

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Avaliação de modelos preditivos de seleção genômica ampla em
testes de progênies e testes clonais de *Eucalyptus***

Carla da Costa Garcia

Dissertação apresentada para obtenção do título de
Mestra em Ciências, Programa: Recursos Florestais.
Opção em: Silvicultura e Manejo Florestal

**Piracicaba
2016**

Carla da Costa Garcia
Engenheira Florestal

**Avaliação de modelos preditivos de seleção genômica ampla em testes de
progênies e testes clonais de *Eucalyptus***

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **ANTONIO NATAL GONÇALVES**

Dissertação apresentada para obtenção do título de
Mestra em Ciências, Programa: Recursos Florestais.
Opção em: Silvicultura e Manejo Florestal

**Piracicaba
2016**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Garcia, Carla da Costa

Avaliação de modelos preditivos de seleção genômica ampla em testes de progênies e testes clonais de *Eucalyptus* / Carla da Costa Garcia. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.

69 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Seleção genômica 2. *Eucalyptus* 3. Acurácias 4. Capacidade preditiva 5. Valor genético genômico I. Título

CDD 634.9734
G216a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Dedico.

A DEUS, meu Pai Querido, por tomar as rédeas da minha vida e me conduzir até aqui...

AOS MEUS PAIS, Fátima e Mario, pelo amor incondicional e pela sabedoria de nos direcionar, minha irmã e eu, para caminhos bons...

A MINHA IRMÃ, Paula, minha melhor amiga que sempre foi o meu exemplo de dedicação e minha parceira de todas as horas...

Ofereço.

AGRADECIMENTOS

A Deus, minha fortaleza e minha calma.

Ao meu Orientador, Prof. Dr. Antônio Natal Gonçalves, pelo apoio e amizade.

Aos Co-orientadores Dr. Dario Grattapaglia e Dr. Marcos Deon Vilela de Resende pela orientação na elaboração deste trabalho, pela referência que são no setor florestal e na Genética, Melhoramento e Biotecnologia de Plantas e pela honra e oportunidade de poder trabalhar com vocês.

À empresa International Paper, pelas incontáveis oportunidades a mim dadas, por meio de Armando Santiago, Luis Fernando Silva e Adriano Almeida. Sempre retribuirei da melhor forma possível.

À Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ-USP) e ao Departamento de Ciências Florestais pela formação profissional.

À minha família. Meu pai, minha mãe, minha irmã e meu cunhado, por serem meu porto seguro, minhas referências e pelos valores de vida a mim ensinados. Sem vocês nada disso faria sentido. Amo vocês.

Ao Engenheiro Florestal Dr. Rafael Tassinari da Universidade Federal de Viçosa, a Embrapa Recursos Genéticos e Biotecnologia, e as empresas Hereditas e Genesseek por todo suporte e infraestrutura fornecidos na elaboração dos dados que compuseram este trabalho.

Ao Engenheiro Florestal Dr. Bruno Marco de Lima pelo trabalho desenvolvido em parceria com a International Paper em sua tese de doutorado, com o qual pude dar continuidade através desta pesquisa.

Aos amigos da Gloriosa ESALQ/USP - Desirée Lopes (Iskiter), Rafaela Carneiro (14-Bis), Guilherme Batistuzzo (Inxada), José Luiz Conti (Pogobol), Thiago Tofoli (Só-Fulia), Daniele Balestrin (Bananiña). Pessoas incríveis de uma valiosa época de aprendizado, amizade e vivência.

Aos amigos colegas de profissão Adriano Almeida, Regiane Estopa, Jupiter Abad, Teotônio Assis por serem minhas referências, por tudo que constantemente aprendo e por compartilharem bons momentos de risadas e histórias.

Ao Grupo Florestal Monte Olimpo (GFMO) por me mostrar constantemente a beleza e a arte da Ciência Florestal e em como vamos mais longe quando se há união e amizade. Orgulho de ser parte desta família.

Aos mestres Teotônio Francisco de Assis, Roland Venkovsky, José Luiz Stape e José Leonardo de Moraes Gonçalves, grandes mestres do melhoramento de plantas e das ciências florestais, os quais eu admiro muito profissional e pessoalmente.

