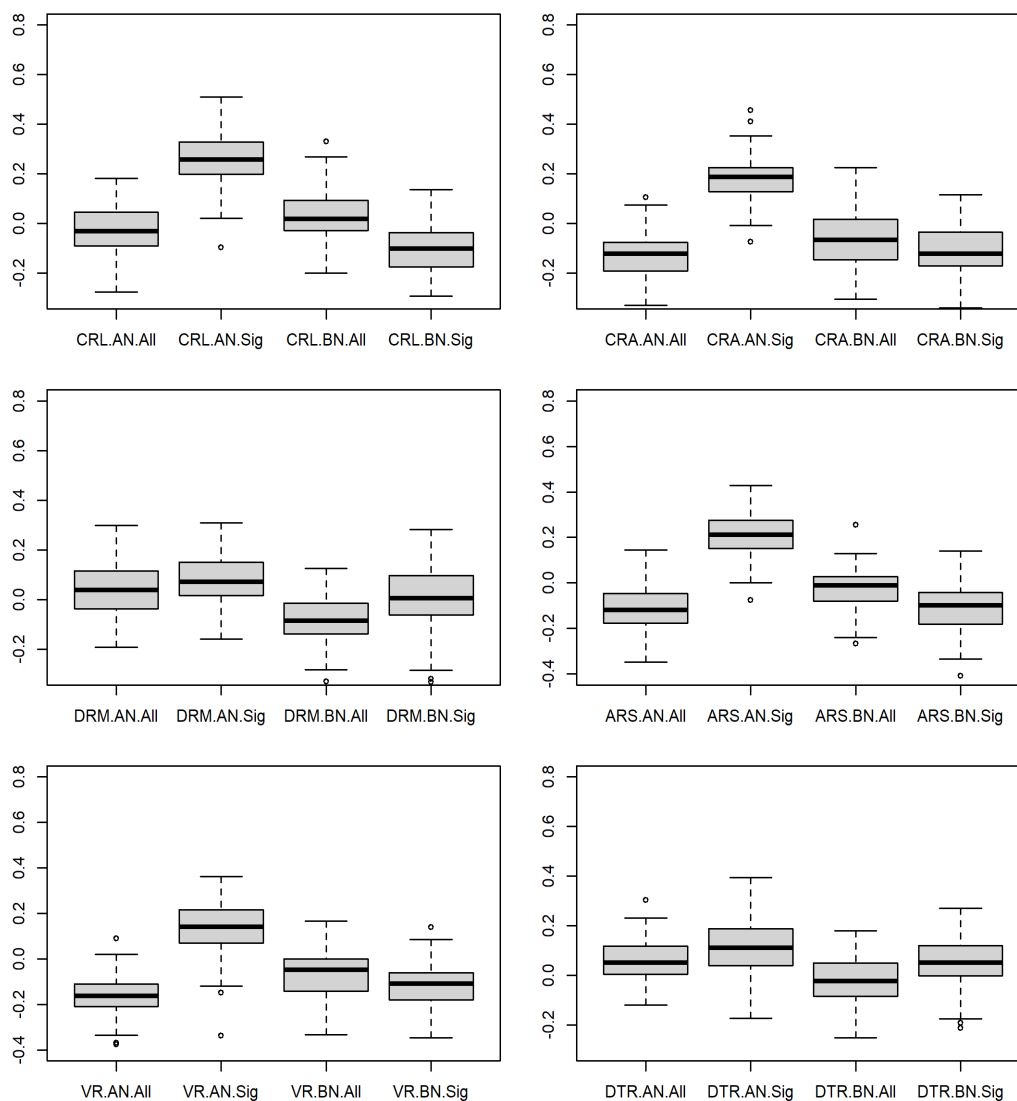


Figura 6. Acurácia da seleção genômica (eixo y) para os caracteres CRL, CRA, DRM, ARS, VR e DTR em condições de alta (AN) e baixa (BN) disponibilidade de nitrogênio usando todos os SNPs (All) e os SNPs significativamente associados aos caracteres (Sig)



