

PEDRO PATRIC PINHO MORAIS

**EFICIÊNCIA NO USO DE NITROGÊNIO AO LONGO DOS ESTÁDIOS  
FENOLÓGICOS E UTILIZAÇÃO DE DADOS PÚBLICOS NA SELEÇÃO GENÔMICA  
DE MILHO TROPICAL**

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

VIÇOSA  
MINAS GERAIS – BRASIL  
2016

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

T

O48e  
2016  
Morais, Pedro Patric Pinho, 1985-  
Eficiência no uso do nitrogênio ao longo dos estádios  
fenológicos e utilização de bancos de dados públicos na seleção  
genômica de linhagens de milho tropical / Pedro Patric Pinho  
Morais. – Viçosa, MG, 2016.  
x, 75f. : il. (algumas color.) ; 29 cm.

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

1. Milho. 2. *Zea mays*. 3. Variabilidade (Genética).  
4. Plantas - Efeito do nitrogênio. I. Universidade Federal de  
Viçosa. Departamento de Fitotecnia. Programa de Pós-graduação  
em Genética e Melhoramento. II. Título.

CDD 22 ed. 633.15

PEDRO PATRIC PINHO MORAIS

**EFICIÊNCIA NO USO DE NITROGÊNIO AO LONGO DOS ESTÁDIOS  
FENOLÓGICOS E UTILIZAÇÃO DE DADOS PÚBLICOS NA SELEÇÃO GENÔMICA  
DE MILHO TROPICAL**

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

APROVADA: 24 de Junho de 2016.

---

Caillet Dornelles Marinho

---

Elizabeth Ann Veasey

---

Fernando Angelo Piotto

---

Roberto Fritsche Neto  
(Coorientador)

---

Aluizio Borém de Oliveira  
(Orientador)

*“A ciência se compõe de erros  
que, por sua vez, são os passos  
até a verdade.”*

*-Júlio Gabriel Verne-*

*Deus e minha família,  
Dedico.*

## **AGRADECIMENTOS**

Agradeço as três mulheres referências na minha vida. A minha mãe Izaíra Pinho e minha avó Prudencia C. Pinho por toda garra e dedicação prestada durante a minha formação como pessoa e cidadão; a minha esposa Carla Meimberg pelo carinho e suporte durante todo o período de doutorado.

À minha família em nome das minhas irmãs Luzara, Bruna e Camila, assim como tios, tias e primos que de alguma forma ou outra me deram apoio.

À Universidade Federal de Viçosa (UFV), em específico ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade da efetivação do doutorado. Ao CNPq e CAPES pelas concessões de bolsas regulares e de doutorado sanduíche (processo BEX 2429/15-2).

À FAPESP (processo 2013/24135-2), Pioneer Hi-Bred International Int. (Johnston, IA - USA), Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP) e ao Laboratório de Alógamas da ESALQ/USP pelo total e irrestrito auxílio financeiro, suporte de pesquisa, condução experimental e conclusão dos trabalhos.

À Cornell University pela disponibilidade do doutorado sanduíche, crescimento pessoal e acadêmico no período vivenciado.

Aos grandes mestres que me orientaram, Professor Aluizio Borém de Oliveira e Roberto Fritsche Neto, essenciais na elaboração desse estudo. Grato também ao Professor Dr. Jean-Luc Jannink (Cornell University) pela oportunidade de trabalho em sua equipe e condução dos primeiros resultados aqui exibidos.

Aos amigos da UFV e do Laboratório de Alógamas da ESALQ/USP: Luciano Andrade, Giovanni Galli, Ítalo Granato, Filipe Couto, Filipe Matias, Leandro Mendonça, Danilo Hottis, Evellyn Couto, Anna Rita, Massaine Bandeira e Miriam Vidotti. Não podendo deixar de mencionar o agradecimento a todos os estagiários e funcionários desses locais, especialmente ao funcionário Miguel Ribeiro de Camargo.

De forma geral, agradeço a todos aqueles que de algum modo contribuíram nesse processo e estudo científico, deixo aqui minha gratidão registrada e meu muito obrigado.

## RESUMO

MORAIS, Pedro Patric Pinho, D.Sc., Universidade Federal de Viçosa, junho de 2016. **Eficiência no uso de nitrogênio ao longo dos estádios fenológicos e utilização de dados públicos na seleção genômica de milho tropical**. Orientador: Aluizio Borém de Oliveira. Coorientadores: Roberto Fritsche Neto e Marcos Deon Vilela de Resende.

Este estudo foi feito com uma população de linhagens de milho tropical, visando inferir sobre a eficiência no uso de nitrogênio (EUN), assim como a aplicação de seleção genômica (GS) nesse conjunto. Para a EUN foram avaliadas 64 linhagens, em duas disponibilidades de adubação nitrogenada (N) (ideal e baixo N), para 16 características. Estas avaliações ocorreram em três estádios fenológicos de oito folhas completamente expandidas (V8), florescimento masculino (VT) e após a maturação fisiológica (MF). Nas análises dos dados foi usado equações de modelos mistos via REML/G-BLUP. Foram observados que há possibilidade de seleção de linhagens superiores quanto a EUN. Mas, os componentes da produtividade não apresentam comportamento diferenciado em função da disponibilidade de N. Características de raiz, reprodutivos e fisiológicas não são atrativas para seleção precoce de genótipos EUN. De forma semelhante, os estádios V8 e VT não se mostraram promissores para avaliação precoce da EUN, devendo essa ser praticada em função da produtividade de grãos. Por outro lado, constatou-se que a seleção indireta para EUN pode ser feita via número de espigas (NE), ou ainda, por meio de índice de seleção considerando NE e peso de espigas (PE). No estudo de GS, visando a aplicação dessa em programas de melhoramento de pequeno porte e, ou então, iniciando suas atividades, foi proposto um panorama hipotético: orçamento limitado e necessidade de acréscimo de variabilidade genética. Para isso, foram definidos 29 conjuntos de população de treinamento (PT) através de painéis de diversidade: USP (painel a ser predito), ASSO e NCRPIS (painéis públicos externos - preditores). Estes conjuntos foram agrupados divididos em quatro cenários distintos, quanto a configuração usada na formação. Nas análises foram usados 28.260 marcadores SNP para 2748 linhagens, via modelo G-BLUP, para características de altura de planta e de espiga. Os resultados indicaram que é possível o uso de informações de bancos públicos para

formação de PT. Além disso, a estrutura populacional influenciou as capacidades preditivas da GS, principalmente quando populações de treinamento e validação (PV) são divergentes. Nos quatro cenários propostos, evidenciou-se que conjuntos demasiadamente pequenos ou grandes não proporcionam as melhores capacidades preditiva para GS. Entretanto, PT compostas por 250 indivíduos, escolhidas de forma otimizada (PTO) a partir de informações de bancos públicos, conduzem a capacidades preditivas satisfatórias (0,32). Contudo, a adição da PV na PTO, é o cenário que conduz às maiores capacidades preditivas (0,59). Sendo seu uso, sempre que possível, o recomendado.

## ABSTRACT

MORAIS, Pedro Patric Pinho, D.Sc., Universidade Federal de Viçosa, June, 2016. **Nitrogen Use Efficiency along growing stages and public dataset in genomic selection for tropical maize.** Adviser: Aluizio Borém de Oliveira. Co-Advisers: Roberto Fritsche Neto and Marcos Deon Vilela de Resende.

This study was proceeded with a population of tropical maize inbred lines, aiming to infer about the nitrogen use efficiency (NUE) as well as the application of genomic selection (GS) in this set. For EUN, 64 inbred lines were evaluated in two nitrogen (N) fertilizing conditions (optimum and low N), taking into account 16 traits. These evaluations were performed in three phenological stages eight fully expanded leaves (V8), tasseling stage (VT) and after physiological maturity (MF). Data analysis was performed using mixed models via REML/G-BLUP. It was observed that there is a possibility of selecting inbred lines for NUE, and the yield components do not show differential behavior according with the N condition. Root, reproductive and physiological traits are not attractive for early selection in efficient nitrogen use genotypes. Similarly, V8 and VT stages do not appear promising for early NUE estimation, it has to be done taking into account the yield. Indirect selection for NUE can be done via number of ear (NE), or with selection index considering NE and ear weight (PE). For the GS study, aiming to use GS in small breeding programs and, or, starting its activities, a hypothetical scenario was proposed: a limited budget and the need for genetic variability. For that, 29 training sets (PT) were developed through diversity panels: USP (panel to be predicted), ASSO and NCRPIS (external public panels - predictors); they were clustered into four scenarios with different PT. In the analysis 28.260 SNP markers for 2748 inbred lines were used, via model G-BLUP for plant height and ear height. Through the study it was found that it is possible to use information from public banks for PT formation. The population structure has the capacity of affecting GS predictive abilities, especially when training set and validation (PV) hold low genetic relationship. In the four proposed scenarios, it was shown that too small or large sets do not provide best predictive abilities. But, training sets with 250 individuals, via use of optimized training set (PTO) from public dataset banks, is enough for getting estimates up to 0.32. However, adding PV into PTO

is the best scenario, it got predictive ability estimates up to 0.59. It should be used whenever possible.

## ÍNDICE

1. INTRODUÇÃO GERAL.....	3
1.1 Eficiência no Uso de Nitrogênio - EUN.....	3
1.2 Seleção Genômica (GS) e a utilização de bancos públicos de informação genômica e fenotípica.....	3
1.3 Referências.....	6
2. OBJETIVO GERAL.....	10
2.1 Objetivos Específicos.....	10
Capítulo 1:.....	10
3. CAPÍTULO 1.....	11
3.1 Resumo.....	12
3.2 Abstract.....	13
3.3 Introdução.....	14
3.4 Objetivos.....	15
3.5 Material e métodos.....	15
3.5.1 Experimentos a campo.....	15
3.5.2 Características avaliadas.....	16
3.5.3 Análises estatístico-genéticas.....	18
3.6 Resultados e discussão.....	21
3.6.1 Análises conjunta, individual e parâmetros genéticos.....	21
3.6.2 Relação fenotípica entre características.....	30
3.7 Conclusões.....	35
3.8 Referências.....	36
4. CAPÍTULO 2.....	42
4.1 Resumo.....	43
4.2 Abstract.....	44
4.3 Introdução.....	45
4.4 Objetivos.....	47
4.5 Material e métodos.....	47
4.5.1 Conjunto genotípico e experimentos a campo.....	47
4.5.2 Características avaliadas.....	48
4.5.3 Valores genéticos preditos (BLUP).....	48
4.5.4 Genotipagem e controle de qualidade.....	49
4.5.5 Parentesco entre linhagens e estrutura populacional.....	50
4.5.6 População de treinamento (PT) e de validação (PV).....	51
4.5.7 Predição genômica.....	53
4.6.1 Componentes de variância e herdabilidades.....	54
4.6.2 Estrutura populacional.....	55
4.6.3.1 CPT1.....	56
4.6.3.2 CPT2.....	56
4.6.3.3 CPT3.....	59
4.6.3.4 CPT4.....	62

4.6.4 Aplicação no melhoramento de plantas e perspectivas.....	64
4.7 Conclusões .....	66
4.8 Referências .....	67
5. TABELAS E FIGURAS SUPLEMENTARES .....	72
6. CONCLUSÃO GERAL .....	75

## **1. INTRODUÇÃO GERAL**

### **1.1 Eficiência no Uso de Nitrogênio - EUN**

Para atender à crescente demanda mundial e brasileira por alimentos e nutrientes, uma fração significativa dos cultivos agrícolas terão que ser realizados em solos de áreas marginais (Giaveno *et al.*, 2007). Os quais, geralmente, possuem baixa fertilidade natural, associados a pouca aplicação de fertilizantes nitrogenados e à ocorrência de estresses hídricos. Destes insumos, o nitrogênio (N) possui grande importância no metabolismo da planta, principalmente na síntese de proteínas, como a clorofila, ácidos nucleicos e biomoléculas (ATP, NADH e NADPH). Assim, o N é fundamental no estabelecimento e manutenção da produtividade e qualidade dos grãos (Richardson *et al.*, 2009).

Dessa forma, com o intuito de alcançar elevadas produtividades há necessidade de aumento do nível tecnológico empregado, por meio do uso de insumos, como corretivos e fertilizantes, o que implica na elevação dos custos de produção e, conseqüentemente, na dificuldade de adoção por parte dos pequenos e médios produtores (Caixeta, 2012). Todavia, ao se realizarem adubações excessivas (além da capacidade de absorção da planta), principalmente de N, resultam na formação de gases nocivos à atmosfera, como exemplo os óxidos nitroso e nítrico, e também na ocorrência de chuvas ácidas. Entretanto, o principal problema vem decorrente da lixiviação do  $\text{NO}_3^-$ , que devido a sua grande mobilidade, tende a contaminar os lençóis freáticos (Ju *et al.*, 2009), podendo causar problemas de saúde pública pela ingestão da água oriunda de fontes com excesso de  $\text{NO}_3^-$  (Andrade *et al.*, 2009).

Concomitantemente a esses aspectos abordados, nos últimos anos, o interesse pela agricultura sustentável, com menor consumo energético e

ecologicamente correta, tem estimulado os programas de melhoramento de plantas a desenvolver cultivares que possuam maior eficiência no uso de nutrientes (Maia *et al.*, 2011).

Outro fator de relevância está condicionado ao milho de segunda safra, pois tendo em vista que a soja é a principal cultura de verão, o produtor tem preferido semear o milho em sucessão. Esse sistema de produção tem predominado nos Estados do Paraná, São Paulo, Minas Gerais, Goiás, Mato Grosso e Mato Grosso do Sul (Bicudo *et al.*, 2009). Sendo que o crescimento da segunda safra tem se intensificado, saindo de uma área de 2,88 milhões de hectares com produção de 6,18 milhões de toneladas na safra 2001/2002 para 9,55 milhões de hectares e produção de 54,59 milhões de toneladas na safra 2014/15 (Conab, 2016). Esses dados indicam um crescimento de 331% na área plantada e de 883% na produção. Entretanto, é verificado que para as condições de safrinha, a adubação usada é menor que no cultivo de verão devido ao efeito residual de adubos utilizados na cultura anterior. Outra questão da baixa adubação é em função do risco de ocorrência de baixas precipitações, que fazem com que as plantas não aproveitem adequadamente o adubo (Bicudo *et al.*, 2009). Adicionado a isso, existe a oscilação dos preços de venda do grão entre as safras, no qual muitas vezes pode não ser atrativo, reduzindo o interesse do produtor em investir em adubação. Assim, o desenvolvimento de cultivares com maior Eficiência no Uso de Nitrogênio (EUN) é de interesse tanto para a maximização da produção em áreas férteis quanto para a exploração de novas áreas com menores teores de N. Adicionalmente, isso possui a capacidade de auxílio em alcançar a segurança alimentar das populações rurais de baixa renda e, ou, de baixa tecnologia de cultivo.

Neste contexto, alguns estudos têm constatado ampla diversidade genética entre plantas quanto a mecanismos fisiológicos, morfológicos e bioquímicos, quando elas são submetidas às condições adversas de fertilidade do solo (DoVale *et al.*, 2012, Gallais e Hirel, 2004; Le Gouis *et al.*, 2000). Nesse sentido, o uso de métodos de avaliação precoce (antes do florescimento) ou de seleção indireta em condições de estresses abióticos, são de grande interesse, pois aceleram o processo de seleção, descartando de imediato os genótipos inferiores e possibilitam concentrar os recursos nos potencialmente superiores (Fritsche-Neto *et al.*, 2012). Para isso, a utilização dos componentes da EUN e,

ou, de características secundárias relacionadas a essa, podem auxiliar na seleção e no desenvolvimento de cultivares de milho com maior adaptação a condições de baixa disponibilidade de N.

## **1.2 Seleção Genômica (GS) e a utilização de bancos públicos de informação genômica e fenotípica**

Para aperfeiçoar a relação custo-benefício, otimizar o tempo e realizar uma seleção acurada, deve-se lançar mão de metodologias precisas, que amostram o genoma e o correlacionem com as características. Para isto, podem-se usar os marcadores moleculares, cuja disponibilidade e baixo custo possibilitaram o amplo emprego dos mesmos nos programas de melhoramento (Bernardo e Yu, 2007), excluindo a necessidade de avaliações fenotípicas contínuas em linhagens previamente fenotipadas. Dentre esses, destacam-se os marcadores genéticos moleculares de Polimorfismo de um Único Nucleotídeo (Single-Nucleotide Polymorphism - SNP). Os SNP se baseiam na detecção de polimorfismos resultantes da alteração de uma única base no genoma. De forma que, são preferidos em relação a outros marcadores genéticos devido à sua ampla distribuição no genoma e facilidade de genotipagem (Resende *et al.*, 2013), principalmente para o milho, ao qual possui plataforma comercial específica para esse fim.

Os esforços na área de pesquisa com marcadores genéticos em *prol* do melhoramento podem ser divididos em duas linhas (Resende *et al.*, 2012): *i*) Detecção e mapeamento de QTL dentro de famílias oriundas de cruzamentos biparentais, seguido do uso dos marcadores nos programas de melhoramento para seleção auxiliada por marcadores (MAS); *ii*) Genotipagem com alta densidade de marcadores no genoma e conseguinte predição de valores genéticos genômicos (GEBV) por meio da Seleção Genômica, ou então para o estudo de Associação Genômica (Genome Wide Association Studies - GWAS). Esse último, permite fazer correlação entre o genótipo e o fenótipo, com o uso de teste de hipóteses individuais para cada marcador em uma determinada característica de interesse. Já o MAS tem se mostrado com maior eficácia para a uso com alelos de grande efeito e associação conhecida com determinado marcador (Hofmann *et al.*, 2013; Zhong *et al.*, 2006). No entanto, pelo seu próprio princípio, essa tem falhado quando aplicada as características poligênicas, ou seja, muitos alelos de pequeno efeito, alocados ao longo de todo o genoma (Bernardo, 2008; Moreau

*et al.*, 2004). Por outro lado, a GS explica proporção muito maior da variação genética, pois concentra um grande número de marcadores, os quais cobrem todo o genoma, de forma a capturar os efeitos dos locos de um caráter quantitativo (QTL, de pequenos e grandes efeitos), usando todos na análise de predição dos GEBVs de determinado genótipo (Meuwissen *et al.*, 2001). Com isso, a superioridade da GS em relação ao MAS alcança valores de 18-43% (Bernardo e Yu, 2007), além de não necessitar realizar testes de hipóteses individuais para marcadores, o que é comum para a GWAS.

Para a implementação da GS, três populações são necessárias: *i*) População de treinamento ou estimação (PT): indivíduos são genotipados e fenotipados, equações usando esses dados são definidas para prever os GEBVs; *ii*) população de validação (PV): as equações de predição dos GEBVs são testadas em uma amostra populacional independente menor, que também são genotipadas e fenotipadas. Nesse ponto, as capacidades preditivas e acurácias do modelo podem ser estimadas; *iii*) população de seleção (PS): necessita apenas de um conjunto de indivíduos genotipados, as equações de predição obtidas na população de treinamento e posteriormente validadas são então usadas na predição dos GEBVs de possíveis candidatos. Dessas, uma determinada população pode exercer duas funções ao mesmo tempo (PT e PV); e ou podem exercer funções separadas, uma população específica para PT, outra para PV.

O ganho esperado via GS se dá em função de  $\Delta_G = \frac{ir\sigma_a}{T}$ , em que  $i$ = intensidade de seleção;  $r$ = acurácia seletiva;  $\sigma_a$  = raiz quadrada da variância genética aditiva;  $T$ = tempo necessário para completar um ciclo de melhoramento. Considerando estimativas realísticas de acurácia seletiva, ciclo de melhoramento e intensidade seletiva, a GS pode aumentar o ganho genético por unidade de tempo comparado ao seleção fenotípica, tanto em melhoramento de plantas como em animal (Heffner *et al.*, 2010, Lorenzana *et al.*, 2009; Schaeffer *et al.*, 2006). Isso só foi possível devido aos rápidos avanços em tecnologia de marcadores genéticos e preços decrescentes de genotipagem, permitindo ao melhoristas de plantas e animais uma abundante quantidade de informações genotípicas (Heffner *et al.*, 2009; Jannink *et al.*, 2010).

Um fator importante no processo de GS é em função da composição da PT. A escolha de indivíduos a serem incluídos na PT, é uma das variáveis mais difíceis para otimizar-se, mas é crucial para atingir elevadas acurácias (Isidro *et al.*, 2015). O seu tamanho e relacionamento com a PS são elementos essenciais para determinar a precisão da predição via GS (Bassi *et al.*, 2016). Todavia, o grande número de indivíduos a serem genotipados e fenotipados, para compor uma PT confiável e representativa da PS, demanda tempo e custo na sua formação, especialmente devido a fenotipagem. Que tem se tornado o grande gargalo e fator de restrição das análises genéticas, sendo ainda um processo manual, de trabalho intensivo e dispendioso (Cobb *et al.*, 2013).

Entretanto, quando há restrições de tamanho e variabilidade genética na população de treinamento, ocorre redução na capacidade de predição dos modelos usados (Estaghvirou *et al.*, 2015). Com isso, é necessário incluir maior ou adequado número de indivíduos na PT, de forma a viabilizar uso da GS em relação a seleção fenotípica. Nesse contexto, quando não há possibilidade de conseguir esse aumento de variabilidade via o próprio programa de melhoramento, o uso de informações genômicas e fenotípicas provindas de bancos de dados públicos é uma solução a ser explorada. Isto poderá aumentar a viabilidade do emprego da GS em programas de melhoramento menores, com menor aporte financeiro e, ou em início de atividades, e com restrição de variabilidade genética para características de interesse.

Instituições como o International Maize and Wheat Improvement Center (CIMMYT) e o USDA-ARS North Central Regional Plant Introduction Station (NCRPIS), possuem juntos mais de 28.000 genótipos nas suas coleções. Tal qual, muitos pesquisadores têm caracterizado genotipicamente e fenotipicamente subconjuntos dessas linhagens de milho com fins curatoriais, de avaliação de diversidade e também para uso em mapeamento associativo (Flint-Garcia *et al.*, 2005; Yang *et al.*, 2010; Mezmouk *et al.*, 2011). Com isso, muitos desses dados são disponíveis e publicados em bancos de informações públicos (ex: Panzea e Gramene), sendo uma excelente fonte de dados genômicos e fenotípicos do germoplasma global de milho.

No entanto, o melhorista deve ter cautela quanto ao uso de conjuntos de dados de populações externas para a predição de populações internas de um determinado programa de melhoramento. Isto porque, ambas populações

possuem diferentes origens, de forma que a relação genética entre essas é um fator importante que tende a gerar distúrbio na precisão da GS (Habier *et al.*, 2007; Hayes *et al.*, 2009; Albrecht *et al.*, 2011; Clark *et al.*, 2011; Pszczola *et al.*, 2012), Deste modo, deve ser considerado a existência da estruturação populacional nas análises de GS e, ou lançar o uso de métodos que possam otimizar a formação da PT por meio de bancos de dados públicos, afim de prever com eficiência a população de interesse.

