

GUSTAVO CÉSAR SANT'ANA

**RELAÇÃO DA ESTRUTURA GENÉTICA COM A EFICIÊNCIA NO USO DE
NITROGÊNIO E COM A ACURÁCIA DE GWAS EM 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*.

VIÇOSA
MINAS GERAIS – BRASIL
2015

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

S232r
2015
Sant'Ana, Gustavo César, 1980-
Relação da estrutura genética com a eficiência no uso de
nitrogênio e com a acurácia de GWAS em milho tropical /
Gustavo César Sant'Ana. – Viçosa, MG, 2015.
ix, 66f. : il. (algumas color.) ; 29 cm.

Orientador: Aluizio Borém de Oliveira.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Milho - Melhoramento genético. 2. Planta - Efeito do
nitrogênio. I. Universidade Federal de Viçosa. Departamento de
Fitotecnia. Programa de Pós-graduação em Genética e
Melhoramento. II. Título.

CDD 22. ed. 631.52

GUSTAVO CÉSAR SANT'ANA

**RELAÇÃO DA ESTRUTURA GENÉTICA COM A EFICIÊNCIA NO USO DE
NITROGÊNIO E COM A ACURÁCIADE GWAS EM 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: 27 de Março de 2015.

Carlos Siqueyuki Sedyama

Rogério Faria Vieira

Eveline Teixeira Caixeta
(Coorientadora)

Roberto Fritsche Neto
(Coorientador)

Aluízio Borém de Oliveira
(Orientador)

*A minha filha Lúza,
Dedico.*

AGRADECIMENTOS

Agradeço à minha namorada Connie Gallardo pelo carinho, compreensão e apoio.

Ao Professor Aluizio Borém de Oliveira pela orientação e por todo apoio e incentivo.

Ao Professor Roberto Fritsche Neto, da ESALQ-USP, pela oportunidade, pelo incentivo e pelo empenho dedicados a este trabalho.

A Doutora Eveline Teixeira Caixeta pela colaboração e pela disponibilidade em ajudar sempre que necessário.

Ao professor Carlos Sigueyuki Sedyamae ao Dr. Rogério Faria Vieira pela colaboração e pelas valiosas sugestões.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade de realização do doutorado.

A Universidade de São Paulo, em especial ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas da Escola Superior de Agricultura “Luiz de Queiroz”, pela oportunidade e pelo grande apoio.

Aos funcionários da USP/ESALQ e UFV, pela colaboração na condução dos experimentos.

Aos colegas de trabalho do Laboratório de Melhoramento de Alógamias da USP/ESALQ, pela colaboração.

Ao Laboratório BioCafé e ao Programa Milho da UFV pelo auxílio.

A Capes e ao CNPq pela concessão da bolsa de estudos.

A FAPESP (processo 2013/24135-2), a FAPEMIG e a Pioneer Hi-Bred International Int. (Johnston, IA - USA) pelo financiamento da pesquisa.

SUMÁRIO

RESUMO	vi
ABSTRACT	viii
1 INTRODUÇÃO GERAL	1
2 CAPÍTULO 1	3
3 Resumo	4
4 Abstract	5
5 Introdução.....	6
6 Objetivos e Hipóteses.....	8
a. Objetivo Geral.....	8
b. Hipóteses Testadas	8
7 Materiais e Métodos	9
a. Material Vegetal e condução dos experimentos em campo.....	9
b. Avaliação da Eficiência no Uso do Nitrogênio (EUN)	9
c. Análise dos Dados Fenotípicos	10
d. Genotipagem	10
e. Análise dos Dados Moleculares.....	11
8 Resultados e Discussão	14
a. Análise Fenotípica da EUN.....	14
b. Análise dos marcadores SNPs	16
c. Diversidade Genética.....	18
d. Análise da estrutura genética da população e da sua relação com a EUN	25
9 Conclusões	33
10 Referências Bibliográficas	34
11 CAPÍTULO 2	38
12 Resumo	39
13 Abstract	40
14 Introdução.....	41
15 Objetivo e Hipóteses testadas	43
a. Objetivo Geral.....	43
b. Hipóteses Testadas	43
16 Materiais e Métodos	44
a. Material Genético e Fenotipagem.....	44
b. Genotipagem	44
c. Análise da estrutura genética da população	45
d. Análises de detecção de QTLs via GWAS.....	45
e. Análise do efeito da estrutura populacional e do Critério de Evanno sobre a acurácia em GWAS	46
17 Resultados e Discussão	48
a. Análise Fenotípica	48
b. Análise da Estrutura Genética da População	49

c. GWAS.....	51
d. Análise do efeito do número de grupos e do critério de Evanno na acurácia em GWAS.....	54
18 Conclusões.....	61
19 Referências Bibliográficas	62

RESUMO

SANT'ANA, Gustavo César, D.Sc. Universidade Federal de Viçosa, Março de 2015. **Relação da estrutura genética com a eficiência no uso de nitrogênio e com a acurácia de GWAS em milho tropical.** Orientador: Aluizio Borém de Oliveira. Coorientadores: Eveline Teixeira Caixeta e Roberto Fritsche Neto.

Os marcadores moleculares SNP (*Single Nucleotide Polymorphism*) têm sido utilizados com sucesso em estudos de diversidade e estrutura genética de populações de milho. Os SNPs também têm sido amplamente empregados em estudos de associação genômica ampla (Genome Wide Association Studies - GWAS). Na primeira parte desse estudo, os objetivos foram analisar a estrutura genética e estudar a relação entre a variação genômica, a estrutura genética e a variação fenotípica para eficiência no uso de nitrogênio (EUN) em linhagens de milho tropical. As linhagens foram avaliadas a campo em relação à EUN, e genotipadas com 768 marcadores moleculares SNPs (*Single Nucleotide Polymorphisms*). O conjunto de 62 linhagens analisado apresentou elevado grau de diversidade genética, comparável à de painéis maiores analisados em outros estudos. Isso indica que essas linhagens podem ser incluídas em painéis úteis para estudos de associação genômica ampla para EUN. Houve forte relação entre o desempenho fenotípico para a EUN e a estruturação genética das linhagens de milho tropical com base nos marcadores SNPs pelo método bayesiano, principalmente quando a população foi subdividida em quatro grupos. Foram observados alelos exclusivos, ausentes ou com grandes diferenças de frequências entre os grupos fenotípicos com diferentes níveis de EUN, e as diferenças alélicas foram maiores entre os grupos mais distintos fenotipicamente. Foi possível identificar genitores superiores quanto à EUN e geneticamente divergentes, o que pode possibilitar a exploração mais eficiente da heterose na produção de híbridos a partir de cruzamentos entre essas linhagens de milho tropical. A segunda parte desse estudo teve como objetivos analisar o efeito da estrutura

genética da população, do número de subgrupos em que a população é estruturada e do número ótimo de grupos com base no critério de Evanno sobre a acurácia em GWAS para características oligogênicas e poligênicas em linhagens de milho tropical. Para isso, 62 linhagens foram genotipadas com 768 SNPs e fenotipadas para altura de planta e produtividade de grãos. Foi realizada a análise da estruturação da população em diferentes números de grupos ($K=1$ a 9), com auxílio do programa STRUCTURE. As análises de GWAS foram realizadas usando o programa TASSEL. O número de grupos e o valor de ΔK influenciaram de forma distinta a acurácia dos modelos de GWAS para as diferentes características. No caso da altura de planta, a inclusão da matriz Q nos modelos levou a um maior ajuste destes, principalmente nos casos em que os números de grupos apresentaram valores de ΔK mais elevados. Para a produtividade de grãos, a inclusão da matriz Q reduziu a acurácia do modelo para todos os valores de K , independentemente dos valores de ΔK . Portanto, os resultados indicam que a escolha do número ótimo de grupos geneticamente estruturados, com base no critério de Evanno, pode não levar a maiores acurácias dos modelos em GWAS. Além disso, ficou evidente que a estruturação influencia de maneira distinta características oligogênicas e poligênicas.

ABSTRACT

SANT'ANA, Gustavo César, D.Sc. Universidade Federal de Viçosa, March, 2015. **Relationship of genetic structure with nitrogen use efficiency and accuracy of GWAS in tropical maize.** Adviser: Aluizio Borém de Oliveira. Co-Advisers: Eveline Teixeira Caixeta and Roberto Fritsche Neto.

Single Nucleotide Polymorphism (SNP) has been used successfully in genetic diversity and populations structure studies of maize. SNPs have also been widely used in genome-wide association studies (Genome Wide Association Studies - GWAS). In the first part of this study, the objectives were to analyze the genetic structure of 62 inbred lines of tropical maize, and to study the relationship between genomic variation, genetic structure and phenotypic variation for nitrogen use efficiency. The lines were phenotyped to NUE and genotyped with 768 SNPs markers. The set of 62 inbred lines showed a high level of genetic diversity comparable to other panels evaluated by other scientists. This indicates that these lines could be included in panels for genome wide association studies for NUE. The existence of a strong relationship between the phenotypic performance for NUE and the genetic structure of tropical maize lines was detected. It was observed the occurrence of exclusive alleles, missing alleles or alleles with large frequency differences between the phenotypic groups of tropical maize lines with different levels of NUE, and the allelic differences were greater among the most phenotypically divergent groups regarding NUE. The data obtained allow identifying superior parents in relation to NUE that are genetically divergent, which may enable more efficient heterosis exploitation in the production of hybrids from crosses among the tropical maize lines. The second part of this study aimed to analyze the effect of the genetic structure of the population, the number of subgroups in the population is structured and the optimal number of groups based on criteria Evanno on accuracy in GWAS for different characteristics in tropical maize lines. To this 62 lines were genotyped with 768 SNPs and phenotyped for plant height (PH) and grain yield (GY). Then the analysis was performed structuring of the population in different numbers of groups ($k = 1$ to 9) by means of Bayesian model, using the STRUCTURE program. The analysis of GWAS were performed using TASSEL

program. The results indicated that the number of groups and the value of ΔK affect differently the accuracy of GWAS models for different characteristics. In the case of PH, the inclusion of Q matrix led to a better adjustment of model, especially in cases where the number of the groups had higher ΔK values. In the case of GY, the inclusion of the Q matrix reduced the accuracy of the model for all K values, regardless of the values of ΔK . Therefore, the results indicate that the choice of the optimal number of genetically structured groups through Evanno criterion cannot lead to higher accuracies of models in GWAS. Furthermore, it was evident that the structure affects differently oligogenic and polygenic traits.

1 Introdução Geral

No mundo, cerca de 160 milhões de hectares são cultivados anualmente com milho (Food and Agriculture Organization of the United Nations, 2014). Aproximadamente 50% desse cultivo é realizado em regiões tropicais, nas quais predominam condições de baixa fertilidade natural dos solos e ocorrência de secas. No Brasil o milho é o principal cereal cultivado, sendo que o país é o terceiro maior produtor mundial. A área plantada na primeira safra no período 2012/2013 foi 6,90 milhões de hectares e, na segunda safra, 8,99 milhões, com produtividades médias de 5.146 kg ha⁻¹ e 3.911 kg ha⁻¹, respectivamente (Companhia Nacional de Abastecimento, 2014). Essa produtividade é baixa, em relação ao potencial produtivo dos cultivares disponíveis no mercado (Borém, Galvão e Pimentel, 2015).

Considerando apenas as áreas aptas ao cultivo, a produção de milho não será suficiente para atender à demanda mundial nas próximas décadas (Fritsche-Neto e Borém, 2012). Além da baixa disponibilidade de água e elevada acidez dos solos, algumas regiões do cerrado possuem limitação à produção agrícola devido à baixa fertilidade do solo. Assim, nos últimos anos, o interesse pela agricultura sustentável, com menor consumo energético e ecologicamente correta, tem estimulado os programas de melhoramento de plantas a desenvolverem cultivares que possuam maior eficiência nutricional, ou seja, que apresentem produtividade satisfatória em condições de estresse nutricional e, portanto, com menor custo de produção e menor impacto ambiental (Fritsche-Neto e Borém, 2012; Abdel-Ghani et al., 2013; Xu et al., 2014).

Entre os diversos tipos de estresses abióticos, a deficiência de nitrogênio (N) é comum nos solos brasileiros, afetando negativamente o desenvolvimento das plantas e reduzindo a produtividade (Fritsche-Neto e Borém, 2012). Assim, o desenvolvimento de cultivares eficientes no uso de N apresenta-se como uma opção economicamente viável e ecologicamente sustentável. Para essa finalidade, o conhecimento da diversidade genética dentro do germoplasma a ser melhorado auxilia na organização da variabilidade desses materiais e potencializa a maximização dos ganhos genéticos (Romay et al., 2013).

Os marcadores moleculares denominados polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphism* – SNP) têm sido utilizados com sucesso em estudos da

diversidade e estrutura genética de populações de milho. Os SNPs também têm sido amplamente empregados em estudos de associação genômica ampla (Genome Wide Association Studies - GWAS). No caso do milho, GWAS têm sido utilizados com sucesso em pesquisas para a detecção de genes relacionados a diversas características quantitativas (Cook et al., 2012; Almeida et al., 2014; Kuma et al., 2014; Pace et al., 2014; Peiffer et al., 2014; Rincent et al., 2014; Sobkowiak et al., 2014).

A estrutura genética da população é um fator importante que pode influenciar consideravelmente os resultados em GWAS (Wang et al., 2014; Li 2014), e pode existir em populações aleatórias ou com *pedigree* definido e conhecido, devido à geografia, à seleção natural ou à seleção artificial (van Heerwaarden et al., 2011). Devido à ocorrência de diferentes frequências alélicas entre subpopulações, a estrutura populacional pode produzir falsas associações entre os marcadores e as características fenotípicas em GWAS (Price, 2010). Consequentemente, essas falsas associações podem inflacionar as estimativas de herdabilidade e reduzir a acurácia das predições genômicas (Riedelsheimer et al 2012; Wray et al 2013).

Neste contexto, métodos estatísticos têm sido desenvolvidos com o objetivo de se evitar essas associações espúrias (Zhang et al., 2010, Li et al., 2014). No entanto, inferir o número de grupos que melhor se ajusta aos dados é um desafio, e uma série de métodos foram desenvolvidos para tal (Pritchard et al., 2000; Evanno et al. 2005; Schwartz e McKelvey, 2009). Um dos métodos mais utilizados para definir o número ótimo de grupos em que a população está estruturada geneticamente é o critério de Evanno (Evanno et al., 2005), o qual é baseado no valor de ΔK , que expressa a taxa de mudança de segunda ordem do logaritmo da probabilidade marginal dos dados (Evanno et al., 2005).

Apesar do grande número de pesquisas utilizando GWAS e a estruturação de populações via modelos bayesianos para detecção de genes relacionados a características de interesse (Stich et al., 2008; Caniato et al., 2011; Westenge et al., 2012), ainda é reduzido o número de estudos que avaliem o efeito do número de grupos em que a população é estruturada e do número ótimo de grupos com base no critério de Evanno (Evanno et al., 2005) sobre a acurácia em GWAS.

2 CAPÍTULO 1

**Caracterização genética e estrutura populacional de linhagens de milho tropical
para a eficiência no uso de nitrogênio**

3 Resumo

O crescente interesse pela agricultura sustentável tem estimulado os programas de melhoramento a desenvolverem cultivares com maior eficiência nutricional. O melhoramento para aumentar a eficiência no uso de nitrogênio (N) requer a disponibilidade de variabilidade genética adequada. O presente trabalho teve como objetivo analisar a estrutura genética e estudar a relação entre a variação genômica, a estrutura genética e a variação fenotípica para eficiência no uso de nitrogênio (EUN) em linhagens de milho tropical. As linhagens foram avaliadas a campo em relação à EUN, e genotipadas com 768 marcadores moleculares SNPs (*Single Nucleotide Polymorphisms*). O conjunto de 62 linhagens analisado apresentou elevado grau de diversidade genética, comparável a de painéis maiores analisados em outros estudos. Isso indica que essas linhagens podem ser incluídas em painéis úteis para estudos de associação genômica ampla para EUN. Houve forte relação entre o desempenho fenotípico para a EUN e a estruturação genética das linhagens de milho tropical com base nos marcadores SNPs pelo método bayesiano, principalmente quando a população foi subdividida em quatro grupos. Foram observados alelos exclusivos, ausentes ou com grandes diferenças de frequências entre os grupos fenotípicos com diferentes níveis de EUN, e as diferenças alélicas foram maiores entre os grupos mais distintos fenotipicamente. Foi possível identificar genitores superiores quanto à EUN e geneticamente divergentes, o que pode possibilitar a exploração mais eficiente da heterose na produção de híbridos a partir de cruzamentos entre essas linhagens de milho tropical.

Palavras-chaves: diversidade genética, polimorfismo de nucleotídeo único, eficiência no uso de nitrogênio.

4 Abstract

Genetic characterization and population structure of tropical maize lines for nitrogen use efficiency

The crescent interest in sustainable agriculture has stimulated the breeding programs to develop cultivars that have higher nutritional efficiency. The breeding for increased the nitrogen use efficiency (NUE) requires the availability of suitable genetic variability. This study aimed to analyze the genetic structure of 62 inbred lines of tropical maize, and to study the relationship between genomic variation, genetic structure and phenotypic variation for nitrogen use efficiency. The lines were phenotyped to NUE and genotyped with 768 SNPs markers. The set of 62 inbred lines showed a high level of genetic diversity comparable to other panels evaluated by other scientists. This indicates that these lines could be included in panels for genome wide association studies for NUE. The existence of a strong relationship between the phenotypic performance for NUE and the genetic structure of tropical maize lines was detected. It was observed the occurrence of exclusive alleles, missing alleles or alleles with large frequency differences between the phenotypic groups of tropical maize lines with different levels of NUE, and the allelic differences were greater among the most phenotypically divergent groups regarding NUE. The data obtained allow identifying superior parents in relation to NUE that are genetically divergent, which may enable more efficient heterosis exploitation in the production of hybrids from crosses among the tropical maize lines.

Keywords: genetic diversity, single nucleotide polymorphism, nitrogen use efficiency.

5 Introdução

Com o intuito de alcançar elevadas produtividades, produtores de milho adotam manejo de suas lavouras com elevadas doses de insumos, como corretivos e fertilizantes, o que implica em altos custos de produção. Segundo o International Plant Nutrition Institute (IPNI, 2014), no Brasil, durante o ano agrícola de 2012, foram utilizados aproximadamente 3,5 milhões de toneladas de nitrogênio. Deste montante, 70% são provenientes de importações. Além disso, pelas estimativas da empresa Petróleo Brasil S/A (PETROBRAS, 2014), o consumo de uréia no país apresenta um crescimento de 3,4% ao ano, e a cultura do milho é a principal consumidora desse fertilizante, consumindo cerca de 1,3 milhões de toneladas por ano.