6.4. DISCUSSÃO

6.4.1. Associação genômica e genes candidatos

O sistema radicular desempenha um papel importante na aquisição de água e nutrientes essenciais, como o nitrogênio, além de desempenhar o papel de sustentação da planta (Pace et al., 2015; Sanchez et al., 2018). A arquitetura e o desenvolvimento das

raízes demonstraram ser um componente-chave em EUN (Hirel et al., 2007). Assim, compreender os mecanismos moleculares que influenciam a arquitetura radicular é importante para aumentar o potencial e a estabilidade produtiva em condições ambientais contrastante. Os caracteres radiculares são difíceis e trabalhosos de serem mensurados na fase adulta e mais ainda em condições de campo, o que faz necessário a mensuração desses em plântulas. De acordo com Andorf et al. (2019), é importante integrar ferramentas genômicas para identificar alelos desejáveis e usá-los para acessar o germoplasma. Wang et al. (2019) avaliaram a associação genômica de 297 linhas consanguíneas de milho e um conjunto de 131.271 SNPs a partir dos modelos GLM e MLM. Um total de 355 e 28 associações marcador-caractere foram identificadas por GLM e MLM, respectivamente, e 96 genes candidatos foram localizados e destes cinco genes candidatos foram identificados como promissores. Zaidi et al. (2016) fenotiparam um subconjunto de 396 linhagens de milho tropical para caracteres de raiz em condições contrastantes de disponibilidade de água. Para análise de GWAS usando MLM em 331.390 SNPs os autores encontraram 167 SNPs associados com funções de raiz, estruturais, biomassa de parte aérea e rendimento de grãos. Eles encontraram entre esses SNPs associações favoráveis no cromossomo 3 para tolerância ao estresse hídrico no milho. Os autores concluíram que identificar e compreender SNPs associados a caracteres de raiz é fundamental para programas de melhoramento especialmente agora com mudanças climáticas recorrentes.

No presente estudo, a GWAS utilizando o MLM com as abordagens Q3 e Loiselle + K, indicados com os melhores modelos pelos gráficos quantil-quantil, que são uma ferramenta comumente usada para escanear a estratificação da população em GWAS. Nós encontramos SNPs significativos para quase todos os caracteres avaliados em alta disponibilidade de nitrogênio, e três caracteres em baixa disponibilidade de N. De acordo com Wang et al. (2019), uma das possíveis razões de não ter sido encontrado associação significativas para alguns caracteres se deve ao tamanho reduzido da população e de marcadores que foi utilizada para aplicação da GWAS.

Comparando as localizações físicas dos SNPs significativos no genoma de referência B73 encontramos 16 genes candidatos associados aos caracteres de interesse. Morosini et al. (2017) identificaram sete genes candidatos relacionados ao controle e regulação da transcrição, detectados para todos os caracteres de morfologia de raiz para EUN avaliados em um painel de linhagens de milho tropical avaliados, e à síntese da

Guanosina Monofosfato Sintetase, enzima diretamente envolvida no fornecimento e reciclagem de nitrogênio. Eles concluirão que a análise de GWAS revelou regiões genômicas em milho tropical associadas com a EUN em condições de campo.

Os genes Zm00001d042574 e Zm00001d042575 foram encontrados no caractere crescimento diário. As proteínas que são codificadas pelo gene Zm00001d042574 são proteínas induzíveis pelo hormônio do crescimento auxina. O hormônio vegetal auxina, representado pelo ácido indol-3-acético (IAA), influencia muitos processos complexos da planta, que inclui dominância apical, alongamento vascular, embriogênese, iniciação lateral da raiz e desenvolvimento de flores e frutos (Woodward e Bartel, 2005; Liu et al., 2011). Edelman e Schopfer (1989) avaliaram o papel da síntese de proteínas inibidoras e RNA na iniciação do crescimento mediado por auxina (IAA) em milho. Os autores concluíram que o início do crescimento pela indução da auxina depende diretamente da síntese de proteínas inibidoras do crescimento (GLP). Lowe et al. (2018) realizaram uma ampla pesquisa do genoma do milho para encontrar promotores induzível por auxina do milho adequados para conduzir a expressão de tecidos. Os autores identificaram um gene com base em sua expressão em folhas, embriões e calos, enquanto era regulado em raízes, meristemas e tecidos reprodutivos. O gene Zm00001d042575 apresenta a função relacionada a tradução de sinal para à resistência das plantas a patógenos causadores da doença anã do milho. Quando uma planta sofre estresses bióticos ou abióticos, a expressão de grupos de genes relacionados à resposta de defesa muda significativamente após a explosão do sinal (Jia et al., 2012).