Aos colegas (e ex-colegas) da International Paper Jim Rakestraw, Cristiane Lemos, Rodrigo Hakamada, Gabriela Moreira, Guilherme Batistuzzo, Jair Gabriel, Danielli Arruda, Sebastião Oliveira que muito contribuíram neste trabalho.

Casa do Estudante Universitário, “patrimônio histórico e cultural de Piracicaba”, por todos estes anos de convívio familiar.

Ao Grupo do Caminho da Comunicação Autêntica. Grupo (de Girafas queridas) mais empático e acolhedor com quem pude formar laços de amizade e que trouxeram leveza, empatia e compaixão para a minha vida.

A todos que de alguma maneira contribuíram para este trabalho, minha eterna e mais profunda gratidão.

SUMÁRIO

RESUMO.....	9
ABSTRACT	11
1 INTRODUÇÃO	13
2 REVISÃO BIBLIOGRÁFICA.....	17
2.1 Melhoramento Florestal Clássico	17
2.2 Seleção Genômica Ampla em espécies Florestais	18
2.3 Valores Genéticos Genômicos (VGGs) e Acurácias de Seleção	21
2.4 Perspectivas e desafios da Seleção Genômica Ampla em espécies Florestais..	27
3 MATERIAL E MÉTODOS.....	29
3.1 População de Treinamento e Modelos Preditivos	29
3.2 Populações utilizadas para avaliação da acurácia dos modelos preditivos	30
3.2.1 Populações Geneticamente Relacionadas.....	30
3.2.2 População Geneticamente Não-Relacionada.....	31
3.3 Fenotipagem das populações de avaliação e estimativa dos parâmetros genéticos.....	34
3.4 Genotipagem das populações de avaliação.....	36
3.5 Acurácias de Seleção e Capacidades Preditivas	36
3.6 Correlação entre VGGs estimados pelos modelos de SGA e dados fenotípicos mensurados.....	40
4 RESULTADOS E DISCUSSÃO.....	41
4.1 Estimativa dos parâmetros genéticos.....	41
4.2 Acurácias de Seleção e Capacidades Preditivas	43
4.2.1 Capacidade preditiva com base no modelo proposto por Lima (2014) para população de Eucalipto da International Paper	43
4.2.2 Capacidade preditiva e acurácia de seleção com base no modelo desenvolvido das populações de avaliação	45
4.3 Comparação do ranking individual utilizando a Seleção Genômica e BLUP para seleção fenotípica	55
5 CONCLUSÕES	57
REFERÊNCIAS.....	59
APÊNDICE	65

RESUMO

Avaliação de modelos preditivos de seleção genômica ampla em testes de progênies e testes clonais de *Eucalyptus*