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. Assim, considerando os dados acima, cada ponto percentual em que o melhoramento genético reduzir a quantidade de uréia necessária no cultivo de milho, significaria uma economia anual de 12,7 mil toneladas do adubo, o que equivale a aproximadamente US\$ 6,8 milhões. Adicionalmente, o aumento da EUN não significa apenas benefícios econômicos, mas também reduz a poluição ambiental. Segundo Cantarella (2007), na cultura do milho, as perdas de N por lixiviação de nitrato chegam a 4% e por volatilização, é de pelo menos 5% do total aplicado. Assim, cada ponto percentual de aumento na EUN reduziria, por ano, 215 toneladas de N nos lençóis freáticos do Brasil e, no mínimo, 269 toneladas de N deixariam de ser liberados na atmosfera sob a forma de amônia. Nesse contexto, o interesse pela agricultura sustentável, com menor consumo energético e ecologicamente correta, tem estimulado os programas de melhoramento a desenvolverem cultivares com maior eficiência nutricional, com um menor custo de produção e menor impacto ambiental, contribuindo para aumentar a segurança alimentar (Fritsche-Neto e Borém, 2012; Abdel-Ghani et al., 2013; Xu et al., 2014).

Existem várias metodologias descritas para mensurar a eficiência nutricional em plantas. Dentre essas, Moll et al. (1982) definiram a eficiência no uso de N (EUN) como a relação entre a produtividade de grãos e a quantidade de nitrogênio disponibilizado para a planta. Todavia, o melhoramento para aumentar a EUN requer variabilidade

genética suficiente (Moll et al., 1985; Anbessa et al., 2009). No caso do milho, a variação e as relações genéticas entre linhagens quanto à EUN é pouco conhecida, o que pode ter impacto significativo na escolha de genitores para a formação de híbridos (Wu et al., 2011).

A utilização eficiente do germoplasma de milho no melhoramento requer a caracterização acurada do desempenho das linhagens a serem utilizadas nos cruzamentos e das relações de parentesco entre elas (Lu et al., 2009; Van Inghelandt et al., 2010; Semagn et al., 2012; Romay et al., 2013; Wu et al., 2014). Para desenvolver populações melhoradas, é desejável a utilização de genitores que apresentem desempenho superior para a característica de interesse, maximize a variância intrapopulacional e preserve o padrão heterótico, de forma a maximizar as respostas heteróticas nos híbridos (Nelson et al., 2008).

Para isso, além de uma acurada caracterização fenotípica, a caracterização detalhada dos acessos com base em dados moleculares pode contribuir estrategicamente para otimizar a utilização do germoplasma disponível em programas de melhoramento de milho (Romay et al., 2013). Neste sentido, marcadores SNPs (*Single Nucleotide Polymorphism*) têm sido desenvolvidos e disponibilizados para diversas espécies de plantas, incluindo o milho (Yan et al., 2010; Elshire et al., 2011; Ganai et al., 2011). Os SNPs apresentam vantagens em relação a outros tipos de marcadores moleculares: é a forma de variação mais abundante no genoma, apresentam alta reprodutibilidade, possibilidade de automação, ampla cobertura do genoma e baixo custo. No caso do milho, esses marcadores vêm sendo amplamente utilizados para a construção de mapas genéticos (Gore et al., 2009), detecção de genes/QTLs (Peiffer et al., 2013), análise de diversidade e estrutura genética populacional (Frascaroli et al., 2013; Wu et al., 2014), verificação de parentesco e análises evolutivas (Van Heerwaarden et al., 2011), análise de arquitetura genética de características complexas (Cook et al., 2012; Peiffer et al., 2014), manutenção da pureza genética (homogeneidade), confirmação da identidade genética de linhagens (Semagn et al., 2012) e análise de associação (Poland et al., 2011; Tian et al., 2011).

6 Objetivos e Hipóteses

a. Objetivo Geral

Analisar a estrutura genética e estudar a relação entre a variação genômica e a variação fenotípica para eficiência no uso de nitrogênio em 62 linhagens endogâmicas de milho tropical.

b. Hipóteses Testadas

H1: A população apresenta diversidade genética suficiente para sua utilização em estudos de associação genômica ampla para EUN em milho tropical.

H2: É possível identificar genótipos superiores em relação à EUN, mas que sejam geneticamente divergentes para explorar melhor a heterose em cruzamentos entre linhagens de milho tropical.

H3: Existem alelos exclusivos, ausentes ou com grandes diferenças de frequências entre os grupos fenotípicos de linhagens de milho tropical com diferentes níveis de EUN.

H4: Há relação entre o desempenho fenotípico para a EUN e a estruturação genética das linhagens de milho tropical.

7 Materiais e Métodos

a. Material Vegetal e condução dos experimentos em campo

Foram avaliadas 64 linhagens endogâmicas de milho tropical, que apresentam variabilidade para EUN (DoVale et al., 2011). Todos os experimentos de campo foram conduzidos na estação experimental Anhembi da Escola Superior de Agricultura Luís de Queiroz (ESALQ), da Universidade de São Paulo (USP) (22°50'51``S, 48°01'06``W, 466h). Para obter sementes em quantidade suficiente e de qualidade para a realização deste estudo, primeiramente foi realizada a multiplicação das linhagens. As parcelas foram constituídas por duas linhas de 4m de comprimento, com espaçamento de 0,80m entre linhas e 0,20m entre plantas. Na semeadura foram colocadas 20 sementes por linha. O manejo e os tratamentos culturais foram realizados de acordo com os procedimentos recomendados para a cultura do milho (Borém, Galvão e Pimentel, 2015).

As linhagens foram distribuídas no campo de acordo com o delineamento experimental látice simples (8x8), com duas repetições. Foi usado um esquema fatorial duplo, constituído por 64 linhas x duas disponibilidades de N (alto e baixo). As parcelas foram constituídas de uma linha com 4m de comprimento, utilizando-se o espaçamento de 0,80m entre linhas e 0,2m entre plantas, totalizando 20 plantas por linha, após desbaste. O plantio foi realizado na segunda safra (safrinha) de 2014.

No alto N (AN) foram aplicados 150 kg.ha⁻¹ de N (27 kg.ha⁻¹ no plantio e 133 kg.ha⁻¹ em cobertura) e no baixo N (BN), 60 kg.ha⁻¹ de N (27 kg.ha⁻¹ no plantio e 33 kg.ha⁻¹ em cobertura). O estabelecimento do baixo nitrogênio (BN) correspondeu à dose em que houve redução de 50% da produtividade de grãos em relação à condição ideal de cultivo, conforme Fritsche-Neto e Borém (2012). O manejo e os demais tratamentos culturais foram realizados de acordo com os procedimentos recomendados para a cultura do milho (Borém, Galvão e Pimentel, 2015).

b. Avaliação da Eficiência no Uso do Nitrogênio (EUN)

Os índices de eficiência nutricional foram obtidos conforme descrito por Moll et al. (1982):

$$EUN = \frac{PG}{N(\text{aplicado})} (\text{Kg mg}^{-1})$$

Em que, *EUN*: eficiência no uso de nitrogênio; *PG*: produtividade de grãos; *N(aplicado)*: quantidade de nitrogênio fornecido à planta.

c. Análise dos Dados Fenotípicos

Primeiramente, os dados de EUN foram submetidos à análise de variância. Em seguida as linhagens foram agrupadas pelo método de Scott Knott (1974) modificado, de acordo com seus valores médios de EUN, a um nível de significância de 0,01. Os grupos formados com base no desempenho fenotípico, foram utilizados nas análises posteriores de estruturação e diversidade genética.

d. Genotipagem

A genotipagem das 64 linhagens foi realizada no laboratório da Pioneer Hi-Bred International Int. (Johnston, IA - USA), utilizando o ensaio de alta performance Illumina GoldenGate (Fan et al., 2003) contendo 768 marcadores SNPs. Os marcadores SNPs passaram por análise de qualidade, por meio do método *Call Rate* (taxa de atendimento, qualificação ou repetibilidade). Essa medida é utilizada para eliminar marcadores com grande quantidade de valores perdidos. Foi adotado um valor de *Call Rate* > 0,85, ou seja, marcadores com mais de 15% de dados perdidos foram excluídos das análises. Além disso, foi avaliada a MAF (menor frequência alélica) de cada marcador, a qual está relacionada com o polimorfismo dos locos na população. Foram selecionados apenas os marcadores com MAF >0,05. Para esse fim utilizou-se o pacote HapEstXXR (Knueppel e Rohde, 2014) do programa R (R Core Team, 2014). Para converter os dados genômicos de bases nitrogenadas (letras) em algarismos numéricos foi utilizado o pacote Scrim (Schwender e Fritsch, 2013).

e. Análise dos Dados Moleculares

Para cada SNP foram estimadas as frequências alélicas, as frequências genotípicas, a heterozigosidade observada, a diversidade gênica (heterozigosidade esperada) e o conteúdo de informação polimórfica (PIC).

O PIC estima o poder discriminatório do marcador ao considerar não só o número de alelos por loco, mas também as frequências relativas desses alelos. É expresso por:

$$PIC = 1 - \sum_{i=1}^a p_i^2 - \sum_{i,j=1}^a \sum_{(i \neq j)} p_i^2 p_j^2$$

Em que a=número de alelos no loco; p_i =frequência alélica estimada do i-ésimo alelo; e p_j = frequência alélica estimada do j-ésimo alelo.

A diversidade gênica (DG), também chamada de heterozigosidade esperada, foi estimada a partir das frequências alélicas, e dada por:

$$DG = 1 - \sum_{j=1}^a P_j^2$$

Em que P_i = frequência alélica estimada do i-ésimo alelo.

A heterozigosidade observada (H_o) foi estimada por:

$$H_o = 1 - \sum_{j=1}^a p_j^2$$

Em que p_i = frequências dos genótipos homozigotos.

Para essas análises foi utilizado o aplicativo computacional PowerMarker(Liu e Muse, 2005).

Para inferir sobre a estrutura populacional das linhagens foi utilizado o método de agrupamento baseado no modelo “admixture”, com o auxílio do aplicativo STRUCTURE 2.3.3 (Pritchard et al., 2000). No STRUCTURE foram realizadas corridas para vários números de grupos (K), variando de 1 a 10. Para cada K foram realizadas 10 corridas com um *burn-in* de 500.000 e 500.000 replicações MCMC (*Markov Chain Monte Carlo*)

após o *burn-in*. As linhagens foram atribuídas aos grupos (populações) de acordo com as médias de probabilidade de associação (Q). Q_{ik} representa a proporção estimada do genoma da linhagem i derivada da população k (Pritchard, 2000). Com os valores de Q para cada linhagem, foi gerado um gráfico de barras que representa a associação da linhagem a cada um dos K grupos. Foram considerados como pertencentes ao mesmo grupo os indivíduos cuja probabilidade de associação a esse grupo foi superior à probabilidade de associação a qualquer outro grupo. O valor mais provável de K (número de grupos) foi estimado de acordo com o critério de Evanno baseado no valor de ΔK (Evanno et al., 2005), com auxílio do aplicativo Structure Harvest (Earl e Vonholdt, 2012). O ΔK é expresso por:

$$\Delta K = \frac{m \left(|L(K+1) - 2L(K) + L(K-1)| \right)}{\sigma [L(K)]}$$

Em que, $L(K)$ é a média do logaritmo neperianos da probabilidade dos dados de cada passo do MCMC, subtraído pela metade da variância para o número de populações K .

Foi construída uma matriz de dissimilaridade baseada na distância genética de Rogers (Rogers, 1972), expressa por:

$$D_R = \frac{1}{m} \sum_j^m \sqrt{\frac{1}{2} \sum_i^{a_j} (p_{ij} - q_{ij})^2}$$

Em que p_{ij} e q_{ij} são as frequências do i -ésimo alelo no j -ésimo loco nas populações X e Y respectivamente, em que j é o número de alelos no j -ésimo loco, e m é o número de locos examinados.

Com base nessa matriz, foi gerado um dendrograma pelo método de agrupamento UPGMA (*Unweighted Pair Group Method Average*) e as árvores resultantes foram visualizadas usando o programa MEGA versão 5.1 (Tamura et al., 2011). A estabilidade dos agrupamentos foi testada por meio de 1.000 reamostragens *bootstrap*. Foram considerados significativos os nós que apresentaram o mesmo padrão de ramificação em pelo menos 50% dos 1000 dendrogramas gerados.

A diversidade genética foi estimada na população total e dentro e entre os grupos fenotípicos obtidos pelo método de Scott Knott baseado em dados fenotípicos de EUN. Para quantificar a diversidade genética existente dentro de cada grupo, foram estimados PIC médio e DG média, conforme descrito anteriormente.

8 Resultados e Discussão

a. Análise Fenotípica da EUN

Com base no resultado da análise de variância das 64 linhagens para EUN em condição de baixa disponibilidade de nitrogênio no solo, não houve efeitos significativos de blocos e de repetições. No entanto, foi observado um efeito significativo de genótipo na EUN (Tabela 1). Diferenças em relação à EUN foram observadas em outros estudos, que também reportaram considerável variação para a EUN entre linhagens de milho, com base na produtividade de grãos (Wu et al., 2011; Abdel-Ghani et al., 2013; Bi et al., 2014). O coeficiente de variação de 23,54% está dentro da faixa aceitável para experimentos de campo com plantas sob condições de estresse, em que a mortalidade delas é geralmente alta. Segundo Gallais e Hirel (2004), para a maioria dos caracteres, na condição de estresse há redução da variabilidade genética, uma vez que os genótipos tendem a ter desempenhos semelhantes. Além disso, as médias e a herdabilidade são menores e os valores de CV são geralmente maiores em relação a essas características e parâmetros avaliados em nível ideal de N, o que pode dificultar o processo de seleção de genótipos superiores (Zhang et al., 2015).

Tabela 1—Quadro de análise de variância para eficiência no uso de Nitrogênio (EUN) em 64 linhagens de milho tropical em condição de baixa disponibilidade de N no solo.

Fontes de Variação	GL	QM
Bloco(rep)	14	46,25 ns
Repetição	1	87,33 ns
Genótipo	63	418,04 **
Resíduo	49	108,89
CV(%)	-	23,54
Media	-	44,33

^{ns}não significativo, ^{**} significativo a 1% de probabilidade pelo teste F, respectivamente.

QM: quadrado médio.

CV: Coeficiente de variação.

O conjunto de 64 linhagens de milho tropical se mostrou bastante diverso em relação à EUN, com valores variando de 17,52 na linhagem L13 a 81,66 na linhagem L18, com média geral de 44,33 (Tabela 2).

As 64 linhagens formaram 3 subgrupos, com base nas médias da EUN. Os grupos foram denominados AE (Alta EUN), IE (Intermediária EUN), e BE (Baixa EUN) (Tabela 2). O AE foi composto por 11 linhagens com valores de EUN variando de 60,03 (L35) a 81,66 (L18), com média de 68,37. O grupo IE foi formado por 30 linhagens com EUN variando de 38,37 (L45) a 57,80 (L14), com média de 47,52. Já o grupo BE foi formado por 21 linhagens com valores de EUN variando de 10,48 (L50) a 37,42 (L43), com média de 27,80. O grupo IE foi o mais homogêneo em relação à EUN, com variância de 32,97. O grupo BE foi o mais heterogêneo quanto a EUN, com variância de 63,74. O grupo AE também apresentou variância relativamente alta (55,72). Portanto, os resultados reiteram a existência de grande variabilidade entre as linhagens quanto a EUN, como descrito por DoValle et al (2012).

Alguns estudos têm mostrado que existe grande variação natural intraespecífica em relação ao metabolismo do N em *Arabidopsis*, incluindo variações na absorção e na remobilização desse macronutriente (Masclaux et al., 2011; Chardon e Masclaux, 2012). No caso do milho, variações genéticas relacionadas à remobilização e à absorção de N após o espigamento, ao metabolismo do N, ao transporte do N e à senescência também têm sido reportadas na literatura (Coque et al., 2008; Canas et al., 2009; Sylvester et al., 2009; Canas et al., 2012).

No Centro Internacional de Melhoramento de Milho e Trigo (CIMMYT) no México e mais recentemente no Zimbábue e no Quênia, são avaliados todos os anos milhares de acessos de milho tropical e temperado. Nessas avaliações, são poucos os genótipos (cerca de 3%) que apresentam potencial para serem explorados em condições de baixa disponibilidade de N no solo (Fritsche-Neto e Borém 2012). Isso ocorre porque a grande maioria dos cultivares modernos e populações de melhoramento de milho foram desenvolvidos em condições de alta fertilidade do solo, com foco em alta produtividade. Assim, é provável que essa seleção de genótipos mais responsivos a adubação tenha reduzido a diversidade alélica associada à tolerância a baixa disponibilidade de N (Wissuwa et al., 2009). Portanto, o germoplasma selecionado em condições ótimas de

cultivo, geralmente não é adequado para ser utilizado em condições de estresses abióticos (Fritsche-Neto e Borém 2012, Abdel-Ghani et al., 2013; Bi et al., 2014).

Tabela 2 – Agrupamento das 64 linhagens de milho tropical com base na EUN em condições de baixa disponibilidade de N no solo.

Número	Linhagens	Grupo	EUN	Número	Linhagens	Grupo	EUN
1	L18	AE	81,66	32	L58	IE	43,86
2	L17	AE	81,50	33	L64	IE	43,85
3	L12	AE	73,28	34	L52	IE	42,89
4	L25	AE	68,04	35	L6	IE	42,15
5	L19	AE	67,98	36	L7	IE	42,01
6	L37	AE	66,49	37	L22	IE	41,37
7	L39	AE	65,31	38	L40	IE	39,58
8	L2	AE	63,23	39	L11	IE	38,89
9	L34	AE	63,06	40	L16	IE	38,81
10	L24	AE	61,53	41	L45	IE	38,37
11	L35	AE	60,03	42	L43	BE	37,42
12	L14	IE	57,80	43	L49	BE	36,63
13	L10	IE	57,61	44	L31	BE	36,17
14	L26	IE	57,12	45	L29	BE	36,03
15	L53	IE	56,11	46	L62	BE	33,74
16	L23	IE	55,92	47	L55	BE	33,53
17	L44	IE	52,56	48	L33	BE	32,77
18	L61	IE	51,95	49	L27	BE	32,48
19	L51	IE	51,77	50	L47	BE	31,68
20	L20	IE	50,82	51	L42	BE	31,32
21	L36	IE	50,63	52	L48	BE	30,99
22	L54	IE	49,05	*	L56	BE	30,94
23	L21	IE	48,62	53	L41	BE	28,60
24	L15	IE	48,61	54	L63	BE	27,76
25	L9	IE	48,42	55	L4	BE	26,75
26	L38	IE	47,62	56	L60	BE	24,27
27	L46	IE	46,98	57	L32	BE	22,68
28	L30	IE	46,50	58	L59	BE	22,21
29	L3	IE	46,12	59	L13	BE	17,52
*	L57	IE	45,95	60	L28	BE	14,60
30	L5	IE	45,60	61	L1	BE	13,09
31	L8	IE	45,50	62	L50	BE	10,48

* Linhagens retiradas das análises posteriores por falhas no processo de genotipagem.

b. Análise dos marcadores SNPs

Do total de 768 marcadores SNPs utilizados neste estudo, 50 falharam na genotipagem de todos os indivíduos e foram eliminados das análises. Dos marcadores restantes, 78 apresentaram uma taxa de dados perdidos superior a 15% (Call Rate = 0,85), e também foram eliminados, juntamente com 101 marcadores que apresentaram MAF < 0,05. Portanto, para as análises posteriores de estrutura e diversidade genética

foram utilizados 539 marcadores SNPs altamente informativos, que atenderam aos critérios de qualidade.