Para os caracteres massa da parte aérea seca e massa de raiz seca foram encontrados os genes Zm00001d053620 e Zm000 01d053619. O gene Zm00001d053620 pertence a família da peptidil-prolil cis-trans isomerase (PPIA), um conjunto de genes de referência que codifica proteínas com funções em respostas imunes (Song et al., 2020). Esse grupo de proteínas desempenham papéis centrais importantes em resposta a sinais fisiológicos e patológicos (Bach-Pages et al., 2020). Romano et al. (2004) identificaram diversos isômeros, além da peptidil-prolil cis-trans que pertencente a família do gene da ciclofilina de *Arabidopsis*. Esses isômeros são expressos em toda a planta e podem desempenhar papéis importantes no processamento de mRNA, degradação de proteínas e transdução de sinal e, portanto, podem ser cruciais durante o desenvolvimento e a capacidade de resposta ao estresse ambiental. A peptidil-prolil cis-trans apresenta papel importante na função do cloroplasto, dada a abundância dessas enzimas no lúmen do etilacóide. E o

gene Zm00001d053619 que codifica a proteína fenilalanina amônia-liase (PAL) catalisa a primeira etapa da biossíntese de lignina nas plantas; a desaminação de l-fenilalanina em ácido cinâmico. Etapas enzimáticas sucessivas levam à formação de três monolignóis, constituindo a estrutura complexa da lignina (Andersen et al., 2007). Jones et al. (2013) avaliaram o papel dos fenilpropanóides no crescimento e desenvolvimento em milho. Para os autores seus resultados fornecem evidências que sustentam e expandem os papéis previamente hipotetizados dos fenilpropanóides na regulação do alongamento e diferenciação celular, adesão intercelular e aumento da resistência à degradação da parede celular. Janas et al. (2000) ressaltam que a atividade da fenilalanina amônia-liase e o acúmulo de fenólicos ocorrem como consequência de estresses bióticos e estresses abióticos como temperatura e disponibilidade de água. Solecka e Kacperska (1995) relatam que estes estresses podem modificar tanto a síntese da enzima, como a sua atividade, em uma ampla variedade de plantas. Tan (1980) observou em seu experimento com maçãs cultivadas em casa de vegetação que a deficiência de nitrogênio levou a um aumento no acúmulo de PAL nas folhas das macieiras.

Os genes Zm00001d016972, Zm00001d016973, Zm00001d016975 e Zm00001d016977 foram encontrados para os caracteres massa da parte aérea seca e massa total seca e apresentam as seguintes funções, o gene Zm00001d016972 que codifica a proteína da família de transportadores (MIPs) de colina na membrana plasmática que é responsável por regular o movimento de solutos e água e manter uma composição de soluto celular muito diferente daquela do ambiente externo. Segundo Su et al. (2004), mais de 50 genes transportadores de aminoácidos distintos foram identificados no genoma de *Arabidopsis*, o que indica que o transporte de aminoácidos através das membranas é uma característica altamente complexa nas plantas. Esses transportadores estão localizados em tecidos jovens que se dividem rapidamente, como folhas jovens e meristema apical da raiz. Chaumon et al., (2000) encontraram resultados que indicam que as células de milho contêm vários MIPs em sua membrana plasmática que têm funções diferentes ou são regulados de forma diferenciada. O gene Zm00001d016973 que codifica a proteína GID2, uma subunidade F-box do complexo SCF E3 que interage especificamente com a proteína SLR1 fosforilada e regula a degradação dependente de giberelina (Gomi et al., 2003), que são hormônios que induzem respostas de crescimento das plantas, alongamento de hipocótilo, alongamento de caule e expansão de folha (Olszewski et al., 2002). Até o momento, várias proteínas F-box

identificadas em *Arabidopsis* e *Antirrhinum* demonstraram desempenhar papéis importantes na transdução de sinal de auxina, formação de órgão floral, floração e senescência foliar (Hongyun et al., 2002). O gene Zm00001d016975 envolvido no reparo por excisão de base de pirimidinas oxidadas ao DNA em cloroplasto que são os locais de fotossíntese nas plantas, são também locais importantes para a produção de espécies reativas de oxigênio, que podem danificar componentes essenciais do cloroplasto (Gutman e Niyogi, 2009). Malhotra e Sowdhamini (2013) utilizaram o genoma totalmente sequenciado de *Arabidopsis thaliana* para sequenciar o genoma para proteínas de ligação ao DNA (DBPs) que ajudará a compreender sua distribuição e as funções associadas em um genoma específico. Os autores encontraram que a DNA-(sítioapurínico ou apirimidínico) liase, tem a função de regulação do gene e reparo do DNA. E o gene Zm00001d016977 que codifica a proteína da família de cadeia leve de dineína tipo 1 que são um complexo de proteínas que move a carga em direção à extremidade negativa dos microtúbulos (King, 2000). A cadeia leve 1 da dineína (LC8), uma proteína altamente conservada, é conhecida por se ligar a uma variedade de polipeptídeos diferentes. Se liga à tubulina e promove a montagem de microtúbulos e induz o agrupamento de microtúbulos reconstituídos *in vitro* (Asthana et al., 2012). De acordo com Havecker et al. (2005), uma compreensão mais aprofundada do papel das cadeias leves de dineína nas plantas é necessário.