### 1.3 Referências

- Albrecht, T.; Wimmer, V.; Auinger, H.J.; Erbe, M.; Knaak, C.; Ouzunova, M.; Simianer, H.; Schön, C.C. 2011. Genome-based prediction of testcross values in maize. **Theoretical and Applied Genetics**, v.123, p.339-350
- Andrade, E. M.; Aquino, D. N.; Crisóstomo, L. A.; Rodrigues J. O.; Lopes, F. B. 2009. Impacto da lixiviação de nitrato e cloreto no lençol freático sob condições de cultivo irrigado. **Ciência Rural**, v. 39, v.01, p.88-95
- Bassi, F.M.; Bentley, A.R.; Charmet, G.; Ortiz, R.; Crossa, J. 2016. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). **Plant Science**, v.42, p.23-36. doi:10.1016/j.plantsci.2015.08.021
- Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, v.48, p.1649-1664
- Bernardo R.; Yu, J. 2007. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, n.47: p.1082-1090
- Bicudo, S.J.; Cruz, S.C.S.; Pereira, F.R.S.; Brachtvogel, E.L.; Madalena, J.A.S. 2009. Níveis de adubação para a cultura do milho safrinha. **Caatinga**, v.22, p.23-30
- Caixeta, D. S. 2012. **Seleção precoce e indireta para eficiência no uso de nitrogênio em milho**. Dissertação de mestrado. Universidade Federal de Viçosa
- Clark, S.; Hickey, J. van der Werf, J. 2011. Different models of genetic variation and their effect on genomic evaluation. **Genetic Selection Evolution**, p.43-18
- Cobb J.N.; DeClerck, G.; Greenberg, A.; Clark, R.; McCouch, S. 2013. Next-generation phenotyping: requirements and strategies for enhancing our

- understanding of genotype–phenotype relationships and its relevance to crop improvement. **Theoretical and Applied Genetics**, v.126, p.1-21.
- Companhia Nacional de Abastecimento. 2016. **Acompanhamento de safra brasileira: grãos, oitavo levantamento, maio 2016** . Companhia Nacional de Abastecimento. - Brasília: Conab, 29p
- DoVale, J. C.; Fritsche-Neto, R.; Bermudez, F. P.; Miranda, G. V. 2012. Efeitos gênicos de caracteres associados à eficiência no uso de nitrogênio em milho. **Pesquisa Agropecuária Brasileira**, v.47, n.3, p.385-392
- Duarte, A.P. 2004. Milho safrinha: Características e sistemas de produção. In: Galvão, J.C.C.; Miranda, G.V. (Eds.). **Tecnologias de produção de milho**. Viçosa: Editora UFV, p.109-138
- Estaghvirou, S.B.O.; Ogutu, J. O.; Piepho, H. 2015. How Genetic Variance and Number of Genotypes and Markers Influence Estimates of Genomic Prediction Accuracy in Plant Breeding. **Crop Science**. V.55, p.1911-1924. doi:10.2135/cropsci2014.09.0620
- Flint-Garcia, S. A., Thuillet, A.-C. ; Yu, J. ; Pressoir, G. ; Romero, S. M.; Mitchell, S.E. et al. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. **The Plant Journal**, v.44, p.1054-1064
- Fritsche-Neto, R.; Resende, M. D. V.; Miranda, G. V; Do Vale, J. C. 2012. Seleção genômica ampla e novos métodos de melhoramento do milho. **Ceres**, v.59, p.794-802
- Gallais, A.; Hirel, B. 2004. An approach to the genetics of nitrogen use efficiency in maize. **Journal of Experimental Botany**, v.396, p.295-306
- Giaveno, C. D.; Ribeiro, R. V.; Souza, G. M.; Oliveira, R. F. 2007. Screening of tropical maize for salt stress tolerance. **Crop Breed Applied Biotechnol.** V.7, p.304-313
- Hofmann, K.; Silvar, C.; Casas, A.M.; Herz, M.; Buttner, B. et al. 2013. Fine mapping of the Rrs1 resistance locus against scald in two large populations derived from Spanish barley landraces. **Theoretical and Applied Genetics**, v.126, p.3091-3102
- Habier, D.; Fernando, R.L.; Dekkers, J.C.M. 2007. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v.77, p.2389-2397

- Heffner, E.L.; Sorrells, M.E.; Jannink, J-L. 2009. Genomic selection for crop improvement. **Crop Science**, v.49, p.1-12
- Heffner, E.L.; Lorenz, A.J.; Jannink, J-L.; Sorrells, M.E. 2010. Plant breeding with genomic selection: gain per unit time and cost. **Crop Science**, v.50, p.1681-1690
- Hayes, B.J.; Visscher, P.M.; Goddard, M.E. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research (Camb)**, v.91, p.47-60
- Jannink, J-L.; Lorenz, A.J.; Iwata, H. 2010. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, 9(2), p.166-177
- Ju, X. T.; Xing, G. X.; Chen, X. P.; Zhang, S. L.; Zhang, L. J. et al. 2009. Reducing environmental risk by improving N management in intensive Chinese agricultural systems. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, p.3041-3046
- Le Gouis, J.; Béghin, D.; Heumez, E.; Pluchard, P. 2000. Genetic differences for nitrogen uptake and nitrogen utilization efficiencies in winter wheat. **European Journal of Agronomy**, v.12, p.163-173
- Lorenzana, R.E., Bernardo, R. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics**, v.120, p.151-161
- Maia, C.; DoVale, J. C.; Fritsche-Neto, R. F.; Cavatte, P. C.; Miranda, G. V. 2011. The difference between breeding for nutrient use efficiency and for nutrient stress tolerance. **Crop Breeding and Applied Biotechnology**, v.11, n. 3, p.270-275
- Mezmouk, S.; Dubreuil, P.; Bosio, M.; Décousset, L.; Charcosset, A.; Praud, S.; Mangin, B. 2011. Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. **Theoretical and Applied Genetics**, v.122, p.1149-1160
- Moreau, L.; Charcosset, A.; Gallais, A. 2004. Experimental evaluation of several cycles of marker-assisted selection in maize. **Euphytica** v.137, p.111-118
- Pszczola, M.; Strabel, T.; Mulder, H.A.; Calus, P.L. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. **Journal of Dairy Science**, v.95, p.389-400

- Resende, M. D.; Silva, F. F.; Resende Jr. 2013. Seleção Genômica Ampla (GWS). In: **Biotecnologia aplicada ao melhoramento de plantas /** Borém, A. V. Fritsche-Neto, R.[editores]. – Visconde de Rio Branco, Suprema, 2013, 336p
- Resende, M.D.V.; Silva, F.F.; Lopes, P.S.; Azevedo, C.F. 2012. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada e Estatística Espacial.** 1. Ed, 291p
- Richardson, A.E.; Barea, J.M.; McNeill, A.M.; Prigent-Combaret, C. 2009. Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. **Plant Soil**, v321, p.305-339. doi:10.1007/s11104-009-9895-2
- Romay; M.; Millard, M.; Glaubitz, J.; Peiffer, J.; Swarts, K. et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biology**, 14:R55
- Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, v.123, p.218-223
- Yang, X.; Gao, S.; Xu, S.; Zhang, Z.; Prasanna, B.M. et al. 2010. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. **Molecular Breeding**, v.28, p.511-526
- Zhong, S.; Toubia-Rahme, H.; Steffenson, B.J.; et al. 2006. Molecular mapping and marker-assisted selection of genes for septoria speckled leaf blotch resistance in barley. **Phytopathology**, v.96, p.993-999

## **2. OBJETIVO GERAL**

- Inferir sobre a eficiência no uso de nitrogênio assim como a utilização de bancos de dados públicos na seleção genômica de linhagens de milho tropical.

### **2.1 Objetivos Específicos**

#### **Capítulo 1:**

- Analisar o comportamento e relações entre características dos componentes de produtividade, florescimento, sistema radicular e fisiológicas em linhagens de milho tropical submetidas a disponibilidades contrastantes de nitrogênio;
- Verificar a possibilidade de seleção precoce e indireta de linhagens de milho tropical para EUN;

#### **Capítulo 2:**

- Testar a viabilidade do uso das informações genômicas e fenotípicas de bancos de dados públicos para predição de linhagens de milho tropical;
- Identificar se a estrutura populacional, o uso de populações de treinamento otimizadas e a quantidade de informação oriundas de bancos de dados públicos influenciam na capacidade preditiva de linhagens de milho tropical.

### **3. CAPÍTULO 1**

**Eficiência no Uso de Nitrogênio em linhagens de milho tropical:  
uma análise para características de produção, reprodutivos, fisiológicas e  
radiculares**

### 3.1 Resumo

Parte significativa dos cultivos de milho ocorrem em condições edafoclimáticas que restringem o pleno desenvolvimento da cultura. No Brasil, isto se deve, principalmente, ao fato de que boa parte do cultivo migrou para a chamada “safrinha” e para áreas de Cerrado. Assim, torna-se importante o desenvolvimento de cultivares com maior eficiência no uso de nitrogênio (EUN) e adaptados a essas condições de cultivo. Para isto, é necessário identificar características que possuam alta relação com EUN, afim de auxiliar e maximizar a acurácia do processo de seleção. Com isto, este trabalho teve por objetivo: *i*) analisar o comportamento e relações entre características dos componentes de produtividade, florescimento, sistema radicular e fisiológicas em linhagens de milho tropical submetidas a disponibilidades contrastantes de nitrogênio; *ii*) verificar a possibilidade de seleção precoce e indireta de linhagens de milho tropical para EUN. Para isso, foram avaliadas 64 linhagens de milho tropical em duas disponibilidades de adubação nitrogenada (N) (ideal e baixo N), para 16 características, dentre essas a EUN. Estas avaliações ocorreram em três estádios fenológicos oito folhas completamente expandidas (V8), florescimento masculino (VT) e após a maturação fisiológica (MF). Na análise dos dados foram usados equações de modelos mistos via REML/G-BLUP. Foi observado que há possibilidade de seleção de linhagens superiores quanto a EUN, e que os componentes da produtividade não apresentam comportamento diferenciado em função da disponibilidade de N. Características de raiz, reprodutivos e fisiológicas não são atrativas para seleção precoce de genótipos EUN. De forma semelhante, os estádios V8 e VT não se mostraram promissores para a avaliação precoce da EUN, devendo essa ser praticada em função da produtividade de grãos. Por outro lado, ao final do ciclo, a seleção indireta para EUN pode ser feita via número de espigas (NE), ou ainda, com o uso de índice de seleção considerando NE e peso de espigas (PE).

**Palavras-chave:** estresse abiótico, seleção indireta, G-BLUP, seleção precoce.

### 3.2 Abstract

A significant portion of maize crops occurs in soil and climatic conditions that limit growing conditions. In Brazil, this is due mainly to the fact that most of the cultivation area has migrated to the second session “safrinha” and to Brazilian Cerrado. Thus, it becomes important to develop cultivars efficient in nitrogen (NUE) and adapted to these growing conditions. For this, it is necessary to identify traits that have relationship with NUE, in order to assist and maximize the accuracy of the selection process. This study aimed to: i) analyze the behavior and relationships between traits for yield components, flowering, root and physiological system in tropical inbred lines under contrasting nitrogen availability; ii) examine the possibility of indirect and early selection for NUE in tropical inbred lines. To that, 64 inbred lines were evaluated in two nitrogen (N) fertilizing conditions (optimum and low N), taking into account 16 traits, among these the NUE. These evaluations were performed in three phenological stages: eight fully expanded leaves (V8), tasseling stage (VT) and after physiological maturity (MF). Data analysis was performed using mixed models via REML/G-BLUP. Results showed that there is a possibility of selecting inbred lines for NUE, and the yield components do not show differential behavior according with the N condition. Root, yield and physiological traits are not attractive for early selection in efficient nitrogen use genotypes. Similarly, V8 and VT stages do not appear promising for early NUE estimation, it has to be done taking into account the yield. On the other hand, in the end of the crop cycle, the indirect selection for NUE can be done via number of ear (NE), or with selection index considering NE and ear weight (PE).

**Keywords:** abiotic stress, indirect selection, G-BLUP, early selection.

### 3.3 Introdução

O milho de segunda safra no Brasil, “safrinha”, teve seu crescimento intensificado na última década (Conab, 2015). Parte desses resultados deve-se à expansão da cultura no Cerrado. Todavia, os solos dessa região, principalmente os Latossolos, apresentam baixa disponibilidade de nitrogênio (N), em função dos seus baixos teores de matéria orgânica (Silva *et al.*, 2005). Além disso, o cultivo de segunda safra é caracterizado, de modo geral, pelo baixo investimento dos produtores em adubação (Bicudo *et al.*, 2009).

Neste cenário, com o intuito de alcançar elevadas produtividades há a necessidade do aumento do uso de insumos, como corretivos e fertilizantes, dentre esses o N; o que implica na elevação dos custos de produção e, conseqüentemente, na dificuldade de adoção por parte dos pequenos e médios produtores. No entanto, por meio do melhoramento genético é possível desenvolver cultivares de milho mais eficientes no uso de nitrogênio, ou seja, genótipos que produzem mais ou a mesma quantidade de grãos, porém, com menor demanda por N (Caixeta *et al.*, 2015).

Para o estudo da eficiência nutricional em plantas, existem várias definições e metodologias descritas. 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 N disponibilizado para a planta (N no solo mais o fornecido com a adubação). Porém, o desenvolvimento da planta ao longo do ciclo é uma questão oportuna a ser verificada quanto a EUN. Quando há a emissão de uma nova folha, parte dos carboidratos são direcionados para o desenvolvimento desse novo órgão (Vessey *et al.*, 1990). Paralelamente, há diminuição na taxa de translocação desses compostos para o sistema radicular, diminuindo a disponibilidade de energia nas raízes, o que reduz a taxa de absorção de nitrogênio. Com isso, cria-se o interesse de identificar o estágio ideal para inferências quanto a EUN, vislumbrando a possibilidade de estimá-la em estádios iniciais da cultura e permitindo a seleção precoce.

Além das características agrônômicas que fazem parte dos componentes da produtividade, pesquisas têm constatado ampla diversidade genética entre plantas quanto a mecanismos fisiológicos e morfológicos da raiz e de parte aérea, quando elas são submetidas às condições adversas de fertilidade do solo (DoVale *et al.*, 2012). Dentre essas, a eficiência fotossintética do fotossistema II

(PSII), devido ao fato de que por meio do PSII há possibilidade de se obter a eficiência no uso da radiação fotoquímica e a assimilação de carbono pelas plantas, auxiliando no diagnóstico da integridade do aparato fotossintético em função de estresses ambientais (Melo *et al.*, 2010; Tester e Bacic, 2005). Adicionalmente, a exploração de um maior desenvolvimento do sistema radicular possui potencial para ser usado na seleção indireta da EUN. Essa característica é altamente responsiva à disponibilidade de nutrientes e sua distribuição no solo (Linkohr *et al.*, 2002). Não menos importante, as características reprodutivas, tal como o Intervalo de florescimento masculino-feminino (IFMF), que é o intervalo entre a antese e a exposição dos estilos-estigmas, têm correlação negativa com produtividade em função de estresse hídrico (Teixeira *et al.*, 2010). Não obstante, a relação dessa característica com ambientes em restrição de N pode ter comportamento semelhante (Jacobs e Pearson, 1991), e influenciar negativamente os componentes da produtividade.

De modo geral, a escolha dessas características visa a viabilização do uso de métodos de avaliação precoce ou de seleção indireta para EUN, pois aceleram o processo de seleção, descartando de imediato os genótipos inferiores e possibilitam concentrar os recursos nos potencialmente superiores (Fritsche-Neto *et al.*, 2012).

### **3.4 Objetivos**

Diante do exposto, este trabalho teve por objetivos: *i*) analisar o comportamento e relações entre características dos componentes de produtividade, florescimento, sistema radicular e fisiológicas em linhagens de milho tropical submetidas a disponibilidades contrastantes de nitrogênio; *ii*) verificar a possibilidade de seleção precoce e indireta de linhagens de milho tropical para EUN.

### **3.5 Material e métodos**

#### **3.5.1 Experimentos a campo**

Foram conduzidos seis experimentos usando 64 linhagens endogâmicas de milho tropical (S7-S9), oriundas da Universidade Federal de Viçosa (UFV) e com comprovada variabilidade para EUN (DoVale *et al.*, 2012). O delineamento

experimental empregado foi o de látice simples (8 x 8), com duas repetições, seguindo o mesmo sorteio experimental nas safrinhas de 2014 e 2015. Duas disponibilidades de N (ideal e baixo N) foram aplicadas. Para “N ideal” a dose considerada padrão às necessidades da cultura para uma produtividade de até no máximo 5,8 t.ha<sup>-1</sup> (Embrapa, 2007). Com isto, para IN foram aplicados 100 kg.ha<sup>-1</sup> de N (12 kg.ha<sup>-1</sup> no plantio e restante em cobertura em V4 e V6 ); e em BN foram aplicados 35 kg.ha<sup>-1</sup> de N (12 kg.ha<sup>-1</sup> no plantio e restante em cobertura em V4 e V6). A adubação de plantio foi constituída de 300 kg ha<sup>-1</sup> do formulado NPK 4-14-8, já a adubação de cobertura foi feito com uréia na linha. O estabelecimento do BN está em função da dose de adubação em que há, em média, a redução de 50% da produtividade de grãos (Fritsche-Neto e Borém, 2012).

Os ambientes considerados nesse estudo foram determinados em função da combinação safra x local, formando assim três ambientes. Sendo esses a área experimental Departamento de Genética da ESALQ-USP, (22°42'23"S, 47°38'14"W, 546 m, Nitossolo Vermelho Eutroférico típico) na safrinha 2015, e a Estação Experimental de Anhumas na cidade de Anhembi-SP (22°84'30"S; 48°02'20"W, 450 m, Latossolo Amarelo Distrófico) nas safrinhas de 2014 e 2015.

As parcelas foram constituídas de uma linha com 4 m de comprimento (safrinha 2014) e de 5 m (safrinha 2015), essa diferença ocorreu devido a necessidade de maior número de plantas para avaliação de características radiculares; em ambos os casos utilizou-se espaçamento de 0,85 m entre linhas e 0,20 m entre plantas, totalizando 20 e 25 plantas por linha respectivamente. O manejo e os tratos culturais foram realizados de acordo com os procedimentos recomendados para a cultura do milho (Borém *et al.*, 2015).

### 3.5.2 Características avaliadas

Para as inferências quanto a EUN em diferentes fases do ciclo das linhagens, foram amostradas plantas nos estádios fenológicos de oito folhas completamente expandidas (V8) e florescimento masculino (VT). As amostras, composta da parte aérea de duas plantas de cada parcela experimental foram separadas do sistema radicular e acondicionadas em sacos de papel. Posteriormente, foi feita a secagem em estufa de circulação forçada de ar a 60° C por 120 horas, até atingir massa constante. Nesta ocasião, foi mensurado, em gramas, a massa seca da parte aérea (MPS). A última estimativa de EUN foi

feita após a maturação fisiológica (MF), quando mensurada a produção individual de cada linhagem. Para obter a EUN nas fases listadas, foi usado a metodologia de Moll *et al.* (1982):

$$EUNV8 = \frac{MPSV8}{N(\text{aplicado})} \quad (1) \quad EUNVT = \frac{MPSVT}{N(\text{aplicado})} \quad (2) \quad EUNP = \frac{PG}{N(\text{aplicado})} \quad (3)$$

em que, EUNV8: eficiência no uso de nitrogênio em estágio V8; EUNVT: eficiência no uso de nitrogênio em estágio VT; EUNP: eficiência no uso de nitrogênio na produção; MPSV8: massa da parte aérea seca em V8, mensurado em gramas (g); MPSVT: massa da parte aérea seca em VT, mensurado em gramas (g); PG: produtividade grãos, mensurado em kg.parcela<sup>-1</sup> de N (aplicado): quantidade de nitrogênio fornecido as plantas da parcela, definido em gramas (g) por meio de aplicações no plantio e cobertura.

As avaliações das características radiculares foram obtidas no estágio V8 pela amostragem de duas plantas por parcela. Foram avaliados os comprimentos de raízes laterais (CRLat) e axiais (CRAXi), mensurados em centímetros (cm); Volume de raízes laterais (VRLat) e axiais (VRAXi), mensurados em centímetros cúbicos (cm<sup>3</sup>). Para avaliação das características citadas foi utilizado o software WinRHIZO Arabidopsis acoplado a um scanner LA2400, com definição de 2400 dpi. As dez classes de diâmetro de raízes fornecidas pelo software foram agrupadas em apenas duas, conforme metodologia proposta por Trachsel *et al.* (2009). Assim, para CRLat e VRLat, foram considerados os fragmentos com diâmetro menor ou igual a 0,5 mm, e para CRAXi e VRAXi, os fragmentos com diâmetro maior que 0,5 mm. Essas características foram avaliadas somente na safrinha de 2015.

As características fisiológica e reprodutiva avaliadas foram a eficiência fotossintética do fotossistema II (EFPSII) e o intervalo florescimento masculino-feminino (IFMF). A EFPSII foi estimada por meio do fluorômetro portátil compacto para medição de fotossíntese JUNIOR-PAN. As informações coletadas para essa característica foram feitas entre às 9 e 12h da manhã, em uma folha por planta (terceira a partir do extrato superior) em duas plantas aleatórias dentro da parcela. Seguindo a metodologia descrita por DaMatta *et al.* (1997), antes da coleta de dados as folhas permaneceram no escuro durante 30 minutos. Para isso foram empregados grampos com áreas cilíndricas opacas que mantêm a área a ser mensurada sob completa ausência de luz. Após os centros de reações

estarem completamente abertos, foram determinados a fluorescência inicial (F0), fluorescência máxima (Fm), fluorescência variável (Fv),  $Fv = Fm - F0$ , e posteriormente EFPSII. Da mesma forma que as características radiculares, EFPSII foi inferido apenas na safrinha 2015.

$$EFPSII_{\phi} = \frac{Fv}{Fm} \quad (4)$$

O IFMF foi estimado em dias, em função das datas em que 50% das plantas da parcela estavam em estádios de antese (DA) e posteriormente em função da exposição dos estilos-estigmas (DF):

$$IFMF = DF_{50\%} - DA_{50\%} \quad (5)$$

As características relativas aos componentes de produção, número de espigas por parcela (NE), comprimento de espiga, (CE), em centímetros (cm); diâmetro de espigas (DE), em centímetros (cm); peso de espiga (PE), em gramas(g), número de fileiras por espiga (NFE) e número de grãos por fileira (NGF) foram inferidos pela amostragem de cinco plantas por parcela. Produtividade (PROD), em quilogramas por hectare ( $kg \cdot ha^{-1}$ ), foi obtida pela extrapolação dos resultados de produtividade por parcela (13% umidade).

### 3.5.3 Análises estatístico-genéticas

Para a obtenção dos componentes de variância e dos valores genéticos individuais, os dados foram submetidos às análises por meio da Máxima Verossimilhança Restrita (REML) e pelo Melhor Preditor Genético Linear Não-Viesado, considerando a matriz genômica de parentesco (GBLUP) (Hayes *et al.* 2009). Na análise conjunta foi considerado o seguinte modelo:

$$y = Jb + Zg + Wu + Vl + So + Te + Cf + Hp + \varepsilon \quad (6)$$

em que,  $y$  é o vetor de observações fenotípicas nas linhagens para as características avaliadas;  $b$  é o vetor do efeito do bloco dentro da repetição, considerado como fixo;  $g$ : vetor do efeito de genótipo (linhagens), considerado como aleatório, em que  $g \sim N(0, G)$ , sendo  $G = K\sigma_a^2$ , sendo  $K$  a matriz de parentesco genômico (Hayes *et al.*, 2009) e  $\sigma_a^2$  a variância genética;  $u$ : vetor do efeito da disponibilidade de N, considerado como fixo;  $l$ : vetor do efeito do ambiente, considerado como fixo;  $o$ : vetor do efeito da interação disponibilidade de N x ambiente, considerado como fixo;  $e$ : vetor do efeito da interação genótipo x disponibilidade de N, considerado como aleatório, em que  $e \sim N(0, E)$ , sendo

$E = I\sigma_e^2$ ;  $f$ : vetor do efeito da interação genótipo x ambiente, considerado como aleatório, onde  $f \sim N(0, F)$ , sendo  $F = I\sigma_f^2$ ;  $p$ : vetor do efeito da interação genótipo x disponibilidade de N x ambiente, considerado como aleatório, em que  $p \sim N(0, P)$ , sendo  $P = I\sigma_p^2$ ;  $\varepsilon$ : vetor de resíduos, onde  $\varepsilon \sim N(0, \mathcal{E})$ , em que  $\mathcal{E} = I\sigma_\varepsilon^2$ .  $I$  é a matriz identidade e  $J, Z, W, V, S, T, C, H$  são matrizes de incidência que relacionam os efeitos dos vetores independentes de cada matriz ao vetor dependente  $y$ .