O grande potencial da genômica em benefício do melhoramento genético aplicado é a utilização direta das informações de marcadores de DNA na seleção, de forma a permitir igual ou maior eficiência seletiva, maior rapidez na obtenção de ganhos genéticos e redução de custos de fenotipagem, em comparação com a seleção tradicional. Esta era genômica está trazendo novas oportunidades para os melhoristas florestais, porém desafios ainda existem para o uso operacional da Seleção Genômica Ampla. Sendo assim, este trabalho teve por objetivo avaliar a capacidade preditiva de características de crescimento volumétrico de modelos de SGA previamente construídos para características de crescimento volumétrico e qualidade da madeira com uma população elite de melhoramento genético da International Paper do Brasil que atuou como população de treinamento. Os modelos preditivos foram aplicados em quatro populações de candidatos à seleção constituídos por testes de progênies e teste clonal de *Eucalyptus* com características contrastantes do ponto de vista de relacionamento genético com a população na qual os modelos foram desenvolvidos: (1) três populações compostas por indivíduos geneticamente relacionados com a população de descoberta e (2) uma população composta por indivíduos geneticamente não relacionados com a população de descoberta. Posteriormente, as quatro populações de avaliação, compostas por 100 indivíduos genotipados em cada uma, foram utilizadas para construir novos modelos preditivos e validações cruzadas realizadas entre estas populações. Foram utilizados para as análises SNPs com frequência de declaração de genótipo (call rate) ≥ 0.90 e MAF (menor frequência alélica) ≥ 0.01 , totalizando 29.090 marcadores. O modelo preditivo previamente elaborado para população elite da International Paper apresentou capacidades preditivas que variaram de -0,296 para o caráter Altura em população geneticamente não relacionada até 0,440 para o caráter DAP em uma população geneticamente relacionada. Com os modelos desenvolvidos com as quatro populações genotipadas, e realizando seleção de marcas com base nos seus efeitos, as maiores capacidades preditivas foram obtidas por um número de SNPs médio que variou de 1371 para volume e IMA a 1467 para DAP. Usando esta abordagem, a capacidade preditiva maximizada para DAP foi de 0,744, 0,727 para altura, 0,751 para volume e 0,752 para IMA. Acurácias preditivas maximizadas, foram em seguida estimadas ao utilizar o menor número de SNPs selecionados. Com 237 SNPs acurácias da ordem de 0,660 para DAP, 0,555 para altura, 0,743 para volume e 0,743 para IMA foram obtidas. Embora estes resultados sugerem que a SGA teria bom resultado somente entre indivíduos geneticamente relacionados, uma análise conjunta dos dados utilizados para o desenvolvimento dos modelos preditivos anteriormente gerados com os dados das quatro populações aqui avaliadas, se faz necessária para alcançar resultados mais conclusivos. Adicionalmente, a abordagem de seleção de marcas para maximizar a capacidade preditiva ou acurácia deverá ser melhor avaliada à luz do seu impacto na medida que a SGA venha a ser praticada em futuras gerações de seleção.

Palavras-chave: Seleção genômica; *Eucalyptus*; Acurácias; Capacidade preditiva; Valor genético genômico

ABSTRACT

Assessment of predictive genome-wide selection models in clonal and progeny trials of *Eucalyptus*

The potential of genomics to the benefit of applied breeding is the direct use of DNA markers information for selection, allowing equal or higher selection efficiency, higher genetic gains per unit time and cost reduction in phenotyping, when compared with traditional selection based on phenotypic data. Researchers have proposed a new selection method called genome-wide selection (GWS), which has been successfully applied in livestock genetics and promises to revolutionize the improvement of perennial species with long life cycles. This genomic era is bringing new opportunities to forest breeders, but major challenges still need to be overcome for the operational use of GWS. Thus, this study aimed to assess the predictive ability of GWS models previously built for volume growth and wood quality traits based on an elite breeding population of International Paper in Brazil who served as discovery or training population. Predictive models were applied to four populations composed by progeny and clonal trials with contrasting characteristics from the standpoint of relatedness to the training population: (1) a population of individuals genetically related to the training population and (2) a population of individuals genetically unrelated to the training population. Subsequently, the four populations evaluated, with 100 genotyped individuals each, were used to build new models which were cross validated among them. Only SNPs with call rate ≥ 0.90 and MAF (minor allele frequency) ≥ 0.01 were used totaling 29,090 markers. Model previously developed yielded predictive abilities ranging from a lower -0.296 for height growth in genetically unrelated population to 0.440 for DBH and 0.219 for height in a genetically related population. With the GWS models built with the four genotyped populations and applying SNP marker selection based on their estimated effect the highest predictive capabilities were obtained when using an average number of SNPs ranging from 1371 for volume and MAI, to 1467 for DBH. The average predictive ability was maximized at 0.744 for DBH, 0.727 for height growth, 0.751 for volume and 0.752 for MAI. Maximized accuracies were obtained using the smallest number of SNPs with highest effects. With 237 selected SNPs accuracies around 0.660 for DBH, 0.555 for height, 0.743 for volume and 0.743 for MAI were obtained. While these results suggest that GWS should work well only across related individuals, a combined analysis of the data from the previous models with those of the four tested populations is necessary to reach more conclusive results. Furthermore, the approach of SNP marker selection to maximize predictive ability or accuracy is one that is still controversial and should be better evaluated in light of its impact as GWS is practiced in further generations of selection