Nelson et al. (2008) usaram esses 768 SNPs para caracterizar a diversidade genética de 92 linhagens norte americanas de milho com prazos de PVPA (*Plant Variety Protection Act*) expirados. Após a seleção dos marcadores de acordo com os critérios de qualidade de *Call Rate* igual a 0,80 e MAF igual a 0,05, foram utilizados nas análises de diversidade e estruturação 614 locos SNPs, que permitiram a diferenciação acurada das linhagens de acordo com o nível de similaridade genética de seus genitores. No presente estudo optou-se por um critério mais rigoroso de *Call Rate* (0,85) para se obter uma estruturação genética mais acurada, o que pode ser crucial para futuros estudos de associação genômica ampla para EUN.

Considerando a população total (62 linhagens), o conteúdo de informação polimórfica (PIC) dos marcadores variou de 0,09 a 0,37, com média de 0,29. Apenas 60 marcadores apresentaram PIC inferior a 0,15, sendo que 300 apresentaram valores de PIC superiores a 0,30 (Figura 1). A MAF variou de 0,05 a 0,5. A maior parte dos marcadores (65,5%) apresentou MAF superior a 0,20, sendo que desses, 32,8% apresentaram MAF superior a 0,35 e 22% apresentaram frequências semelhantes dos dois alelos (MAF próxima de 0,5). Esse grau de polimorfismo indica alta informatividade do conjunto de marcadores utilizados, considerando que se tratam de marcadores bialélicos, que normalmente apresentam uma informatividade baixa por loco em comparação com marcadores multialélicos, como os microssatélites (Van Inghelandt et al., 2010; Frascaroli et al., 2013).

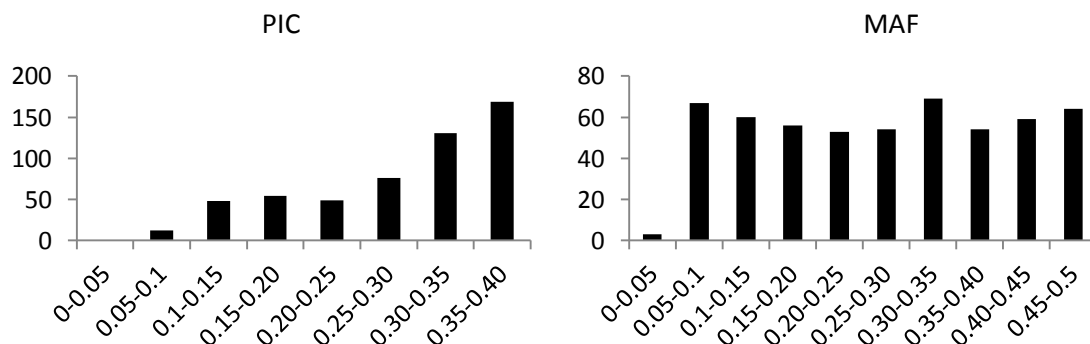


Figura 1 – Variação dos valores de PIC (Polymorphism Information Content) e MAF (Minor Allele Frequency) dos 539 marcadores SNPs em 62 linhagens de milho tropical.

c. Diversidade Genética

A diversidade genética entre as 62 linhagens foi estimada por meio da diversidade gênica (DG), a qual apresentou um valor médio de 0,36. Considerando apenas as 11 linhagens com alta EUN, a DG média foi de 0,34. Já o valor médio do PIC para esse grupo foi de 0,27, variando de 0 a 0,37. No grupo IE, a DG média foi igual a 0,36. O PIC nesse grupo variou de 0 a 0,37, com média de 0,29. Já no grupo BE, a DG média foi de 0,34. O PIC variou de 0 a 0,375 com média de 0,27. (Tabela 3). Nelson et al (2008) caracterizaram geneticamente 92 linhagens de milho americanas com prazo de PVPA (*Plant Variety Protection Act*) expirados, e observaram um valor de DG médio igual a 0,35. Em outro estudo, Inghelandt et al. (2010) caracterizaram um grande painel composto por 1537 linhagens elite de milho utilizando 8.244 SNPs, e observaram uma DG média igual a 0,32. Já Westengen et al (2012) observaram em um painel africano, composto por 48 acessos de origens variadas, um valor médio de DG igual a 0,254. Wen et al. (2012) caracterizaram 498 acessos de milho, incluindo 321 acessos tropicais da raça Tuxpeño utilizando 1433 SNPs. Considerando apenas o conjunto de 321 acessos Tuxpeño, a DG média observada pelos autores foi igual a 0,29. Recentemente, Wu et al. (2014) caracterizaram um painel de 367 linhagens chinesas de milho, e observaram valores médios de DG e PIC iguais a 0,36 e 0,29 respectivamente.

Os valores de DG e PIC obtidos no presente estudo foram semelhantes aos observados em outros estudos, nos quais foram analisados painéis maiores. Isso indica que o conjunto de 62 linhagens analisadas quanto a EUN, apesar de pouco numeroso, apresenta altos níveis de diversidade genética. A alta diversidade genética observada, tanto a nível fenotípico quanto a nível de genoma, pode estar relacionada com a origem da maioria dessas linhagens, as quais foram obtidas de acessos de milho tropical, que possuem base genética mais ampla em relação aos germoplasmas de origem temperada (Yan et al., 2009, Rincent et al., 2014). Além disso, essas linhagens se encontram em estágio intermediário de melhoramento, e vêm sendo conduzidas de forma a manterem um nível elevado de diversidade genética em relação à EUN. A diversidade é fundamental para a utilização eficiente das linhagens em futuros estudos de associação genômica ampla (GWAS). Tais estudos poderão contribuir para a elucidação da

arquitetura genética dessa característica complexa, possibilitando a planejamento mais eficiente de métodos de melhoramento para aumentar a EUN.

A heterozigosidade observada entre as 62 linhagens variou de 0 a 0,26, com média de 0,03. Considerando cada grupo separadamente, os valores de heterozigozidade média também foram baixos: 0,04 no grupo AE, 0,03 no grupo IE e 0,04 no grupo BE. Esses valores estão dentro da faixa de valores de heterozigosidade residual observados em linhagens endogâmicas de milho (Yan et al., 2009).

Tabela 3 – Valores de diversidade gênica (DG), conteúdo de informação polimórfica (PIC) e heterozigosidade observada na população total e nos grupos fenotípicos (Alta EUN, Intermediária EUN e Baixa EUN).

População	Número de indivíduos	Diversidade gênica (DG)	PIC	Heterozigosidade observada
Alta EUN(AE)	11	0,34	0,27	0,036 (0 – 0,40)
Intermediária EUN (IE)	30	0,36	0,29	0,029 (0 – 0,28)
Baixa EUN (BE)	21	0,34	0,27	0,04 (0 – 0,22)
População Total	62	0,36	0,29	0,03 (0 – 0,26)
Limite Teórico	-	0,5	0,375	1

*PIC:conteúdo de informação polimórfica

Considerando a população total, 132 (24,49%) locos apresentaram uma heterozigozidade observada igual a 0. Em 398(73,84%)locos ela variou entre 0 e 0,1 e em 539 (1,48%)locos variou de0,11 a 0,14. No grupo AE, em 377 (69,94%) locos foi observada uma heterozigosidade igual a 0. Em 119 (22,08%) locos a heterozigozidade observada variou de 0a 0,1. Nos 43 (8,35%) locos restantes a heterozigozidade variou de 0,11 a 0,4. No grupo IE, em 268 (49,72%) locos a heterozigozidade foi igual 0. Em 240 (44,53%) locos a heterozigozidade variou de0 e 0,10, e em 30 (5,57%) locos variou de 0,11 e 0,18. Já no grupo BE a heterozigozidade foi igual a 0 em 280 (51,95%) locos, e variou de 0,001 a 0,10 em 259 (48,05). Nos 40 (7,42%) locos restantes variou de 0,11 a 0,22 (Figura 2A).

O loco PHM11985.27 apresentou uma heterozigozidade superior á média e á heterozigozidade individual de todos os outros locos, tanto na população total como nos grupos fenotípicos (figura 2B). Além disso, a heterozigozidade desse loco foi maior no grupo AE, e menor no grupo BE. No grupo AE a proporção de locos com heterozigozidade observada maior que 0,1 foi superior em relação aos outros grupos e

em relação à população total. Portanto, considerando somente o loco PHM11985.27, a heteroziguidade foi proporcional a EUN média de cada grupo. Todavia, apesar da relação entre o grau de heteroziguidade nesse loco e a EUN nos grupos fenotípicos, esse marcador está posicionado em uma região genômica cuja função ainda é desconhecida. Esse SNP se encontra posicionado na região pericentromérica do cromossomo 6. Alguns estudos anteriores constataram que regiões pericentroméricas apresentam taxas de recombinação suprimidas, mas por outro lado apresentam elevada heteroziguidade residual em relação ao restante do genoma (McMullen et al., 2009; Wu et al., 20014).

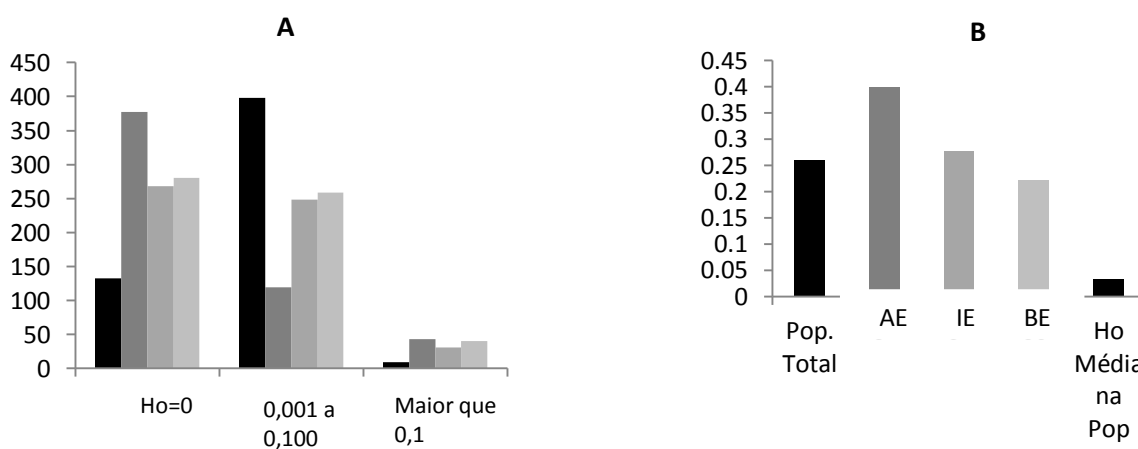


Figura 2 – Variação da heteroziguidade observada (Ho) considerando todos os locos (A) na população total (barra preta), e nos grupos AE, IE e BE (barras cinza escura, cinza intermediário e cinza claro, respectivamente), e apenas o locoPHM11985.27(B).

O dendrograma gerado pelo método de agrupamento UPGMA, com base no índice de dissimilaridade de Rogers (Figura 3), permitiu inferir sobre o grau de parentesco e divergência genética entre as linhagens. Essas inferências são importantes no direcionamento de cruzamentos entre genitores que possuam alta EUN, mas que sejam geneticamente divergentes, o que pode permitir a exploração mais eficiente da heterose na obtenção de híbridos altamente produtivos e com altas EUN.

Por exemplo, apesar de as linhagens L02 e L37 apresentarem altas médias de EUN, não é indicado o cruzamento entre elas para a produção de híbridos, pois são pouco divergentes geneticamente. No entanto, as combinações entre L25 e L37 ou entre

Para estimar a diversidade alélica entre os grupos fenotípicos, foram estimados o número de alelos privados e o número de alelos ausentes em cada grupo (Tabela 4). No grupo AE, não foram observados alelos privados. Já no grupo IE, foram observados dois alelos privados, um no loco PHM4757.14 e um no loco PZA00525.1. No grupo BE, foi observado um alelo privado no loco PHM15871.1.

No grupo AE, foram observados 42 alelos ausentes e no grupo BE foram observados 19. No grupo IE ocorreu apenas um alelo ausente, no loco PHM15871.1. Isso indica que existem muito mais alelos comuns entre os grupos AE e IE e entre os grupos IE e BE, do que entre os grupos AE e BE, mostrando que a diferença na EUN está relacionada com a diversidade alélica entre os grupos, ou seja, os grupos AE e IE e os grupos IE e BE são mais relacionados geneticamente e compartilham mais alelos entre si do que o AE com o BE. O maior número de alelos ausentes no grupo AE em relação ao grupo BE pode ser explicado pelo fato de o grupo AE ser constituído por 11 linhagens enquanto o grupo BE é formado por 21 linhagens.

Tabela 4 – Alelos privados e alelos ausentes nos três grupos fenotípicos (AE, IE e BE).

Grupo Fenotípico	Alelos privados	Alelos ausentes
Alta EUN (AE)	0	42
Intermediária EUN (IE)	2	1
Baixa EUN (BE)	1	19

Para analisar as diferenças de frequências alélicas entre os grupos AE, IE e BE, foi realizada uma análise comparativa entre cada par de grupos (Tabela 5). Considerando os contrastes entre os pares de grupos, em relação aos 10 locos que exibiram as maiores diferenças de frequências entre grupos, foi observado que entre o AE e BE, esta apresentou média de 0,53, um valor significativamente maior do que os observados entre AE e IE (0,46) e entre IE e BE (0,36). Esse resultado evidencia a existência de uma relação entre a diversidade fenotípica para EUN e a diversidade genômica entre as linhagens avaliadas.

Tabela 5 – Diferenças de frequências alélicas entre pares de grupos fenotípicos (AE, IE e BE), para os 10 locos que apresentaram as maiores diferenças.

Loco	Base	AE	IE	Diferença
PZA03043.14	A	0,778	0,267	0,511
PHM5526.25	A	0,909	0,400	0,509
pza02746.2	T	0,909	0,423	0,486
pza00755.2	A	0,909	0,433	0,476
PHM13823.7	C	0,722	0,259	0,463
PHM11114.7	T	0,900	0,448	0,452
PZA02299.16	G	0,818	0,383	0,435
PHM4955.12	C	0,600	0,167	0,433
PZA02519.7	G	0,682	0,267	0,415
PHM12706.14	C	0,682	0,283	0,399
Média				0,458

Loco	Base	IE	BE	Diferença
PZA00399.10	A	0,577	0,079	0,498
PZA00243.24	G	0,724	0,300	0,424
PHM1745.16	G	0,444	0,079	0,366
pza01884.1	A	0,517	0,167	0,351
PHM10525.9	G	0,867	0,524	0,343
PZA02731.1	T	0,397	0,737	0,340
PZA00439.6	G	0,640	0,306	0,334
PHM3463.18	C	0,857	0,524	0,333
PHM5805.19	T	0,621	0,294	0,327
PHM5798.39	G	0,483	0,158	0,325
Média				0,364

Loco	Base	AE	BE	Diferença
pza00755.2	A	0,909	0,238	0,671
PZA03043.14	A	0,778	0,150	0,628
PHM13823.7	C	0,722	0,143	0,579
PZA02519.7	G	0,682	0,167	0,515
PZA00627.1	T	0,636	0,125	0,511
PZA00439.6	A	0,200	0,694	0,494
PZA01359.1	C	0,800	0,313	0,488
PZA00243.24	G	0,773	0,300	0,473
PHM4353.31	A	0,545	1,000	0,455
PHM635.23	C	1,000	0,548	0,452
Média				0,527

Entre os grupos AE e IE, nove SNPs apresentaram diferenças de frequências alélicas superiores a 0,4. Entre os grupos IE e BE, apenas dois SNPs apresentaram diferenças superiores a 0,4. Entre os grupos AE e BE, 26 SNPs apresentaram diferenças

superiores a 0,4, indicando uma diversidade alélica acentuadamente superior às observadas entre os demais pares de grupos, o que corrobora com os dados de divergência fenotípica em relação à EUN (Figura 4 A, B e C).

Considerando todos os 539 SNPs utilizados, as frequências alélicas do grupo AE apresentaram maior correlação com as do grupo IE ($R^2=0,68$) do que com as do grupo BE ($R^2=0,63$). O grupo IE apresentou maior correlação com o grupo BE ($R^2= 0,78$). Esses resultados indicam a existência de uma relação entre as diferenças nas frequências alélicas observadas entre os três grupos fenotípicos e os diferentes níveis de EUN apresentados pelas linhagens de milho tropical (Figura 4 D, E e F).