Para o caractere relação raiz/parte aérea o gene Zm00001d009743 que contém uma repetição de pentatricopeptídeo (PPR) restaura a fertilidade de plantas estéreis masculinas citoplasmáticas. A esterilidade masculina citoplasmática é codificada por genes mitocondriais aberrantes que impedem o desenvolvimento do pólen (Bentolila et al., 2002). A PPR é direcionada para mitocôndrias ou cloroplastos, liga-se a um ou vários transcritos e influencia sua expressão alterando a sequência de RNA. Sua ação combinada tem efeitos profundos na biogênese e função das organelas e, conseqüentemente, na fotossíntese, respiração, desenvolvimento das plantas e respostas ambientais (Barkan e Small, 2014).

Para o caractere diâmetro do colmo o gene Zm00001d045069 codifica a proteína Ribossomal L19. Os genes RP demonstraram ser regulados diferencialmente por fatores ambientais, tanto abióticos quanto bióticos, que afetam diretamente o crescimento da planta e a regulação da transcrição dos genes RP e, em última instância, a biogênese do ribossomo (Fromont-Racine et al., 2003). Segundo Xiong et al. (2020), existem dois

genes em *Arabidopsis thaliana* que codifica a proteína ribossômica (RP). Um dos genes impediu o crescimento e levou a anormalidades morfológicas nas plantas. Moin et al. (2017) sugeriram que os genes RP têm papéis não apenas no crescimento e desenvolvimento, mas também na capacidade de resposta e tolerância ao estresse abiótico e biótico. Para os autores a suprarregulação significativa e imediata desses genes após o início do estresse pode ser uma necessidade celular para manter a integridade do complexo ribossômico, de modo que a tradução de outras proteínas não seja prejudicada, conferindo, assim, uma defesa precoce à planta contra o estresse iminente. Hernández-Hermenegildo et al. (2018) identificaram diferenças na abundância de proteínas ribossômicas (RP) em embriões de milho em diferentes estágios de desenvolvimento. Eles concluíram que a identificação de ribossomos heterogêneos pode indicar um importante nível de regulação da expressão gênica em plantas, supostamente devido à sua necessidade de se adaptar rapidamente às mudanças ambientais. Mais estudos são necessários para elucidar totalmente esse tipo de regulação.

Para o caractere altura de planta os genes Zm00001d026398, Zm00001d026394, Zm00001d026395, Zm00001d026396, Zm00001d026397 e Zm00001d026401, que codificam para fator de transcrição TGA4, provável galacturonosiltransferase 10, proteína semelhante a poliproteína de histona ligante micronuclear, Proteína fator de padrão epidérmico, formação da proteína 1 das junções cristas e a proteína de ligação a RNA AKIP 1% 3B e por fim, a proteína de ligação a RNA AKIP1 isoforma 2, respectivamente. Os membros da família TGA4 são fatores de transcrição que estão ligados as respostas de defesa/estresse, eles foram identificados por Maofu et al. (2015), na maioria dos tecidos, com alta expressão em folhas e baixa expressão em botões florais de *A. thaliana*. Os resultados encontrados por esses autores sugerem que o TGA4 pode atuar como regulador no controle da floração em plantas.

6.4.2. Seleção genômica

A seleção genômica está se tornando uma ferramenta importante para prever genótipos de materiais não fenotipados, aumentar a eficiência do melhoramento por seleção precoce e acelerar o ganho genético (Crossa et al., 2010; Edriss et al., 2017). Diversos fatores podem afetar a acurácia da seleção genômica, como o tamanho da população de treinamento, a relação genética, a herdabilidade, a densidade do marcador e os métodos estatísticos (Technow et al., 2013; Wang et al., 2018; Cerrudo et al., 2018;