Para a construção da matriz de parentesco (K), as 64 linhagens foram genotipadas via plataforma Illumina GoldenGate (Fan *et al.*, 2003), contendo 768 marcadores SNPs. Procedimento este realizado pela Pioneer Hi-Bred International Int. (Johnston, IA). Foram feitas análises de qualidade no conjunto de marcadores, sendo eliminadas as linhagens com dados perdidos acima de 20% (Dao *et al.*, 2014), Call Rate  $\geq 0,85$  e MAF  $\leq 0,05$ . De forma que, ao final do controle de qualidade, permaneceram 539 marcadores e 62 linhagens. Para a construção da matriz de parentesco genômico entre as linhagens foi usado o método de VanRaden (2008).

Para as análises individuais, ou seja, dentro de cada nível nutricional, foram considerados sub modelos da equação (6). Entretanto, foram descartadas as fontes de variação das disponibilidades de N e as suas interações simples e tripla com genótipo e ambiente.

Para os testes dos efeitos aleatórios foi empregado o *Likelihood Ratio Test* (LRT):

$$LRT = 2.[\log L_2 - \log L_1] \sim \chi_{r_2-r_1}^2 \quad (7)$$

em que,  $H_0: \sigma^2 = 0$  e  $H_1: \sigma^2 \neq 0$ ;  $\log L_2$ : estimativa *Likelihood* do modelo sem o efeito aleatório de interesse;  $\log L_1$ : estimativa *Likelihood* do modelo com o efeito aleatório de interesse.

Para os testes dos efeitos fixos foi utilizado o teste de Wald:

$$W = \frac{\hat{\beta}}{DP(\hat{\beta})} \quad (8)$$

em que,  $H_0: \beta = 0$  e  $H_1: \beta \neq 0$ ;  $\hat{\beta}$ : estimativa *Likelihood* para o efeito fixo de interesse;  $DP(\hat{\beta})$ : desvio padrão de  $\hat{\beta}$ .

Com o objetivo de obter a relação entre as respostas das linhagens de milho sob disponibilidade de BN e IN, foi considerado o desempenho de cada linhagem na mesma característica em ambas as disponibilidades de N. Para isso, foi utilizado o índice de Correlação de *Spearman*:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (9)$$

sendo,  $d_i^2$  : a diferença entre os ranques de cada linhagem nas duas disponibilidades de N;  $n$  : número de pares de valores considerados.

A herdabilidade no sentido amplo foi estimada para cada característica nas duas disponibilidades de N:

$$\hat{h}^2 = \frac{\sigma_g^2}{\sigma_f^2} \quad (10)$$

em que,  $\sigma_g^2$  : variância genotípica  $\sigma_f^2$  : variância fenotípica.

Foi também estimada a matriz de coeficientes de correlação linear de *Pearson* ( $r$ ) entre as características estudadas e as diferentes fases de estimacão da EUN (V8, VT e com base na produtividade), em função das disponibilidades de N (BN e IN). Posteriormente a matriz de correlação fenotípica foi dividida em dois grupos, o primeiro em função apenas das características envolvidas com os componentes da produtividade, e o segundo em função das características de raízes, fisiológicas e reprodutivos. Para cada grupo foi procedido o diagnóstico da multicolinearidade baseado em Montgomery e Peck (1992), ao qual leva em consideração o Número de Condições da matriz de correlação. Portanto:

$$NC = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (11)$$

em que,  $\lambda_{\max}$  : maior autovalor da matriz de correlação;  $\lambda_{\min}$  : menor autovalor da matriz de correlação. Dessa forma, temos que  $NC < 100$  : multicolinearidade fraca;  $100 < NC < 1000$  : multicolinearidade moderada a forte;  $NC > 1000$  : multicolinearidade severa. Em seguida, as correlações dentro dos agrupamentos (variáveis explicativas) e EUNV8, EUNVT e EUNP (variáveis dependentes) foram desdobradas em efeitos diretos e indiretos por meio da análise de trilha

(Wright, 1921), estabelecendo assim as relações de causa e efeito entre as características de interesse (Cruz e Carneiro, 2006):

$$\begin{aligned} r_{VD.VI_1} &= p_{VI_1} + p_{VI_2} r_{VI_1.VI_2} + \dots + p_{VI_n} r_{VI_1.VI_n} \\ r_{VD.VI_2} &= p_{VI_1} r_{VI_1.VI_2} + p_{VI_2} + \dots + p_{VI_n} r_{VI_2.VI_n} \\ r_{VD.VI_n} &= p_{VI_1} r_{VI_1.VI_n} + p_{VI_2} r_{VI_2.VI_n} + \dots + p_{VI_n} \end{aligned} \quad (12)$$

em que,  $VD$  : variável dependente;  $VI$  : variável independente ( $1 - n$ ).

O coeficiente de determinação foi calculado seguindo a expressão:

$$R^2 = p_{VI_1}^2 + p_{VI_2}^2 + \dots + p_{VI_n}^2 + 2p_{VI_1}p_{VI_2}r_{VI_1.VI_2} + \dots + 2p_{VI_{n-1}}p_{VI_n}r_{VI_{n-1}.VI_n} \quad (13)$$

O efeito da variável residual sobre a variável dependente foi estimado por:

$$P_\varepsilon = (1 - R^2)^{1/2} \quad (14)$$

As análises conjunta e individuais dos modelos foram realizadas por meio do pacote ASReml (Gilmour *et al.*, 2009) no software R (R Core Team, 2013). Para a obtenção dos coeficientes de trilha foi usado o software GENES (Cruz, 2006). As análises gráficas em função das interações entre os fatores genótipos, ambientes e disponibilidades de nitrogênio foram feitas via software GGEbiplot (Yan e Tinker, 2005).

### 3.6 Resultados e discussão

#### 3.6.1 Análises conjunta, individual e parâmetros genéticos

Para proceder inferências por meio da análise conjunta verificou-se, primeiramente, a diferença entre a magnitude dos resíduos das análises individuais. Sendo que as diferenças entre o maior e o menor valor residual para as dosagens de N, bem como os ambientes estudados, ficaram dentro da margem aceitável. Segundo Pimentel-Gomes (1990), essa relação entre os resíduos quando menor que sete permite a viabilidade da análise conjunta.

Os valores dos coeficientes de variação (CV%) apresentaram maiores diferenças entre as disponibilidades BN e IN apenas para EUNVT (Tabela 1). Para características com esse perfil, estudos com tratamentos em déficit de nutrientes, mostram que é normal observar CV's mais elevados, pois sob estresse as médias geralmente são menores e os quadrados médios e os desvios padrão são maiores (Soares *et al.*, 2011). Entretanto, no geral, as

estimativas de CV% foram altas. Uma das principais razões para isso se deve a grande variabilidade dos dados em torno da média em função da ação do fator ambiente, o qual oscilou consideravelmente as estimativas da maioria das características. Isso é avalizado pela significância dos fatores ambiente (A) e nitrogênio x ambiente (NxA) (Tabela 1).

Conforme o exposto anterior, um dos principais pontos a ser considerado nesta questão é que as estimativas das médias dos níveis de N para EUNP, NE, CE, DE, PE, NFE, NGF, EUNVT e CRAxi, na população estudada, interagem com ambiente devido ao ordenamento ou magnitude das estimativas (Figura 1). Indicando que conforme os ambientes considerados, esses podem trazer instabilidade dentro do próprio nível; fato que poderia ser explicado quando há falta de correlação entre os ambientes. Essa questão é de grande interesse no melhoramento de plantas, pois quando se possui grande quantidade de ambientes, é primordial o agrupamento dos mais correlacionados, ao qual é um facilitador para a recomendação de cultivares. Logo, a performance agrícola dessas cultivares tende a ser próxima em ambientes similares do que em divergentes.

Por meio da análise conjunta foi observado que houve diferenças significativas entre o comportamento das linhagens para as características EUNP, EUNV8, EUNVT, NFE, CRLat, CRAxi, VRAXi e EFPSII em função das disponibilidades de N (Tabela 1). Assim, pode se inferir que as doses contrastantes de N aplicadas foram suficientes. Para corroborar essas afirmações, as médias de EUNP e de características de raiz, dentro de cada nível de N, obtiveram maiores estimativas quando submetidas ao déficit do nutriente. Neste, o incremento foi de 6,3%, 6,8%, 6,9% e 175% para CRAxi, VRAXi, CRLat e EUNP, respectivamente. Estudos indicam que, a extensão e natureza da variação genética do sistema radicular, possuem grande implicação na obtenção e absorção de água e nutrientes para o aumento de produtividade sob estresse hídrico ou baixa fertilidade do solo (Toaldo *et al.*, 2013), principalmente em função do N (Worku *et al.* 2007; Dawson *et al.*, 2008). Isto porque, o crescimento espacial do sistema radicular determina a habilidade da planta em explorar recursos que estão mal distribuídos, sendo proporcionalmente positivo a relação do comprimento e área radicular com o volume ocupado e explorado do solo (Rich e Watt, 2013).

Tabela 1 - Teste de Wald para os efeitos fixos do intercepto, repetição (R), bloco dentro de repetição, disponibilidade de nitrogênio (N), ambientes (A), disponibilidade de nitrogênio x ambientes (NxA). Estimativas da média ( $\bar{x}$ ) e coeficiente de variação (CV%) em linhagens de milho tropical em duas disponibilidades de nitrogênio

Características <sup>1</sup>										
FV	GL	EUNP	NE	CE	DE	PE	NFE	NGF	PROD	IFMF
Intercepto	1	1485,39***	3332,50***	28010,30***	41173,00***	6119,20***	3588,20***	17870,90***	2427,75***	655,64***
Repetição (R)	1	12,30***	1,50 <sup>ns</sup>	7,30**	7,00*	38,30***	0,10 <sup>ns</sup>	1,90 <sup>ns</sup>	21,69***	5,82*
Bloco/R	14	9,46 <sup>ns</sup>	21,00 <sup>ns</sup>	14,50 <sup>ns</sup>	10,00 <sup>ns</sup>	19,40 <sup>ns</sup>	9,00 <sup>ns</sup>	21,70 <sup>ns</sup>	16,76 <sup>ns</sup>	23,73*
Nitrogênio (N)	1	539,17***	1,10 <sup>ns</sup>	1,40 <sup>ns</sup>	1,00 <sup>ns</sup>	0,20 <sup>ns</sup>	8,00**	3,40 <sup>ns</sup>	0,50 <sup>ns</sup>	0,11 <sup>ns</sup>
Ambiente (A)	2	420,17***	52,60***	68,40***	206,00***	165,00***	15,00***	121,90***	908,57***	1,62 <sup>ns</sup>
NxA	2	132,08***	8,50*	23,10***	43,00***	11,50***	7,00*	9,00*	2,26 <sup>ns</sup>	0,64 <sup>ns</sup>
$\bar{x}$	BN	0,11	17,69	11,58	3,81	60,46	13,20	20,91	2060,13	1,80
	IN	0,04	18,01	11,68	3,83	60,73	13,42	21,22	2037,47	1,83
CV%	BN	56,92	37,62	14,69	11,63	28,87	12,31	17,96	68,07	92,86
	IN	60,59	38,05	15,31	13,00	30,95	11,94	18,57	67,39	89,55
FV	GL	CRLat	CRAxi	VRLat	VRAXi	EFPSII	EUNV8	EUNVT		
Intercepto	1	3501,30***	4240,40***	2859,04***	3165,60***	37178,00***	5077,90***	5852,10***		
Repetição (R)	1	4,50*	5,20*	4,24*	0,20 <sup>ns</sup>	7,00**	7,80**	1,50 <sup>ns</sup>		
Bloco/R	14	17,20 <sup>ns</sup>	12,40 <sup>ns</sup>	10,65 <sup>ns</sup>	12,70 <sup>ns</sup>	26,00*	14,00 <sup>ns</sup>	23,30 <sup>ns</sup>		
Nitrogênio (N)	1	5,00*	6,10*	2,83 <sup>ns</sup>	4,80*	129,00***	1282,10***	1678,50***		
Ambiente (A)	1	63,10***	37,80***	61,41***	77,40***	5,00*	305,20***	26,30***		
NxA	1	1,10 <sup>ns</sup>	3,80*	1,19 <sup>ns</sup>	2,00 <sup>ns</sup>	1,00 <sup>ns</sup>	126,10***	18,00***		
$\bar{x}$	BN	878,28	645,08	0,43	42,43	0,66	0,01	0,07		
	IN	821,15	606,69	0,41	39,70	0,74	0,03	0,22		
CV%	BN	41,37	32,55	43,46	40,33	12,50	35,06	30,74		
	IN	43,28	34,36	42,21	41,91	9,81	37,41	22,59		

<sup>ns</sup>= não significativo, \*\*\* = 0,001, \*\* = 0,01, \* = 0,05; <sup>1</sup> Características: eficiência no uso de nitrogênio na produção (EUNP), número de espigas (NE), comprimento de espiga (CE), diâmetro de espigas (DE), peso de espigas (PE), número de fileiras por espiga (NFE), número de grãos por fileira (NGF), produtividade (PROD), intervalo de florescimento masculino-feminino (IFMF), comprimento de raiz lateral (CRLat), comprimento de raiz axial (CRAxi), volume de raiz lateral (VRLat), volume de raiz Axial (VRAXi), eficiência fotossintética do fotossistema II (EFPSII), eficiência no uso de nitrogênio em estágio V8 (EUNV8) e eficiência no uso de nitrogênio em estágio VT (EUNVT)

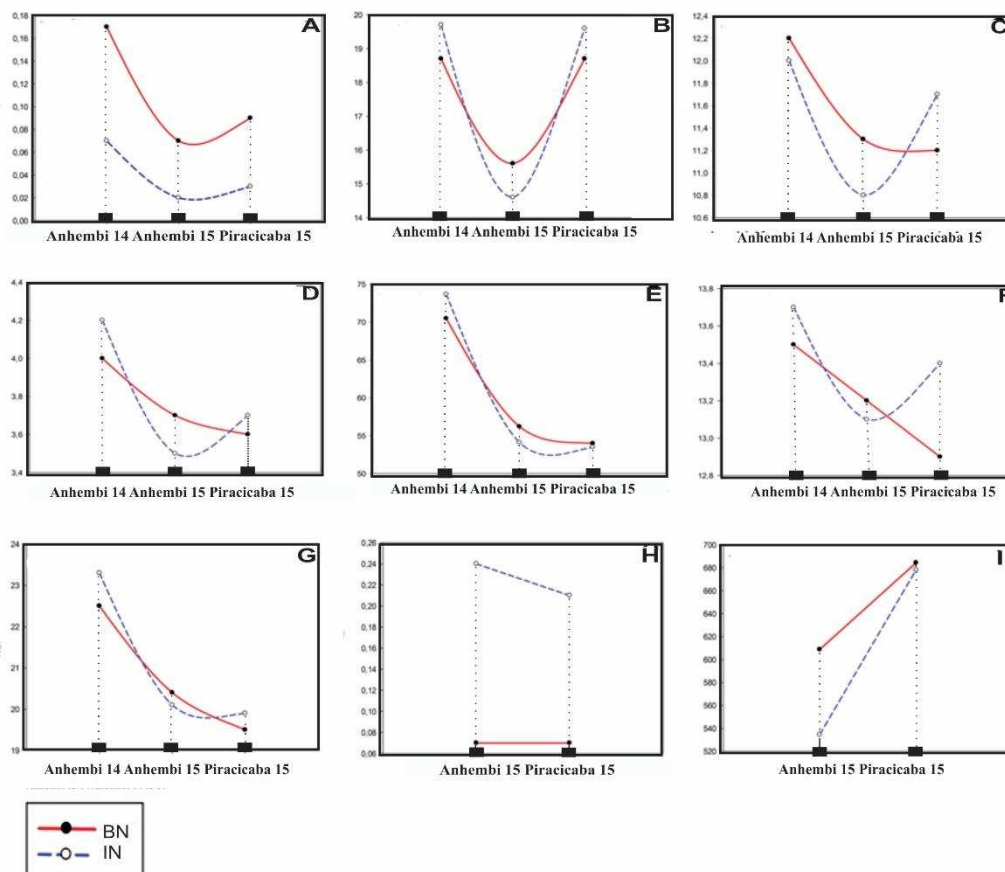


Figura 1 - Médias das disponibilidades de nitrogênio (baixo - BN e ideal - IN) nos ambientes Anhembi 2014, Anhembi 2015 e Piracicaba 2015 para as características eficiência no uso de nitrogênio na produção – EUNP (A), número de espigas - NE (B), comprimento de espiga - CE (C), diâmetro de espiga - DE (D), peso de espiga - PE (E), número de fileiras por espiga - NFE (F), número de grãos por fileira – NGF (G), eficiência de uso de nitrogênio em estádios VT - EUNVT (H) e comprimento de raiz axial - CRAXi (I)

Diferentemente EFPSII, EUNV8 e EUNVT apresentaram maiores estimativas (0,74, 0,03 e 0,22) respectivamente (Tabela 1), quando submetidas a adubações mais elevadas. Sendo 12,2%, 300% e 314,3% os incrementos, quando comparadas com as mesmas características em BN. Para as demais características, as linhagens não apresentaram diferenças significativas de comportamento em função de N, principalmente PROD. Resultados observados por Soares *et al.* (2011), Barbieri *et al.* (2008) e Presteri *et al.* (2003) indicam que NE, PE e PROD sofrem redução quando avaliadas em ambientes deficitários em N, sendo esta redução de 24 a 37%. Isto ocorre devido ao N possuir grande importância no metabolismo da planta, principalmente na síntese de proteínas,

como a clorofila, os ácidos nucleicos e biomoléculas como o ATP, NADH e NADPH, sendo fundamental no estabelecimento da produtividade (Richardson *et al.*, 2009). No entanto, a resposta paralela entre BN e IN para PROD, leva a hipótese de que o estresse não foi suficiente para gerar diferenças entre as disponibilidades de N. Isso pode ser devido ao N residual no solo provindo de adubações anteriores nas áreas experimentais e também devido a sucessão com leguminosas. Exemplo disso é o valor de N residual na área experimental de Piracicaba, que varia de 36-42 kg.ha<sup>-1</sup> de N (Tabela Suplementar 1), possuindo potencial de mascarar as dosagens utilizadas.

As linhagens não apresentaram diferenças significativas para IFMF sob as diferentes condições de N, não havendo portanto nenhum efeito negativo (atraso, má formação ou asincronia) no desenvolvimento e crescimento das inflorescências masculinas e femininas. Tal fato possibilita obter condições necessárias para o desenvolvimento dos grãos nas duas condições de adubação, sendo isso visível em um dos componentes da produtividade, NGF, que não apresentou diferença significativa entre BN e IN, assim como não houve ausência de falhas no ápice das espigas avaliadas.

A característica EFPSII apresentou maiores estimativas de média quando avaliada em IN. Porém, a maior eficiência não resultou em maior produtividade ou em acréscimos em MPS, justificado, quando analisado o baixo efeito direto de PROD com EUNV8 e EUNVT (Tabela 3). As maiores estimativas de MPS em IN são o reflexo das doses aplicadas, que permitiram maior crescimento vegetativo. Isso pode ser constatado por meio dos elevados valores de EUNV8 e EUNVT em IN, ao qual são totalmente dependentes de MSP mensuradas em estágio V8 e VT. A partir disso, do aumento do volume vegetativo, pode se inferir que a translocação de nutrientes, do colmo e das folhas para a espiga, que normalmente ocorre a partir do estágio R2 (grão bolha d'água) não resulta necessariamente no aumento da capacidade de incrementar o PE e conseqüentemente a PROD em IN. Condizente a essa posição, estudos prévios indicam que a translocação de nutrientes, principalmente do N, possuem os mesmos padrões tanto para BN como para doses mais elevadas, ou então maiores taxas de translocação para BN (Gallais e Coque, 2005).

Através da análise conjunta para os efeitos aleatórios do modelo estatístico-genético, os resultados indicaram diferenças significativas entre as

linhagens avaliadas, para praticamente todas as características, com exceção de EFPSII (Tabela 2). Resultado esse que demonstra a existência de variabilidade genética na população, o que se presume possibilidades de ganho genético por meio de seleção. Esse fato, quando analisado para EUN, trará ganhos também para os híbridos gerados a partir dessas linhagens, pois os efeitos aditivos nessa características são de maior importância em relação ao não aditivos (DoVale *et al.*, 2012). Os efeitos significativos de interação genótipo x nitrogênio (GxN) para EUNP, IFMF, EUNVT demonstraram que determinadas linhagens respondem de forma diferenciada quando submetidas a BN e IN. Esses resultados, de certa forma já eram esperados, pois as EUN são índices que se baseiam na relação entre produção de grãos ou matéria seca, submetidos por um denominador específico para cada condição de N. Assim, devido a não significância de PROD para o fator nitrogênio (N) em conjunto com maiores estimativas de MSP em IN, resultam em respostas contrastantes para genótipo x disponibilidade de N (GxN). Já para IFMF, a magnitude dessa interação não propiciou nenhum efeito que pudesse restringir a produtividade final. Para as demais características não foram obtidos valores significativos, possibilitando que o conjunto de linhagens possa ser avaliado em um único ambiente de forma que as inferências tomadas para um ambiente será o mesmo para o outro (DoVale e Frischie-Neto, 2013).

Com o mesmo padrão de comportamento, o qual modificou as médias do fator N, os ambientes utilizados também imprimiram efeito considerável nas respostas das linhagens, evidenciando o efeito da GxA (Tabela 2). Ilustrando esse comportamento, os gráficos via GGEbiplot (Figura 2), indicam grande dispersão das linhagens para EUNP, PE, NGF e PROD em função dos três ambientes; NE e CE possuem dispersões intermediárias. Para DE, NFE e CRAXi, as linhagens tendem a apresentar menor diferença de comportamento ao longo dos ambientes avaliados, ou seja, são mais estáveis. A estabilidade é visualizada em relação ao centro do gráfico. Logo, quanto maior for a distância em relação a esse ponto, menor é a estabilidade (Gabriel, 1971).