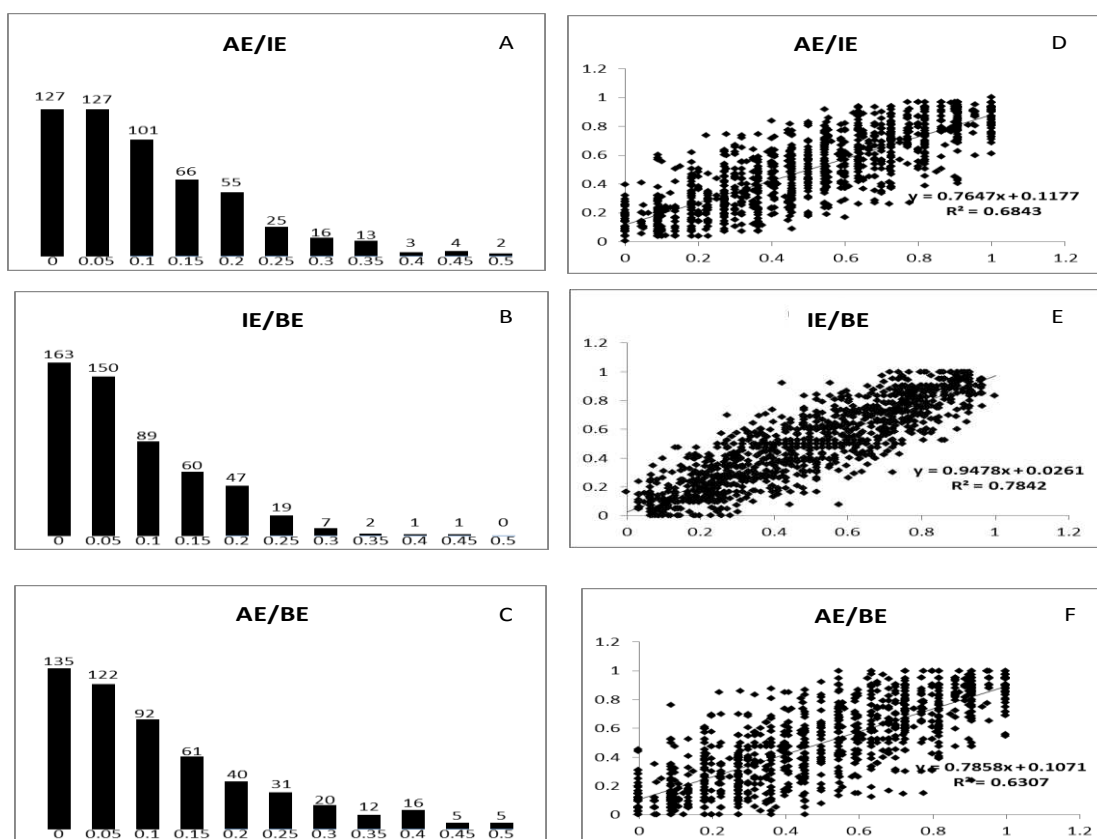


Figura 4 - Número de marcadores SNPs que apresentaram diferentes frequências alélicas entre os diferentes pares de grupos fenotípicos (A, B e C); e correlação entre frequências alélicas entre pares de grupos fenotípicos (D, E, F).

d. Análise da estrutura genética da população e da sua relação com a EUN

De acordo com o critério de Evanno, foram observados três picos de ΔK em função dos valores de K (Figura 5A). O número de grupos (K) mais provável em que a população pode ser geneticamente estruturada foi $K=2$. No entanto, as estruturações com $K=8$ e $K=4$ apresentaram valores de ΔK muito próximos ao observado para $K=2$.

Considerando a estruturação genética com $K=2$ (Figura 5B), das 62 linhagens, 54 foram alocadas no grupo 2.1 (predominantemente vermelho) e oito no grupo 2.2 (predominantemente verde). O grupo 2.1 foi mais numeroso e diverso, com 10 (das 11) linhagens do grupo fenotípico AE, 28 (das 30) do grupo IE e 14 (das 21) do grupo BE. A média da EUN nesse grupo foi de 45,67 e a variância de 219,64. Entre as oito linhagens do grupo 2.2, uma pertence ao grupo fenotípico AE (3), duas ao grupo IE (12 e 21) e as cinco restantes ao grupo BE (54, 55, 56, 57 e 62). Estas últimas estão entre as linhagens com menor EUN dentre as avaliadas nesse estudo. As linhagens do grupo 2.2 apresentaram EUN média de 36,71 e variância de 456,40. O fato de 10 das 11 linhagens do grupo AE terem sido atribuídas ao grupo 2.1 indica que nesse grupo está concentrada a maior parte dos alelos favoráveis para EUN presentes na população. Além disso, a maior parte das linhagens do grupo IE também foram atribuídas a esse grupo. No entanto, o grupo 2.1 também contém alelos desfavoráveis em relação à EUN, pois 14 linhagens do grupo BE também foram atribuídas a esse grupo.

Na estruturação com $K=4$ (Figura 5C), o grupo 4.1 (predominantemente vermelho) foi formado por 15 linhagens, das quais nove pertencem ao grupo fenotípico BE, quatro ao grupo IE e uma ao grupo AE. As duas linhagens com menor EUN pertencem a esse grupo, que apresentou EUN média de 35,83 e variância de 196,59. Portanto, podemos concluir que esse grupo possui alelos desfavoráveis para a EUN em frequências relativamente elevadas. O grupo 4.2 (predominantemente verde) foi constituído por sete linhagens: uma do grupo fenotípico AE, duas do grupo IE e quatro do grupo BN. A EUN média observada nesse grupo foi igual a 40,45 e a variância igual a 401,46. As linhagens 3, 12 e 21 desse grupo apresentam EUN bem superiores aos demais membros desse grupo (54, 55, 56, 57). Essas três linhagens, apesar de terem sido alocadas no grupo 4.2, possuem uma porção significativa do genoma em comum com as do grupo 4.4, no qual está alocada a maioria das linhagens do grupo AE.

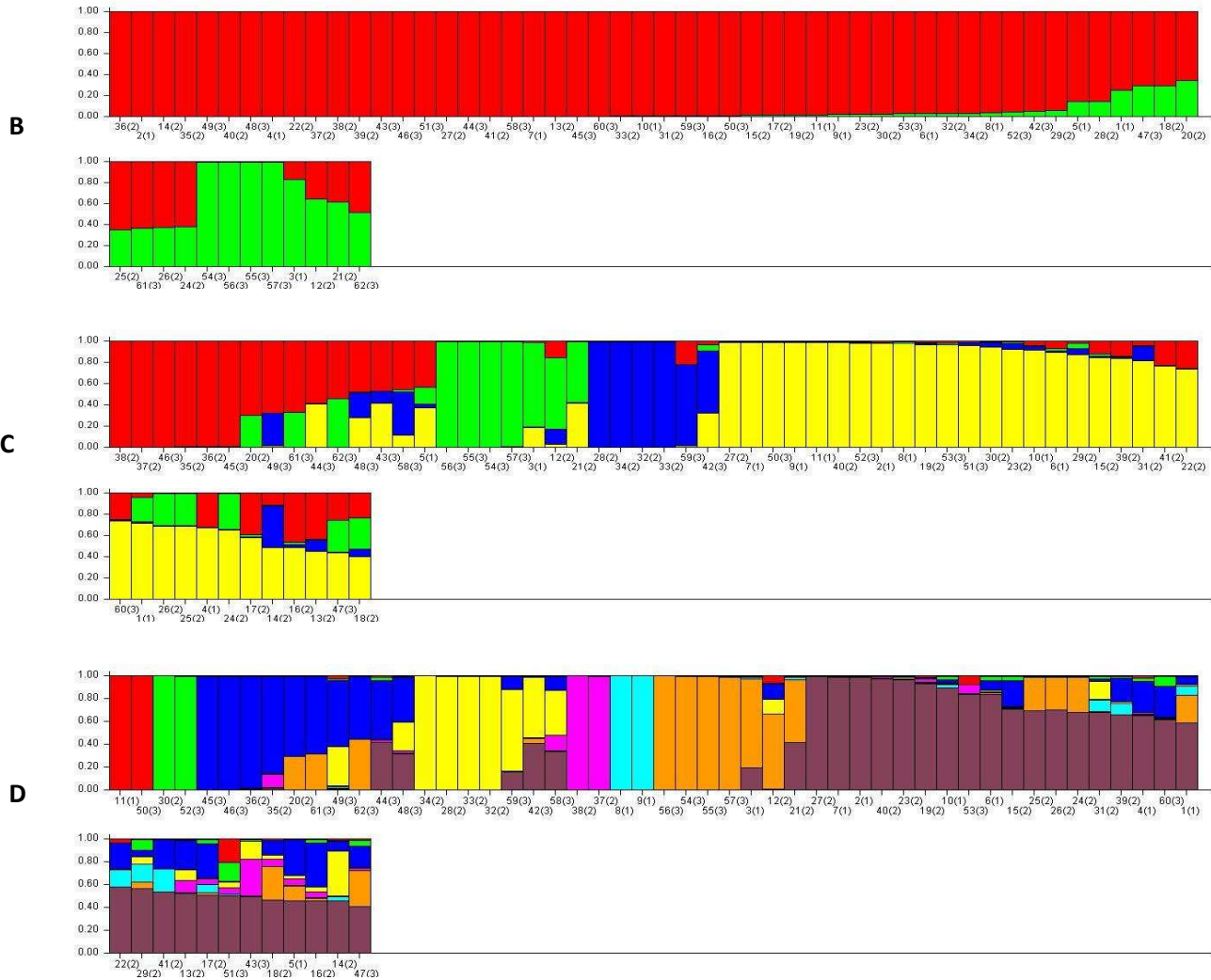
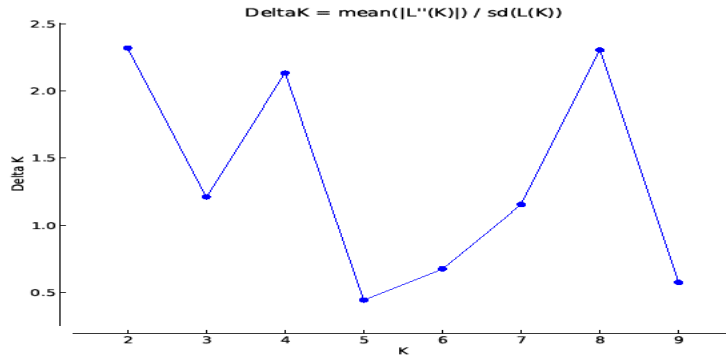


Figura 5 – Valores de delta K em função do número de grupos (A). Gráficos gerados pelo STRUCTURE mostrando a atribuição de cada linhagem a cada grupo genotípico (cores) para números de grupos (K) iguais a 2, 4 e 8 (B, C e D, respectivamente).

Isso que indica que elas possuem uma quantidade razoável dos alelos favoráveis que estão em frequências mais elevadas no grupo 4.4. O grupo 4.3 (predominantemente azul) foi composto por seis linhagens, das quais quatro pertencem ao grupo fenotípico BE e duas ao IE. As quatro linhagens com maiores valores de EUN grupo 4.3 (28, 32, 33 e 34) apresentam coeficiente de atribuição a esse grupo igual a 1. As outras duas (42 e 59), que apresentam valores de EUN bem inferiores aos das demais, possuem parte do genoma atribuída aos grupos 4.1(Vermelho), sendo que a linhagem 59 apresenta coeficiente de atribuição ao grupo 4.1 maior do que a linhagem 42. Portanto há maior concentração de alelos desfavoráveis em relação à EUN oriundos do grupo 4.1. O grupo 4.3 apresentou uma EUN média igual a 38,67 e variância igual a 116,35. O grupo 4.4 (predominantemente amarelo) foi formado por 34 linhagens. Nesse grupo, foram alocadas nove das 11 linhagens do grupo fenotípico AE, o que indica que os alelos favoráveis em relação à EUN estão presentes em frequências relativamente elevadas nesse grupo. Também fazem parte desse grupo 20 linhagens do grupo fenotípico IE e seis do grupo BE. As linhagens 50, 51, 52 e 53 possuem coeficientes de atribuição ao grupo 4.4 próximos de 1, o que indica que esse grupo, apesar de concentrar a maior parte dos alelos favoráveis a EUN da população, também contém alelos desfavoráveis em frequências relativamente elevadas. A EUN média desse grupo foi de 50,95 e a variância de 179,51.

Na estruturação obtida com $K=8$ (Figura 5D), o grupo 8.1 (predominantemente vermelho) foi formado por duas linhagens (11 e 50) com valores de EUN muito diferentes entre si. Uma linhagem pertence ao grupo fenotípico AE e a outra ao BE, o que indica que esse grupo possui alelos favoráveis e desfavoráveis para a EUN em frequências semelhantes. Esse grupo apresentou valor médio de EUN de 45,86 e variância de 60,03. O grupo 8.2 (predominantemente verde) foi constituído por duas linhagens com valores de EUN relativamente baixos (30 e 52). Uma delas pertence ao grupo fenotípico IE e a outra ao BE, indicando que esse grupo possui alelos desfavoráveis em frequências mais elevadas do que de alelos favoráveis para EUN. A EUN média observada nesse grupo foi de 38,30 com variância de 106,73. O grupo 8.3 foi formado por 10 linhagens, das quais três pertencem ao grupo fenotípico IE e sete ao grupo BE. As duas linhagens com menores valores de EUN (61 e 62) pertencem a esse grupo, e ambas apresentam boa

parte do genoma atribuído ao grupo 8.7 (predominantemente laranja), o qual parece concentrar alelos desfavoráveis em frequências elevadas. O grupo 8.3 apresentou uma EUN média de 32,97 e variância de 156,23. O grupo 8.4 (predominantemente amarelo) foi composto por sete linhagens, quatro delas pertencentes ao grupo fenotípico IE e três ao grupo BE. As duas linhagens desse grupo com valores mais baixos de EUN (58 e 59) possuem uma parte do genoma atribuída ao grupo 8.3 (azul), o qual possivelmente contém alelos desfavoráveis a EUN em frequências relativamente elevadas. As quatro linhagens que possuem coeficiente de atribuição a esse grupo próximos de 1 (28, 32, 33 e 34) são pertencentes ao grupo fenotípico IE e apresentam valores de EUN muito próximos entre si. A EUN média desse grupo foi de 36,32 e a variância de 135,68. O grupo 8.5 (predominantemente rosa) foi formado por duas linhagens do grupo fenotípico IE, e possuem valores de EUN muito semelhantes entre si, com média 40,48 e variância 1,60. O grupo 8.6 foi composto por duas linhagens, ambas pertencentes ao grupo fenotípico AI e com valores semelhantes de EUN (8 e 9), o que indica presença de alelos favoráveis em altas frequências nesse grupo, o qual apresentou média de EUN de 63,15 e variância de 0,01. O grupo 8.7 (predominantemente laranja) foi formado por sete linhagens. Uma delas pertence ao grupo fenotípico AE (3), duas ao grupo IE (12 e 21) e quatro ao grupo BE (54, 55, 56 e 57). A linhagem 3 apesar de alocada nesse grupo, possui uma porção significativa do genoma atribuída ao grupo 8.8, no qual parece ocorrer a maior frequência de alelos favoráveis. As quatro linhagens desse grupo com menores valores de EUN (54, 55, 56 e 57) possuem coeficientes de atribuição a esse grupo próximos de 1, indicando que nesse grupo existem alelos desfavoráveis para EUN em frequências relativamente elevadas. A EUN média desse grupo foi igual a 40,45 e a variância igual a 401,46. O grupo 8.8 (predominantemente roxo) foi constituído por sete das 11 linhagens do grupo fenotípico AE, o que indica a presença de alelos favoráveis a EUN nesse grupo. Também fazem parte desse grupo 19 linhagens do grupo fenotípico AE e apenas três do grupo BE, o que indica a presença de alelos favoráveis em altas frequências, além de alelos desfavoráveis em frequências relativamente menores que na maioria dos outros grupos. A maioria das linhagens do grupo fenotípico IE alocadas no grupo 8.8 é composta pelas que tiveram os maiores valores de EUN dentro do grupo IE. A EUN média do grupo 8.8 foi de 50,35, com variância de 219,91.

Por fim, considerando as três possíveis estruturas genética da população ($K=2$, $K=4$ e $K=8$), de acordo com as médias e variâncias relativas à EUN, pode-se concluir que a divisão em quatro grupos foi a mais coerente com os dados fenotípicos de EUN, os quais evidenciara a formação de três grupos fenotípicos significativamente distintos em relação à EUN.

No milho, os marcadores moleculares têm sido utilizados com sucesso em vários estudos que analisaram a relação entre a estrutura populacional e características fenotípicas. Ao analisar a estrutura genética em função da distribuição espacial de acessos de milho cultivados em regiões da África (Leste, Saariana, e Oeste), Wentengen et al. (2012) observaram que os acessos do leste se mostraram claramente divergentes dos acessos do oeste. Ademais, o grupo de acessos da região saariana apresentou características de um grupo misto, coincidindo com o padrão de distribuição geográfica dos acessos. Essa coincidência entre o padrão de distribuição geográfica e a estrutura genética refletiu a existência de diferenças nas características fenotípicas de cada grupo, e que é resultado de diferenças adaptativas que foram fixadas ao longo do tempo nas diferentes populações, em função das diferentes pressões seletivas existentes nos diferentes ambientes, às quais os acessos estiveram submetidos.

Para observar como os três grupos fenotípicos de linhagens com diferentes níveis de EUN se estruturaram geneticamente e comparar as frequências de marcadores associados a alelos possivelmente favoráveis e desfavoráveis em relação a essa característica em cada um dos grupos, foi estimada a soma dos coeficientes de atribuição (Q) de cada linhagem de um mesmo grupo fenotípico (AE, IE e BE) a cada um dos grupos gerados pela análise do STRUCTURE (cores), para $K=2$, $K=4$ e $K=8$ (figura 6). Os grupos AE, IE e BE apresentaram médias fenotípicas de EUN iguais a 63,37, 47,52 e 32,79, respectivamente, e as variâncias iguais a 55,72, 32,97 e 63,73.

Para $K=2$, as somas dos coeficientes de atribuição das 11 linhagens do grupo fenotípico AE ao grupo 1 do STRUCTURE (vermelho) foi de 88% e ao grupo 2 foi de 12%. Esses resultados sugerem que, em média, as linhagens desse grupo fenotípico compartilham a maior parte dos seus genomas com o grupo genotípico 1 (STRUCTURE). No caso do grupo fenotípico IE os valores foram idênticos aos do AE. No grupo fenotípico BE, a soma dos coeficientes de atribuição das 30 linhagens grupo fenotípico AE ao grupo

1 do STRUCTURE (vermelho) foi igual a 74% e ao grupo 2 igual a 26%. As linhagens do grupo fenotípico BE, apesar de também possuírem em média grande parte do genoma em comum com o grupo 1 (74%), possuem uma proporção do genoma em comum com o grupo 2 maior do que os grupos AE e BE. Portanto, com base nessa estruturação, constatamos que a maioria dos alelos favoráveis em relação à EUN ocorrem em frequências mais altas no grupo 1 (vermelho) do que no grupo 2 (verde).

Para $k=4$, as somas dos coeficientes de atribuição das linhagens do grupo fenotípico AE aos grupos 1 (vermelho), 2 (verde), 3 (azul) e 4 (amarelo) foram 9%, 11%, 1% e 79%, respectivamente. No grupo fenotípico IE, os valores das somas dos coeficientes de atribuição aos grupos 1, 2, 3 e 4 foram 24%, 10%, 24% e 49%, respectivamente. Já no grupo fenotípico BE, esses valores foram 31%, 25%, 12% e 32%.