Edwards et al., 2019; Zhang et al., 2019). Neste estudo, a seleção genômica apresentou baixa acurácia de predição. As acurácias do modelo que utilizaram apenas os SNPs associados aos caracteres foram significativamente maiores do que as acurácias obtidas no modelo que incorporaram todos os SNPs. De acordo com Wang et al. (2019) a utilização dos resultados da análise de associação (GWAS) nos modelos de predição genômica podem melhorar a precisão da seleção, pois a variação da genética explicada pelos marcadores é um fator importante que afeta a acurácia da seleção genômica. Liu et al. (2015) concluiu em seu trabalho que apenas os SNPs associados aos caracteres eram suficientes para obter boa acurácia. Chaikam et al. (2019) em seus estudos de associação genômica para identificação de regiões genômicas que influenciam a fertilidade espontânea em haploides (HF) de milho, observaram uma acurácia moderada para os caracteres avaliados. Eles afirmaram que estes resultados são devido à natureza quantitativa dos caracteres e que a fertilidade em haploides de milho parece ser um caractere mais complexo, pois apresentou menor acurácia de predição (0,37) ou pode estar associado à ausência de alelos favoráveis para HF na maioria das linhagens no painel avaliado.

6.5. CONCLUSÃO

A análise de GWAS evidenciaram associações significativas para os caracteres de morfologia de raiz relacionados a eficiência no uso de nitrogênio em milho tropical. Alguns SNPs detectados estão localizados em regiões já apresentadas na literatura e contém genes candidatos ou modelos de genes com funções celulares relacionados a proteção das plantas a estresses ambientais, incluindo estresse de nitrogênio. Os resultados da GWAS nos modelos de GWS podem aumentar a acurácia e melhorar a precisão da seleção. Os genes encontrados podem ser mais estudados para ajudar a entender a base genética do desenvolvimento radicular e auxiliar no desenvolvimento de cultivares de milho mais eficientes no uso de nitrogênio.

6.6. REFERÊNCIA BIBLIOGRÁFICA

Andersen, J. R.; Zein, I.; Wenzel, G.; Krütfeldt, B. High levels of linkage disequilibrium and associations with forage quality at a Phenylalanine Ammonia-Lyase locus in European maize (*Zea mays* L.) inbreds. *Theoretical and Applied Genetics* 114(2):307-19, 2007.

- Andorf, C.; Beavis, W. D.; Hufford, M.; Smith, S.; Suza, W. P.; Wang, K.; Woodhouse, M.; J Yu, J.; Lübberstedt, T. Technological advances in maize breeding: past, present and future. *Theoretical and Applied Genetics*, v.132, p.817–849, 2019.
- Asthana, J.; Kuchibhaatla, A.; Jana, S. C.; Ray, K.; Panda, D. Dynein Light chain 1 (LC8) association enhances microtubule stability and promotes microtubule bundling. *The journal of biological chemistry*, v.287, n.48, p.793-805, 2012.
- Bach-Pages, M.; Chen; Sanguankiattichai, N.; Soldan, R.; Kaschani, F.; Kaiser, M.; Mohammed, S.; Hoorn, R. A. L.; Castello, A.; Preston, G. M. Proteome-wide Profiling of RNA-Binding Protein Responses to flg22 Reveals Novel Components of Plant Immunity. *BioRxiv preprint*, 2020.
- Bentolila, S.; Alfonso, A. A.; Hanson, M. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Plant biology*, v.99, n.16, p.10887-10892, 2002.
- Barkan, A.; Small, I. Pentatricopeptide Repeat Proteins in Plants. *Plant Biology*, v. 65, p.415-442, 2014.
- Cerrudo, D.; Cao, S.; Yuan Y., et al. Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a Maize Doubled Haploid Population Across Water Treatments. *Front Plant Sci* 9, 2018.
- Clark, S. A.; Hickey, J. M.; van der Werf, J. H. Different models of genetic variation and their effect on genomic evaluation. *Genet Sel Evol*, v.43, p.18, 2011.
- Chaikam, V.; Gowda, M.; Nair, S. K.; Melchinger, A. E. Prasanna M. Boddupalli Genome-wide association study to identify genomic regions influencing spontaneous fertility in maize haploids. *Euphytica*, 215:138, 2019.
- Chaumont, F.; Barrieu, F.; Jung, R.; Chrispeels, M. J. Plasma Membrane Intrinsic Proteins from Maize Cluster in Two Sequence Subgroups with Differential Aquaporin Activity. *Plant Physiology*, v. 122, p.1025–1034, 2000.