Tabela 2 - Teste de máxima verossimilhança (LTR) para os efeitos aleatórios de genótipo (G), genótipos x disponibilidade de nitrogênio (GxN), genótipos x ambiente (GxA), genótipos x disponibilidade de nitrogênio x ambientes (GxNxA). Estimativas do resíduo, herdabilidades ( $h^2$ ) e correlação de Spearman entre genótipo em função das disponibilidades de nitrogênio ( $r_{INxLN}$ ) em linhagens de milho tropical

<sup>ns</sup>= não significativo, \*\*\* = 0,001, \*\* = 0,01, \* = 0,05; <sup>1</sup> eficiência no uso de nitrogênio na produção (EUNP),

FV		Características <sup>1</sup>			
		EUNP	NE	CE	DE
Genótipo (G)		16,20***	60,81***	93,57***	74,2***
GxN		13,38***	0,00 <sup>ns</sup>	0,00 <sup>ns</sup>	0,00 <sup>ns</sup>
GxA		18,36***	43,65***	6,01**	44,7***
GxNxA		21,41***	0,64 <sup>ns</sup>	5,41*	0,00 <sup>ns</sup>
Resíduo		4,3.10 <sup>-4</sup>	16,42	1,08	6,2.10 <sup>-2</sup>
$h^2$	BN	0,70	0,66	0,73	0,68
	IN	0,67	0,71	0,69	0,73
$r_{INxLN}$	-	0,87	0,82	0,86	0,89
		PE	NFE	NGF	PROD
Genótipo		70,05***	93,92***	92,93***	38,5***
GxN		0,65 <sup>ns</sup>	0,00 <sup>ns</sup>	0,00 <sup>ns</sup>	0,00 <sup>ns</sup>
GxA		102,34***	20,27***	24,75***	95,6***
GxNxA		0,00 <sup>ns</sup>	1,83 <sup>ns</sup>	0,02 <sup>ns</sup>	0,00 <sup>ns</sup>
Resíduo		70,36	0,85	5,07	219929,70
$h^2$	BN	0,79	0,78	0,76	0,68
	IN	0,69	0,81	0,69	0,65
$r_{INxLN}$	-	0,90	0,89	0,87	0,87
		IFMF	CRLat	CRAXi	VRLat
Genótipo		45,27***	51,92***	23,79***	45,19***
GxN		8,87***	0,00 <sup>ns</sup>	0,00 <sup>ns</sup>	0,00 <sup>ns</sup>
GxA		0,05 <sup>ns</sup>	0,06 <sup>ns</sup>	2,89*	0,61 <sup>ns</sup>
GxNxA		0,00 <sup>ns</sup>	0,00 <sup>ns</sup>	0,00 <sup>ns</sup>	0,10 <sup>ns</sup>
Resíduo		1,51	75971,67	28747,95	0,02
$h^2$	BN	0,62	0,45	0,36	0,47
	IN	0,69	0,51	0,28	0,49
$r_{INxLN}$	-	0,62	0,81	0,71	0,78
		VRAXi	EFPSII	EUNV8	EUNVT
Genótipo		20,34***	0,00 <sup>ns</sup>	4,89*	5,07*
GxN		0,00 <sup>ns</sup>	0,10 <sup>ns</sup>	0,00 <sup>ns</sup>	3,93*
GxA		0,70 <sup>ns</sup>	0,21 <sup>ns</sup>	0,90 <sup>ns</sup>	0,00 <sup>ns</sup>
GxNxA		0,12 <sup>ns</sup>	0,01 <sup>ns</sup>	0,00 <sup>ns</sup>	0,25 <sup>ns</sup>
Resíduo		184,08	5,6.10 <sup>-3</sup>	5,2.10 <sup>-5</sup>	1,0.10 <sup>-3</sup>
$h^2$	BN	0,46	0,11	0,31	0,08
	IN	0,14	0,00	0,12	0,37
$r_{INxLN}$	-	0,71	0,15	0,59	0,36

número de espigas (NE), comprimento de espiga (CE), diâmetro de espigas (DE), peso de espigas (PE), número de fileiras por espiga (NFE), número de grãos por fileira (NGF), produtividade (PROD), intervalo de florescimento masculino-feminino (IFMF), comprimento de raiz lateral (CRLat), comprimento de raiz axial (CRAXi), volume de raiz lateral (VRLat), volume de raiz Axial (VRAXi), eficiência fotossintética do fotossistema II (EFPSII), eficiência no uso de nitrogênio em estágio V8 (EUNV8) e eficiência no uso de nitrogênio em estágio VT (EUNVT)

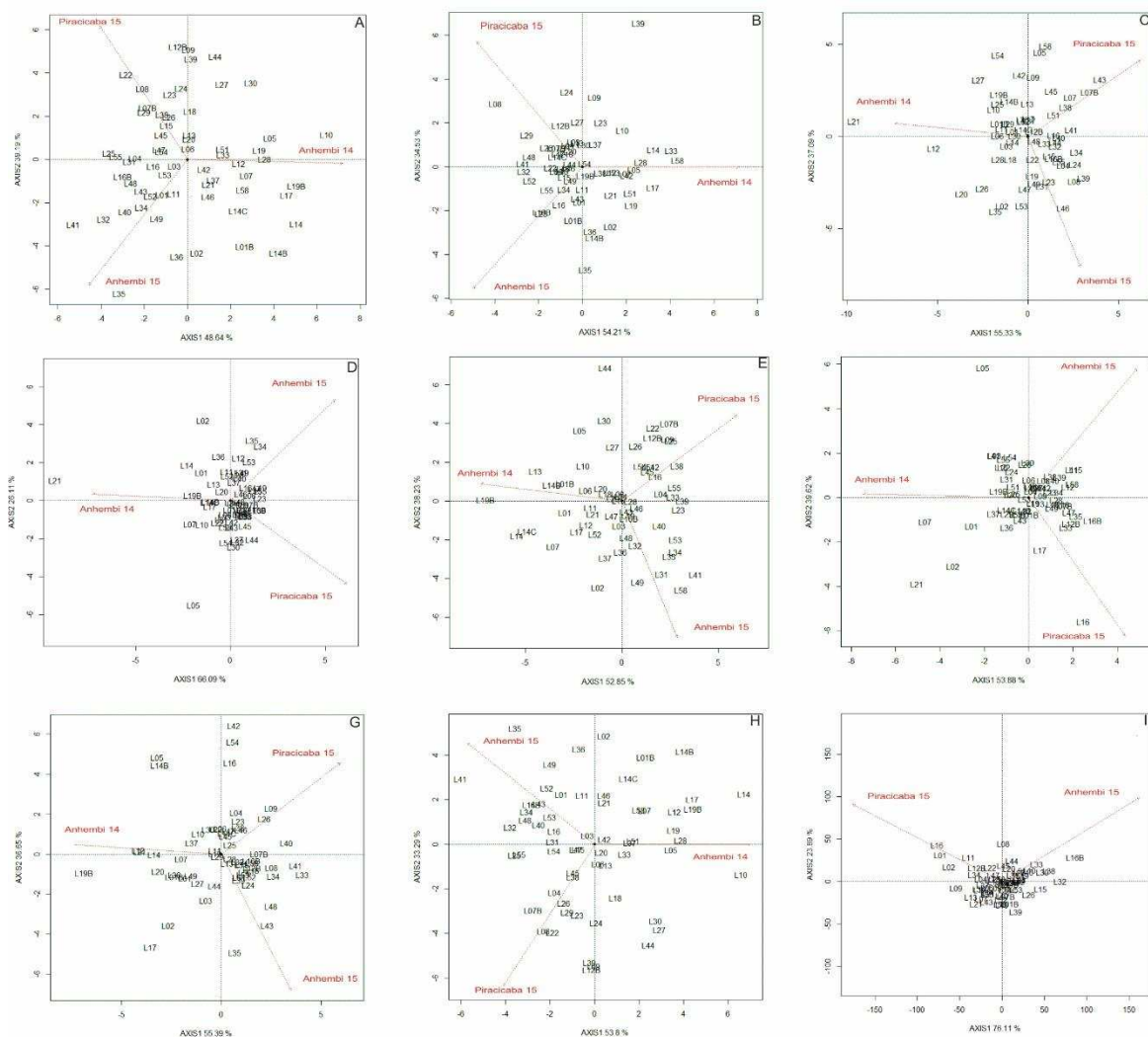


Figura 2 - Análise GGEbiplot da interação genótipo x ambiente para 62 linhagens avaliadas em função dos ambientes Anhembi/2014, Anhembi/2015 e Piracicaba/2015 para as características eficiência no uso de nitrogênio na produção - EUNP (A), número de espigas - NE (B), comprimento de espiga - CE (C), diâmetro de espiga - DE (D), peso de espiga - PE (E), número de fileiras por espiga - NFE (F), número de grãos por fileira - NGF (G), produtividade - PROD (H). Comprimento de raiz axial - CRAxi (I) está em função de Anhembi/2015 e Piracicaba/2015

Já para a interação de linhagens, nitrogênio e ambiente (GxNxA), apenas a característica EUNP e CE apresentaram valores significativos (Tabela 2). As linhagens quando analisadas graficamente para essas mesmas características (Figura 3), indicam grande variabilidade na dispersão dos pontos em EUNP e menor dispersão em CE. Presterl *et al.* (2003) relatam que as variâncias para as interações dos genótipos com locais e anos são geralmente maiores em relação as variâncias entre genótipos e disponibilidades de N.

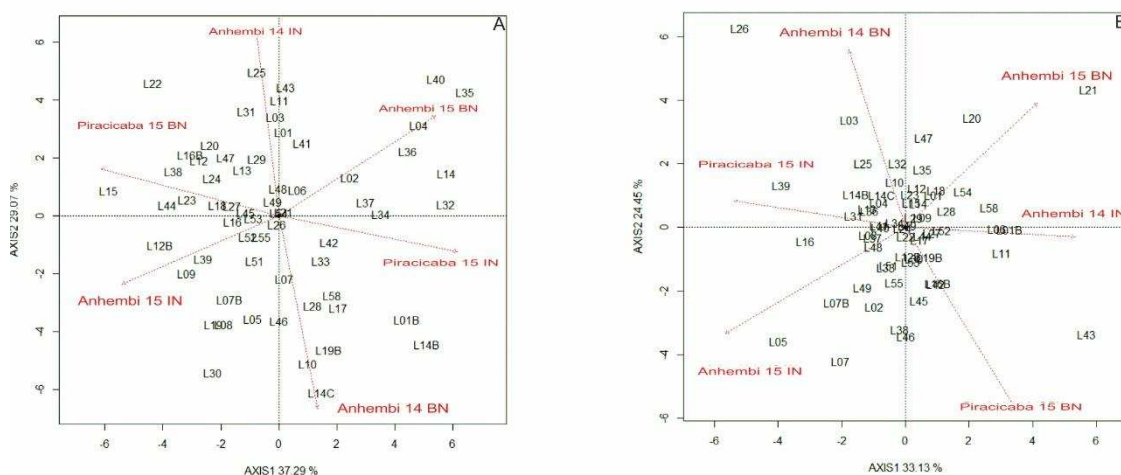


Figura 3 - Análise GGEbiplot para genótipo x disponibilidade de nitrogênio x ambiente com 62 linhagens avaliadas em função dos ambientes Anhembi/2014, Anhembi/2015 e Piracicaba/2015 em duas disponibilidades de nitrogênio, baixo (BN) e ideal (IN), para as características eficiência no uso de nitrogênio na produção - EUNP (A) e comprimento de espiga - CE (B)

Fica assim evidente, quando analisamos os efeitos fixos e aleatórios envolvidos com os ambientes, que há necessidade da exploração desses mesmos ambientes em mais de uma safra a fim de estimar o efeito que pode ocorrer e a resposta dos fatores estudados a essa mudança. Além disso, quando existe a interação entre esses, há possibilidades da classificação e indicação das linhagens EUN em BN e IN para ambientes específicos. As herdabilidades estimadas para a maioria das características estudadas são de intermediárias a altas, indicando relativo equilíbrio entre as condições de BN e IN; com diferença menor que 10 unidades entre ambos, com exceção de VRAXi, EFPSII, EUNV8 e EUNVT. Todavia, em estudos com estresse por N, conforme Coque e Gallais (2006), a variabilidade genética é reduzida e conseguinte a herdabilidade, de forma que os genótipos tendem a apresentar desempenhos semelhantes. Não obstante, as estimativas próximas de herdabilidades entre BN e IN aqui mostradas, são em função das respostas paralelas entre as linhagens nas duas condições de N, não restringindo a expressão da variabilidade genética em BN.

Por meio do índice de Correlação de Spearman (Tabela 2), observou-se que as estimativas para a grande maioria das características consideradas foram de média a alta magnitude nas duas disponibilidades de N, reiterando alguns resultados semelhantes já reportados por Bertin e Gallais (2000) e Presterl *et al.* (2003). Segundo DoVale e Fristche-Neto (2013), tal condição reflete a consistência da superioridade das genótipos, pois, linhagens que tem

desempenho superior em IN tendem a responder da forma semelhante em BN. Por outro lado, EFPSII e EUNVT, nas quais a  $r_{AN \times BN}$  apresentou as menores estimativas, as linhagens tendem a não ser previsíveis para estas características em função da mudança na disponibilidade de N.

### 3.6.2 *Relação fenotípica entre características*

O uso de correlações simples entre as características envolvidas com EUN traz grande contribuição para os melhoristas, pois permitem direcionar e prever comportamento de determinada característica pela seleção de outra. Entretanto, esse procedimento não distingue efeitos diretos e indiretos envolvidos nas estimativas de correlação. Logo, por meio da análise de trilha, há a possibilidade de se obter o desdobramento das estimativas de correlação em efeitos diretos e indiretos, permitindo ao mesmo tempo inferir de forma mais clara a importância de um conjunto de características secundárias sobre uma determinada característica principal (Dewey e Lu, 1959; Silva *et al.*, 2009).

Assim, para implementação da análise de trilha, foi feita a análise de multicolinearidade; tal qual os NC para o agrupamento de características relacionada com os componentes da produtividade, em BN e IN, ficaram no limite entre fraco e moderado. Com base nos resultados, foi necessário a exclusão da característica PROD para as análises com EUNP, pois esse índice é derivado exclusivamente de PROD, obtendo correlação quase que perfeita entre os mesmos. Para o agrupamento de características reprodutivas, fisiológicas e de raízes, os NC observados foram considerados fracos para ambas disponibilidades de N.

A primeira constatação para os componentes da produtividade é que os coeficientes explicaram satisfatoriamente as variações em EUNP em BN, com  $R^2=0,93$  e efeito residual de 0,25 e em IN com  $R^2=0,87$  e efeito residual de 0,36 (Tabela 3). O mesmo não aconteceu para EUNV8 e EUNVT nas duas disponibilidades de N. Para essas duas características (variáveis dependentes) não foram evidenciados efeitos direto e nem indireto nas demais características (variáveis explicativas) que possam superar as estimativas residuais. Isto indica a ineficiência das mesmas em determinar as variações em EUNV8 e EUNVT. Com isto, o principal resultado é a inviabilidade de seleção indireta para EUNV8 e EUNVT. Além disso, ganhos em PROD via EUNV8 e EUNVT não tendem a

ser promissores, devido as baixas herdabilidades e efeitos diretos intermediários ou quase nulos dessas características com PROD.

Tabela 3 - Estimativas dos efeitos diretos e indiretos em função das disponibilidades de nitrogênio (baixo- BN e ideal - IN) para os componentes de produção em função das características eficiência no uso de nitrogênio na produção (EUNP), eficiência no uso de nitrogênio em estágio V8 (EUNV8) e eficiência no uso de nitrogênio em estágio VT (EUNVT) em linhagens de milho tropical

BN					IN				
Caract. <sup>1</sup>	Efeitos	EUNP	EUNV8	EUNVT	Caract.	Efeitos	EUNP	EUNV8	EUNVT
NE					NE				
Direto		0,77	0,07	0,10	Direto		0,77	0,11	-0,34
Indireto					Indireto				
	CE	0,01	0,01	0,03		CE	0,00	0,00	-0,03
	DE	0,00	-0,01	0,00		DE	0,02	0,05	0,06
	PE	-0,08	0,02	0,02		PE	-0,05	0,00	0,02
	NFE	0,00	0,00	0,02		NFE	0,00	-0,01	0,01
	NGF	-0,02	0,02	-0,01		NGF	-0,02	-0,02	0,30
	PROD	-	0,03	-0,06		PROD	-	0,04	0,16
	Total	0,68	0,14	0,10		Total	0,72	0,17	0,18
CE					CE				
Direto		0,07	0,05	0,21	Direto		-0,02	-0,03	-0,14
Indireto					Indireto				
	NE	0,13	0,01	0,01		NE	0,17	0,02	-0,07
	DE	0,02	0,18	0,11		DE	0,07	0,20	0,22
	PE	0,33	-0,07	-0,06		PE	0,24	0,01	-0,09
	NFE	0,00	0,00	0,01		NFE	0,00	0,00	0,01
	NGF	-0,03	0,03	-0,01		NGF	-0,02	-0,02	0,31
	PROD	-	0,02	-0,05		PROD	-	0,02	0,11
	Total	0,52	0,22	0,22		Total	0,44	0,20	0,35
DE					DE				
Direto		0,07	0,53	0,32	Direto		0,13	0,38	0,41
Indireto					Indireto				
	NE	-0,02	0,00	0,00		NE	0,12	0,02	-0,05
	CE	0,02	0,01	0,07		CE	-0,01	-0,02	-0,07
	PE	0,46	-0,10	-0,09		PE	0,32	0,01	-0,11
	NFE	0,00	-0,01	-0,07		NFE	0,01	0,05	-0,04
	NGF	-0,01	0,01	-0,01		NGF	-0,01	-0,01	0,20
	PROD	-	0,02	-0,05		PROD	-	0,03	0,13
	Total	0,52	0,46	0,17		Total	0,56	0,46	0,47
PE					PE				
Direto		0,62	-0,14	-0,13	Direto		0,54	0,02	-0,19
Indireto					Indireto				
	NE	-0,10	-0,01	-0,01		NE	-0,08	-0,01	0,03
	CE	0,03	0,02	0,11		CE	-0,01	-0,01	-0,06
	DE	0,05	0,39	0,24		DE	0,08	0,23	0,25
	NFE	0,00	0,00	-0,04		NFE	0,01	0,04	-0,03
	NGF	-0,02	0,02	-0,01		NGF	-0,02	-0,02	0,21
	PROD	-	0,02	-0,05		PROD	-	0,03	0,12
	Total	0,58	0,30	0,11		Total	0,52	0,28	0,33
NFE					NFE				
Direto		0,00	-0,03	-0,20	Direto		0,03	0,17	-0,14
Indireto					Indireto				

(Continuação)

BN					IN				
Caract. <sup>1</sup>	Efeitos	EUNP	EUNV8	EUNVT	Caract.	Efeitos	EUNP	EUNV8	EUNVT
	NE	-0,08	0,00	-0,01		NE	-0,06	-0,01	0,02
	CE	0,00	0,00	-0,01		CE	0,00	0,00	0,00
	DE	0,03	0,20	0,12		DE	0,04	0,11	0,12
	PE	0,11	-0,03	-0,02		PE	0,14	0,01	-0,05
	NGF	0,00	0,00	0,00		NGF	-0,01	-0,01	0,09
	PROD	-	0,00	0,00		PROD	-	0,01	0,03
	Total	0,04	0,14	-0,12		Total	0,14	0,28	0,07
NGF					NGF				
Direto		-0,04	0,05	-0,03	Direto		-0,04	-0,04	0,52
Indireto					Indireto				
	NE	0,34	0,03	0,04		NE	0,44	0,06	-0,19
	CE	0,04	0,03	0,13		CE	-0,01	-0,02	-0,08
	DE	0,02	0,13	0,08		DE	0,05	0,15	0,16
	PE	0,25	-0,05	-0,05		PE	0,21	0,01	-0,08
	NFE	0,00	0,00	0,00		NFE	0,01	0,03	-0,02
	PROD	-	0,02	-0,05		PROD	-	0,04	0,15
	Total	0,61	0,21	0,12		Total	0,66	0,23	0,46
PROD					PROD				
Direto		-	0,04	-0,08	Direto		-	0,06	0,22
Indireto					Indireto				
	NE	-	0,05	0,07		NE	-	0,08	-0,24
	CE	-	0,03	0,11		CE	-	-0,02	-0,06
	DE	-	0,27	0,17		DE	-	0,22	0,23
	PE	-	-0,08	-0,08		PE	-	0,01	-0,10
	NFE	-	0,00	0,00		NFE	-	0,02	-0,02
	NGF	-	0,03	-0,02		NGF	-	-0,02	0,36
	Total	-	0,33	0,17		Total	-	0,35	0,39
	R <sup>2</sup>	0,93	0,24	0,11		R <sup>2</sup>	0,87	0,25	0,33
	P <sub>ε</sub>	0,25	0,87	0,95		P <sub>ε</sub>	0,36	0,86	0,81

R<sup>2</sup>: coeficiente de determinação; P<sub>ε</sub>: efeito da variável residual; <sup>1</sup> Características: comprimento de espiga (CE), diâmetro de espiga (DE), peso de espiga (PE), número de fileiras por espiga (NFE), número de grãos por fileira (NGF) e produtividade (PROD)

Para a EUNP, as estimativas de grande parte das variáveis explicativas superam o efeito residual, tanto em BN como em IN. As duas disponibilidades avaliadas apresentam comportamento semelhante diante da análise de trilha, de forma que NE e PE são as que mais contribuíram na determinação de EUNP, propiciando o uso da seleção indireta (Tabela 3). Essas mesmas características também possuem grande importância no CE, DE e NGF. Por outro lado, as estimativas de efeito direto de NFE e NGF foram, em geral, baixas ou negativas, de forma que essas características não são boas opções para seleção indireta para EUNP, tanto em BN como em IN. Além disso, essas são características dispendiosas em tempo e custo para avaliação quando comparadas em relação as avaliações de NE, PE e até mesmo para característica principal (EUN).

Em função dos resultados, a seleção indireta para EUNP via NE tende a ser mais vantajosa em relação a PE, devido as suas maiores estimativas (efeito direto e total) em BN e IN pela análise de trilha. Neste sentido, Moll *et al.* (1982) em seu trabalho mostraram justamente que NE é uma característica de relevância para EUN, com estimativas de correlação acima dos 70%. No entanto, para seleção via NE, deve ser levado em consideração a prolificidade das linhagens, índice de acamamento e quebra de colmo, ao quais podem alterar significativamente o número o estande, número de espigas úteis na parcela e consequentemente a produtividade (Kang *et al.*, 1999). Complementarmente, deve ser levado em consideração que NE e PE por meio do coeficiente de trilha possuem sentidos contrários, logo, a seleção para NE tende a diminuir o PE e vice-versa.

Entretanto, essa relação negativa entre NE e PE é que pode dificultar o uso de índices de seleção, mas não o impossibilitam. Segundo Rangel *et al.* (2011), os índices de seleção permitem obter ganhos simultâneos mesmo para características com efeitos adversos, havendo também estudos que atestam o sucesso do emprego dos índices de seleção sob essas condições (Freitas *et al.*, 2009; Freitas *et al.*, 2013).

Para as análises considerando as características de raízes, reprodutivas e fisiológicas, os coeficientes de trilha não explicaram satisfatoriamente as variações em EUNP, EUNV8 e EUNVT (Tabela 4). Outro resultado interessante é que apesar dos incrementos em BN para algumas características radiculares, e por estas se apresentarem como opção viável para a seleção em EUN, devido a capacidade de absorção de água e nutrientes (Lynch, 2013; Lambers *et al.*, 2006) e exploração de camadas mais profundas e compactadas no perfil do solo (Peng *et al.*, 2012; Waisel e Eshel, 2002), as mesmas não possuem efeito direto ou total superior à estimativa residual da análise, em nenhuma das disponibilidades de N (Tabela 4).