Considerando o grupo fenotípico AE, o coeficiente de atribuição ao grupo 4 (amarelo) do STRUCTURE foi maior que no grupo IE, e maior ainda que o grupo BN. Isso indica que o grupo 4 concentra alelos favoráveis para a EUN em frequências relativamente altas. A atribuição ao grupo 1 (vermelho) foi crescente entre os grupos AE, IE e BE respectivamente, evidenciando que nesse grupo existem alelos desfavoráveis em frequências relativamente elevadas. Os coeficientes de atribuição ao grupo 3 (azul) foram maiores no grupos fenotípicos IE e BE do que no grupo fenotípico AE, indicando a predominância de alelos desfavoráveis para a EUN no grupo 3.

Para $K=8$, os coeficientes de atribuição do grupo fenotípico AE aos grupos 8.6 (azul claro) e 8.8 (roxo) foram decrescentes entre os grupos fenotípicos AE, IE e BE, evidenciando que nos grupos 8.6 e 8.8 existem alelos favoráveis à EUN em frequências relativamente altas. No grupo BE, os coeficientes de atribuição aos grupos 8.2 (verde), 8.3 (azul), 8.4 (amarelo), 8.5 (rosa) e 8.7 (laranja) foram maiores do que os observados no grupo fenotípico AE. Isso sugere que nesses cinco grupos existem alelos desfavoráveis para a EUN em frequências superiores as dos grupos 6 e 8.

Nas estruturações com números de grupos (K) igual a 2, 4 e 8, o conjunto de marcadores SNPs usados refletiu claramente as diferenças fenotípicas relativas à EUN entre os três grupos fenotípicos. No entanto, a estruturação obtida com $K=4$ foi a que melhor refletiu a relação entre a estrutura genética e os dados fenotípicos.

A estrutura genética de uma população reflete a história evolutiva de uma espécie (Platt et al., 2010), e representa a presença de subgrupos geneticamente diferenciados em relação a população original. Vários fatores podem levar a divergência genética em uma população, incluindo a adaptação local, a seleção e a deriva genética. Tais fatores podem resultar em uma distribuição não aleatória de importantes características agronômicas. Por exemplo, ao analisar a relação entre a estrutura populacional e a tolerância ao alumínio em um banco de germoplasma de sorgo, Caniato et al. (2011) constataram que o padrão de diversidade genética foi influenciado pela origem racial e pela origem geográfica, resultando em subgrupos distintos em relação a essa característica.

Considerando que o surgimento de genes mutantes que conferem tolerância a estresses abióticos é um evento raro, alguns estudos têm sido realizados para ampliar o entendimento da diversidade de mecanismos de tolerância a estresses em plantas, o que não tem sido uma tarefa simples, uma vez que os pesquisadores precisam avaliar milhares de acessos em busca de algum que possua variantes alélicas relacionados a características desejáveis (Prada, 2009; Caniato et al., 2011). Para dar direcionamento a esses esforços, é importante ampliar o conhecimento sobre a relação entre estrutura genética da população e a tolerância ao estresse.

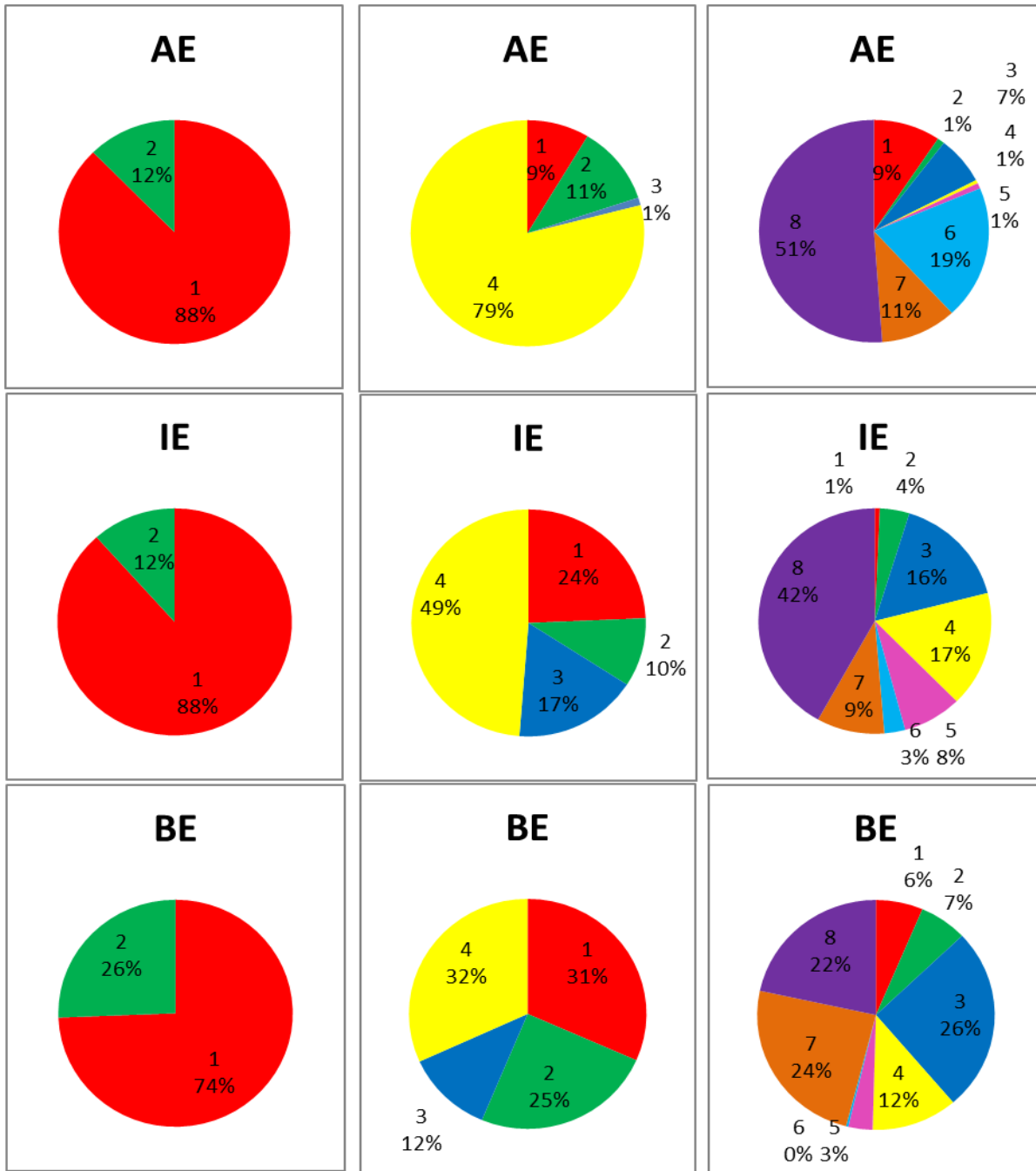


Figura 6 - Soma dos coeficientes de atribuição das linhagens de cada grupo fenotípico (AE, IE e BE) a cada grupo do STRUCTURE para K = 2, K = 4 e K = 8.

9 Conclusões

1 - Este painel de linhagens apresenta elevado grau de diversidade genética, comparável a de grandes painéis analisados em outros estudos.

2 - Há genótipos superiores em relação à EUN e geneticamente divergentes, possibilitando a melhor exploração da heterose.

3- Existem alelos exclusivos, ausentes ou com grandes diferenças de frequências entre os grupos fenotípicos de linhagens de milho tropical com três níveis de EUN. As diferenças alélicas foram maiores entre os grupos mais distintos fenotipicamente em relação à EUN

4 - Há forte relação entre o desempenho fenotípico em relação à EUN e a estruturação genética das linhagens de milho tropical.

10 Referências Bibliográficas

- Abdel-Ghani A, Kumar B, Reyes-Matamoros J, Gonzalez-Portilla P, Jansen C, Martin J, Lee M, Lübberstedt T (2013) Genotypic variation and relationships between seedling and adult plant traits in maize (*Zea mays* L.) inbred lines grown under contrasting nitrogen levels. **Euphytica** 189:123-133
- Anbessa, Y, Juskiw P, Good A, Nyachiro, J, Helm, Genetic variability in nitrogen use efficiency of Spring Barley. **Crop Science**, 2009. 49: p. 1259 - 1269.
- Bi Y-M, Meyer A, Downs G, Shi X, El-kereamy A, Lukens L, Rothstein S (2014) High throughput RNA sequencing of a hybrid maize and its parents shows different mechanisms responsive to nitrogen limitation. **BMC Genomics** 15:77
- Borém A, Galvão JCC e Pimentel MA (2015) Milho: do Plantio à Colheita. Viçosa. Editora UFV. 351
- Caniato FF, Guimarães CT, Hamblin M, Billot C, Rami J-F, Hufnagel B, Kochian LV, Liu J, Garcia AAF, Hash CT, Ramu P, Mitchell S, Kresovich S, Oliveira AC, de Avellar G, Borém A, Glaszmann J-C, Schaffert RE, Magalhaes JV (2011) The Relationship between Population Structure and Aluminum Tolerance in Cultivated Sorghum. **PLoS ONE** 6:e20830
- Cantarella, H. Nitrogênio. in: Novais, r.f.; Alvarez V., V.H.; Barros, N.F.; Fontes, R.L.F.; Cantarutti, R.B.; Neves, J.C.L., eds. (2007) **Fertilidade do solo**. Viçosa, MG, Sociedade Brasileira de Ciência do Solo, p.375-470
- Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2012) Genetic Architecture of Maize Kernel Composition in the Nested Association Mapping and Inbred Association Panels. **Plant Physiology** 158:824-834
- Dovale JC, Fritsche-Neto R, Bermudez FP, Miranda GV (2012) Efeitos gênicos de caracteres associados à eficiência no uso de nitrogênio em milho. **Pesquisa Agropecuária Brasileira**, 47: 385-392
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. **PLoS ONE** 6:e19379
- Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. **Conserv Genet Resour** 4:359-361
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology** 14:2611-2620

- Fan, J. B. et al. (2003) Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology, v. 68, p. 69-78
- Frascaroli E, Schrag T, Melchinger A (2013) Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. **Theor Appl Genet** 126:133-141
- Fritsche-Neto, R.; Borém, A. (2012). Plant Breeding for Abiotic Stress Tolerance. 1th ed. Heidelberg: Springer, 2012. 172p
- Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, Le Paslier M-C, McMullen MD, Montalent P, Rose M, Schön C-C, Sun Q, Walter H, Martin OC, Falque M (2011) 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** 6:e28334
- Gore MA, Chia J-M, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A First-Generation Haplotype Map of Maize. **Science** 326:1115-1117
- International Plant Nutrition Institute. **Evolução do consumo aparente de N, P, K e Total de NPK no Brasil**. Disponível em: <<http://brasil.ipni.net/article/BRS-3132>>. Acesso em 04 de novembro de 2014.
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. **Bioinformatics** 21:2128-2129
- Lu Y, Yan J, Guimarães C, Taba S, Hao, Z.; Gao, S. (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. **Theor Appl Genet**, 120: 93-115
- McMullen, M. D. et al. Genetic Properties of the Maize Nested Association Mapping Population. **Science** 325:737-740
- Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization1. **Agron J** 74:562-564
- Nelson PT, Coles ND, Holland JB, Bubeck MB, Smith S, and Goodman JMM (2008) Molecular Characterization of Maize Inbreds with Expired U.S. Plant Variety Protection. **Crop Science** 48:1673-1685
- Pace J, Lee N, Naik HS, Ganapathysubramanian B, Lübberstedt T (2014) Analysis of Maize *Zea mays* L.) Seedling Roots with the High-Throughput Image Analysis Tool ARIA (Automatic Root Image Analysis). **PLoS ONE** 9:e108255
- Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradbury PJ, Buckler ES (2014) The Genetic Architecture Of Maize Height. **Genetics** 196:1337-1356

- Petróleo Brasileiro S.A. **Análise do Mercado de Fertilizantes**. Disponível em: <<http://www.cetem.gov.br/agrominerais/livros/AnaliseMercado.pdf>>. Acesso em 03 de novembro de 2014
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. **Genetics** 155:945-959
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014. URL <http://www.R-project.org/>.
- Rincent R, Nicolas S, Bouchet S, Altmann T, Brunel D, Revilla P, Malvar RA, Moreno-Gonzalez J, Campo L, Melchinger AE, Schipprack W, Bauer E, Schoen CC, Meyer N, Ouzunova M, Dubreuil P, Giauffret C, Madur D, Combes V, Dumas F, Bauland C, Jamin P, Laborde J, Flament P, Moreau L, Charcosset A (2014) Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production. **Theor Appl Genet** 127:2313-2331
- Rogers, JS (1972). Measures of genetic similarity and genetic distance. *Stud Genet VII Univ Texas Publ.* 7213:145-153
- Romay M, Millard M, Glaubitz J, Peiffer J, Swarts K, Casstevens T, Elshire R, Acharya C, Mitchell S, Flint-Garcia S, McMullen M, Holland J, Buckler E, Gardner C (2013) Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biology** 14:R55
- SAS. 2007. SAS 9.1.3 (TS1M3) for Windows Microsoft. SAS Institute Inc., Cary, USA, sem paginação
- Semagn K, Magorokosho C, Vivek B, Makumbi D, Beyene Y, Mugo S, Prasanna B, Warburton M (2012) Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. **BMC Genomics** 13:113
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. **Molecular Biology and Evolution** 24:1596-1599
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. **Nat Genet** 43:159-162
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. **Proceedings of the National Academy of Sciences** 108:1088-1092
- Van Inghelandt D, Melchinger A, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and

SNP markers. **Theor Appl Genet** 120:1289-1299

- Westengen OT, Berg PR, Kent MP, Brysting AK (2012) Spatial Structure and Climatic Adaptation in African Maize Revealed by Surveying SNP Diversity in Relation to Global Breeding and Landrace Panels. **PLoS ONE** 7:e47832
- Wu X, Li Y, Shi Y, Song Y, Wang T, Huang Y, Li Y (2014) Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. **Theor Appl Genet** 127:621-631
- Wu Y, Liu W, Li X, Li M, Zhang D, Hao Z, Weng J, Xu Y, Bai L, Zhang S, Xie C (2011) Low-nitrogen stress tolerance and nitrogen agronomic efficiency among maize inbreds: comparison of multiple indices and evaluation of genetic variation. **Euphytica** 180:281-290
- Xu Y, Wang R, Tong Y, Zhao H, Xie Q, Liu D, Zhang A, Li B, Xu H, An D (2014) Mapping QTLs for yield and nitrogen-related traits in wheat: influence of nitrogen and phosphorus fertilization on QTL expression. **Theor Appl Genet** 127:59-72
- Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch J, Xu Y (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. **Mol Breeding** 25:441-451
- Zhang X, Perez-Rodriguez P, Semagn K, Beyene Y, Babu R, Lopez-Cruz MA, San Vicente F, Olsen M, Buckler E, Jannink JL, Prasanna BM, Crossa J (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. **Heredity** 114:291-299

11 CAPÍTULO 2

Influência da estrutura genética da população pelo critério de Evanno na acurácia de GWAS em milho tropical

12 Resumo

Estudos de associação genômica ampla (GWAS) vêm sendo utilizados com sucesso para identificar genes associados a características quantitativas de interesse. Devido à ocorrência de diferentes frequências alélicas entre subpopulações, a estrutura populacional pode produzir falsas associações entre marcadores e características fenotípicas. Métodos estatísticos têm sido desenvolvidos para evitar essas associações espúrias. O objetivo deste estudo foi analisar o efeito da estrutura genética da população, do número de subgrupos em que a população é estruturada e do número ótimo de grupos com base no critério de Evanno (Evanno et al., 2005) sobre a acurácia em GWAS para características oligogênicas e poligênicas em linhagens de milho tropical. Para isso, 62 linhagens foram genotipadas com 768 SNPs e fenotipadas para altura de planta (AP) e produtividade de grãos (PG). Foi realizada a análise da estruturação da população em diferentes números de grupos (K=1 a 9), com auxílio do programa STRUCTURE. As análises de GWAS foram realizadas usando o programa TASSEL. O número de grupos e o valor de ΔK influenciaram de forma distinta a acurácia dos modelos de GWAS para as diferentes características. No caso da AP, a inclusão da matriz Q nos modelos levou a um maior ajuste destes, principalmente nos casos em que o número de grupos apresentaram valores de ΔK mais elevados. Para PG, a inclusão da matriz Q reduziu a acurácia do modelo para todos os valores de K, independentemente dos valores de ΔK . Portanto, os resultados indicam que a escolha do número ótimo de grupos geneticamente estruturados, com base no critério de Evanno, pode não levar a maiores acurácias dos modelos em GWAS. Além disso, ficou evidente que a estruturação influencia de maneira distinta características oligogênicas e poligênicas.

Palavras-chave:- estrutura genética, acurácia, GWAS

13 Abstract

Influence of structure population and Evanno criterion in the accuracy of GWAS in tropical maize

Genome-wide association studies (GWAS) have been successfully used in agricultural research to identify genes associated with quantitative traits of interest. Due to the occurrence of different allelic frequencies among subpopulations, the population structure can produce false associations between markers and phenotypic traits in genomic association studies. Statistical methods have been developed in order to avoid spurious associations in GWAS. The objective of this study was to analyze the effect of the genetic structure of the population, the number of subgroups in the population is structured and the optimal number of groups based on criteria Evanno (Evanno et al., 2005) on accuracy in GWAS for different characteristics in tropical maize lines. To this 62 lines were genotyped with 768 SNPs and phenotyped for plant height (PH) and grain yield (GY). Then the analysis was performed structuring of the population in different numbers of groups ($k = 1$ to 9) by means of Bayesian model, using the STRUCTURE program. The analysis of GWAS were performed using TASSEL program. The results indicated that the number of groups and the value of ΔK affect differently the accuracy of GWAS models for different characteristics. In the case of PH, the inclusion of Q matrix led to a better adjustment of model, especially in cases where the number of the groups had higher ΔK values. In the case of GY, the inclusion of the Q matrix reduced the accuracy of the model for all K values, regardless of the values of ΔK . Therefore, the results indicate that the choice of the optimal number of genetically structured groups through Evanno criterion cannot lead to higher accuracies of models in GWAS. Furthermore, it was evident that the structure affects differently oligogenic and polygenic traits.