- Chun, L.; Mi, G.; Li, J.; Chen, F.; Zhang, F. Genetic analysis of maize root characteristics in response to low nitrogen stress. *Plant and Soil*, v.276, p.369-382, 2005.
- Crossa J, Campos G de los, Pérez P, et al. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186:713–724, 2010.
- Edelmann, H.; Schopfer, P. Role of protein and RNA synthesis in the initiation of auxin-mediated growth in coleoptiles of *Zea mays* L. *Planta*, v.179, p.475–485, 1989.
- Edwards SM, Buntjer JB, Jackson R, et al. The effects of training population design on genomic prediction accuracy in wheat. *Theor Appl Genet* 132:1943–1952, 2019.
- Edriss V, Gao Y, Zhang X, et al. Genomic Prediction in a Large African Maize Population. *Crop Sci* 57:2361, 2017.
- Fernando, R. L.; Habier, D.; Stricker, C.; Dekkers, J. C. M.; Totir, L. R. Genomic selection. *Acta Agriculturae Scandinavica*, v.57, n.4, p.109-195,2007.
- Fromont-Racinea, M.; BrunoSenger, B.; Saveanu, C.; Fasiolo, F. Ribosome assembly in eukaryotes. v.313, p.17-42, 2003.
- Gallais, A.; Hirel, B. An approach to the genetics of nitrogen use efficiency in maize. *Journal of Experimental Botany*, v. 55, n. 396, p. 295-306, 2004.
- Garnett, T.; Conn, V.; Kaiser, B.N. Root based approaches to improving nitrogen use efficiency in plants. *Plant, Cell and Environment*, v.32, p.1272-83, 2009.
- Gomi, K.; Yamamoto, H.; Akimitsu, K. Epoxide hydrolase: a mRNA induced by the fungal pathogen *Alternaria alternata* on rough lemon (*Citrus jambhiri* Lush). *Plant Molecular Biology*, v.53, p.189–199, 2003.
- Gutman, B. L.; Niyogi, K. K.; Evidence for base excision repair of oxidative DNA damage in chloroplasts of *Arabidopsis thaliana*. *Journal of biological chemistry*, 2009.

- Havecker, E. R.; Gao, X.; Voytas, D. F. The sireviruses, a plant-specific lineage of the Ty1/copia retrotransposons, interact with a family of proteins related to dynein light chain 8. *Plant Physiology*, v.139, p.857-868, 2005.
- Hernández-Hermenegildo, R. A.; Bernal, L.; Jiménez-Pérez, L.V.; Bernal-Lugo, I.; Jiménez, E. S. Ribosomal Heterogeneity of Maize Tissues: Insights of Biological Relevance. *Plant Molecular Biology Reporter*, v.36:491–499, 2018.
- Hirel, B.; Bertin, P.; Quillere, I. et al. Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. *Am Soc Plant Physiol* 125:1258–1270, 2001.
- Hongyun, W.; Jian, H.; Zhao, L.; Yongbiao, X. F-box proteins in flowering plants. *Chinese Science Bulletin*, v. 47, 2002.
- Horst, W. J.; Kamh, M.; Jibrin, J. M.; Chude, V. O. Agronomic measures for increasing P availability to crops. *Plant Soil*, v.237, p.211-233, 2001.
- Janas, K. M.; Cvikrová, M.; Pałagiewicz, A.; Eder, J. Alterations in phenylpropanoid content in soybean roots during low temperature acclimation. *Plant Physiology and Biochemistry*, v.38, p.587-593, 2000.
- Jones, A.M.P.; Shukla, M.R.; Chattopadhyay, A.; Zoń, J.; Saxena, P.K. Investigating the roles of phenylpropanoids in the growth and development of *Zea mays* L. In *In Vitro Cellular & Developmental Biology - Plant* v.49, p.765–772, 2013.
- Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011.
- Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS. *Plant Methods*, v.9, p.29, 2013.
- Laperche, A.; Aigu, Y.; Jubault, M. Clubroot resistance QTL are modulated by nitrogen input in *Brassica Napus*. *Theor Appl Genet*, v.130, p.669–684, 2017.