Keywords: Genome-wide selection; *Eucalyptus*; Accuracies; Predictive ability; Genetic breeding value

1 INTRODUÇÃO

A silvicultura hoje é um importante setor da economia do Brasil, possuindo em 2014 aproximadamente 7,74 milhões de hectares de florestas plantadas. Em toda essa área se destaca a cultura do eucalipto, com mais de 5,5 milhões de hectares plantados, principalmente nos Estados de Minas Gerais, São Paulo, Mato Grosso do Sul e Bahia. A madeira do eucalipto é destinada aos mais distintos usos, como na produção de painéis de madeira, como fonte energética renovável e principalmente na indústria de celulose e papel. Vale lembrar que seu uso como fonte energética renovável é uma vertente que tende a crescer, diante de um cenário de insuficiência de combustíveis fósseis para os próximos anos (INDÚSTRIA BRASILEIRA DE ÁRVORES - IBÁ, 2015).

Originário da Austrália, para a maioria das espécies, o eucalipto (*Eucalyptus* spp.) encontrou no Brasil condições ideais para seu desenvolvimento, alcançando no país produtividades não encontradas em qualquer outro lugar no mundo (ELDRIDGE et al., 1993). Aliadas a essas vantagens climáticas, foram geradas tecnologias silviculturais adaptadas à cultura e desenvolveram-se programas de melhoramento genético (FERREIRA; SANTOS, 1997). Esses fatos colocaram o Brasil na vanguarda da silvicultura mundial, apresentando um enorme potencial ainda a ser explorado, principalmente no tocante ao melhoramento. Mesmo possuindo um ciclo relativamente curto, se comparado a outras espécies florestais, um ciclo de melhoramento do eucalipto pode levar anos, o que ainda dificulta a obtenção de genótipos elite. Muitas das características de importância econômica e silvicultural só podem ser avaliadas ao final de seu ciclo. Existe, portanto, uma necessidade de acelerar os programas de melhoramento de *Eucalyptus* visando suprir estas demandas em espaços de tempo mais curtos, disponibilizando preferencialmente clones elites ou variedades de sementes superiores (FERREIRA; SANTOS, 1997).

A eficiência do melhoramento genético depende basicamente de duas ações do geneticista: a criação e a identificação de genótipos superiores. Em ambas as ações, a seleção desempenha papel fundamental, na definição dos cruzamentos a serem realizados, visando à criação de novos genótipos e na indicação dos indivíduos superiores a serem usados comercialmente. A seleção genética tem sido praticada pelo procedimento REML/BLUP com base em dados fenotípicos avaliados a campo (FONSECA et al., 2010).

O grande potencial da genética molecular em benefício do melhoramento genético aplicado é a utilização direta das informações de marcadores de DNA na seleção, de forma a permitir alta eficiência seletiva, rapidez na obtenção de ganhos genéticos com a seleção e menor custo, em comparação com a seleção tradicional baseada em dados fenotípicos (RESENDE et al., 2008).

Visando a esses objetivos, Meuwissen et al. (2001) propuseram um novo método de seleção denominado seleção genômica (GS) ou seleção genômica ampla (SGA), a qual vem sendo aplicada com sucesso na área de genética animal e promete revolucionar o melhoramento de plantas, incluindo as espécies perenes de longos ciclos de vida. A seleção genômica é definida como a seleção simultânea para dezenas ou centenas de milhares de marcadores, os quais cobrem o genoma de uma maneira densa, de forma que todos os genes de um caráter quantitativo estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores. Por probabilidade, cada QTL estará em desequilíbrio de ligação com pelo menos um marcador e somente os marcadores em desequilíbrio de ligação com os QTL's é que explicarão os fenótipos e a variação genética. A SGA pode apresentar alta acurácia seletiva baseada exclusivamente em marcadores uma vez que seus efeitos genéticos são estimados a partir da regressão de dados fenotípicos sobre dados genotípicos em uma população dita "de descoberta" ou "de treinamento".