Keywords: genetic structure, accuracy, GWAS

14 Introdução

Em milho, estudos de associação genômica ampla (GWAS) têm sido realizados com sucesso em pesquisas para detectar genes relacionados a diversas características quantitativas como produção de biomassa (Rincent et al., 2014), altura de planta (Peiffer et al., 2014), arquitetura de espiga e grão (Chen et al., 2014), arquitetura foliar (Tian et al., 2011), caracteres de raiz (Kuma et al., 2014; Pace et al., 2014;) e resistência a doenças (Hao et al., 2014; Liu et al., 2014; Wongkaew et al., 2014). Também têm sido usados para estudos quanto à tolerância a diferentes tipos de estresses abióticos como hídrico (Almeida et al., 2014), baixas (Sobkowiak et al., 2014) e altas temperaturas (Westengen et al., 2012), e para o entendimento da expressão da heterose (Pea et al., 2013; Thiemann et al., 2014).

A estrutura genética da população pode influenciar as predições genômicas e os resultados em GWAS (Zhang et al., 2010; Wang et al., 2014; Li 2014). A estruturação ocorre em populações aleatórias ou com pedigree conhecido, devido à geografia, à seleção natural ou à seleção artificial (Price et al 2010; van Heerwaarden et al., 2011). Devido à ocorrência de diferentes frequências alélicas entre subpopulações, a estrutura populacional pode produzir falsas associações entre os marcadores e as características fenotípicas em GWAS (Price 2010). Consequentemente, essas falsas associações podem inflacionar as estimativas de herdabilidade e reduzir a acurácia das predições genômicas (Makowsky et al; 2011; Visscher et al 2012; Riedelsheimer et al 2012; Wray et al 2013).

Métodos estatísticos têm sido desenvolvidos com o objetivo de se evitar essas associações espúrias em GWAS (Stich et al., 2008; Zhang et al., 2010; Li et al., 2014). Nesses métodos, a população pode ser dividida em subpopulações e, em seguida, podem ser realizados testes de associação dentro das subpopulações, ou serem usadas estimativas de adesão à população como covariável em um modelo linear ou logístico (Pritchard et al., 2000). Associações espúrias também podem ser causadas por diferenças de parentesco entre pares de indivíduos. Essas associações podem ser reduzidas pelo uso de um modelo linear geral (GLM) para estimar a proporção de identidade por descendência entre qualquer par de indivíduos, excluindo indivíduos aparentados (Voight e Pritchard, 2005). Como alternativa, a estrutura da população e o

parentesco desigual podem ser simultaneamente contabilizados em um modelo linear misto (MLM). Neste sentido, os *subpopulation memberships* (Matriz Q), baseados nos marcadores moleculares, podem ser tratados como efeitos fixos e o parentesco pode ser utilizado para definir a estrutura da variância e da covariância dos efeitos aleatórios individuais (Yu et al., 2006). Este método tem se mostrado mais eficiente do que o MLM e outros métodos em relação ao poder estatístico (Zhang et al., 2010).

Para análise de estrutura genética de populações, o método baseado em modelos bayesianos, realizado com auxílio do programa computacional STRUCTURE (Prichard et al., 2000), tem sido um dos mais utilizados (Earl e vonHoldt, 2012). Nessa análise, a população é dividida em grupos e cada indivíduo é atribuído a cada grupo de acordo com suas frequências alélicas. No entanto, inferir o número de grupos que melhor se ajusta aos dados é um desafio, e alguns métodos foram desenvolvidos com essa finalidade (Pritchard et al., 2000; Evanno et al. 2005; Travaet al., 2006; Hubisz et al., 2009; Schwartz e McKelvey, 2009). Um dos métodos mais utilizados para definir o número ótimo de grupos em que a população está estruturada geneticamente é o critério de Evanno (Evanno et al., 2005), o qual é baseado no valor de ΔK , que expressa a taxa de mudança de segunda ordem do logaritmo da probabilidade marginal dos dados (Evanno et al., 2005).

Apesar do grande número de pesquisas utilizando GWAS e a estruturação de populações via modelos bayesianos para o entendimento da arquitetura genética de características complexas e para detecção de genes relacionados a características de interesse (Stich et al., 2008; Caniato et al., 2011; Westenge et al., 2012), ainda é reduzido o número de estudos que avaliem o efeito do número de grupos em que a população é estruturada e do número ótimo de grupos com base no critério de Evanno sobre a acurácia em GWAS.

15 Objetivo e Hipóteses testadas

a. Objetivo Geral

Analisar o efeito da estrutura genética da população, do número de subgrupos em que a população é estruturada e do número ótimo de grupos com base no critério de Evanno sobre a acurácia em GWAS, para características oligogênicas e poligênicas em linhagens de milho tropical.

b. Hipóteses Testadas

H1: Há efeito da estruturação genética via inferência bayesiana na acurácia em GWAS.

H2: O número de grupos formados pela estruturação genética via inferência bayesiana e pelo critério de Evanno afeta a acurácia em GWAS.

H3: A estruturação populacional afeta de maneira diferenciada a acurácia em GWAS para caracteres oligogênicos e poligênicos.

16 Materiais e Métodos

a. Material Genético e Fenotipagem

Foram avaliadas 62 linhagens endogâmicas de milho tropical, em campo, na estação experimental Anhembi da Escola Superior de Agricultura Luís de Queiroz (ESALQ), da Universidade de São Paulo (22° 50' 51" S, 48° 01' 06" W, 466h).

As linhagens foram distribuídas no campo de acordo com o delineamento experimental látice simples (8x8), com duas repetições. As parcelas foram constituídas de uma linha com 4m de comprimento, utilizando-se o espaçamento de 0,85m entre linhas e 0,20m entre plantas, totalizando 20 plantas por linha.

O manejo e demais tratamentos culturais foram realizados de acordo com os procedimentos recomendados para a cultura do milho (Borém, Galvão e Pimentel, 2015).

Foram avaliadas duas características, sendo uma oligogênica (altura de planta, AP) e uma poligênica (produtividade de grãos, PG). AP foi mensurada em cinco plantas por parcela, em centímetros, da base do caule à base da folha bandeira. A PG foi mensurada em gramas/parcela e corrigida para 13% de água nos grãos. Nesse caso, não houve correção para o estande, uma vez que seu efeito foi significativo pela análise de variância.

Para a análise de variância foi considerado o seguinte modelo:

$$y_{il(j)} = \mu + t_i + r_j + (b/r)_{l(j)} + e_{il(j)}$$

Em que: μ =média geral; t_i =efeito do tratamento i (genótipo); r_j = efeito da repetição j ; $(b/r)_{l(j)}$ =efeito do bloco incompleto l dentro da repetição j ; $e_{il(j)}$ = erro aleatório associado a $y_{il(j)}$; $e_{il(j)} \sim N(0, \sigma^2)$; $y_{il(j)} \sim N(\mu + t_i + r_j + (b/r)_{l(j)}, \sigma^2)$.

Essas análises foram realizadas por meio do programa SAS 9.3.1 (SAS, 2007).

b. Genotipagem

A genotipagem foi realizada utilizando o ensaio de alta performance Illumina GoldenGate (Fan et al., 2003) contendo 768 marcadores SNPs distribuídos por todo o genoma do milho. Esses marcadores foram disponibilizados pela Pioneer Hi-Bred International Int. (Johnston, IA, USA). Os marcadores SNPs passaram por um processo de análise de qualidade, via *Call Rate* (taxa de atendimento, qualificação ou

repetibilidade). Foi adotado um valor de *Call Rate* > 0,85. Também foi avaliada a MAF (menor frequência alélica) de cada marcador, e selecionados apenas os marcadores com $MAF \geq 0,05$. Para essas análises foi utilizado o pacote HapEstXXR (Knueppel e Rohde, 2014) do programa R (R Core Team, 2014). Para converter os dados genômicos de bases nitrogenadas (letras) em algarismos numéricos foi utilizado o pacote *Scrim* (Schwender e Fritsch, 2013).

c. Análise da estrutura genética da população

A estrutura populacional das 62 linhagens foi analisada por meio do método bayesiano, com o auxílio do aplicativo STRUCTURE versão 2.3.3 (Pritchard et al., 2000). Foram realizadas corridas para vários números possíveis de grupos (K), variando de 1 a 9. Para cada K foram realizadas 10 corridas com um período de *burn-in* de 500.000 e 500.000 replicações MCMC (Markov Chain Monte Carlo) após o *burn-in*. As linhagens foram atribuídas aos grupos (populações) de acordo com os seus valores médios de probabilidade de associação (Q). Q_{ik} representa a proporção estimada do genoma da linhagem *i* derivada da população *k* (Pritchard, 2000). Foram considerados como pertencentes a um mesmo grupo os indivíduos cuja probabilidade de associação a esse grupo foi superior a probabilidade de associação a qualquer outro grupo. O valor mais provável de K foi estimado utilizando o critério de Evanno baseado no valor de ΔK (Evanno et al., 2005), com auxílio do aplicativo Structure Harvest (Earl and Vonholdt 2012). ΔK foi estimado para cada K por meio da expressão “

$$\Delta K = \frac{m(|L(K+1) - 2L(K) + L(K-1)|)}{\sigma[L(K)]}$$

”, em que, $L(K)$ é a média dos logaritmos neperianos

da probabilidade dos dados de cada passo do MCMC e subtraído pela metade da variância para o número de populações K .

d. Análises de detecção de QTLs via GWAS

O conjunto de marcadores, juntamente com as matrizes de estrutura populacional (Q) e Kinship (que estabelece a relação de parentesco entre as linhagens utilizadas) foram analisados pelo seguinte modelo misto:

$$y = X\beta + Zv + \varepsilon$$

Em que: y é o vetor de observações fenotípicas para AP e PG; β é o vetor contendo os efeitos fixos, incluindo a média geral de efeito dos marcadores genéticos e a matriz de estrutura populacional (Q); v vetor dos efeitos aleatórios genéticos aditivos de múltiplos SNPs e a matriz de parentesco (*Kinship*); ε vetor residual. Os vetores v e ε são assumidos como tendo distribuição normal, média zero e variância:

$$\text{Var} \begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$$

Em que: $G = \sigma_a^2.K$, e σ_a^2 é a variância genética aditiva; $R = I.\sigma_e^2$, e σ_e^2 é a variância residual, assumida como homogênea.

Para essas análises de GWAS, foram considerados diferentes cenários. Primeiramente, foi considerado um modelo sem estruturação da população (K=1). Em seguida foram considerados modelos com diferentes números de grupos formados pela estruturação genética da população (K=2 a 9). Os modelos acima foram analisados por meio do software TASSEL 5.3 (Bradbury *et al.*, 20015). As matrizes Q e K foram geradas pelos softwares STRUCTURE (Pritchard *et al.*, 2000) e TASSEL 5.3, respectivamente (Bradbury *et al.*, 2015).

Para os testes de significância dos p-valores, foi usado o método de Bonferroni (Bonferroni, 1936), em que o limiar para que o efeito de um SNP seja considerado como significativo é $0,05/n^\circ$ de marcadores, em que 0,05 é o nível de significância global. Para os 539 SNPs avaliados, os seus efeitos foram considerados significativos quando $p \geq 9,27 \times 10^{-5}$. O teste foi realizado por meio do software TASSEL 5.3.

e. Análise do efeito da estrutura populacional e do Critério de Evanno sobre a acurácia em GWAS

Para analisar o efeito da estrutura populacional, do número de grupos e do número ótimo de grupos pelo critério de Evanno sobre o ajuste do modelo, os p-valores esperados e observados foram plotados em gráficos *Quantili-Quantile (Q-Q Plot)*. Nesses gráficos, o ajuste dos modelos para diferentes números de grupos (K) foi avaliado por meio da observação da dispersão dos pontos em torno da linha de tendência. Além disso, para quantificar essa dispersão foi utilizada a medida MSD (*Mean Square Difference*) entre os p-valores esperados e observados, (Yu *et al.*, 2006, Stich *et*

al, 2008), dado que os p-valores dos marcadores aleatórios seguem uma distribuição uniforme (Mamidi et al., 2011). O MSD para cada cenário foi estimado como a soma dos valores absolutos dos desvios entre os p-valores observados e esperados para cada marcador. Foi estimado também o MSD considerando apenas o segundo quadrante do Q-Q Plot para cada cenário (MSD2), pois a análise do segundo quadrante pode refletir melhor o grau de desvio de cada modelo, uma vez que é justamente nele que são observadas as maiores variações entre os p-valores esperados e observados.

Foram analisados também o número de coincidências entre os 5% dos marcadores de maiores efeitos em cada um dos cenários avaliados com e sem estrutura populacional, tanto para altura de planta quanto para produtividade de grãos, indiferentemente da significância do efeito do SNP.

17 Resultados e Discussão

a. Análise Fenotípica

Houve efeito significativo de genótipo pelo sobre a AP e a PG (Tabela 1). Entre as 62 linhagens, a altura variou de 127,5 (L44) a 201,5 cm (L50), com média de 160,88 e variância de 576,63. Para essa característica, o coeficiente de variação (CV%) foi de 5,58% e a herdabilidade estimada foi de 0,82. A produtividade de grãos variou de 2112,5 a 3194,66 gramas por parcela, com média 2519,62 e variância 64135,38. Para essa característica o CV% foi de 7,05% e a herdabilidade estimada de 0,71.

Tabela 1 – Quadro de análise de variância de 62 linhagens de milho tropical em relação a altura de planta (AP) e produtividade de grãos (PG).

Fontes de Variação	GL	AP	PG
		QM	QM
Bloco(rep)	14	673,49*	69090,75 ns
Repetição	1	3,51ns	273957,34 *
Genótipo	63	478,05*	125324,13 *
Resíduo	49	85,31	39796,26
CV(%)	-	5,62	7,97
Media	-	164,62	2519,62
Herdabilidade		0,82	0,71

^{ns}não significativo, * significativo a 1% de probabilidade pelo teste F, respectivamente.

QM: quadrado médio.

CV: Coeficiente de variação.

De acordo com Fritsche-Neto et al. (2012), os valores de CV% obtidos neste trabalho para AP e PG são considerados baixos, o que indica alta precisão experimental.

A AP é uma das características mais herdáveis e facilmente mensuráveis em milho. De acordo com Peiffer et al. (2014), em um dado conjunto de genótipos de milho em que seja conhecido o *pedigree* ou as estimativas genômicas de identidade por estado, a altura pode ser predita com bastante precisão. No caso da PG, a precisão das predições tem sido menores, provavelmente devido ao maior número de genes envolvidos na expressão e à arquitetura genética mais complexa dessa característica (Wang et al., 2015). Logo, a AP pode ser considerada uma característica oligigênica e a PG poligênica (Hallauer et al, 2010).

b. Análise da Estrutura Genética da População

Do total de 768 marcadores SNPs inicialmente utilizados neste estudo, 50 falharam na genotipagem de todos os indivíduos e foram eliminados. Dos marcadores restantes 78, apresentaram uma taxa de dados perdidos superior a 15% (*Call Rate* = 0,85), e também foram eliminados. Foram eliminados também 101 marcadores que apresentaram MAF inferior a 0,05. Na análise da estrutura genética foram utilizados 539 marcadores SNPs altamente informativos, que atenderam aos critérios de qualidade estabelecidos.

De acordo com o critério de Evanno, foram observados três picos de probabilidade (ΔK), com K variando de 1 a 9 (Figura 1). O número de grupos mais provável em que a população pode ser geneticamente estruturada foi K=2. No entanto, as estruturas com K=4 e K=8 apresentaram valores de ΔK muito próximos do valor observado para K=2. As matrizes Q formadas pelos coeficientes de atribuição de cada indivíduo a cada grupo, obtidas para diferentes números de grupos (2 a 9), foram incorporadas aos modelos para permitir a análise comparativa do efeito do número de grupos e dos valores ΔK sobre a acurácia do modelo, sendo que as matrizes Q para K=2, 4 e 8 foram as mais prováveis de acordo com o critério de Evanno para a estruturação genética da população e as mais indicadas para análises em GWAS.

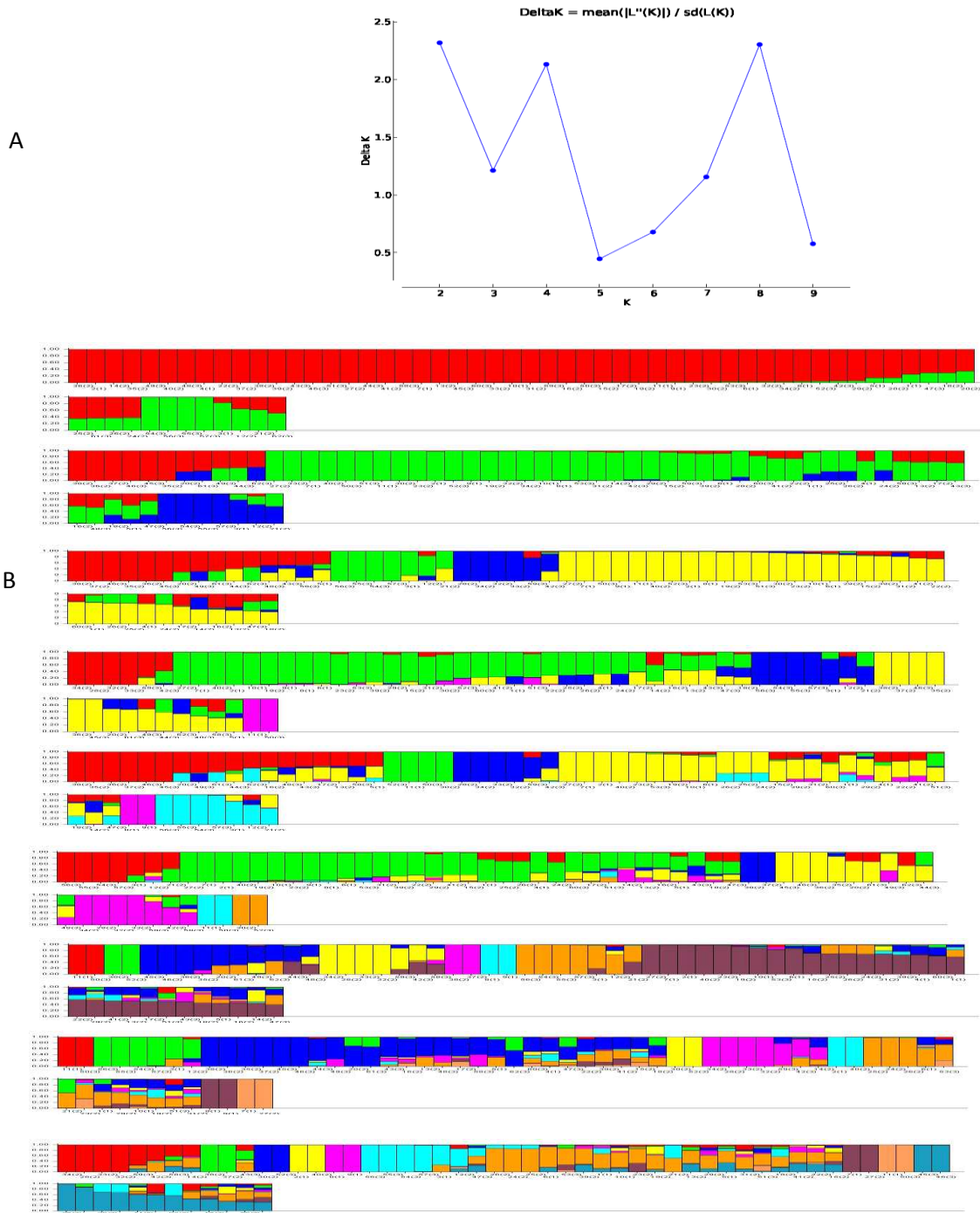


Figura 1 – Valores de ΔK em função do número de grupos K (A). Gráficos de barras gerados pelo STRUCTURE mostrando a atribuição de cada linhagem a cada grupo genotípico (cores) para números de grupos (K) variando de 2 a 9 (B).

c. GWAS

Para altura de plantas, utilizando o modelo sem incluir a matriz Q ($K=1$), nenhum SNP foi detectado como significativo. Quando foi utilizada a matriz Q da estruturação das linhagens em três grupos ($K=3$), foram detectados três SNPs significativos. Um está localizado no cromossomo 5 (PHM4185.17) e dois no cromossomo 1 (PZA00664.3 e PZA01887.1). Quando as linhagens foram estratificadas em sete grupos ($K=7$), um SNP posicionado no cromossomo 2 foi detectado como significativo (PHM3309.8). Para todos os outros valores de K ($K=1; 2; 4; 5; 6; 8$ e 9), nenhum SNP com efeito significativo foi detectado (Figura 2).