- Jia, M., Li, Y.; Lei, L.; Di, D.; Miao, H.; Fan, Z. Alteration of gene expression profile in maize infected with a double-stranded RNA fijivirus associated with symptom development. *Molecular Plant Pathology*, v.13(3), p.251–262, 2012.
- Ma, A.; Qingb, C; Freia, U; Shenb, Y; Lübberstedta, T. Association mapping for root system architecture traits under two nitrogen conditions in germplasm enhancement of maize doubled haploid lines. *The Crop Journal*, v.8; p.213–226, 2020.
- Malhotra, S.; Sowdhamini, R. Genome-wide survey of DNA-binding proteins in *Arabidopsis thaliana*: analysis of distribution and functions. *Nucleic Acids Research*, 1–8, 2013.
- Maofu, L.; Wang, H.; Yang, Y.; Jin, W. Capsella rubella TGA4, a bZIP transcription factor, cause delayed flowering in *Arabidopsis thaliana*. *Archives of biological Sciences*, v68:141, 2015.
- Meuwssen, T. H. E.; Hayes, B. J.; Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps, *Genetics*, v.157, p.1819-1829, 2001.
- Moin, M.; Bakshi, A.; Madhav, M. S.; Kirti, P. B. Expression Profiling of Ribosomal Protein Gene Family in Dehydration Stress Responses and Characterization of Transgenic Rice Plants Overexpressing RPL23A for Water-Use Efficiency and Tolerance to Drought and Salt Stresses. *Frontiers Chemistry*, 2017.
- Morosini, J. S.; Mendonça, L. F.; Lyra, D. H.; Galli, G.; Vidotti, M. S.; Fritsche-Neto, R. Association mapping for traits related to nitrogen use. *Plant Soil*, v.421, p.453–463, 2017.
- Olszewski, N.; Sun, T.; Gubler, F. Gibberellin Signaling: Biosynthesis, Catabolism, and Response Pathways. *The Plant Cell*, 2002.
- Pace, J.; Gardner, C.; Romay, C.; Ganapathysubramanian, B.; Lubberstedt, T. Genome-wide association analysis of seedling root development in maize (*Zea mays* L.). *BMC Genomics*, v.16 p.47, 2015a.

- Pace, J.; Yu, X.; Lübberstedt, T. Genomic prediction of seedling root length in maize (*Zea mays* L.). *Plant J* 83:903–912, 2015b.
- Romano, P.G.N.; Horton, P.; Gray, J.E. The Arabidopsis Cyclophilin Gene Family. *Plant Physiology*, v.134, p.1268–1282, 2004.
- Saengwilai, P.; Tian, X.; Lynch, J. P. Low crown root number enhances nitrogen acquisition from low-nitrogen soils in maize. *Plant Physiol*, v.166, p.581–589, 2014.
- Sanchez, D.L., Liu, S.; Ibrahim, R.; Blanco, M.; Lübberstedt, T. Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (*Zea mays* L.). *Plant Sci*, v.268, p.30–38, 2018.
- Solecka, D.; Kacperska, A. Phenylalanine ammonia-lyase activity in leaves of winter oilseed rape plants as affected by acclimation of plants to low temperature. *Plant physiology and biochemistry*, 1995.
- Song, H.; Mao, W.; Duan, Z.; Que, W.; Zhou, W.; Chen, X.; Li, P. Selection and validation of reference genes for measuring gene expression in *Toona ciliata* under different experimental conditions by quantitative real-time PCR analysis. *BMC plant biology*, v.20, 2020.
- Su, Y. H.; Frommer, W.B.; Ludewig, U. Molecular and Functional Characterization of a Family of Amino Acid Transporters from Arabidopsis. *Plant Physiology*, v.136, p. 3104–3111, 2004.
- Tan, S.C. Phenylalanine Ammonia-Lyase and the Phenylalanine Ammonia-Lyase Inactivating System: Effects of Light, Temperature and Mineral Deficiencies, *Australian journal of plant physiology*, 7(2):159, 1980.
- Torres, L. G.; Caixeta, D. G.; Rezende, W. M.; Schuster, A.; Silva, F. F.; DeLima, R. O. Genotypic variation and relationships among traits for root morphology in a panel of tropical maize inbred lines under contrasting nitrogen levels. *Euphytica*, v.215, p.51, 2019.