Esse método permaneceu discreto por cerca de cinco anos desde a proposição de Meuwissen et al. em 2001, devido ao fato do número de marcadores moleculares disponíveis à época ser muito limitado e de custo elevado. Recentemente, com o desenvolvimento e redução progressiva do custo de marcadores tipo SNP (Single Nucleotide Polymorphism), polimorfismos de base individual, o método tornou-se viável e muito atrativo. Geneticistas e melhoristas renomados e adeptos de métodos tradicionais têm demonstrado e confirmado a superioridade e exequibilidade prática do método em benefício do melhoramento genético clássico (RESENDE et al., 2008)

Portanto, neste momento são grandes as expectativas da SGA aplicada ao melhoramento genético de espécies perenes e especificamente de *Eucalyptus*, pois se abre uma perspectiva concreta de realizar seleção direta e com isso aumentar o ganho genético por unidade de tempo, permitindo a instalação imediata de testes clonais a partir de tecido juvenil de mudas selecionadas potencialmente eliminando a necessidade dos testes de progênie, e, com isso, podendo reduzir para a metade o

tempo necessário para seleção de clones superiores (GRATTAPAGLIA; RESENDE, 2011).

Neste contexto, Lima (2014) realizou um estudo em que foram desenvolvidos modelos preditivos para diversas características de crescimento e qualidade da madeira para uma população elite de *Eucalyptus* da International Paper. Este trabalho teve por objetivo avaliar a capacidade preditiva para as características de crescimento volumétrico de modelos previamente construídos naquele estudo para características de crescimento volumétrico e qualidade da madeira com base em uma população elite de melhoramento da International Paper do Brasil que atuou como população de descoberta. Os modelos preditivos foram aplicados em quatro populações instaladas em testes de progênies e teste clonal de *Eucalyptus* com características contrastantes para o grau de relacionamento genético: (1) três populações compostas por indivíduos geneticamente relacionados com a população de descoberta e (2) uma população composta por indivíduos geneticamente não relacionados com a população de descoberta.

Dada esta estrutura experimental procurou-se avaliar as seguintes hipóteses de trabalho:

- i) Para populações geneticamente relacionadas, a capacidade preditiva e acurácia de seleção serão equivalentes àsquelas obtidas com a seleção fenotípica tradicional, pois os modelos preditivos capturam seja o componente de relacionamento genético, além do desequilíbrio de ligação a longa e curta distâncias e o efeito de substituição alélica;
- ii) Para populações geneticamente não relacionadas, a capacidade preditiva e acurácia de seleção serão consideravelmente menores em relação àsquelas obtidas para populações relacionadas, pois dependerão da existência de desequilíbrio de ligação a curta distância entre os marcadores SNPs analisados e os polimorfismos causais da variação nas características quantitativas.

2 REVISÃO BIBLIOGRÁFICA

2.1 Melhoramento Florestal Clássico

Os programas de melhoramento florestal são comumente desenvolvidos em ciclos repetidos de seleção e recombinação. As estratégias de melhoramento estabelecem como estes ciclos serão organizados para produzir material genético melhorado a serem utilizados nos plantios comerciais. A estrutura básica de uma estratégia de melhoramento é composta da população base, da população de melhoramento, dos métodos para avaliar e selecionar árvores geneticamente superiores, dos métodos a serem utilizados na recombinação destas árvores para regenerar populações de melhoramento, que serão submetidas a ciclos repetidos de seleção, e dos métodos de multiplicação para prover material genético melhorado em quantidade para plantios comerciais (ASSIS, 1986).

Segundo Santos et al. (2006), têm-se adotado nos programas de melhoramento duas etapas na seleção e avaliação de clones de *Eucalyptus*. A primeira ocorre com grande número de indivíduos, em que se adotam parcelas pequenas e em vários locais, utilizando a seleção precoce na identificação de indivíduos superiores para as características de interesse. Na segunda etapa, os indivíduos selecionados na etapa inicial são avaliados em experimentos maiores (escala piloto), visando avaliar o desempenho representativo em plantio comercial. Essa estratégia tem fornecido bons resultados na área florestal, e sido recomendada como adequada em trabalhos recentes desenvolvidos com dados de campo de empresas florestais (SILVA, 2001).