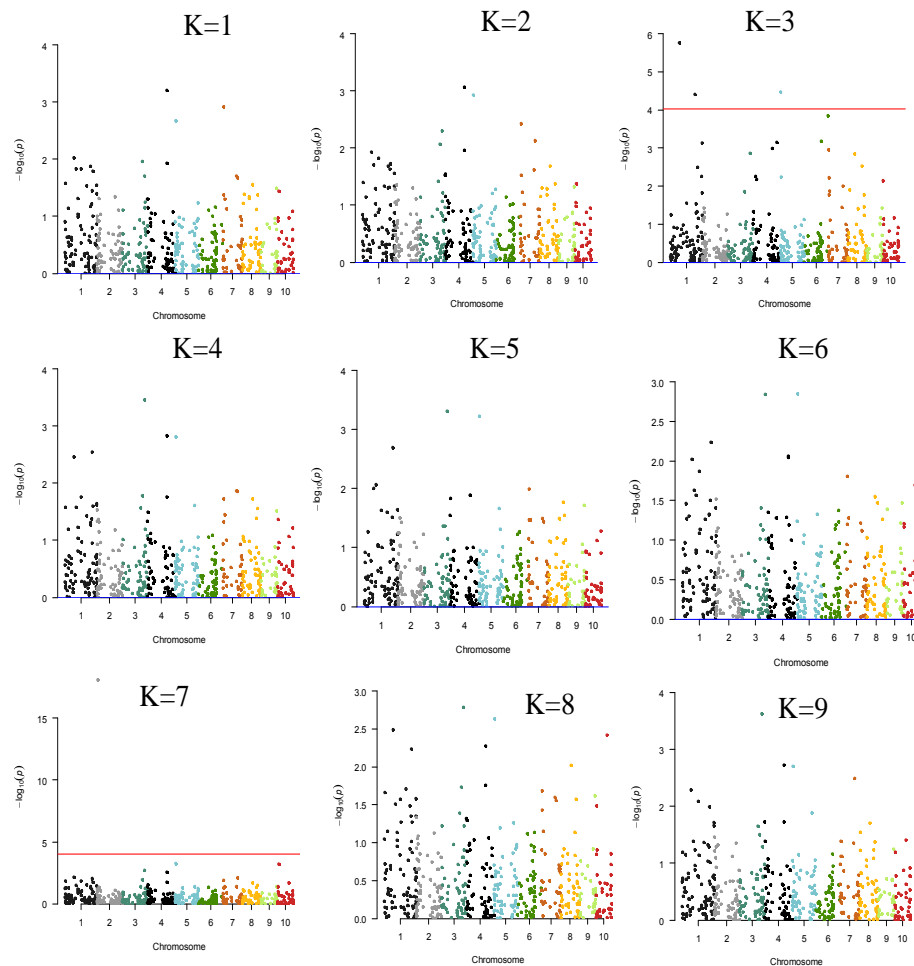


Figura 2 –Manhattan Plot de GWAS para altura de plantas em 62 linhagens de milho tropical genotipadas com 539 SNPs, sem estruturação ($K=1$) e com estruturação em diferentes números de grupos ($K=2$ a 9).

Para produtividade de grãos, utilizando o modelo sem incluir a matriz Q, nenhum SNP foi significativo. No caso da estruturação em dois grupos ($K=2$), ocorreram 151 SNPs significativos, distribuídos por todos os 10 cromossomos. Nas estruturações com K igual a 4, 5, 6 e 8 foram detectados 7, 3, 6 e 4 SNPs significativos, respectivamente. Com $K=9$ não foi identificado nenhum SNP significativo (Figura 3).

Para essa característica, apesar do alto poder de detecção de SNPs significativos para vários modelos com diferentes valores de K , principalmente para $K=2$, há grande probabilidade de serem falsos positivos. A ocorrência de falsos positivos (erro tipo I) em estudos de associação genômica ampla pode ser controlada, mas isso só é possível à custa da redução do poder de detecção de verdadeiros positivos ou do poder estatístico (Li et al., 2014). Por exemplo, o estabelecimento de um limiar rigoroso para o teste de associação é uma maneira eficaz de controlar a taxa de falsos positivos. Porém, isso também faz com que o número de verdadeiros positivos detectados seja reduzido. Uma solução desejável consistiria em reduzir os falsos positivos, sem comprometer poder estatístico da análise. Esta solução é crítica, porque a herança da maioria das características agronomicamente importantes é controlada por muitos genes, os quais individualmente têm efeitos pequenos ou alelos raros (Buckler et al., 2009; Yang et al., 2010).

A inclusão da estrutura populacional modificou significativamente os resultados em GWAS. Em alguns casos ela aumentou o poder de detecção de SNPs significativos, e em outros casos ela o reduziu. Além disso, nos diferentes cenários em que foram detectados SNPs significativos, estes não foram coincidentes, o que indica uma inconsistência e grande variação nos resultados a serem obtidos de acordo com o K utilizado, sendo K o mais provável ou não pelo critério de Evanno.

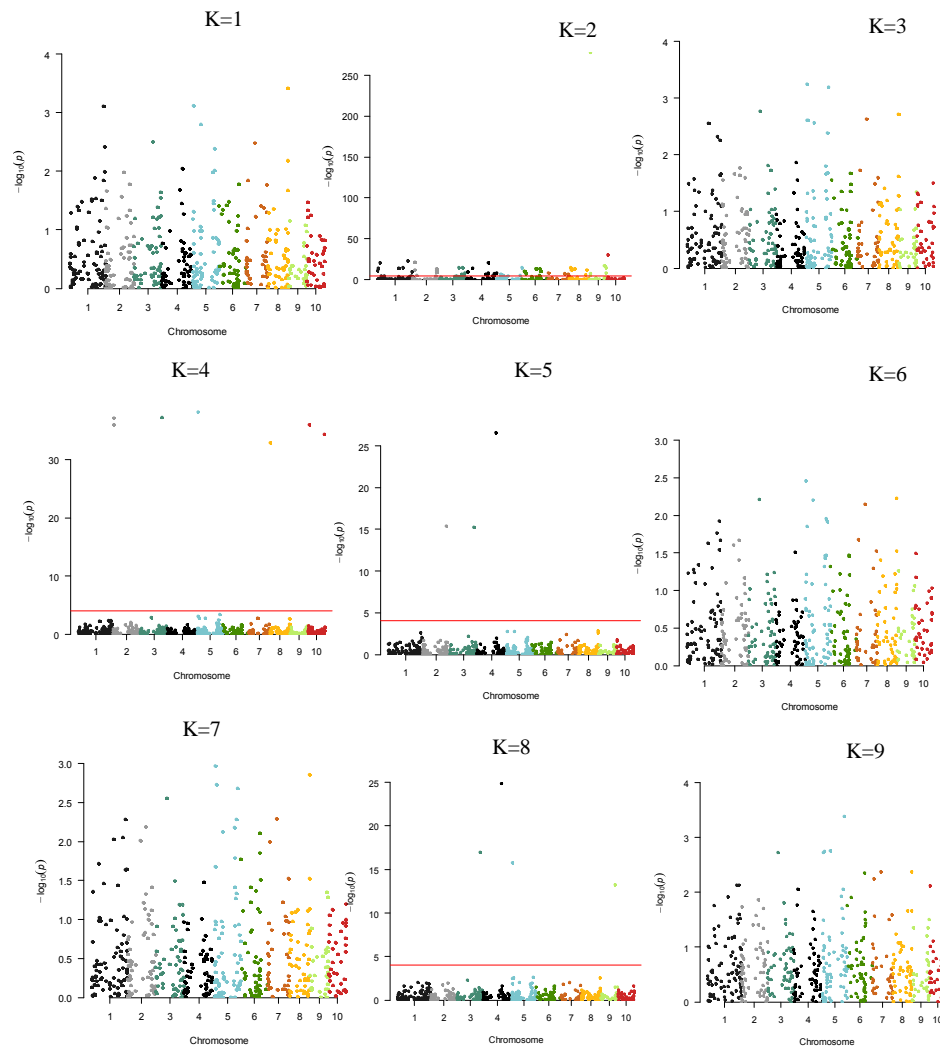


Figura 3 - Manhattan Plot de GWAS para produtividade de grãos em 62 linhagens de milho tropical genotipadas com 539 SNPs, sem estruturação ($K=1$) e com estruturação em diferentes números de grupos ($K=2$ a 9).

Considerando o conjunto dos 5% dos SNPs com maior efeito na AP, independentemente de serem ou não significativos, nos modelos sem estruturação e com estruturação, para todos os números de grupos, os SNPs com maiores efeitos foram os mesmos em todos os casos ($K=1$ a 9). Diferentemente, no caso da PG observou-se uma grande variação dos SNPs identificados como os de maiores efeitos para os diferentes valores de K . Portanto, apesar do baixo poder de detecção de SNPs

significativos, os modelos de GWAS para AP, com ou sem estruturação, em geral apresentaram maior acurácia em relação à observada no caso da PG.

Tabela 5 – Número de SNPs coincidentes entre as 5% de maior efeito (27) nos diferentes modelos de GWAS sem estrutura (K=1) e com estrutura populacional (K=2 a 9) para as características altura de planta (AP – diagonal superior) e produtividade de grãos (PG – diagonal inferior).

PG/AP	K1	K2	K3	K4	K5	K6	K7	K8	K9
K1	-	27	27	27	27	27	27	27	27
K2	3	-	27	27	27	27	27	27	27
k3	27	3	-	27	27	27	27	27	27
K4	1	2	1	-	27	27	27	27	27
K5	0	2	0	15	-	27	27	27	27
K6	0	2	0	18	15	-	27	27	27
k7	1	2	1	20	14	21	-	27	27
k8	1	2	1	18	13	17	20	-	27
K9	1	2	1	15	10	15	16	14	-

d. Análise do efeito do número de grupos e do critério de Evanno na acurácia em GWAS

Para a AP, o menor desvio dos p-valores observados em relação aos esperados ocorreu na estruturação da população em oito grupos (K=8), a qual na análise de estruturação apresentou o segundo maior valor de ΔK , quase idêntico ao maior valor, que foi observado para K=2 (Figura 4). Isso indica que o modelo incluindo a matriz Q da estruturação com K=8 foi o que mostrou o maior grau de ajuste. As estruturações com K=5 e K=6 apresentaram valores relativamente baixos e próximos entre si (0,015 e 0,018, respectivamente). Para K= 1, 2, 4 e 9 os valores de MSD foram um pouco maiores do que os anteriores e também muito próximos entre si (0,025, 0,027, 0,027 e 0,024 respectivamente). As estruturações com K=3 e K=7, que apresentaram valores intermediários de ΔK , foram as que mais levaram a um desajuste do modelo, apresentando valores de MSD bem superiores em relação às outras estruturações (0,066 e 0,069 respectivamente). Portanto, os únicos casos em que a inclusão da matriz Q no modelo levou a uma melhora no ajuste deste, foram nas estruturações com K=5, K=6 e principalmente K=8. Destas, a única coerente com o critério de Evanno foi K=8, o que

indica uma baixa relação entre este critério e a melhor estruturação a ser usada para GWAS.

Entretanto, considerando apenas o segundo quadrante do gráfico Q-Q Plot, onde se observa uma maior variação entre os p-valores esperados e observados (MSD2), ficou clara a relação positiva entre os valores ótimos de ΔK e o aumento na acurácia do modelo. Nesse caso, as estruturações com os maiores valores de ΔK foram exatamente as que apresentaram maiores ajustes dos modelos e, portanto, menores valores de MSD2. Os modelos sem estrutura ($K=1$) e com estrutura da população em dois grupos e em 8 grupos ($K=2$ e $K=8$) apresentaram valores de MSD muito próximos entre si (0,11; 0,11 e 0,10, respectivamente). Isso indica que, a inclusão das matrizes Q com $K=2$ e $K=8$, que foram os números ótimos de grupos geneticamente estruturados pelo critério de Evanno, no modelo de GWAS, dentre os modelos com estrutura, foram os mais acurados. No entanto, em relação a $K=1$ não houve aumento significativo da acurácia. Os valores de MSD2 para modelos com estruturações em números de grupos com os mais baixos valores de ΔK ($K=5, 6$ e 9) também foram relativamente baixos (0,13; 0,09 e 0,17, respectivamente). O menor valor de MSD2 foi observado quando $K=6$. Os valores de MSD2 para os modelos incluindo estruturações com valores intermediários de ΔK ($K=3$ e $K=7$), foram extremamente elevados em relação aos demais casos (1,28 e 1,19, respectivamente), indicando uma baixa acurácia desses modelos de GWAS. Portanto, os resultados evidenciam que para AP, os modelos mais acurados são o sem estruturação ($K=1$) e aqueles cujas estruturações apresentam valores extremos de ΔK , tanto os mais altos ($K=2, 4$ e 8) quanto os mais baixos ($K=5, 6$ e 9).

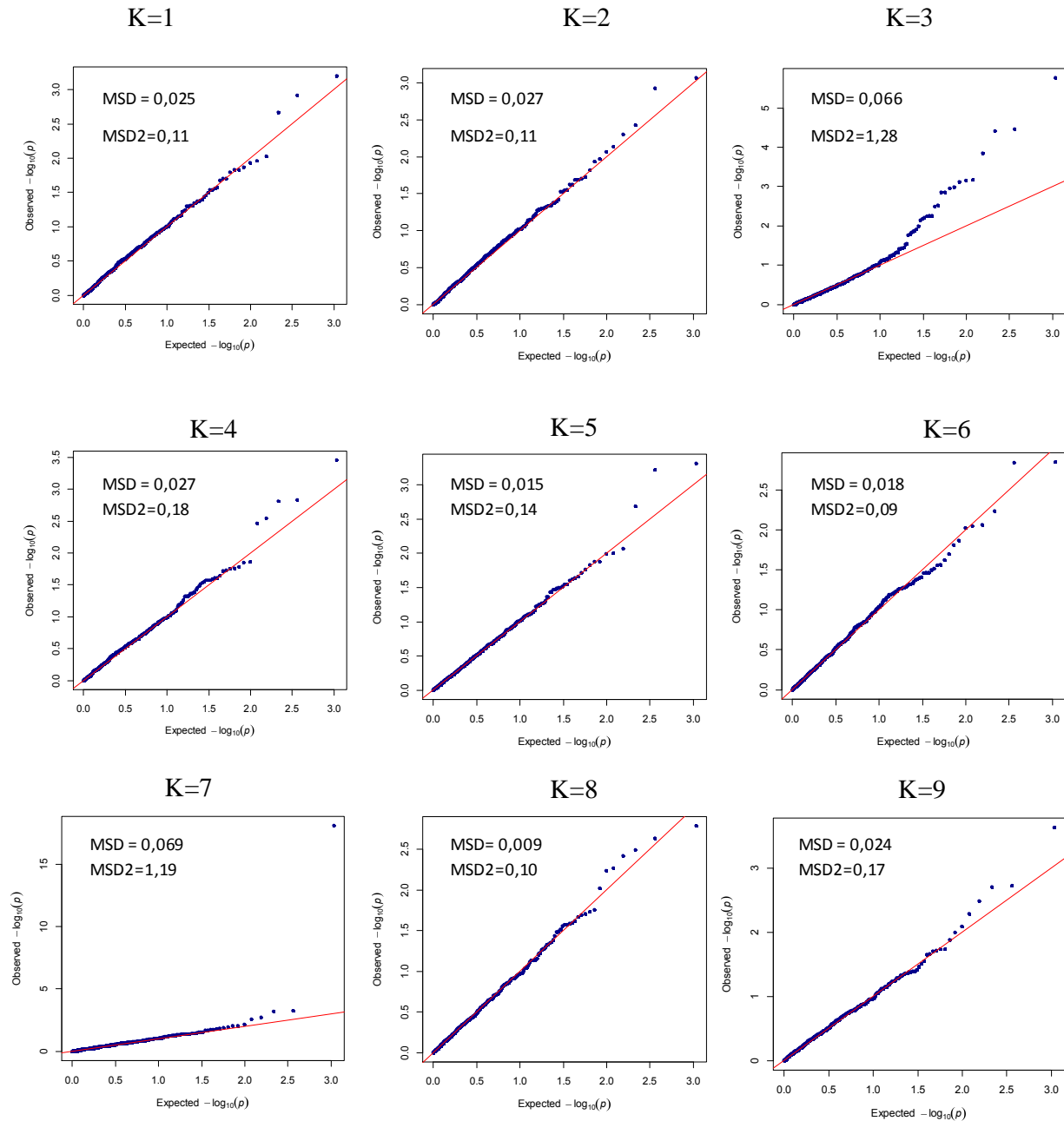


Figura 4 – *Quantile-Quantile Plot* para a variação dos p-valores esperados e observados (*Mean Square Difference*) para 539 SNPs usados em GWAS para altura de planta em 62 linhagens de milho tropical sem estruturação (K=1) e estruturadas em diferentes números de grupos (K=2 a 9). O MSD considera todos os marcadores e MSD2 refere-se a apenas aos marcadores do segundo quadrante do gráfico.