Tabela 4 - Estimativas dos efeitos diretos e indiretos em função das disponibilidades de nitrogênio (baixo- BN e ideal - IN) para os componentes de raiz, fisiológico e reprodutivo em função das características eficiência no uso de nitrogênio na produção (EUNP), eficiência no uso de nitrogênio em estágio V8 (EUNV8) e eficiência no uso de nitrogênio em estágio VT (EUNVT) em linhagens de milho tropical

		BN			IN				
Caract. <sup>1</sup>	Efeitos	EUNP	EUNV8	EUNVT	Caract.	Efeitos	EUNP	EUNV8	EUNVT
CRLat					CRLat				
Direto		-0,33	0,18	0,52	Direto		-0,56	0,30	0,52
Indireto					Indireto				
	CRAxi	-0,19	0,00	0,10		CRAxi	-0,28	-0,23	0,06
	VRLat	0,39	-0,21	-0,52		VRLat	0,82	-0,10	-0,59
	VRAXi	0,14	0,34	0,09		VRAXi	0,10	0,36	0,18
	EFPSII	0,00	0,00	0,00		EFPSII	0,01	0,00	0,00
	IFMF	0,00	0,00	0,00		IFMF	0,00	0,00	0,00
	Total	0,01	0,31	0,19		Total	0,09	0,33	0,17
CRAxi					CRAxi				
Direto		0,26	0,00	0,14	Direto		-0,40	-0,31	0,09
Indireto					Indireto				
	CRLat	-0,24	0,14	0,39		CRLat	-0,40	0,21	0,37
	VRLat	0,37	-0,20	-0,48		VRLat	0,76	-0,10	-0,55
	VRAXi	0,20	0,47	0,12		VRAXi	0,14	0,48	0,25
	EFPSII	-0,01	0,00	-0,01		EFPSII	0,02	0,00	0,00
	IFMF	-0,01	0,00	-0,01		IFMF	0,01	-0,02	0,00
	Total	0,57	0,41	0,15		Total	0,13	0,26	0,16
VRLat					VRLat				
Direto		0,43	-0,23	-0,56	Direto		0,88	-0,12	-0,64
Indireto					Indireto				
	CRLat	-0,30	0,17	0,48		CRLat	-0,52	0,28	0,48
	CRAxi	-0,22	0,00	0,12		CRAxi	-0,34	-0,27	0,07
	VRAXi	0,16	0,38	0,10		VRAXi	0,12	0,40	0,21
	EFPSII	0,00	0,00	0,00		EFPSII	0,01	0,00	0,00
	IFMF	-0,01	0,00	-0,01		IFMF	0,00	-0,01	0,00
	Total	0,06	0,32	0,13		Total	0,15	0,28	0,12
VRAXi					VRAXi				
Direto		0,28	0,68	0,18	Direto		0,20	0,68	0,35
Indireto					Indireto				
	CRLat	-0,16	0,09	0,27		CRLat	-0,29	0,16	0,27
	CRAxi	-0,18	0,00	0,09		CRAxi	-0,28	-0,22	0,06
	VRLat	0,24	-0,13	-0,32		VRLat	0,52	-0,07	-0,37
	EFPSII	-0,02	0,00	-0,02		EFPSII	0,03	0,00	0,00
	IFMF	0,00	0,00	0,00		IFMF	0,00	0,00	0,00
	Total	0,16	0,64	0,20		Total	0,18	0,55	0,31
EFPSII					EFPSII				
Direto		0,11	-0,01	0,09	Direto		0,13	0,00	0,00
Indireto					Indireto				
	CRLat	0,01	0,00	-0,02		CRLat	-0,06	0,03	0,06
	CRAxi	0,02	0,00	-0,01		CRAxi	-0,07	-0,05	0,01
	VRLat	-0,01	0,00	0,01		VRLat	0,09	-0,01	-0,06
	VRAXi	-0,06	-0,14	-0,04		VRAXi	0,04	0,14	0,07
	IFMF	-0,01	0,00	-0,01		IFMF	0,00	0,00	0,00
	Total	-0,06	-0,15	0,02		Total	0,13	0,11	0,08
IFMF					IFMF				
Direto		-0,07	0,00	-0,04	Direto		0,02	-0,12	0,00
Indireto					Indireto				
	CRLat	-0,03	0,02	0,05		CRLat	-0,02	0,01	0,02
	CRAxi	-0,04	0,00	0,02		CRAxi	-0,07	-0,06	0,01
	VRLat	0,06	0,01	-0,08		VRLat	0,10	-0,01	-0,07

(Continuação)

BN					IN				
Caract. <sup>1</sup>	Efeitos	EUNP	EUNV8	EUNVT	Caract.	Efeitos	EUNP	EUNV8	EUNVT
	VRAXi	0,00	0,01	0,00		VRAXi	0,00	-0,01	0,00
	EFPSII	0,02	0,00	0,01		EFPSII	0,00	0,00	0,00
	Total	-0,06	-0,02	-0,04		Total	0,03	-0,19	-0,04
	R <sup>2</sup>	0,06	0,42	0,09		R <sup>2</sup>	0,08	0,39	0,14
	P <sub>ε</sub>	0,96	0,76	0,95		P <sub>ε</sub>	0,95	0,78	0,93

R<sup>2</sup>: coeficiente de determinação; P<sub>ε</sub>: efeito da variável residual; <sup>1</sup> Características: comprimento de raiz lateral (CRLat), comprimento de raiz axial (CRAxi), volume de raiz lateral (VRLat), volume de raiz axial (VRAxi), eficiência fotossintética do fotossistema II (EFPSII) e intervalo de florescimento masculino-feminino (IFMF)

De modo semelhante, Abdel-Ghani *et al.* (2013) trabalhando com linhagens em BN e IN concluíram que características de sistema radicular não se correlacionam com produtividade. Além disso, outros estudos relatam que essa correlação tende até mesmo ser negativa (Garnett *et al.*, 2009), devido ao fato que o crescimento vigoroso do sistema radicular tem o custo de competir pelo carbono a ser destinado para a parte aérea, o qual seria destinado a produção de grãos.

Neste contexto, ressalta-se que a característica secundária ideal é aquela que seja correlacionada geneticamente com a característica principal (EUN), apresente alta herdabilidade e variabilidade genética, seja fácil e rápida de avaliar, estável durante o período de medição e que possa ser observada antes do florescimento (O'Neill *et al.*, 2006).

### 3.7 Conclusões

- Existe a possibilidade de seleção de linhagens EUN para ambiente com restrição de N e também obter resultados satisfatórios para ambientes em condições ideais de N.
- As características de produtividade não possuem comportamento diferencial frente as variações de disponibilidades de N.
- Características de raiz, reprodutivas ou fisiológicas não são atrativas para seleção precoce de genótipos EUN.
- Os estádios V8 e VT não se mostram promissores para avaliação precoce da EUN, devendo essa ser praticada em função da produtividade de grãos.
- É recomendável o uso de índice de seleção incluindo NE e PE, afim de obter ganhos em prolificidade, produtividade e conseqüentemente em EUN.

### 3.8 Referências

- Abdel-Ghani, A.H.; Kumar, B.; Reyes-Matamoros, J.; Gonzalez- Portilla, P.J., Jansen, C.; San Martin, J.P.; Lee, M.; Lu¨bberstedt, 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**, v.189, p.123–133. doi:10.1007/s10681-012-0759-0
- Barbieri, P.A.; Echeverria, H.E.; Sainz Rozas, H.R.; Andrade, F.H. 2008. Nitrogen use efficiency in maize as affected by nitrogen availability and row spacing. **Agronomy Journal**, v.100, p.1094-1100
- Bertin, P.; Gallais, A. 2000. Physiological and genetic basis of nitrogen use efficiency in maize. I. Agrophysiological results. **Maydica**, v.45, p.53-66.
- Bicudo, S.J.; Cruz, S.C.S.; Pereira, F.R.S.; Brachtvogel, E.L.; Madalena, J.A.S. 2009. Níveis de adubação para a cultura do milho safrinha. **Caatinga**, v.22, p.23-30
- Borém A.; Galvão, J.C.C., Pimentel, M.A. 2015. **Milho: do Plantio à Colheita**. Viçosa. Editora UFV. 351p
- Caixeta, S.D.; Fritsche-Neto, R.; Granato, I.S.C; Oliveira, L.R.; Galvão, J.C.C. 2015. Early indirect selection for nitrogen use efficiency in maize. **Revista Ciência Agronômica**, v.46(2), p.369-378
- Cruz, C.D., Carneiro, P.C.S. 2006. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, v.2, 506p
- Cruz, C.D. 2013. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum Agronomy**, Maringá, v.35, n. 3, p. 271-276
- Companhia Nacional de Abastecimento. 2013. Acompanhamento de safra brasileira: **12º Levantamento - Safra 2014/15**. Companhia Nacional de Abastecimento. - Brasília: Conab, 29p
- Coque, M.; Gallais, A. 2006. Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. **Theoretical and Applied Genetics**, v.112, n.7, p.1205-1220
- Damatta, F. M.; Maestri, M.; Mosquim, P.R.; Barros, R.S. 1997. Photosynthesis in coffee (*Coffea arabica* and *C. canephora*) as affected by winter and summer conditions. **Plant Science**, v.128, n.01, p.43-50

- Dao, A.; Sanou, J.; Mitchell, S.E.; Gracen, V.; Danquah, E.Y. 2014. Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. **BMC Genetics**, p.115:127. doi:10.1186/s12863-014-0127-2
- Dawson, J.C.; Huggins, D.R.; Jones, S.S.; 2008. Characterizing nitrogen use efficiency in natural and agricultural ecosystems to improve the performance of cereal crops in low-input and organic agricultural systems. **Field Crops Research**, v.107, p.89-101
- Dewey, D.R.; Lu, K.H. 1959. A correlation and path analysis of components of crested wheatgrass seed production. **Agronomy Journal**, v.51, p.515-518.
- DoVale, J.C.; Fritsche-Neto, R.; Bermudez, F.P.; Miranda, G.V. 2012. Efeitos gênicos de caracteres associados à eficiência no uso de nitrogênio em milho. **Pesquisa Agropecuária Brasileira**, v.47, n.3, p.385-392
- DoVale, J.C.; Fritsche-Neto, R. 2013. Genetic control of traits associated with phosphorus use efficiency in maize by reml/blup. **Revista Ciência Agrônômica**, v44, n3, p.554-563
- Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA. Centro Nacional de Pesquisa de Milho e Sorgo. Manejo da Adubação Nitrogenada na Cultura do Milho. Sete Lagoas, 2007. 11p
- Fan, J.B.; Oliphant, A.; Shen, R.; et al. 2003. Highly parallel SNP genotyping. **Cold Spring Harbor Symposia on Quantitative Biology**, v.68, p.69-78
- Freitas Júnior, S. de P.; Amaral Júnior, A.T. do; Rangel, R.M.; Viana, A.P. 2009. Predição de ganhos genéticos na população de milho-pipoca UNB-2U sob seleção recorrente utilizando-se diferentes índices de seleção. **Semina: Ciências Agrárias**, v.30, p.803-814. doi: 10.5433/1679-0359.2009v30n4p803
- Freitas, I.L.J. al. 2013. Ganho genético avaliado com índices de seleção e com REML/Blup em milho-pipoca. **Pesquisa Agropecuária Brasileira**, v.48, n.11, p1464-1471. <http://dx.doi.org/10.1590/S0100-204X2013001100007>
- Fritsche-Neto, R; Borém, A. 2012. **Plant Breeding for Abiotic Stress Tolerance**. Springer. 176p. doi 10.1007/978-3-642-30553-5

- Fritsche-Neto, R.; Resende, M.D.V.; Miranda, G.V.; DoVale, J.C. 2012. Seleção genômica ampla e novos métodos de melhoramento do milho. **Ceres**, v.59, p.794-802
- Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, v.58, n.3, p.453-467
- Gallais, A.; Coque, M. 2005. Genetic variation and selection for nitrogen use efficiency in maize: a synthesis. **Maydica**, v.50, p.531-547
- Garnett, T.; Conn, V.; Kaiser, B.N. 2009. Root based approaches to improving nitrogen use efficiency in plants. **Plant, Cell & Environment**, v.32, p.1272-1283
- Giaveno, C. D.; Ribeiro, R. V.; Souza, G. M.; Oliveira, R. F. 2007. Screening of tropical maize for salt stress tolerance. **Crop Breeding and Applied Biotechnology**, v.7, p.304-313
- Gilmour, A.R.; Gogel, B.J.; Cullis, B.R.; Thompson, R. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK [www.vsni.co.uk](http://www.vsni.co.uk)
- Hayes, B.J., Visscher, P.M.; Goddard, M.E. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research (Camb)**, v.91, p.47-60
- Jacobs, B.J., Pearson, C.J. 1991. Potential yield of maize. determined by rates of growth and development of ears. **Field Crops Research**, Amsterdam, v.27, n.2, p.281-298
- Kang, M.S.; Din, A.K.; Zhang, Y.D.; Magari, R. 1999. Combining ability for rind puncture resistance in maize. **Crop Science**, v.39, p.368-371
- Lambers H.; Shane, M.W.; Cramer, M.D.; Pearse, S.J., Veneklaas, E.J. 2006. Root structure and functioning for efficient acquisition of phosphorus: Matching morphological and physiological traits. **Annal of Botany**, v.98, p.693-713. doi:10.1093/aob/mcl114
- Linkohr, B. I.; Williamson, L. C.; Fitteri, A. H.; Leyser, H. M. O. 2002. Nitrate and phosphate availability and distribution have different effects on root system architecture of Arabidopsis. **The Plant Journal**, v.29, n.6, p.751-760
- Lynch, J.P. 2013. Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. **Annal of Botany**, v.112, p.347-357. doi: 10.1093/aob/mcs293 PMID: 23328767

- McCully, M.E.; Canny, M.J. 1988. Pathways and processes of water and nutrient movements in roots. **Plant Soil**, v.111, p.159-170.
- Melo, A. S.; Suassuna, J. F.; Fernandes, P. D.; Brito, M. E. B.; Suassuna, A. F.; Aguiar Netto, A. O. Crescimento vegetativo, resistência estomática, eficiência fotossintética e rendimento do fruto da melancia em diferentes níveis de água. **Acta Scientiarum Agronomy**, Maringá, v.32, n. 1, p.73-79
- Moll, R.H.; Kamprath, E.L.; Jackson, A. 1982. Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. **Agronomy Journal**, v.74, 562-564
- Montgomery, D.C.; Peck, E.A. 1981. **Introduction to linear regression analysis**. 504p
- O'Neill, P.M.; Shanahan, J.F.; Schepers, J.S. 2006. Use of chlorophyll fluorescence assessments to differentiate corn hybrid response to variable water conditions. **Crop Science**, v.46, p.681-687
- Peng, Y.F.; Li, X.X.; Li, C.J. 2012. Temporal and spatial profiling of root growth revealed novel response of maize roots under various nitrogen supplies in the field. **Plos One**, v.7, p.1-11
- Pimentel-Gomes, F. 1990. **Curso de Estatística Experimental**. 13.ed. Piracicaba: Nobel, 468p
- Presterl, T.; Seitz, G.; Landbeck, M.; Thiemt, E.M.; Schmidt, W.; Geiger, H.H. 2003 Improving nitrogen use efficiency in European maize: estimation of quantitative genetic parameters. **Crop Science**, v.43, p.1259-1265
- R Core Team. R: **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing. 2013. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Rangel, R.M.; Amaral Júnior, A.T. do; Gonçalves, L.S.; Freitas Júnior, S. de P.; Candido, L.S. 2011. Análise biométrica de ganhos por seleção em população de milho-pipoca de quinto ciclo de seleção recorrente. **Revista Ciência Agrônômica**, v.42, p.473-481
- Richardson, A.E.; Barea, J.M.; McNeill, A.M.; Prigent-Combaret, C. 2009. Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. **Plant Soil**, v.321, p.305-339

- Rich, S.M.; Watt, M. 2013. Soil conditions and cereal root system architecture: review and considerations for linking Darwin and Weaver. **Journal of Experimental Botany**, v.64, p.1193-1208. doi:10.1093/jxb/ert043
- Silva, E.D.; Ferreira, S.M; Silva, G.P.; Assis, R.L.; Guimarães, G.L. 2005. Épocas e formas de aplicação de nitrogênio no milho sob plantio direto em solo de cerrado. **Revista Brasileira de Ciência do Solo**, v.29, p.725-733. <http://dx.doi.org/10.1590/S0100-06832005000500008>
- Silva, F.L.; Pedrozo, C.A.; Barbosa, M.H.P.; Resende, M.D.V.; Peternelli, L.A.; Costa, P.M.A.; Vieira, M.S. 2009. Análise de trilha para componentes de produção de cana-de-açúcar via BLUP. **Revista Ceres**, v.56, p.308-314
- Souza, L. V.; Miranda, G. V.; Galvão, J. C. C.; Eckert, F. R.; Mantovani, E. E.; Lima, R. O.; Guimarães, L. J. M. 2008. Genetic control of grain yield and nitrogen use efficiency in tropical maize. **Pesquisa Agropecuária Brasileira**. Brasília, v.43, n.11, p.1517-1523
- Soares, M.O.; Miranda, G.V.; Guimarães, L.J.M.; Marriel, I.E.; Guimarães, C.T. 2011. Parâmetros genéticos de uma população de milho em níveis contrastantes de nitrogênio. **Revista Ciência Agronômica**, v. 42, n.1, p.168-174
- Teixeira, F.F.; Gomide, R.L.; Albuquerque, P.E.P.; Andrade, C.L.T.; Leite, C.E.P.; Parentoni, S.N.; Guimarães, P.E.O.; Guimarães, L.J.M.; Silva, A.R.; Bastos, E. A.; Cardoso, M.J. 2010. Evaluation of maize core collection for drought tolerance. **Crop Breeding and Applied Biotechnology**, v.10, n.4, p.312-320
- Tester, M., Bacic, A. 2005. Abiotic stress tolerance in grasses. From model plants to crop plants. **Plant Physiology**, v.137, p.791-793
- Toaldo, D.; Pinho-Morais, P. P.; Battilana, J.; Coimbra, J. L. M.; Guidolin, A. F. 2013. Selection in early generations and the occurrence of heterosis for the character root distribution. **Euphytica**, v.190, p.335-344
- Trachsel, S.; Messmer, R.; Stamp, P.; Hund, A. 2009. Mapping of QTLs for lateral and axile root growth of tropical maize. **Theoretical and Applied Genetics**, v.119, p. 1413-1424
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v.91, p.4414-4423

- Vessey, J. K.; Raper, C. D.; Henry, L. T. 1990. Cyclic variations in nitrogen uptake rate in soybean plants: uptake during reproductive growth. **Journal of Experimental Botany**, v.41, p.1579-1584
- Waisel, Y.; Eshel, A. 2002. Functional diversity of various constituents of a single root system. In: Waisel, Y.; Eshel, A.; Kafkafi, U. **Plant Roots: The Hidden Half (3rd ed)**, CRC Press. pp. 243-268
- Worku, M.; Bänziger, M.; Schulte Erley, G.; Friesen, D.; Diallo, A.O.; Horst, W.J. 2007. Nitrogen uptake and utilization in contrasting nitrogen efficient tropical maize hybrids. **Crop Science**, v.47, p.519-528
- Wright, S. 1921. Correlation and causation. **Journal of Agricultural Research**, v.20, p.557-585
- Yan, W.; N.A. Tinker. 2005. An integrated system of biplot analysis for displaying, interpreting, and exploring genotypeby-environment interactions. **Crop Science**, v.45, p.1004-1016

## **4. CAPÍTULO 2**

**Uso de bancos públicos para predição genômica de linhagens de milho tropical**

#### 4.1 Resumo

A seleção genômica (GS) tem-se mostrado capaz de incrementar os ganhos de seleção por unidade de tempo, sendo também um processo dinâmico e com custos decrescentes. Todavia, o custo para formação inicial da população de treinamento (PT) ainda é alto. Isso pode tornar a GS inacessível para programas de pequeno porte com limitado orçamento e carência de grande variabilidade genética. Para superar essa etapa, o uso de informações de bancos públicos de dados pode ser uma das soluções. Nesse contexto, foi proposto um panorama hipotético: uso de orçamento limitado e necessidade de acréscimo de variabilidade genética. Com isso, para implementar a GS, os objetivos foram: *i*) testar a viabilidade do uso das informações genômicas e fenotípicas de bancos de dados públicos para predição de linhagens de milho tropical; *ii*) identificar se a estrutura populacional, o uso de populações de treinamento otimizadas (PTO) e a quantidade de informação oriundas de bancos de dados públicos influenciam na capacidade preditiva de linhagens de milho tropical. Para isso, foram definidos 29 conjuntos de população de treinamento através de painéis de diversidade: USP (painel a ser predito), ASSO e NCRPIS (painéis públicos externos - preditores); divididos em quatro cenários com diferentes configurações de PT. Foi utilizado nas análises 28.260 marcadores SNP em 2748 linhagens. Essas foram preditas via modelo G-BLUP para características de altura de planta e de espiga. Por meio desse estudo verificou-se que é possível o uso de informações de bancos públicos para formação de PT. E que a estrutura populacional tem capacidade de modificar significativamente as capacidades preditivas da GS, principalmente quando PT e de validação (PV) tem baixo relacionamento genético. Nos quatro cenários propostos, evidenciou-se que conjuntos demasiadamente pequenos ou grandes não proporcionam as melhores capacidades preditivas para GS. No entanto, conjuntos de PTO compostos por 250 indivíduos, oriundos de bancos públicos, são suficientes para obter estimativas satisfatórias (0,32). Contudo, a adição de indivíduos da população de validação nas PTO é o melhor cenário encontrado, obtendo estimativas de capacidade preditivas de até 0,59. Sendo esse procedimento, sempre que possível, o mais recomendado.

**Palavras-chave:** capacidade preditiva, estrutura de população, otimização de população de treinamento, painéis de diversidade, G-BLUP.

## 4.2 Abstract

Genomic selection (GS) has shown to be able to step up selection gains per unit time, also it has being a dynamic and decreasing cost process. However, the initial cost of training set (PT) establishment may still be high, taking into account large number of individuals genotyped and phenotyped, and also large number of markers. That can make GS inaccessible for small programs with limited budget and lack of genetic variability; to overcome it, the use of information from public databases might be one of the solutions. Therefore, a hypothetical scenario was proposed in which there is a limited budget and the need for genetic variability. Thus, the aims of this study were: *i*) test the feasibility of genomic and phenotypic information from public databases for predicting tropical maize inbred lines; *ii*) identify how the population structure, the use of optimized training set (PTO) and the amount of information coming from these databases influence the predictive ability of tropical maize inbred lines. For that, 29 training sets were developed through diversity panels: USP (panel to be predicted), ASSO and NCRPIS (external public panels - predictors); divided into four scenarios with different settings of training sets. In the analysis 28.260 SNP markers for 2748 inbred lines were used, via model G-BLUP for plant height and ear height. Through this study it was found that it is possible to use information from public banks for PT formation. The population structure has the capacity for modifying GS predictive abilities significantly, especially when training set and validation (PV) hold low genetic relationship. In the four proposed scenarios, it was shown that too small or large sets do not provide best predictive abilities. But, PTO with 250 individuals, from public dataset banks, is enough to reach estimates up to 0.32. However, the addition of PV information into PTO is the best scenario, it got predictive ability estimates up to 0.59. It should be used whenever possible.

**Keywords:** predictive ability, population structure, optimized training set, diversity panels, G-BLUP.

### 4.3 Introdução

Proposta no início do século 21 por Meuwissen *et al.* (2001), a seleção genômica (GS) surgiu como um método estatístico para melhorar a predição de características quantitativas. Tal método utiliza informações de marcadores espalhados ao longo de todo o genoma da espécie em estudo para prever o desempenho dos materiais em questão. A GS tem a capacidade de aumentar a taxa de ganho de seleção por unidade de tempo e as acurácias dos modelos de predição (Meuwissen *et al.*, 2001; Wong e Bernardo, 2008). Para isso, a GS baseia-se em duas fases, a de treinamento e a de seleção (Jonas e De Koning, 2013). Na fase de treinamento são estimados os efeitos dos marcadores. Assim, são utilizadas duas populações nas quais os indivíduos são fenotipados e genotipados, a população de treinamento (PT) e a de validação (PV). Na primeira são feitas as estimativas dos efeitos dos marcadores através de modelos de predição, já na segunda estes efeitos são testados com o intuito de se verificar as acurácias das equações de predições utilizadas (Newell e Jannink, 2014). Na fase de seleção, são feitas as predições dos valores genéticos genômicos (GEBVs) dos materiais, a partir dos efeitos estimados dos marcadores na fase de treinamento (Meuwissen *et al.*, 2001). Nessa população, é feita apenas a genotipagem dos indivíduos.

Um fator importante no processo de GS é em função da composição da PT. A escolha de indivíduos a serem incluídos na PT, é uma das variáveis mais difíceis para otimizar-se, mas é crucial para atingir elevadas acurácias (Isidro *et al.*, 2015). O seu tamanho, e seu relacionamento com a PS são elementos essenciais para determinar a precisão da predição via GS (Bassi *et al.*, 2016). Todavia, o grande número de indivíduos a serem genotipados e fenotipados, para compor uma PT confiável e representativa da PS, demanda tempo e custo na sua formação, especialmente devido a fenotipagem. Esse tem sido o grande gargalo e fator de restrição das análises genéticas, sendo ainda um processo manual, de trabalho intensivo e dispendioso (Cobb *et al.*, 2013).

Entretanto, quando há restrições de tamanho e variabilidade genética na população de treinamento, ocorre redução na capacidade de predição dos modelos usados (Estaghvirou *et al.*, 2015). Com isso, é necessário incluir maior ou adequado número de indivíduos na PT, de forma a viabilizar uso da GS em relação a seleção fenotípica. Nesse contexto, quando não há possibilidade de

conseguir esse aumento de variabilidade via o próprio programa de melhoramento, o uso de informações genômicas e fenotípicas providas de bancos de dados públicos é uma solução a ser explorada. Isto poderá aumentar a viabilidade do emprego da GS em programas de melhoramento menores, com menor aporte financeiro e, ou em início de atividades, e com restrição de variabilidade genética para características de interesse.

Assim, pensando na aplicabilidade da GS em programas de melhoramento público ou privado de pequeno porte e, ou então, iniciando suas atividades, criamos um panorama hipotético em que é comum justamente o uso de orçamentos limitados e necessidade de incremento em variabilidade genética para formar população base da GS. Neste cenário, a aplicação da GS se dá via o uso de PT maiores e com boa variabilidade genética providas de bancos públicos de informações genômicas e fenotípicas de milho. Gerando as predições dos indivíduos, já previamente genotipados, no programa de interesse.