Fonseca et al. (2010) descreve que, na seleção de genótipos, em uma primeira etapa os programas de melhoramento genético baseiam-se na avaliação e seleção de genótipos superiores de espécies puras, visando utilizar para recombinação e hibridação. Na sequência, testes de progênies de espécies puras e híbridas são instalados para dar continuidade ao melhoramento das espécies puras e obtenção de híbridos superiores para utilização nos plantios comerciais.

A seleção para propagação vegetativa deve basear-se nos valores genotípicos dos candidatos (RESENDE et al., 1995). A utilização de recursos estatísticos permitiu o desenvolvimento de métodos baseados na estimativa do valor genético das famílias e de indivíduos dentro de famílias. Após a seleção, os indivíduos remanescentes são recombinados para contribuir com alelos para a próxima geração. Este procedimento é denominado REML/BLUP (melhor predição

linear não viesada) e a vantagem deste método é que em geral proporciona ganhos genéticos superiores à seleção entre e dentro de famílias, além de promover a manutenção da variabilidade genética (PIRES et al., 2011). Devido ao longo ciclo das espécies florestais, a demora na etapa de avaliação é um dos entraves na recomendação de novos clones. Deste modo, a seleção precoce pode diminuir o tempo requerido para avaliação e seleção, maximizando os ganhos genéticos por unidade de tempo, além de proporcionar maior facilidade para tomada de dados e maior adaptabilidade às mudanças de objetivos (RESENDE et al., 1994).

Desta forma, as avaliações silviculturais e tecnológicas aliadas ao tempo de florescimento das árvores para cruzamentos, avaliações e clonagem fazem com que o tempo decorrido do cruzamento até a obtenção de clones híbridos para plantio seja de 14 anos, impactando diretamente no ganho por unidade de tempo (Fonseca et al., 2010).

2.2 Seleção Genômica Ampla em espécies Florestais

Os programas de melhoramento florestal enfrentam um desafio comum: longos ciclos de reprodução, que podem durar décadas, além da expressão tardia de grande quantidade de características importantes, o que significa grandes investimentos. Na perspectiva de acelerar o melhoramento florestal por seleção ultra-precoce é que tecnologias moleculares ganharam força entre os geneticistas através da seleção assistida por marcadores (MAS). Estes marcadores também foram um meio de aumentar a intensidade de seleção, reduzir esforços de testes de campo e aumentar a precisão na seleção de características de baixa herdabilidade, como o crescimento, e de características de expressões tardias, como qualidade da madeira (GRATTAPAGLIA, 2014).

Neste contexto, a Seleção Genômica Ampla (SGA) foi proposta há 11 anos, como um novo paradigma baseado na seleção assistida de todo o genoma (MAS), capaz de aumentar substancialmente a taxa de ganho genético em animais e plantas, especialmente se combinado com técnicas de reprodução para encurtar o intervalo de gerações (MEUWISSEN et al., 2001; GODDARD; HAYES, 2009). Este conceito ganhou força com avanços das técnicas de genotipagem e de seus custos decrescente em conjunto com novos métodos estatísticos para grandes bases de dados.

A SGA pode ser definida como sendo a seleção simultânea para centenas ou milhares de marcadores, cobrindo todo o genoma, de forma que todos ou a grande maioria dos alelos de interesse estarão em desequilíbrio de ligação (DL) com pelo menos um ou mais marcadores genotipados e, portanto, seus efeitos devidamente capturados nos modelos preditivos (RESENDE et al., 2008).

Entretanto para a aplicação de SGA no melhoramento de plantas alguns requisitos se fazem necessários. Além de dispor de plataformas de genotipagem em larga escala e custo reduzido, possível hoje com SNPs e algumas outras tecnologias emergentes de genotipagem por sequenciamento, as populações de melhoramento devem ter preferencialmente tamanhos efetivos menores do que $N_e = 100$ o que resulta em extensões mais longas de DL. No caso de *Eucalyptus* esta situação é plenamente satisfeita com as estratégias mais atuais de grupos reprodutivos envolvendo apenas algumas dezenas de genitores que são inter-cruzados gerando amplas progênies nas quais é praticada seleção individual seguida de instalação de testes clonais. Esta estratégia busca explorar o vigor híbrido e a combinação de características e identificar indivíduos excepcionais, que consolidem diversas características desejáveis (GRATTAPAGLIA; RESENDE, 2011).