No caso da PG, o menor valor de MSD foi observado para K=6 (0,024). Em seguida, os modelos que apresentaram maior grau de ajuste foram os com K=1, K=3, e K=7 (0,083; 0,092 e 0,056, respectivamente). Para K=5, K=8 e K=9 os valores de MSD

foram próximos entre si e superiores aos anteriores (0,154; 0,173 e 0,184). As estruturas que levaram aos maiores valores de MSD foram as com $K=2$ e $K=4$. No caso do $K=2$, cujo valor de ΔK foi maior que os demais, o valor de MSD foi extremamente superior aos demais (2,08), inclusive que do $K=4$ (0,53). Portanto, para essa característica, a inclusão da matriz Q no modelo só levou ao aumento do ajuste no caso de $K=6$ e $K=7$. Para $K=3$, o ajuste do modelo foi muito semelhante ao do modelo sem a matriz Q ($K=1$). Além disso, as estruturas com os valores de K que apresentaram os três picos de ΔK ($K=2$, $K=4$ e $K=8$) mostraram alto grau de desajuste do modelo em relação ao modelo sem matriz Q. A estruturação com $K=2$, cujo valor de ΔK foi o maior, apresentou um grau de ajuste do modelo muito inferior aos demais (Figura 5) evidenciando que, para características poligênicas, a estruturação pelo critério de Evanno não representou o melhor cenário para estudos GWAS, diferente do que foi observado para a característica oligogênica AP.

Considerando apenas o segundo quadrante do gráfico Q-Q Plot, os menores valores de MSD2 foram observados no modelo sem estruturação ($K=1$) e nos modelos com estruturas com $K=3$, 6, 7 e 9 (MSD2=0,53; 0,07; 0,43 e 0,47, respectivamente). O maior MSD2 foi observado com $K=2$, ou seja, para essa característica, o resultado reitera que o modelo incluindo a matriz Q obtida a partir da estruturação da população no número ótimo de grupos pelo critério de Evanno foi o menos acurado, diferentemente do que foi observado no caso da AP. Além disso, as estruturas com $K=4$ e $K=8$, que também apresentaram picos de ΔK , foram as que, após $K=2$, apresentaram os valores mais elevados de MSD2.

De modo geral, os resultados indicam que o número de grupos e o valor de ΔK influenciam de forma distinta a acurácia dos modelos de GWAS para as características oligogênicas e poligênicas. No caso da AP, a inclusão da matriz Q no modelo levou a uma maior acurácia dos modelos, principalmente nos casos em que os números de grupos apresentaram valores de ΔK mais elevados ($K=2$, 4 e 8). No caso da característica PG, a inclusão da matriz Q reduziu a grau de ajuste do modelo para todos os valores de K, independentemente dos valores de ΔK .

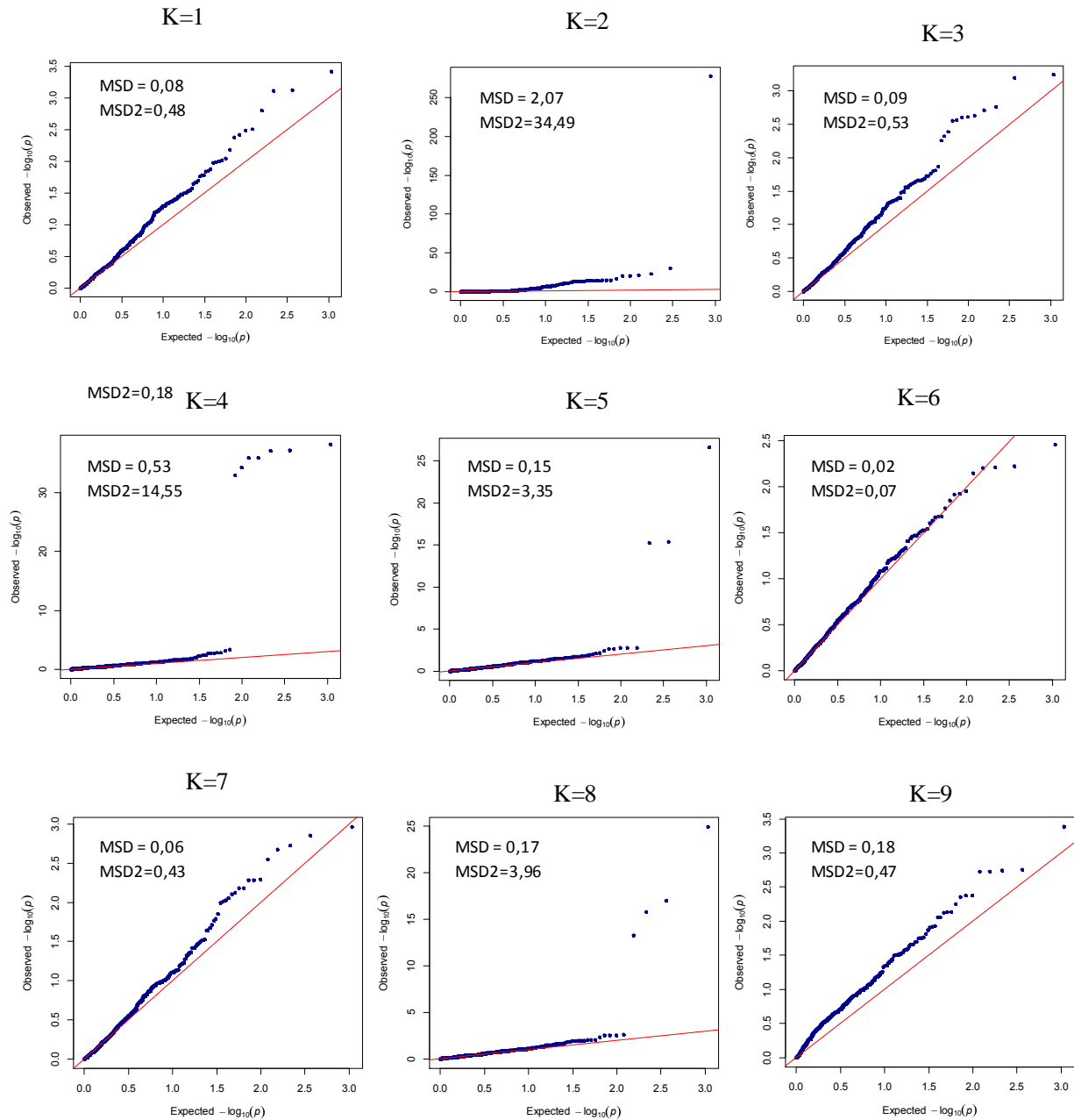


Figura 5 — *Quantile-Quantile Plot* para a variação dos p-valores esperados e observados (*Mean Square Difference*) para 539 SNPs usados em GWAS para produtividade de grãos em 62 linhagens de milho tropical sem estruturação (K=1) e estruturadas em diferentes números de grupos (K=2 a 9). O MSD considera todos os marcadores e MSD2 refere-se a apenas aos marcadores do segundo quadrante do gráfico.

Estes resultados corroboram com os obtidos por Wang et al. (2015), que analisando a predição genômica em populações biparentais de milho tropical, concluíram que a herdabilidade e a arquitetura genética da característica de interesse influenciam o desempenho das predições de desempenho com base em dados genômicos. Segundo esses autores, a precisão da predição de características complexas como a produção de grãos foi consistentemente menor do que a observada no caso de características geneticamente mais simples, como tempo de antese e altura de planta. O efeito da estruturação também pode variar dependendo da característica analisada. Características altamente correlacionadas com a estrutura populacional apresentam problemas para GWAS, independentemente da metodologia (Yu et al., 2006; Cheng et al. 2010). Além disso, as diferenças na estrutura da população entre conjuntos de germoplasma com altos níveis de diversidade genética e aqueles estreitamente relacionados também podem contribuir para as diferenças observadas no poder estatístico e na acurácia em GWAS (Guo et al., 2014).

Em um estudo comparando vários métodos para o mapeamento de associação em trigo utilizando um conjunto de dados empíricos, Stich et al. (2008) avaliaram nove métodos baseados em modelos mistos para mapeamento de associação. Em dois métodos foi utilizada a matriz Q obtida por meio do programa STRUCTURE. Para decidir sobre o número mais provável de grupos foi utilizado o critério de Evanno, desconsiderando-se, portanto, todas as outras possibilidades de número de grupos em que a população pode ser geneticamente estruturada. No entanto, como observado no presente trabalho, a definição do número ótimo de grupos pelo critério de Evanno pode não levar a um melhor ajuste do modelo em GWAS, dependendo da característica analisada. Portanto, apesar de Stich et al. (2008) terem constatado que o método incluindo a matriz de parentesco obtida por RELM foi mais apropriado para GWAS com respeito ao poder de detecção de marcadores significativos e de controle do erro tipo I, os resultados podem ter sido influenciados por uma escolha inadequada do número de grupos em que a população foi geneticamente estruturada para a obtenção da matriz Q.

Apesar da ampla utilização do critério de Evanno para determinar o número mais provável de grupos e utilização da matriz Q obtida dessa estruturação em GWAS, os próprios autores que descreveram o método ressaltam que apesar do ΔK ser útil para a

identificação do número correto de grupos na maioria das situações, ele não deve ser usado exclusivamente (Evanno et al., 2005). Os autores ressaltam ainda que esse critério deve ser usado em conjunto com outras informações fornecidas pelo STRUCTURE, como o $L(K)$ em si, o valor de α e os padrões de trabalho individuais (ver seção 5 em Pritchard e Wen, 2003). Também é importante ressaltar que os resultados obtidos por esse método são sensíveis ao tipo de marcador molecular utilizado, ao número de locos analisados, ao número de populações amostradas e ao número de indivíduos em cada amostra.

Portanto, os resultados do presente trabalho indicam que a escolha do número ideal de grupos geneticamente estruturados por meio do critério de Evanno pode não levar a melhores ajustes dos modelos em GWAS. Além disso, fica evidente que a estruturação influencia de maneira completamente distinta as características oligogênicas e as poligênicas.

18 Conclusões

- 1) A estruturação genética populacional via inferência bayesiana tem efeito sobre a acurácia dos modelos de GWAS.
- 2) O uso do número ótimo de grupos pelo critério de Evanno para estrutura populacional pode não conduzir a maiores acurácias em GWAS.
- 3) A inclusão da estruturação genética populacional influencia de maneira diferenciada a GWAS para características oligogênicas e poligênicas.

19 Referências Bibliográficas

- Almeida G, Nair S, Borém A, Cairns J, Trachsel S, Ribaut J-M, Bänziger M, Prasanna B, Crossa J, Babu R (2014) Molecular mapping across three populations reveals a QTL hotspot region on chromosome 3 for secondary traits associated with drought tolerance in tropical maize. **Mol Breeding** 34:701-715
- Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-Wide Association Mapping in Arabidopsis Identifies Previously Known Flowering Time and Pathogen Resistance Genes. **PLoS Genet** 1:e60
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics** 23:2633-2635
- Bonferroni, C. E., (1936) Teoria statistica delle classi e calcolo delle probabilità. Istituto Superiore di Scienze Economiche e Commerciali di Firenze.
- Borém, A., Galvão, J.C.C., Pimentel, M.A. 2015. Milho: do plantio à colheita. Viçosa: Editora UFV. 351p.
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Sofia da Silva H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The Genetic Architecture of Maize Flowering Time. **Science** 325:714-718
- Caniato FF, Guimarães CT, Hamblin M, Billot C, Rami J-F, Hufnagel B, Kochian LV, Liu J, Garcia AAF, Hash CT, Ramu P, Mitchell S, Kresovich S, Oliveira AC, de Avellar G, Borém A, Glaszmann J-C, Schaffert RE, Magalhaes JV (2011) The Relationship between Population Structure and Aluminum Tolerance in Cultivated Sorghum. **PLoS ONE** 6:e20830
- Chen Z, Wang B, Dong X, Liu H, Ren L, Chen J, Hauck A, Song W, Lai J (2014) An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F2 maize population. **BMC Genomics** 15:433
- Cheng R, Lim JE, Samocha KE, Sokoloff G, Abney M, Skol AD, Palmer AA (2010) Genome-Wide Association Studies and the Problem of Relatedness Among

- Advanced Intercross Lines and Other Highly Recombinant Populations. **Genetics** 185:1033-1044
- Cheng R, Lim JE, Samocha KE, Sokoloff G, Abney M, Skol AD, Palmer AA (2010) Genome-Wide Association Studies and the Problem of Relatedness Among Advanced Intercross Lines and Other Highly Recombinant Populations. **Genetics** 185:1033-1044
- Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359-361
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology** 14:2611-2620
- Guo Z, Tucker D, Basten C, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G (2014) The impact of population structure on genomic prediction in stratified populations. **Theor Appl Genet** 127:749-762
- Hao D, Cheng Y, Chen G, Lu H, Shi M, Zhang Z, Huang X, Mao Y, Xue L (2014) Identification of significant single nucleotide polymorphisms for resistance to maize rough dwarf disease in elite maize (*Zea mays* L.) inbred lines. **Euphytica**:1-12
- Kumar B, Abdel-Ghani AH, Pace J, Reyes-Matamoros J, Hochholdinger F, Lübberstedt T (2014) Association analysis of single nucleotide polymorphisms in candidate genes with root traits in maize (*Zea mays* L.) seedlings. **Plant Science** 224:9-19
- Li M, Liu X, Bradbury P, Yu J, Zhang Y-M, Todhunter R, Buckler E, Zhang Z (2014) Enrichment of statistical power for genome-wide association studies. **BMC Biology** 12:1-10
- Liu C, Weng J, Zhang D, Zhang X, Yang X, Shi L, Meng Q, Yuan J, Guo X, Hao Z, Xie C, Li M, Ci X, Bai L, Li X, Zhang S (2014) Genome-wide association study of resistance to rough dwarf disease in maize. **Eur J Plant Pathol** 139:205-216
- Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, Duarte CW, Allison DB, de los Campos G (2011) Beyond Missing Heritability: Prediction of Complex Traits. **PLoS Genet** 7:e1002051
- Pace J, Lee N, Naik HS, Ganapathysubramanian B, Lübberstedt T (2014) Analysis of Maize (*Zea mays* L.) Seedling Roots with the High-Throughput Image Analysis Tool (Automatic Root Image Analysis). **PLoS ONE** 9:e108255

- Pea G, Aung HH, Frascaroli E, Landi P, Pe ME (2013) Extensive genomic characterization of a set of near-isogenic lines for heterotic QTL in maize (*Zea mays* L.). **BMC Genomics** 14:61
- Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradbury PJ, Buckler ES (2014) The Genetic Architecture Of Maize Height. **Genetics** 196:1337-1356
- Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Ågren J, Bossdorf O, Byers D, Donohue K, Dunning M, Holub EB, Hudson A, Le Corre V, Loudet O, Roux F, Warthmann N, Weigel D, Rivero L, Scholl R, Nordborg M, Bergelson J, Borevitz JO (2010) The Scale of Population Structure in *Arabidopsis thaliana*. **PLoS Genet** 6:e1000843
- Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. **Nature reviews Genetics** 11:459-463
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. **Genetics** 155:945-959
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association Mapping in Structured Populations. **American Journal of Human Genetics** 67:170-181
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014. URL <http://www.R-project.org/>.
- Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. **Nat Genet** 44:217-220
- Rincent R, Nicolas S, Bouchet S, Altmann T, Brunel D, Revilla P, Malvar RA, Moreno-Gonzalez J, Campo L, Melchinger AE, Schipprack W, Bauer E, Schoen CC, Meyer N, Ouzunova M, Dubreuil P, Giauffret C, Madur D, Combes V, Dumas F, Bauland C, Jamin P, Laborde J, Flament P, Moreau L, Charcosset A (2014) Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production. **Theor Appl Genet** 127:2313-2331
- Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences* 103:18656-18661

- SAS. 2007. SAS 9.1.3 (TS1M3) for Windows Microsoft. SAS Institute Inc., Cary, USA, sem paginação
- Sobkowiak A, Jończyk M, Jarochovska E, Biecek P, Trzcinska-Danielewicz J, Leipner J, Fronk J, Sowiński P (2014) Genome-wide transcriptomic analysis of response to low temperature reveals candidate genes determining divergent cold-sensitivity of maize inbred lines. **Plant Mol Biol** 85:317-331
- Thiemann A, Fu J, Seifert F, Grant-Downton R, Schrag T, Pospisil H, Frisch M, Melchinger A, Scholten S (2014) Genome-wide meta-analysis of maize heterosis reveals the potential role of additive gene expression at pericentromeric loci. **BMC Plant Biology** 14:88
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. **Nat Genet** 43:159-162
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. **Proceedings of the National Academy of Sciences** 108:1088-1092
- Visscher Peter M, Brown Matthew A, McCarthy Mark I, Yang J (2012) Five Years of GWAS Discovery. **The American Journal of Human Genetics** 90:7-24
- Voight BF, Pritchard JK (2005) Confounding from Cryptic Relatedness in Case-Control Association Studies. **PLoS Genet** 1:e32
- Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, Branham KE, Heckenlively J, The FS, Fulton R, Wilson RK, Mardis ER, Lin X, Swaroop A, Zollner S, Abecasis GR (2014) Ancestry estimation and control of population stratification for sequence-based association studies. **Nat Genet** 46:409-415
- Westengen OT, Berg PR, Kent MP, Brysting AK (2012) Spatial Structure and Climatic Adaptation in African Maize Revealed by Surveying SNP Diversity in Relation to Global Breeding and Landrace Panels. **PLoS ONE** 7:e47832
- Wongkaew A, Phumichai C, Chunwongse J, Jampatong S, Grudloyma P, Pulam T, Dounghan W (2014) Detection of candidate R genes and single nucleotide polymorphisms for downy mildew resistance in maize inbred lines by association analysis. **Euphytica** 197:109-118

- Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM (2013) Pitfalls of predicting complex traits from SNPs. **Nat Rev Genet** 14:507-515
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. **Nat Genet** 42:565-569
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nat Genet** 38:203-208
- Zhang X, Perez-Rodriguez P, Semagn K, Beyene Y, Babu R, Lopez-Cruz MA, San Vicente F, Olsen M, Buckler E, Jannink JL, Prasanna BM, Crossa J (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. **Heredity** 114:291-299
- Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. **Nat Genet** 42:355-360