- Thornberry, J.M.; Goodman, M.M.; Doebley, J.; Kresovich, S.; Nielsen, D. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* v.28, p.286–89, 2001.
- Wang, X.; Xu, Y.; Hu, Z.; Xu, C. Genomic selection methods for crop improvement: Current status and prospects. *Crop J.* v.6, p.330–340, 2018.
- Wang, H.; Wei, J.; Li, P.; Wang, Y.; Ge, Z.; Quian, J.; Fan, Y.; Ni, J.; Xu, Y.; Yang, Z.; Xu, C. Integrating GWAS and Gene Expression Analysis Identifies Candidate Genes for Root Morphology Traits in Maize at the Seedling Stage. *Genes (Basel)* v.10, p.773, 2019.
- Qing-Jun Wang, QJ; Yuan, Y; Liao, Z; Jiang, Y; Wang, Q; Zhang, L; Gao, S.; Wu, F; Li, M; Xie, W; Liu, T; Xu, J; Liu, Y; Feng, X; Lu, Y. Genome-Wide Association Study of 13 Traits in Maize Seedlings under Low Phosphorus Stress. *The plant genome*, v12, no. 3, 2019.
- Wang, N.; Liu, B.; Liang, X.; Zhou, Y. Song, J.; Yang, J. Yong, H.; Weng, J.; Zhang, D.; Li, M.; Nair, S.; Vicente, F. S.; Hao, Z.; Zhang, X.; Li, X. Genome-wide association study and genomic prediction analyses of drought stress tolerance in China in a collection of off-PVP maize inbred lines. *Mol Breeding*, v.39, p.113, 2019.
- Woodward, A.; Bartel, B. The Arabidopsis Peroxisomal Targeting Signal Type 2 Receptor PEX7 Is Necessary for Peroxisome Function and Dependent on PEX5. *Molecular Biology of the Cell*, v. 16, n.2, 2005.
- Wu, Y.; Vicente, F. S.; Huang, K.; Dhliwayo, T.; Costich, D. E.; Semagn, K.; Sudha, N.; Olsen, M.; Prasanna, B. M.; Zhang, X.; Babu, R. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theor Appl Genet*, v.129, p.753-765, 2016.
- Yang, L.; Li, T.; Tian, X.; Yang, B.; Lao, Y.; Wang, Y.; Zhang, X.; Xue, J.; Xu, S. Genome-wide association study (GWAS) reveal genetic basis of ear-related traits in maize. *Euphytica*, v.216, p.172, 2020.

Xiong, W.; Chen, X.; Zhu, C.; Zhang, J. Arabidopsis ribosomal proteins RPL23aA and RPL23aB are functionally equivalent.

Zhang, N.; Gibon, Y.; Wallace, J.G. Genome-wide Association of Carbon and Nitrogen Metabolism in the maize nested association mapping population. *Plant Physiol*, v.168, p.575–583, 2015.

Zaidi, P.H., Seetharam K, Krishna G, et al. Genomic regions associated with root traits under drought stress in tropical maize (*Zea mays* L.). *PLoS One* 11:1–18, 2016.

Zaidi, P. H.; Seetharam, K.; Krishna, G.; Krishnamurthy, L.; Gajanan, S.; Babu, R.; Zerka, M.; Vinayan, M. T.; Vivek B. S. Genomic Regions Associated with Root Traits under Drought Stress in Tropical Maize (*Zea mays* L.). *PLoS ONE*, 11(10), 2020.

7. CONCLUSÃO GERAL

Os resultados encontrados no Capítulo 1 mostram uma grande variação fenotípica e genotípica no conjunto de linhagens endogâmicas de milho tropical do programa de melhoramento de milho da UFV que pode ser explorada para fins de melhoramento de milho. Além disso, foram identificados três grupos heteróticos, através da análise de estrutura populacional e PCA, que juntamente com a análise de agrupamento serão úteis para auxiliar na seleção das melhores combinações parentais para novos cruzamentos, mapeamento de populações, cruzamentos orientados para obtenção de híbridos altamente heteróticos e produtivos e na predição de híbridos não testados no programa de melhoramento de milho da UFV.

Os resultados também mostraram que há baixos coeficientes de parentesco relativo e alta distância genética (com base em dados fenotípicos e genotípicos) entre a maioria dos pares de linhagens endogâmicas, juntamente com curta distância de decaimento de LD em todo o conjunto de linhagens. Isso indica claramente a singularidade da maioria das linhagens endogâmicas que podem contribuir com novos alelos para o programa de melhoramento da UFV e, além disso, podem compreender um painel de linhagens endogâmicas de milho tropical que é adequado para o estudo de associação genômica ampla abordado no Capítulo 2.

Os resultados encontrados no Capítulo 2 mostra que o painel de linhagens endogâmicas de milho do programa de melhoramento de milho da UFV é fonte de novos genes para selecionar os caracteres de interesse. A análise de GWAS evidenciaram associações significativas para os caracteres de morfologia de raiz relacionados a eficiência no uso de nitrogênio em milho tropical. Alguns SNPs detectados estão localizados em regiões já apresentadas na literatura e contém genes candidatos ou modelos de genes com funções celulares relacionados a proteção das plantas a estresses ambientais, incluindo estresse de nitrogênio. Os resultados da GWAS nos modelos de GWS podem aumentar a acurácia e melhorar a precisão da seleção. Os genes encontrados podem ser mais estudados para ajudar a entender a base genética do desenvolvimento radicular e auxiliar no desenvolvimento de cultivares de milho mais eficientes no uso de nitrogênio.