Grandes painéis de diversidade de cunho público, os quais representam uma ampla amostragem de linhagens e grupos heteróticos da cultura do milho, podem ser acessados livremente. Como exemplo o Nested Association Mapping (NAM) (McMullen *et al.*, 2009; Hung *et al.*, 2012), Maize 282 Association Panel (Flint-Garcia *et al.*, 2005) e a coleção da USDA Ames - North Central Regional Plant Introduction Station (NCRPIS) (Romay *et al.*, 2013).

Contudo, se a PT e a população a ser testada e validada (PV) apresentam diferentes origens, a relação genética entre essas é um fator importante que afeta a precisão da GS (Habier *et al.*, 2007; Hayes *et al.*, 2009; Albrecht *et al.*, 2011; Clark *et al.*, 2011; Pszczola *et al.*, 2012). Adicionalmente, fatores como a herdabilidade da característica (Heffner *et al.*, 2009) e o tamanho da população de treinamento (Jannink *et al.*, 2010), também são capazes de modificar a acurácia da GS.

Por outro lado, recentemente, com o desenvolvimento de algoritmos que auxiliam na escolha de indivíduos para a formação de populações de treinamento otimizadas (PTO) tem atraído atenção do melhoramento animal e de plantas (Rincent *et al.*, 2012). Isso por que o aumento da capacidade preditiva dos modelos testados é um ponto chave para alocar eficientemente os recursos de um programa de melhoramento e ter o melhor balanço entre variabilidade

genética, custo e acurácia. Além disso, esse procedimento tem apresentado bons resultados quanto ao aumento da acurácia na GS em estudos recentes (Isidro *et al.*, 2015, Akdemir *et al.*, 2015), o que a torna interessante sua aplicação no panorama aqui proposto.

#### **4.4 Objetivos**

Partindo do que foi exposto, os objetivos desse estudo foram: *i*) testar a viabilidade do uso das informações genômicas e fenotípicas de bancos de dados públicos para predição de linhagens de milho tropical; *ii*) identificar se a estrutura populacional, o uso de populações de treinamento otimizadas e a quantidade de informação oriundas de bancos de dados públicos influenciam na capacidade preditiva de linhagens de milho tropical.

#### **4.5 Material e métodos**

##### *4.5.1 Conjunto genotípico e experimentos a campo*

Neste estudo foram utilizados três painéis compostos por linhagens de milho: *i*) Universidade de São Paulo (USP), ao qual possui o conjunto de linhagens de interesse a serem preditas; *ii*) Nested Association Mapping (NAM) em conjunto com o Maize 282 Association Panel. Sendo denominado para esse estudo como painel Associativo (ASSO); *iii*) Department of Agriculture - Agricultural Research Service (USDA-ARS) North Central Regional Plant Introduction Station (NCRPIS).

O conjunto de linhagens pertencente à USP foi formado por 64 linhagens tropicais, conduzidas em experimentos delineados em látice simples (8 x 8), com duas repetições, seguindo o mesmo sorteio experimental nas safrinhas de 2014 e 2015. Esses foram alocados na área experimental do Departamento de Genética (22°42'23"S, 47°38'14"W, 546 m, Nitossolo Vermelho Eutroférico típico) na safrinha 2015, Piracicaba-SP, Brasil. E na Estação Experimental de Anhumas na cidade de Anhembi-SP, Brasil (22°84'30"S; 48°02'20"W, 450 m, Latossolo Amarelo Distrófico), nas safrinhas de 2014 e 2015. As parcelas foram constituídas de uma linha com 4 m de comprimento (safrinha 2014) e de 5 m (safrinha 2015), essa diferença ocorreu devido a necessidade de maior número de plantas para avaliação de características radiculares (não sendo objeto dessa

pesquisa); em ambos os casos utilizou-se espaçamento de 0,85 m entre linhas e 0,20 m entre plantas, totalizando 20 e 25 plantas por linha respectivamente. O manejo e os tratos culturais foram realizados de acordo com os procedimentos recomendados para a cultura do milho (Borém *et al.*, 2015).

Os conjuntos NAM, Maize 282 Association Panel foram avaliados em 10 ambientes no Estados Unidos (USA), sob sistema convencional de cultivo. Os ambientes foram: Aurora - NY, Clayton - NC, Columbia - MO e Urbana - IL, nas safras de 2006 e 2007; em Aurora - NY na safra de 2008; e na safra de 2009 em Columbia - MO.

O conjunto NCRPIS foi avaliado em 2010 em parcelas de uma única linha em três ambientes nos USA: Aurora - NY; Clayton - NC e Columbia - MO. Em Aurora as linhagens foram plantadas em parcelas de 12 plantas na Muskgrave Research Station. Já em Clayton o experimento foi instalado na Central Crops Research Station e em Columbia na South Farm, com 15 plantas por parcela. Nesses experimentos, o conjunto NCRPIS foi estratificado em nove grupos de maturação, e as linhagens foram aleatoriamente designadas em blocos incompletos de 19 linhagens, tendo como testemunhas as linhagens B73, IL14H, KI111, P39, SA24 e TX303.

#### 4.5.2 Características avaliadas

Nos três painéis as características avaliadas foram a altura de planta (AP) e altura de espiga (AE). AP foi medida em centímetros em relação a distância a partir da linha do solo até a base da folha bandeira após estágio de pendoamento (VT); AE foi mensurada em centímetros a partir do solo até o nó primário da espiga, no mesmo estágio de avaliação de AP.

#### 4.5.3 Valores genéticos preditos (BLUP)

Os BLUPs oriundos dos experimentos acima citados para os conjuntos NAM, Maize 282 e NCRPIS foram obtidos por Peiffer *et al.* (2014) (disponível em [www.panzea.org](http://www.panzea.org)).

Para a obtenção dos componentes de variância e dos BLUPs das linhagens USP, os dados foram submetidos à análise de Máxima Verossimilhança Restrita (REML) e pelo Melhor Preditor Genético Linear Não-Viesado (BLUP), conforme o seguinte modelo:

$$y = Jb + Zg + Vs + Tf + \varepsilon \quad (1)$$

em que,  $y$  é o vetor de observações fenotípicas das linhagens para as características avaliadas;  $b$  é o vetor do efeito do bloco dentro da repetição, considerado como fixo;  $g$  é vetor do efeito de genótipo (linhagens), considerado como aleatório, em que  $g \sim N(0, G)$ , sendo  $G = I\sigma_g^2$  e  $\sigma_g^2$  a variância genética;  $s$  é o vetor do efeito de ambiente, considerado como fixo;  $f$  é o vetor do efeito da interação genótipo x ambiente, considerado como aleatório, onde  $f \sim N(0, F)$ , sendo  $F = I\sigma_f^2$ ;  $\varepsilon$  é o vetor de erros, onde  $\varepsilon \sim N(0, \mathcal{E})$ , sendo  $\mathcal{E} = I\sigma_\varepsilon^2$ .  $I$  é a matriz identidade e  $J, Z, V, T$  são matrizes de incidência que relacionam os efeitos dos vetores independentes de cada matriz ao vetor dependente  $y$ .

Para os testes de significância dos efeitos aleatórios dos genótipos foi empregado o teste de razão de verossimilhança (*Likelihood Test Ratio* - LTR):

$$LTR = 2.[\log L_2 - \log L_1] \sim \chi_{r_2 - r_1}^2 \quad (2)$$

em que,  $H_0: \sigma^2 = 0$  e  $H_1: \sigma^2 \neq 0$ ;  $\log L_2$ : estimativa *Likelihood* do modelo sem o efeito aleatório de interesse;  $\log L_1$ : estimativa *Likelihood* do modelo com o efeito aleatório de interesse.

As análises das equações de modelos mistos foram realizadas por meio do pacote ASReml (Gilmour *et al.*, 2009), no software R (R Core Team, 2013).

A herdabilidade no sentido amplo foi estimada para cada característica por:

$$\hat{h}^2 = \frac{\sigma_g^2}{\sigma_f^2} \quad (3)$$

em que,  $\sigma_g^2$ : variância genética;  $\sigma_f^2$ : variância fenotípica.

#### 4.5.4 Genotipagem e controle de qualidade

Para cada uma das amostras do painel USP foi efetuada a extração do DNA genômico. Posteriormente as linhagens foram genotipadas utilizando a plataforma Affymetrix® Axiom® Maize Genotyping Array (Unterseer *et al.*, 2014) com aproximadamente 616 mil marcadores do tipo de Polimorfismo de um Único Nucleotídeo (Single-nucleotide Polymorphism - SNP). As informações de genotipagem para os painéis ASSO e NCRPIS foram obtidas por meio de Genotyping by Sequencing (GBS) - ZeaGBSv2.7 (Glaubitz *et al.*, 2014). Nesse conjunto de dados, estão contidos, incluindo os dois painéis citados, 17.280 mil

linhagens públicas, genotipadas com 955.690 SNPs (disponível em [www.panzea.org](http://www.panzea.org) em versão parcialmente imputada em arquivo HDF5).

O controle de qualidade para o conjunto de linhagens levou em consideração a eliminação de genótipos com dados perdidos e/ou heterogeneidade  $\geq 0,10$ , afim de manter linhagens com grande quantidade de informação genômica e homocigotas. Para realizar o controle de qualidade dos marcadores, foi empregado *Call Rate*  $\geq 0,90$ , que consiste em eliminar SNPs com grande quantidade de valores perdidos e a *Minor Allele Frequency* ( $MAF \leq 0,05$ ), que está relacionada com o polimorfismo dos loci marcadores na população.. Desse modo, esse processo de controle de qualidade dos dados genômicos resultou ao final em um total de 63 linhagens selecionadas e 409 mil SNPs no Painel USP; e no conjunto de dados GBS, o processo resultou em e 12.149 linhagens para 359 mil marcadores

Após a limpeza dos dados foram feitos dois processos de pareamento: *i*) de informações do conjunto de marcadores obtidos via GBS e via Affymetrix, considerando para isso informações do número de referência do cromossomo e de posição física exata do marcador (bp); *ii*) do conjunto de linhagens genotipadas via GBS com as linhagens do conjunto de dados de Peiffer *et al.* (2014) (NCRPIS = 2815, NAM = 4982, Maize 282 = 282). Como consequência, foram encontrados 28.260 SNPs em comum entre os dois conjuntos de genotipagem e 2685 linhagens pareadas. Dessas, 2237 linhagens pertencentes ao Painel NCRPIS, 448 ao ASSO (NAM=166, Maize 282= 282). Com isso, o conjunto de dados final quando adicionado às 63 linhagens do painel USP ficou em 2748 indivíduos, genotipados para os mesmos 28.260 marcadores SNP.

Todo o processo foi realizado utilizando TASSEL Quality Control (Bradbury *et al.*, 2007) e o software R via scripts próprios.

#### 4.5.5 Parentesco entre linhagens e estrutura populacional

A matriz de relação genética foi estimada utilizando o modelo descrito por VanRaden (2008). Em seguida foi procedido a análise de agrupamento via Componentes Principais - PCA, definindo a *a priori*  $k=5$ . Essa definição segue os dados publicados por Romay *et al.* (2013), em que traz a definição dos agrupamentos das linhagens em NCRPIS, sendo assim adotado como critério para estruturação populacional do conjunto genotípico total aqui estudado. Para ambas as análises foram considerados o arranjo de 28.260 SNP e 2748

linhagens. Dessa forma, definiu-se os agrupamentos *tropical*, *popcorn*, *non-stiffstalk*, *stiffstalk* e *sweetcorn*. Porém, o agrupamento *popcorn* foi retirado para evitar sua entrada na formação das populações de treinamento e nas análises de predição, pois este possui características agronômicas e aptidão agrícola (Franzoni *et al.*, 2012) que diferem dos almejados em milho para produção de grãos.

#### 4.5.6 População de treinamento (PT) e de validação (PV)

Foram definidos 29 conjuntos de população de treinamento (CPT) divididos em 4 cenários (Tabela 1).

Tabela 1: Conjuntos de população de treinamento (PT) e validação (PV) e seus respectivos tamanhos,  $N_t$  e  $N_v$ , utilizados para os cenários CPT1, CPT2, CPT3 e CPT4

Conjunto	PT	$N_t$	PV	$N_v$
CPT1	USP	10	USP	53
		20		43
		30		33
CPT2	NCRPIS + ASSO + USP	2465	USP	32
	NCRPIS + ASSO	2434		32
	NCRPIS	2046		32
	ASSO	388		32
	NCRPIS (agrupamento USP)	512		32
	ASSO (agrupamento USP)	136		32
CPT3	PTO1 - PTR1	50	USP	32
	PTO2 - PTR2	250		32
	PTO3 - PTR3	500		32
	PTO4 - PTR4	1000		32
	PTO5 - PTR5	1500		32
CPT4	PTO6 - PTR6	81	USP	32
	PTO7 - PTR7	281		32
	PTO8 - PTR8	531		32
	PTO9 - PTR9	1031		32
	PTO10 - PTR10	1531		32

USP: painel da Universidade de São Paulo; NCRPIS: painel US Department of Agriculture–Agricultural Research Service (USDA–ARS) North Central Regional Plant Introduction Station; ASSO: painel Associação; PTO: população de treinamento otimizada; PTR: população de treinamento aleatória

- **CPT1** – Considerou apenas o painel USP para a formação da PT. Esse cenário visa inferir o comportamento da capacidade preditiva quando analisada dentro do próprio painel sem adição de genótipos ou painéis de diversidade externos. Assim, a PT foi determinada em função de três conjuntos de treinamento ( $k=3$ ), cada um considerando diferentes proporções de PT e PV dentro do painel USP. Logo, no primeiro, segundo e terceiro conjunto, a PT possui 10, 20, 30 indivíduos alocados, respectivamente; e a PV possui 53, 43 e 33 indivíduos a serem validados. A designação das linhagens incluídas na PT foi

definido aleatoriamente, incorporando posteriormente as linhagens remanescentes na PV.

- **CPT2** – Foi usado os painéis USP, NCRPIS e ASSO para a formação de PT e utilizou-se informação de estrutura populacional para isso. Nesse cenário e nos demais descritos adiantes, foi considerado uma situação comumente encontrada no melhoramento de plantas, em que o ambiente usado na fenotipagem do conjunto de treinamento difere daqueles usados pelo conjunto de validação. Além disso, as linhagens a serem preditas possuem baixo parentesco com o conjunto de treinamento (Windhausen *et al.*, 2012), assim como a existência de estrutura populacional. Para isso, foram criados seis diferentes conjuntos de PT considerando: *i*) painel individual (NCRPIS ou ASSO); *ii*) junção de painéis (NCRPIS + ASSO + USP ou NCRPIS + ASSO); *iii*) seleção de linhagens pertencentes aos painéis NCRPIS ou ASSO que foram alocadas para o mesmo agrupamento das linhagens USP via PCA, de modo que essas linhagens selecionadas tenham maior relação genética com o painel USP (PV).

Especificamente, em um dos conjuntos formados em que há participação do painel USP na formação da PT (NCRPIS + ASSO + USP). As linhagens remanescentes desse painel, que não foram usadas no conjunto de validação, foram então realocadas para compor a PT. Como exemplo, a PV em CPT2 é composto por 32 linhagens, tomadas aleatoriamente no painel USP, as demais linhagens (31) foram então incorporadas a PT juntamente com as linhagens dos painéis NCRPIS e ASSO.

- **CPT3** – Para a determinação PT nesse cenário aplicou-se o método de otimização de população de treinamento (PTO) proposto por Akdemir *et al.* (2015), com tamanho de população predefinidos. Nesse metodologia a seleção de linhagens exige-se apenas informações genotípica sobre os indivíduos presentes em um conjunto de candidatos (NCRPIS + ASSO) e no conjunto a ser testado (USP). Posteriormente, com base nessas informações, um algoritmo genético faz uma aproximação da variância do erro (PEV) a partir do uso de componentes principais via matriz de marcadores, e seleciona determinadas linhagens que formarão a PTO. Foram determinados nesse cenário cinco conjuntos com diferentes tamanhos de PTO (50, 250, 500, 1000 e 1500). O algoritmo para formação da PTO foi implementado via pacote STPGA (Akdemir, 2014) no software R.

Para poder inferir e comparar sobre a eficácia da metodologia proposta, foi definido um segundo conjunto de treinamento determinado aleatoriamente (PTR), e com o mesmo tamanho dos conjuntos nas PTO. Assim, as PTR com 50, 250, 500, 1000 e 1500 indivíduos foram formadas via amostragem aleatória dentro do conjunto de candidatos, sem critério de proporcionalidade. Já a PV em CPT3 foi composta por 32 linhagens tomadas aleatoriamente do painel USP.

- **CPT4** – Nesse cenário acrescentou-se ao CPT3 a informação adicional do painel USP para formação da PT. A inclusão de linhagens USP na PT visou simular uma condição na qual o programa de melhoramento fenotipa parte das linhagens do conjunto total (previamente genotipados) e as inclui aos dados externos para predição das demais linhagens. Portanto, foram formados aqui 10 conjuntos de tamanhos diferentes de PT, sendo uma via PTO e outra via PTR. Nesses foi adicionado as linhagens remanescentes do painel USP (N=31), que não foram designadas como PV.

#### 4.5.7 Predição genômica

Em todos os cenários acima, o desempenho das linhagens do painel USP para AP e AE foi predito via GBLUP (Hayes *et al.*, 2009):

$$y = Xu + Zg + \varepsilon \quad (4)$$

em que  $y$  é o vetor de valores genotípicos obtidos via REML/BLUP,  $u$  é o vetor de efeito fixo (média da população),  $g$  é vetor de efeito aleatório dos valores genômicos e  $\varepsilon$  é o vetor de resíduos. A variância dos efeitos aleatórios de  $g$  é  $var(g) = K\sigma_g^2$ , sendo que  $K$  é a matriz de relações genômicas (parentesco) e  $\sigma_g^2$  representa a variação genética. A variância de resíduos é dada por  $var(\varepsilon) = I\sigma_\varepsilon^2$ .  $Z$  é uma matriz de incidência que relacionam os efeitos do vetor independente ao vetor dependente  $y$ ,  $X$  é uma matriz coluna  $n \times 1$  e  $I$  a matriz identidade. Para análise do modelo GBLUP, foi utilizado o pacote rrBLUP (Endelman, 2011) do R.

A capacidade preditiva ( $r_{(\hat{y},g)}$ ) foi obtida por meio do coeficiente de correlação de Pearson ( $r$ ) entre valor genético predito e os GEBV na população de validação de interesse. Para cada um dos diferentes cenários testados, foram feitas 50 repetições, gerando assim 50 valores de capacidade preditivas. Esses,

posteriormente, foram transformados em variáveis, submetidos a análise de variância conforme o modelo a seguir:

$$y_{ij} = \mu + T_i + \varepsilon_{ij} \quad (5) \qquad y_{ij} = \mu + T_i + Ti(N) + \varepsilon_{ij} \quad (6)$$

em que,  $j = 1, \dots, r$ ;  $i = 1, \dots, k$ ;  $y_{ij}$  =  $j$ -ésima observação do nível  $i$  do fator tratamento;  $\mu$  = média geral ;  $T_i$  = efeito de tratamento (população de treinamento);  $N$  = número de indivíduos dentro da PT  $\varepsilon_{ij}$  = resíduo. O modelo via equação (5) é referente aos cenários CPT1 e CPT2. Já a equação (6) designa aos cenários CPT3 e CPT4. A partir desses dados também foram estimados os desvios padrões ( $\sigma$ ) e as médias comparadas pelo teste Tukey com  $\alpha = 0,05$ . Todos os procedimentos foram feitos via software R.

## 4.6 Resultados e discussão

### 4.6.1 Componentes de variância e herdabilidades

Via teste LTR (Tabela 2), foi verificado que as linhagens do painel USP possuem variabilidade, o que indica a possibilidade de ganho com a seleção para as características avaliadas. As herdabilidades foram consideradas de moderadas a altas, (0,60 a 0,94), para os três painéis (USP, ASSO e NCRPIS). Esses resultados são interessantes, pois sabe-se que a capacidade preditiva e a acurácia são influenciadas pela herdabilidade, existindo uma tendência claramente positiva entre essas, com resposta significativa na seleção de características com alta herdabilidade (Moranty *et al.*, 2015; Cavalcanti *et al.*, 2012).

Tabela 2 - Teste de máxima verossimilhança (LTR) para os efeito aleatório de linhagem e herdabilidades para as características altura de planta (AP) e altura de espiga (AE)

	FV	Características	
	-	AP	AE
USP	Linhagem	126,84***	78,74***
USP		0,71	0,60
ASSO <sup>1</sup>	h <sup>2</sup>	0,93	0,94
NCRPIS <sup>1</sup>		0,87	0,86

<sup>1</sup>herdabilidades obtidas em Peiffer *et al.* (2014); USP: painel da Universidade de São Paulo; ASSO: Maize 282 Association Panel em conjunto com Nested Association Mapping; NCRPIS: painel US Department of Agriculture–Agricultural Research Service (USDA–ARS) North Central Regional Plant Introduction Station

#### 4.6.2 Estrutura populacional

A análise PCA revelou estrutura compatível com os dados acessados e pareados providos de Romay et al. (2013). Assim, a análise de agrupamento se comporta de forma que todos os conjuntos podem ser separados nos dois primeiros eixos de componentes principais (PC), apresentando estimativas de 4,9% e 3,3% da variância genética, respectivamente (Figura 1). Tais resultados são condizentes com estudos prévios para a cultura do milho (Guo et al., 2014).

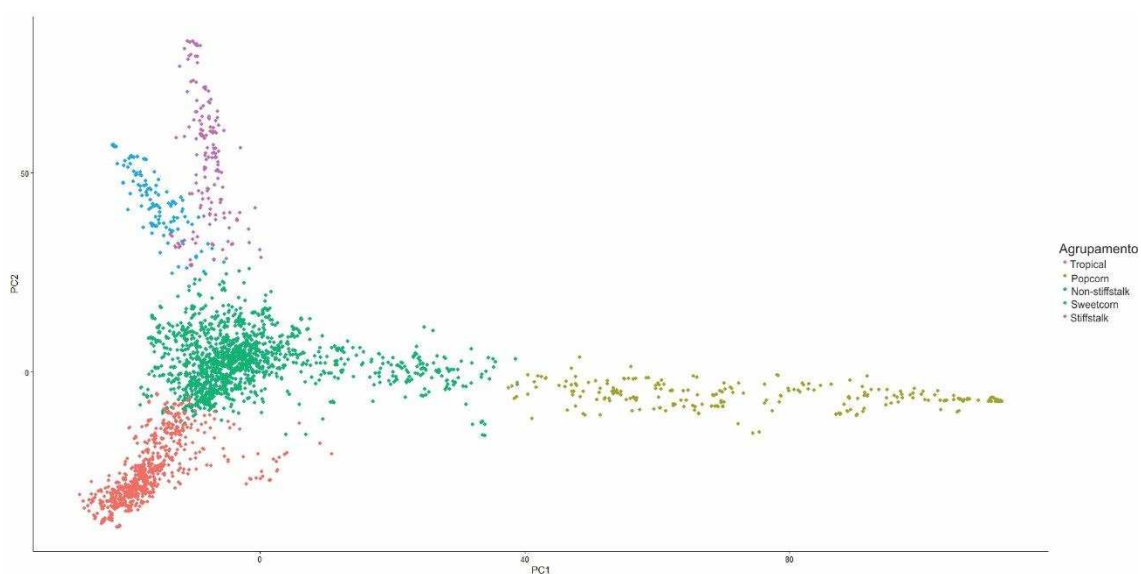


Figura 1 - Estruturação populacional via Análise de Componentes Principais (PCA), com k=5 agrupamentos, utilizando 28.260 marcadores SNP e 2748 linhagens de milho. Cada círculo sólido representa uma linhagem

Considerando os painéis USP, ASSO e NCRPIS, o número de linhagens alocadas nos agrupamentos obtidos seguiu a seguinte distribuição: i) *Tropical*: USP=63, ASSO=136 e NCRPIS=512; *Popcorn*: ASSO=60 e NCRPIS=91; *Non-stiffstalk*: ASSO=207 e NCRPIS=1272; *Stiffstalk*: ASSO=15 e NCRPIS=150; *Sweetcorn*: ASSO=30 e NCRPIS=112. Como descrito anteriormente, as 151 linhagens pertencentes o agrupamento *popcorn* não entraram na formação das populações de treinamento nos diferentes cenários propostos.

#### 4.6.3 Métodos de formação da população de treinamento e capacidades preditivas

#### 4.6.3.1 CPT1

Para esse primeiro cenário, as capacidades preditivas alcançaram valores de 0,16 a 0,20 para AP e de 0,02 a 0,05 para AE (Tabela 3). Portanto, apesar desse painel apresentar variabilidade genética entre as linhagens para as características consideradas, o número de indivíduos incluídos na PT indica ser um fator limitante em termos de uso para GS. Assim, previsões dentro de conjuntos muito pequenos limitam a capacidade preditiva do modelo, não alterando, de forma significativa, suas estimativas mesmo em função do aumento do tamanho da população de treinamento ( $N_t$ ). Adicionalmente, foi verificado que o coeficiente de parentesco médio ( $r_{xy}$ ), via matriz de parentesco genômico, foi de 0,19 (Figura Suplementar 1) nesse painel, o que é relativamente alto, pois é próximo ao parentesco médio de meio-irmãos. Com isto, nota-se a necessidade de inclusão de maior variabilidade genética para efetivação da GS.

Tabela 3 - Média e desvio-padrão ( $\sigma$ ) da capacidade preditiva ( $r_{(\hat{y},g)}$ ) das previsões genômicas obtidas em CPT1 para diferentes tamanhos de população de treinamento ( $N_t$ ) e validação ( $N_v$ ) nas características de altura de planta (AP) e altura de espiga (AE)

Características	$N_t$	$N_v$	$r_{(\hat{y},g)}^1$	$\sigma$
AP	10	53	0.16 <sup>a</sup>	± 0.20
	20	43	0.20 <sup>a</sup>	± 0.13
	30	33	0.22 <sup>a</sup>	± 0.15
AE	10	53	0.02 <sup>a</sup>	± 0.11
	20	43	0.03 <sup>a</sup>	± 0.09
	30	33	0.05 <sup>a</sup>	± 0.10

<sup>1</sup> Médias seguidas pela mesma letra, na coluna não diferem entre si pelo teste de Tukey a 5%

#### 4.6.3.2 CPT2

A partir desse cenário, no qual são utilizados painéis diferentes para formação da PT e PV, as capacidades preditivas obtidas indicam que a adição de linhagens USP dentro das PT resulta em incrementos nessa estimativa, para ambas as características avaliadas. No entanto, esse aumento para altura de planta não foi significativo (Tabela 4). Outro fator observado é que, de forma geral, os painéis NCRPIS e ASSO predizem igualmente as linhagens do painel USP, independente se estão ou não combinados. A exceção foi ASSO para AP, que proporcionou uma capacidade preditiva média inferior a 0,1. Analisando a inferência anterior e tomando em consideração a diferença no tamanho da população de treinamento entre os painéis, NCRPIS = 2046 e ASSO=388,

percebe-se que mesmo com um menor tamanho de população de treinamento, o painel ASSO é capaz de prever semelhante ao NCRPIS. Isto indicaria maior grau de relacionamento das linhagens do painel ASSO com as linhagens do painel USP. Todavia, os dois painéis possuem  $r_{xy}$  de 0,02 e 0,017 para NCRPIS e ASSO, respectivamente, em relação ao painel USP, o que é muito baixo em termos de parentesco genético.

Um terceiro ponto verificado é a clara formação e separação do agrupamento *tropical* dentro dos painéis NCRPIS e ASSO (Figura 2 - A1 e B1). Separando-se esse agrupamento e relacionando-o com o painel USP (Figura 2 - A2 e B2; Figura Suplementar 2), se obtém valores de  $r_{xy}$  maiores que aqueles observados quando se considerou todos os indivíduos dentro de cada painel, sendo 0,071 para NCRPIS e 0,08 em ASSO, e com valores máximos de 0,263 e 0,184 para cada painel, respectivamente. Nesse contexto, o grupo *tropical* do painel ASSO propicia maiores capacidades preditivas quando utilizado como conjunto de treinamento do que o painel ASSO completo (Tabela 4). Tal resultado não foi observado para o painel NCRPIS. Neste caso, observou-se a tendência de queda da capacidade preditiva quando foi empregado o grupo *tropical* como população de treinamento.

Tabela 4 - Média e desvio-padrão ( $\sigma$ ) da capacidade preditiva ( $r_{(\hat{y},g)}$ ) das predições genômicas obtidas em CPT2 para diferentes tamanhos de população de treinamento ( $N_t$ ) e de acordo com os painéis USP, NCRPIS e ASSO para as características de altura de planta (AP) e altura de espiga (AE)

Características	Painel usado na PT	$N_t$	$r_{(\hat{y},g)}^1$	$\sigma$
AP	USP + NCRPIS + ASSO	2465	0.29 <sup>a</sup>	±0.13
	NCRPIS + ASSO	2434	0.20 <sup>a</sup>	±0.13
	NCRPIS	2046	0.24 <sup>a</sup>	±0.11
	ASSO	388	0.08 <sup>b</sup>	±0.17
	NCRPIS ( <i>tropical</i> )	512	0.21 <sup>a</sup>	±0.11
	ASSO ( <i>tropical</i> )	136	0.20 <sup>a</sup>	±0.13
AE	USP + NCRPIS + ASSO	2465	0.44 <sup>a</sup>	±0.11
	NCRPIS + ASSO	2434	0.15 <sup>cd</sup>	±0.11
	NCRPIS	2046	0.18 <sup>c</sup>	±0.10
	ASSO	388	0.11 <sup>cd</sup>	±0.15
	NCRPIS ( <i>tropical</i> )	512	0.09 <sup>d</sup>	±0.09
	ASSO ( <i>tropical</i> )	136	0.30 <sup>b</sup>	±0.11

População de Validação (PV): 32 linhagens (Painel USP) determinadas aleatoriamente; <sup>1</sup> Médias seguidas pela mesma letra, na coluna não diferem entre si pelo teste de Tukey a 5%

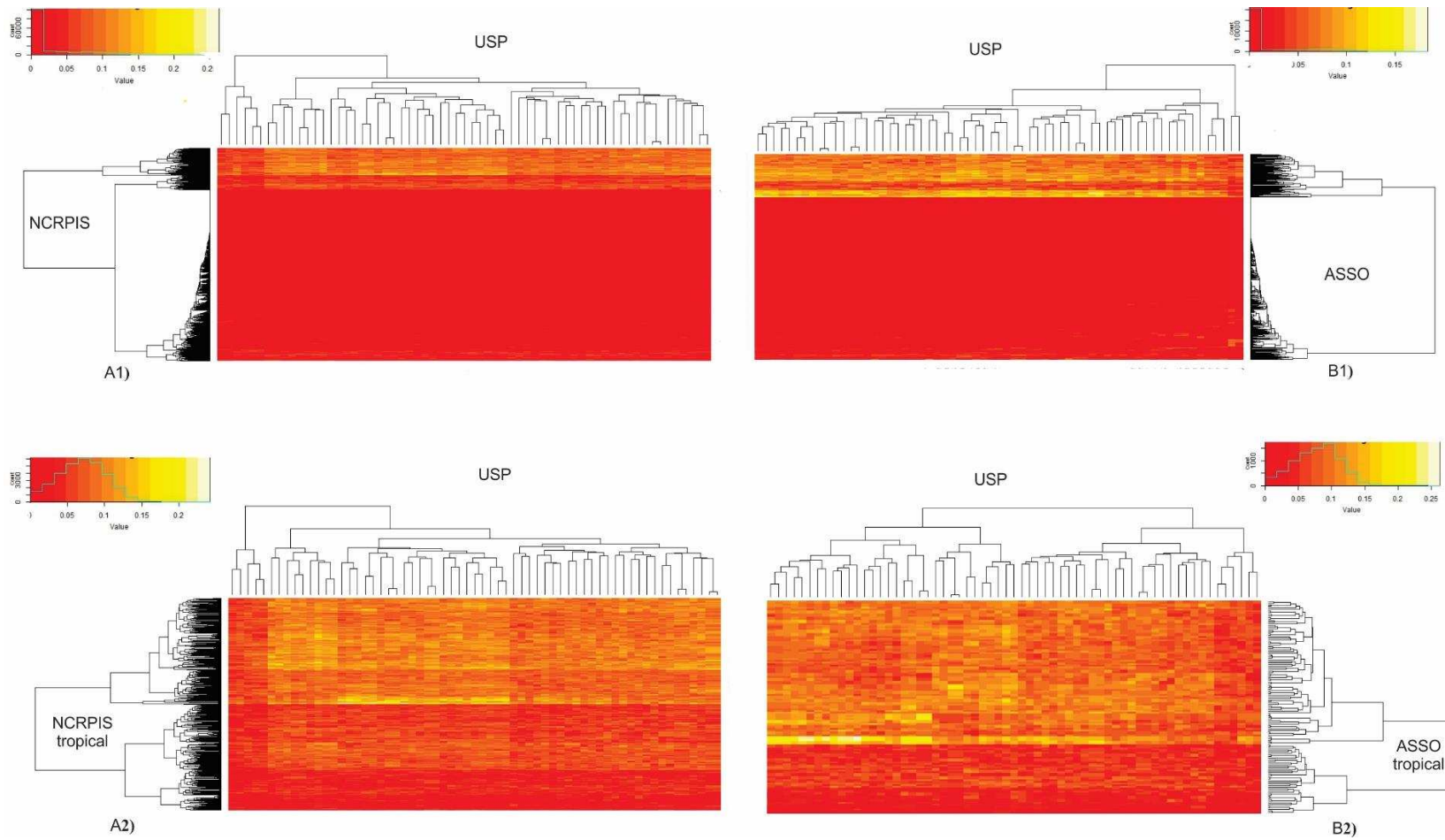


Figura 2 - *Heat map* da matriz de parentesco (K), utilizando 28.260 marcadores SNP, para relação direta entre os painéis USP e NCRISP (A1), USP e ASSO (B1), USP e NCRISP dentro do agrupamento *tropical* (A2) e USP e ASSO dentro do agrupamento *tropical* (B2)

Já é estabelecido que a estrutura da população é um fator crítico, que tende a afetar as predições na GS (Guo *et al.*, 2014). Portanto, é necessário levá-la em consideração na formação da PT ou na definição dos modelos para evitar estimativas irrealista e espúrias de capacidades preditivas e acurácias (Riedelsheimer *et al.*, 2013; Wray *et al.*, 2013). Ratificando os resultados aqui encontrados, Albrecht *et al.* (2011) e Guo *et al.* (2014) observaram que conjuntos de PT e PV formados a partir de um mesmo agrupamento ajudam a melhorar a capacidade preditiva na PV. Por outro lado, quando PT e PV divergem ou quando os genótipos são oriundos de diferentes cruzamentos ou famílias, as capacidades preditivas e acurácias dos modelos de seleção genômica tendem a reduzir.

#### 4.6.3.3 CPT3

Os resultados obtidos em CPT3 indicam que maiores valores de capacidade preditiva podem ser obtidos para altura de plantas, quando se utiliza populações de treinamento otimizadas (PTO) com  $N_t$  de 250, 500 ou 1000 indivíduos. Nestes casos, os valores observados variaram de 0,31 a 0,32 (Tabela 5), não havendo diferenças significativas entre os valores obtidos. Para altura de espiga, maiores estimativas de capacidade preditiva também podem ser vistas em PTO com  $N_t= 250$  indivíduos (Tabela 5).

Tabela 5 - Média e desvio-padrão ( $\sigma$ ) de capacidade preditiva,  $r_{(\hat{y},g)}$ , das predições genômicas obtidas em CPT3 para diferentes tamanhos de população de treinamento otimizadas ( $N_t$ ) obtidas via painel NCRPIS e ASSO para as características de altura de planta (AP) e altura de espiga (AE)

Características	$N_t$	$r_{(\hat{y},g)}^1$		$\sigma$	
		PTO	PTR	PTO	PTR
AP	50	0,17 <sup>b, B</sup>	0,14 <sup>b, B</sup>	±0,14	±0,16
	250	0,32 <sup>a, A</sup>	0,17 <sup>b, B</sup>	±0,10	±0,19
	500	0,31 <sup>a, A</sup>	0,18 <sup>b, B</sup>	±0,19	±0,19
	1000	0,31 <sup>a, A</sup>	0,22 <sup>b, AB</sup>	±0,10	±0,18
	1500	0,18 <sup>b, B</sup>	0,20 <sup>b, AB</sup>	±0,15	±0,11
AE	50	0,15 <sup>a, BC</sup>	0,06 <sup>a, AB</sup>	±0,14	±0,19
	250	0,32 <sup>a, A</sup>	0,07 <sup>b, AB</sup>	±0,11	±0,16
	500	0,22 <sup>a, B</sup>	0,12 <sup>b, AB</sup>	±0,11	±0,18
	1000	0,06 <sup>a, CD</sup>	0,16 <sup>a, A</sup>	±0,12	±0,16
	1500	0,03 <sup>a, D</sup>	0,12 <sup>a, AB</sup>	±0,12	±0,11

População de Validação (PV): 32 linhagens (Painel USP) determinadas aleatoriamente; <sup>1</sup> Médias seguidas pela mesma letra, minúscula na linha e maiúscula na coluna não diferem entre si pelo teste de Tukey a 5%; PTO: população de treinamento otimizada; PTR: população de treinamento aleatória

Assim, é verificado que pode-se empregar PTO constituída por 250 indivíduos, sem risco de penalizar a estimativa da capacidade preditiva, desde que a escolha destes seja feita pela metodologia aqui testada. Como prova da eficiência do método, os resultados indicaram que nos conjuntos de PTR a capacidade preditiva aumentou conforme a adição de indivíduos na população de treinamento. No entanto, esse aumento não possui efeito prático e estatístico. Isto porque a variação observada foi insuficiente para gerar diferenças entre as capacidades preditivas (Tabela 5). Além disso, os valores observados foram inferiores àqueles obtidos quando utilizado PTO constituída por 250 indivíduos. Justamente, a melhor comparação entre PTR e PTO podem ser resumidas quando analisadas as diferenças das capacidades preditivas com  $N_t = 250$ . Nesse conjunto a PTO apresentou estimativas de 46,8 a 457% superiores a PTR para AP e AE, respectivamente (Tabela 5).

A superioridade da PTO se deve ao método de otimização utilizado. Esse é baseado em derivação de métodos computacionais, que encontram-se disponíveis em Akdemir *et al.* (2015). Esse método faz uma aproximação eficiente da predição da variância do erro (PEV), com base nos componentes principais via marcadores nas linhagens candidatas e testadas. A estimação da PEV é feita no conjunto de teste (Painel USP) em vez de ser feita no conjunto de candidatos (NCRPIS + ASSO). Conseqüentemente, a seleção de linhagens para o conjunto otimizado se dá em função daquelas que minimizam a PEV no painel USP. Essa tendência pode ser observada claramente para conjuntos menores,  $N_t = 50$  e 250, ao qual a seleção de linhagens para PTO é retirada, na sua grande maioria, justamente do agrupamento *tropical* (Figura 3).

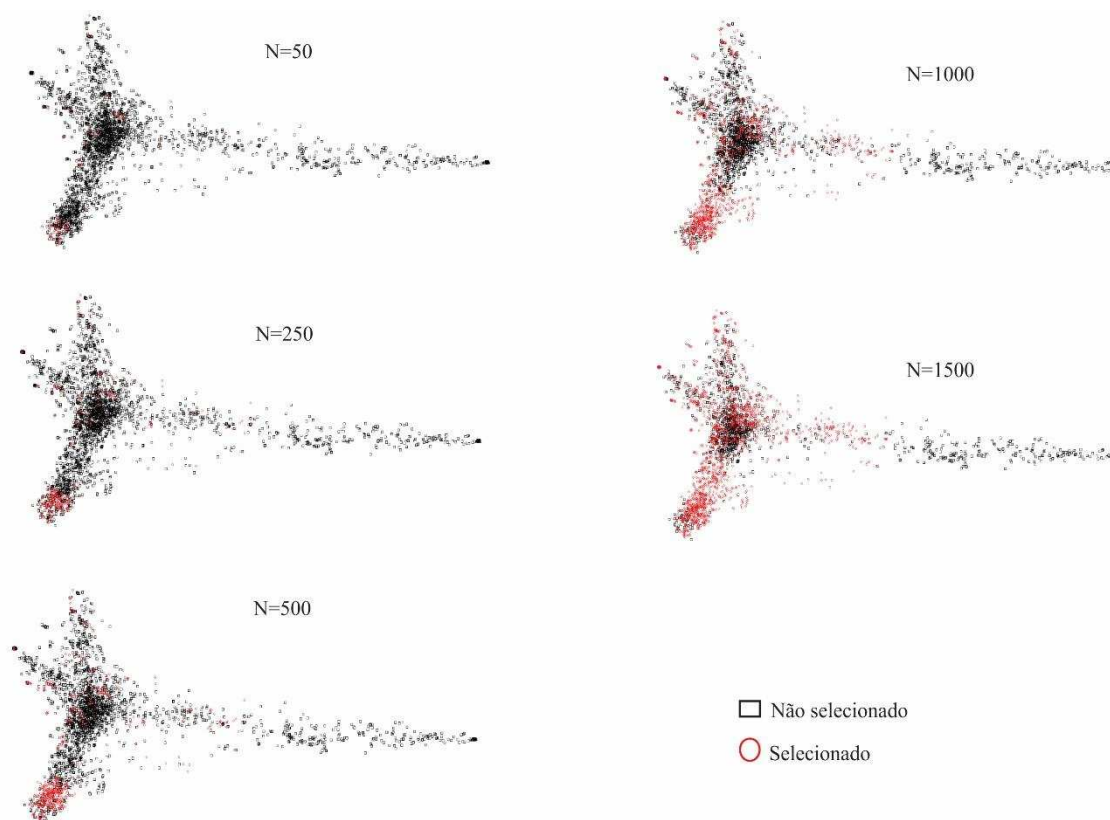


Figura 3 - Estruturação populacional das linhagens selecionadas para compor a população de treinamento otimizado (PTO) conforme o seu tamanho ( $N_t$ ) em relação ao cenário CPT3

De forma contrária, à medida que o tamanho do conjunto selecionado aumenta, são inclusas na população de treinamento linhagens divergentes e sem parentesco algum com o painel USP (Figura 3). Assim, a partir de um determinado ponto há um decréscimo da capacidade preditiva,  $N_t=1500$  e  $500$ , para AP e AE, respectivamente (Tabela 5). Isso demonstra que a estrutura da população desempenha importante papel na otimização das populações de treinamento, pois quando o efeito populacional é menor, as PTOs são capazes de prever de maneira mais acurada as linhagens do painel USP; e sob forte estruturação populacional, as PTO são menos acuradas. Não obstante, valores muito baixos,  $N_t=50$ , também limitam as previsões, pois não incluem variabilidade genética suficiente. Assim, não deve-se levar em consideração apenas o tamanho efetivo da população, mas também o grau de relacionamento desta com a população a ser predita.

#### 4.6.3.4 CPT4

A adição de linhagens pertencentes ao painel USP nos conjuntos de PTO e PTR, proporcionou o aumento nas estimativas da capacidade preditiva para todos os  $N_t$  testados (Tabela 6). O aumento médio em relação ao CPT3 foi de 31,2% para PTO e 52,5% para PTR em AP. Já para AE esses valores foram de 71,8% em PTO e 396% em PTR. Ambos os tipos de população de treinamento testados (PTO e PTR) forneceram estimativas de capacidades preditivas satisfatórias, de forma que com conjuntos de  $N_t$  com 81 ou 281 indivíduos, já é adicionado variabilidade genética suficiente para se alcançar patamares de 0,59 de capacidade preditiva (Tabela 6), no modelo testado e para a população de interesse. Porém, como mostrado anteriormente, conjuntos muito pequenos não são efetivos nas predições. De forma que, se trabalhado com PV maiores do que aqui testadas ( $32 > N_v < 63$ ), o tamanho de  $N_t$  irá diminuir ( $50 > N_t < 81$ ;  $250 > N_t < 281$ ). Podendo ter como resultado mais provável a redução da capacidade preditiva para  $N_t$  próximos de 50 indivíduos. No entanto, o uso de  $N_t$  próximos a 250 indivíduos tenderá manter os patamares das estimativas de capacidades elevadas, pois mostra ser em outros cenários o tamanho de conjunto mais robusto.

Tabela 6 - Média e desvio-padrão ( $\sigma$ ) de capacidade preditiva,  $r_{(\hat{y},g)}$ , das predições genômicas obtidas em CPT4 para diferentes tamanhos de população de treinamento otimizadas ( $N_t$ ) obtidas via painel USP, NCRPIS e ASSO para as características de altura de planta (AP) e altura de espiga (AE)

Características	$N_t$	$r_{(\hat{y},g)}$ <sup>1</sup>		$\sigma$	
		PTO	PTR	PTO	PTR
AP	81	0,43 <sup>a, A</sup>	0,41 <sup>a, A</sup>	±0,10	±0,13
	281	0,44 <sup>a, A</sup>	0,40 <sup>a, A</sup>	±0,08	±0,13
	531	0,41 <sup>a, A</sup>	0,37 <sup>a, A</sup>	±0,09	±0,14
	1031	0,40 <sup>a, A</sup>	0,39 <sup>a, A</sup>	±0,09	±0,11
	1531	0,33 <sup>a, B</sup>	0,33 <sup>a, B</sup>	±0,10	±0,11
AE	81	0,57 <sup>a, A</sup>	0,58 <sup>a, A</sup>	±0,09	±0,09
	281	0,59 <sup>a, A</sup>	0,55 <sup>a, A</sup>	±0,10	±0,13
	531	0,51 <sup>a, B</sup>	0,53 <sup>a, AB</sup>	±0,14	±0,13
	1031	0,42 <sup>b, C</sup>	0,53 <sup>a, AB</sup>	±0,15	±0,12
	1531	0,42 <sup>b, C</sup>	0,48 <sup>a, B</sup>	±0,14	±0,11

População de Validação (PV): 32 linhagens (Painel USP) determinadas aleatoriamente; <sup>1</sup> Médias seguidas pela mesma letra, minúscula na linha e maiúscula na coluna não diferem entre si pelo teste de Tukey a 5%; PTO: população de treinamento otimizada; PTR: população de treinamento aleatória

Outro ponto que é válido inferir é sobre a igualdade de condições entre PTO e PTR. Nesses, a formação do conjunto via PTR depende de uma amostragem aleatória das linhagens. Com isto, podem ser selecionadas linhagens pertencentes à uma única subpopulação e altamente aparentadas, o que resultaria em perda da diversidade no programa de melhoramento genético (Isidro *et al.*, 2015), assim como na capacidade preditiva dos modelos de GS. Por outro lado, a PTO segue um padrão na seleção dos indivíduos que irão compô-la, pelo qual é controlado o parentesco e variabilidade genética dos materiais selecionados (Figura 3). A otimização apenas dentro de um conjunto aparentado não melhora a correlação dos GEBVs com as linhagens não fenotipadas, porém, a associação dessas informações com a informação de distância genética, pode levar à seleção de uma amostra estável e geneticamente representativa de linhagens a serem fenotipadas (Schmidt *et al.*, 2016), otimizando assim o processo de formação da população de treinamento.

Estudos prévios, desenvolvidos por Windhausen *et al.* (2012), indicam que quando 50% dos genótipos no conjunto de validação são incluídos no conjunto de treinamento, a capacidade preditiva aumenta para todas as características. Sendo os maiores aumentos para características de alta herdabilidade. Ademais, os resultados obtidos via CPT3 e CPT4 indicam que a capacidade preditiva pode ser melhorada se as linhagens que compõem o conjunto de treinamento forem corretamente selecionadas por meio de metodologias eficientes, como a proposta por Akdemir *et al.* (2015), viabilizando resultados satisfatórios principalmente via conjuntos menores de indivíduos. Simulações também sugerem que, em alguns casos, pequenos conjuntos de treinamento podem ser tão precisos quanto aqueles de maior proporção (Habier *et al.*, 2009). Fato esse que tem sido confirmado em populações de melhoramento reais (Isidro *et al.*, 2015; Wong e Bernardo, 2008) e nesse estudo. Em geral, PT maiores seriam indicadas para características controladas por um grande número de genes de pequeno efeito e com grande influência do ambiente na expressão do fenótipo (Goddard e Hayes, 2009), como por exemplo a produtividade de grãos. Porém, neste trabalho, as características utilizadas possuem identificados um pouco mais de 40 genes de grande efeito que atuam no seu controle genético (Multani *et al.*, 2003), e nas quais o ambiente tem menor

efeito na expressão fenotípica (Peiffer *et al.*, 2014), o que na prática pode estar permitindo o uso de populações de treinamento menores.

Os desvios padrões ( $\sigma$ ) possuem estimativas muito próximas na maioria dos cenários, com exceção de CPT3 (Tabela 5), que apresenta maiores desvios para PTR. Pontualmente, o conhecimento dos desvios padrões das médias das capacidades preditivas é importante em conjuntos de treinamento pequenos, onde alguns indivíduos com alelos raros e/ou com blocos de haplótipos podem ter uma influência maior e, portanto, podem contribuir mais para o aumento da capacidade preditiva (Schmidt *et al.*, 2016). No entanto, menores desvios caracterizam maior estabilidade da capacidade preditiva o que eleva a confiabilidade no uso de determinado conjunto de treinamento.

#### 4.6.4 Aplicação no melhoramento de plantas e perspectivas

Levando em consideração a temática central do problema aqui proposto, no qual os recursos empregados são limitados e existem as possibilidades de uso de informação genotípica em toda a população do programa, a utilização de painéis externos para predição de populações menores são animadoras. Tal procedimento foi recentemente explorado com sucesso por Jarquin *et al.* (2016) na cultura da soja, alcançando capacidades preditivas de até 0,92 para óleo e proteína e 0,79 para produtividade. Esse resultado e os divulgados aqui por esse estudo, orientam e permitem que os melhoristas acessem bancos públicos de dados genômicos e fenotípicos, podendo usá-los no seu programa de GS. Não necessitando grande número de genótipos na PT formada, quando empregado a otimização de população. Para isso, é necessário garantir um grande conjunto de dados iniciais, com ampla variabilidade genética, para que o processo de otimização da PT possa ser efetivo.

Além disso, é recomendável o uso combinado de informações externas e de parte das linhagens do programa na formação da PT; o que proporcionará em estágios futuros a seleção dos melhores indivíduos, reduzindo a quantidade de material a serem levados nos experimentos de campo, minimizando custo e maximizando ganho com a seleção. Contudo, isso é válido para características com alta herdabilidade; sendo necessário estudos adicionais para características com menores herdabilidades. Ressalta-se que outros fatores não foram testados no presente trabalho, como os modelos estatísticos empregados (Heslot *et al.*, 2012), o número e tipo de marcadores (Chen e Sullivan, 2003; Poland e Rife,

2012) e o desequilíbrio de ligação (Habier *et al.*, 2007). Esses fatores são capazes de influenciar a capacidade preditiva e acurácias da GS e devem ser objetos de estudo dentro do panorama proposto.

Seguindo uma simulação de como aplicar esse procedimento em um programa de melhoramento. Sugerimos um quadro onde é desejável trabalhar com uma PT de 500 linhagens para predição de linhagens Duplo Haplóides (DH). Nesse contexto, pode-se definir uma PV com  $N_v=50$  (10% da PT) e PS constituídas de pelos menos 1000 indivíduos, gerados anualmente pelo programa de melhoramento. A estratégia nesse procedimento se dá via genotipagem e fenotipagem de apenas uma parte da população interna, a cada ciclo ou a cada ano, e uso dos demais dados provindos de bancos públicos. Ao passo que ao longo dos anos haja uma substituição das linhagens externas pelas internas do programa. Com isso, o capital monetário necessário para compor uma PT é minimizado ao longo dos anos. Essa estratégia pode ser resumida da seguinte forma: 1º ano: *i*) genotipagem de 96 indivíduos e fenotipagem dos mesmos; *ii*) dos 96 indivíduos, 50 formarão a PV, os demais integrarão a PT juntamente com linhagens de bancos de dados públicos, ou seja, 46 linhagens internas e 454 externas (PTO com base na PS); *iii*) procede-se a validação e a predição dos GEBVs da PS. 2º ano: *i*) procede-se genotipagem e fenotipagem de 96 novos indivíduos; *ii*) define-se a PV formada apenas com base nas linhagens internas. Assim, das 192 linhagens internas alocadas nessa fase (96 do 1º ano + 96 do 2º ano), 50 integram a PV e 142 a PT, necessitando adicionais 358 linhagens externas selecionadas via procedimentos de otimização; *iii*) procede-se a validação e a predição dos GEBVs na nova PS. Esse procedimento é repetido até o 5º ano, em que praticamente toda a PT será substituída por linhagens internas do programa, considerando introdução de dados de 96 linhagens anualmente. E como mostrado aqui, esse procedimento tem o potencial de manter adequada capacidade preditiva do modelo preditivo.

Uma das vantagens da abordagem acima é a possibilidade de um melhor planejamento técnico operacional e financeiro, pois esse período pode ser anual, bi-anual ou então conforme os recursos disponíveis ao longo do tempo. Tão logo, a substituição das linhagens públicas pelas internas respeita a limitação técnica e econômica de cada programa de melhoramento. Em função dessa abordagem, financeira, o valor empregado para gerar uma PT com 500 indivíduos utilizando

uma *array* de três mil marcadores do tipo SNP, ao qual o custo total (amostra+serviço) ficam em U\$39. A fenotipagem dessa mesma população está em torno de U\$20/parcela (dado cedido por Sementes Biomatrix®). Por meio desses valores, o custo da genotipagem e fenotipagem (considerando apenas um ambiente e duas repetições) é de U\$19.500,00 e U\$20.000, respectivamente, totalizando assim U\$39.500,00. Ao passo que se utilizado o procedimento aqui descrito há a possibilidade de fazer pequenos investimentos a cada ano, amortizando o valor total ao longo do tempo, necessitando investimento de U\$7.584,00 a cada nova genotipagem e fenotipagem de 96 indivíduos.

Outro fato, é que apesar do ambiente ter menor efeito sobre AP e AE, comparando com a produtividade por exemplo. Ainda assim é necessário ser testado a capacidade preditiva de modelos de seleção genômica, seguindo o panorama aqui aplicado, utilizando ambientes relacionados para PT e PV. Pois, há evidências de que essa decresce levemente quando PT e PV divergem nos ambientes testados (Windhausen *et al.*, 2012). Por isso, é proposto e incentivado a ideia de criação de bancos públicos de dados de milho na América do Sul, principalmente no Brasil, podendo ser fomentado por dados de programas de melhoramento de universidades, instituições públicas e programas privados parceiros. Tal ação, possibilitaria o progresso mútuo e diminuição dos custos para a implementação de programas de seleção genômica, principalmente daqueles programas de melhoramento recém formados e que contam com baixa variabilidade genética em seu germoplasma.

#### **4.7 Conclusões**

- O uso de bancos públicos de informações genômicas e fenotípicas é viável, quanto a capacidade preditiva, como fonte de criação de populações de treinamento para implementação da seleção genômica em programas de melhoramento com recursos limitados.
- Ao utilizar esses bancos de informações, a estrutura populacional intrínseca às populações de treinamento e validação interferem nas capacidades preditivas. Com isso, é recomendado o uso de metodologia de otimização de população de treinamento e a adição de informação da população de validação. Desta

forma, conjuntos pequenos de indivíduos ( $N_t=250$ ) são suficientes para obtenção de capacidades preditivas satisfatórias de seleção genômica.

#### 4.8 Referências

- Albrecht, T.; Wimmer, V.; Auinger, H.J.; Erbe, M.; Knaak, C.; Ouzunova, M.; Simianer, H.; Schön, C.C. 2011. Genome-based prediction of testcross values in maize. **Theoretical and Applied Genetics**, v.123, p.339-350
- Akdemir, D.; Sanchez, J.I.; Jannink, J-L. 2015. Optimization of genomic selection training populations with a genetic algorithm. **Genetic Selection Evolution**, p.47-38
- Akdemir, D. 2014. **STPGA: Selection of training populations by genetic algorithm. R package version 1.0.** <http://CRAN.R-project.org/package=STPGA>
- Bassi, F.M.; Bentley, A.R.; Charmet, G.; Ortiz, R.; Crossa, J. 2016. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). **Plant Science**, v.42, p.23-36. doi:10.1016/j.plantsci.2015.08.021
- Borém A.; Galvão, J.C.C., Pimentel, M.A. 2015. **Milho: do Plantio à Colheita.** Viçosa. Editora UFV. 351p
- Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v.23, p.2633-2635
- Cavalcanti, J.J.V.; Resende, M.D.V de; SANTOS, F.H.C.; PINHEIRO, C.R. 2012. Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. **Revista Brasileira de Fruticultura** [online], v.34, p.840-846. <http://dx.doi.org/10.1590/S0100-29452012000300025>
- Chen, X.; Sullivan, P.F. 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. **Pharmacogenomics Journal**, v.3, p.77-96
- Clark, S.; Hickey, J. van der Werf, J. 2011. Different models of genetic variation and their effect on genomic evaluation. **Genetic Selection Evolution**, p.43-18

- Crossa, J.; Pérez P.; Hickey, J.M.; Burguenó, J.; Ornella, L. Cerón-Rojas, J.; Zhang, X.; Dreisigacker, S.; Babu, R.; Li, Y.; Bonnett, D.; Mathews, K. 2014 Genomic prediction in CIMMYT maize and wheat breeding programs. **Heredity**, v.112, p.48-60
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, v.4, p.250-255
- Flint-Garcia, S. A., Thuillet, A.-C. ; Yu, J. ; Pressoir, G. ; Romero, S. M.; Mitchell, S.E. et al. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. **Plant Journal**, v.44, p.1054-1064
- Franzoni, J.; Scapim, C.A.; Beviláqua, M.R.R. Machado, M.F.P.S.; Pacheco, C.A.P.; Mangolin, C.A. 2012. Application of microsatellite markers to evaluate the heterozygosity from the popcorn composite CMS-43 (*Zea mays* L.) during eight cycles of selection. **Plant Breeding**, v.131, p.479-485. doi:10.1111/j.1439-0523.2012.01981.x
- Gilmour, A.R.; Gogel, B.J.; Cullis, B.R.; Thompson, R. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK www.vsn.co.uk
- Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J; Elshire, R.J.; Sun, Q.; Buckler, E.S. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. **PloS One**, 9(2): e90346. doi: 10.1371/journal.pone.0090346
- Guo; Z.; Tucker; D.M.; Basten; C.J.; 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*, v.127, p.749-762
- Habier, D.; Fernando, R.L.; Dekkers, J.C.M. 2007. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v.77, p.2389-2397
- Hayes; B.J.; Visscher, P.M.; Goddard, M.E. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research (Camb)**, v.91, p.47-60
- Heffner, E.L.; Sorrells, M.E.; Jannink, J-L. 2009. Genomic selection for crop improvement. **Crop Science**, v.49, p.1-12
- Heslot, N.; Yang, H.P.; Sorrells, M.E.; Jannink J-L. 2012. Genomic selection in plant breeding: a comparison of models. **Crop Science**, v.52, p.146–160

- Hung; H. Y.; Browne, C.; Guill, K.; Coles, N.; Eller, M. et al. 2012 The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. **Heredity**, v.108, p 490-499
- Isidro; J.; Jannink, J-L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M.E. 2015. Training set optimization under population structure in genomic selection. **Theoretical and Applied Genetics**, v.128, p.145-158. doi:10.1007/s00122-0142418-4
- Jarquín, D.; Specht, J.; Lorenz, A. 2016. Prospects of genomic prediction in the USDA Soybean Germplasm Collection: Historical data creates robust models for enhancing selection of accessions. **bioRxiv**. doi:10.1534/g3.116.031443
- Jonas, E.; De Koning, D.J. 2013. Does genomic selection have a future in plant breeding? **Trends in biotechnology**, v. 31,p.497-504
- Jannink, J-L.; Lorenz, A.J.; Iwata, H. 2010. Genomic selection in plant breeding: from theory to practice. **Brief Funct Genomics**, 9(2), p.166-177
- McMullen; M. D.; Kresovich, S.; Villeda, H. S.; Bradbury, P.J.; Li, H. et al. 2009. Genetic properties of the maize nested association mapping population. **Science**, v.325, p.737-740
- Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v.157, p.1819-1829
- Multani, D.S.; Briggs, S.P.; Chamberlin, M. A.; Blakeslee, J.J.; Murphyyetal, A.S. 2003 Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. **Science**, v.302, p.81-84
- Muranty, H.; Troggio, M.; Sadok, I.B.; Mehdi, A.R.; Annemarie, A. et al. 2015. Accuracy and responses of genomic selection on key traits in apple breeding. **Horticulture Research**, v.2 (15060). doi:10.1038/hortres.2015.60
- Newell, M. A.; Jannink, J-L. 2014. Genomic Selection in Plant Breeding. In: FLEURY, D.; WHITFORD, R. (Eds.). **Crop Breeding**. Berlin/Heidelberg: Springer, p.117-130

- Peiffer, J.A.; Romay, M.C.; Gore, M.A.; Flint-Garcia, S.A.; Zhang, Z. et al. 2014. The genetic architecture of maize height. **Genetics**, v.196, p.1337-1356. doi: 10.1534/genetics.113.159152
- Poland, J.; Rife, T.W. 2012. Genotyping-by-sequencing for plant breeding and genetics. **The Plant Genome**, v.5, p.92-102
- Pszczola, M.; Strabel, T.; Mulder, H.A.; Calus, P.L. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. **Journal of Dairy Science**, v.95, p.389-400
- R Core Team. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing. 2013. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Rincent, R.; Laloë, D.; Nicolas, S.; Altmann, T.; Brunel, D. et al. 2012. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of Maize Inbreds (*Zea mays* L.). **Genetics**, v.192, p.715-728
- Riedelsheimer, C.; Endelman, J.B.; Stange, M.; Sorrells, M.E.; Jannink, J-L.; Melchinger, A.E. 2013. Genomic predictability of interconnected bi-parental maize populations. **Genetics**, v.194, p.493-503 doi:10.1534/genetics.113.150227
- Romay, M.; Millard, M.; Glaubitz, J.; Peiffer, J.; Swarts, K. et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biology**, 14:R55
- Schmidt, M.; Kollers, S.; Maasberg-Prelle, A.; Großer, J., Schinkel B.; Tomerius, A.; Graner, A.; Korzun, V. 2016. Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. **Theoretical and Applied Genetics**, 129(2), p.203-213. doi: 10.1007/s00122-015-2639-1
- Unterseer, S.; Bauer, E.; Haberer, G.; Seidel, M.; Knaak, C. et al. 2014. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. **BMC Genomics**, v.15, p.823.
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. **Journal of Dairy Science**; v.91, p.4414-4423
- Windhausen, V.S.; Atlin, G.N.; Hickey, J.M.; Crossa, J.; Jannink, J-L.; Sorrells, M. et al. 2012. Effectiveness of genomic prediction of maize hybrid

performance in different breeding populations and environments. **G3** (Bethesda), v2, p.1427-1436

Wray, N.R.; Yang, J.; Hayes, B.J.; Price, A.L.; Michael, E.; Goddard, M.E.; Visscher, P.M. 2013. Pitfalls of predicting complex traits from SNPs. **Nature Reviews Genetics**, v.14, p507-515

Zhao; Y.; Gowda, M.; Liu, W.; Würschum, T.; Maurer, H. P. et al. 2012. Accuracy of genomic selection in European maize elite breeding populations. **Theoretical and Applied Genetics**, v.124, p.769-776

## 5. TABELAS E FIGURAS SUPLEMENTARES

Tabela Suplementar 1: Análise de solo das áreas experimentais na cidade de Piracicaba-SP

Determinações		Unidades	Amostras <sup>1</sup>				
			1	2	3	4	5
-	-	-					
Ph	CaCl <sub>2</sub> 0,01mol/L	-	4,7	5,3	5,1	4,9	5,3
P	Resina trocadora de íons	mg/dm <sup>3</sup>	39	46	57	53	47
K	Resina trocadora de íons	mmolc/dm <sup>3</sup>	1,8	3,8	5	5,0	3,5
Ca	Resina trocadora de íons	mmolc/dm <sup>3</sup>	24	43	27	18	28
Mg	Resina trocadora de íons	mmolc/dm <sup>3</sup>	8	22	9	6	9
Cl	Resina trocadora de íons	mmolc/dm <sup>3</sup>	2	2	2	2	2
H+Al	Alumínio trocável KCL 1mol/L	mmolc/dm <sup>3</sup>	31	22	28	38	25
SB	Acidez Potencial – SMP	mmolc/dm <sup>3</sup>	33,8	68,8	41	29,5	40,5
CTC	-	mmolc/dm <sup>3</sup>	64,8	90,8	69	67,5	65,5
V	-	%	52	76	59	44	62
m	-	%	6	3	5	6	5
M.O.	-	g/dm <sup>3</sup>	18	19	18	21	21

<sup>1</sup>Data da emissão:18/09/2014

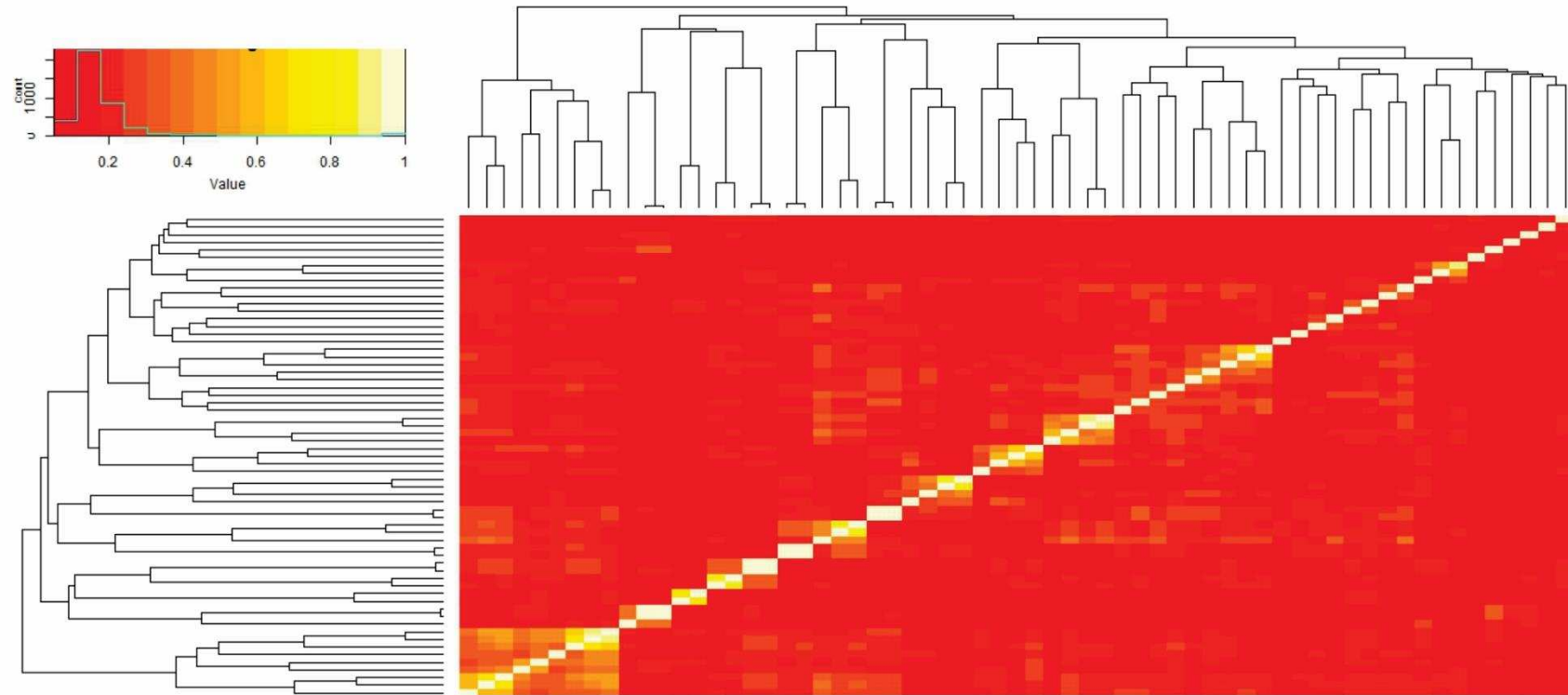


Figura Suplementar 1 - *Heat map* da matriz de parentesco genômico, genotipadas com 28.260 marcadores SNP, entre 63 linhagens de milho tropical do painel USP

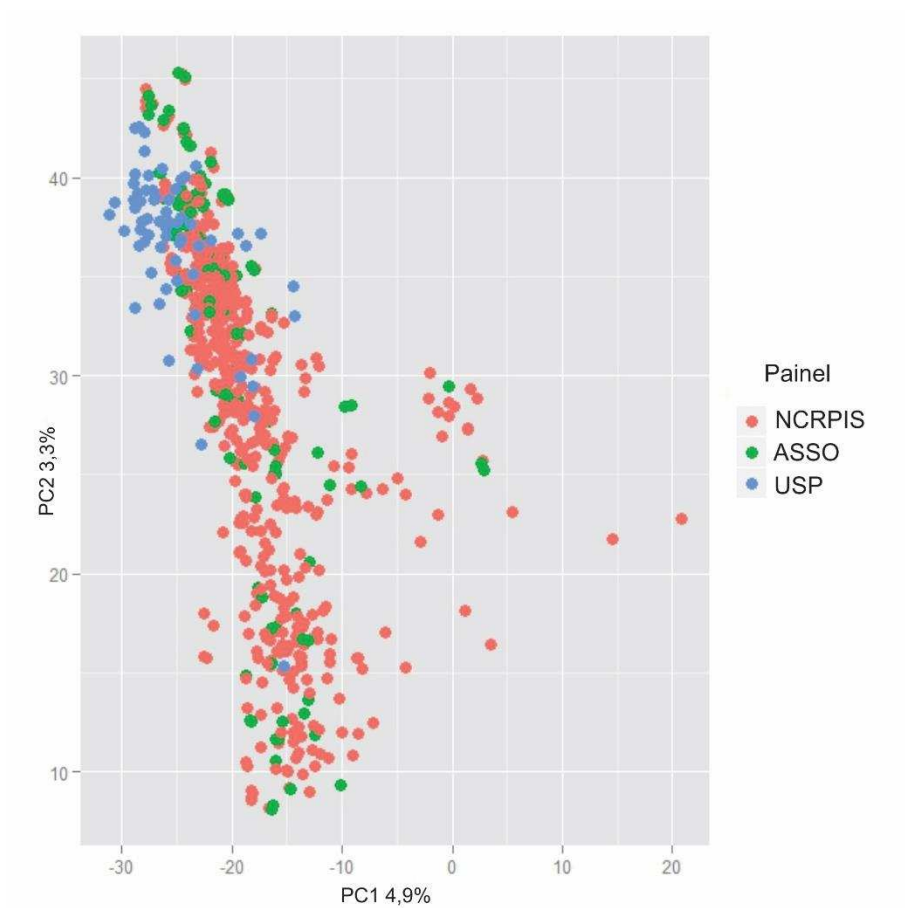


Figura suplementar 2 - Estruturação populacional via Análise de Componentes Principais (PCA) para o agrupamento *tropical*, genotipadas com 28.260 marcadores SNP e com 711 linhagens de milho oriundas dos painéis USP, ASSO e NCRPIS. Cada círculo sólido representa um genótipo

## **6. CONCLUSÃO GERAL**

Os estudos procedidos na população avaliada de linhagens de milho tropical indicam a possibilidade de seleção para eficiência no uso de nitrogênio. Tal como é possível também predizer essa população usando informações de bancos públicos de dados genômicos e fenotípicos para características de alta herdabilidade, obtendo satisfatórias capacidades preditivas via seleção genômica.