Segundo Resende et al. (2008), na prática da seleção genômica ampla, três populações ou conjuntos de dados são necessários:

População de Descoberta ou Treinamento. Esse conjunto de dados contempla um grande número de marcadores SNPs avaliados em um número moderado de indivíduos, os quais devem ter seus fenótipos avaliados para os vários caracteres de interesse. Equações de predição de valores genéticos genômicos são obtidas para cada caráter de interesse. Essas equações associam a cada intervalo marcador o seu efeito (predito por BLUP) no caráter de interesse.

População de Validação. Esse conjunto de dados é menor do que aquele da população de descoberta e contempla indivíduos avaliados para os marcadores SNPs e para os vários caracteres de interesse. As equações de predição de valores genéticos genômicos são testadas para verificar suas acurácias nessa amostra independente. Para computar essa acurácia, os valores genéticos genômicos são preditos (usando os efeitos estimados na população de descoberta) e submetidos à análise de correlação com os valores fenotípicos observados.

População de Seleção. Esse conjunto de dados contempla apenas os marcadores SNPs avaliados nos candidatos à seleção. Essa população não necessita ter os seus fenótipos avaliados. As equações de predição derivadas na população de descoberta são então usadas na predição dos fenótipos futuros dos candidatos à seleção. Mas, a acurácia seletiva associada refere-se àquela calculada na população de validação.

A SGA é ampla porque atua em todo genoma, capturando todos os efeitos genômicos que afetam um caráter quantitativo sem a necessidade prévia de identificar os marcadores com efeitos significativos e de mapear QTL's. Valores genéticos genômicos associados a cada marcador ou alelo são usados para fornecer o valor genético genômico global de cada indivíduo. A diferença básica na predição de valores genéticos tradicionais e na predição de valores genéticos genômicos é que nos primeiros, informações fenotípicas são utilizadas visando inferências sobre os efeitos dos genótipos nos indivíduos e o último, informações genotípicas são usadas visando inferências sobre os valores fenotípicos futuros (valores genéticos genômicos preditos) dos indivíduos (RESENDE, 2008).

O processo de SGA usa associações de um grande número de marcadores SNPs em todo genoma com fenótipo, capitalizando no desequilíbrio de ligação entre os marcadores e QTLs proximamente ligados, sem uma prévia escolha de marcadores com base nas significâncias de suas associações com o fenótipo. Predições são então obtidas para os efeitos dos haplótipos de marcadores ou dos alelos em cada marcador. Essas predições derivadas de dados fenotípicos e de genótipos de SNPs em alta densidade numa geração são então usadas para obtenção de valores genéticos genômicos (VGG) dos indivíduos de gerações subsequentes, tendo por base os seus próprios genótipos marcadores. Sob seleção genômica, todos os candidatos à seleção (indivíduos sem observação fenotípica) poderão ser avaliados para quaisquer ambientes, desde que tais ambientes possuam equações de predição derivadas para os próprios e com alta acurácia (RESENDE, 2008).

Na prática, em SGA, a população de descoberta (ou treinamento) envolve a genotipagem de centenas ou milhares de indivíduos e um painel de marcadores para características fenotípicas de interesse. A partir destes conjuntos de dados, modelos de predição são construídos e validados na população de "validação". Este

modelo é subsequentemente utilizado para calcular o valor genético genômico dos candidatos à seleção para a qual apenas os genótipos são registrados. Validações experimentais de SGA em espécies florestais são extremamente necessárias para fornecer o necessário apoio para a sua recomendação no melhoramento aplicado (GODDARD; HAYES, 2009; GRATTAPAGLIA, 2014).

Em um experimento com duas populações elite de Eucalipto, Resende et al. (2012) mostraram que os ganhos esperados com a SGA nestas populações derivam da possibilidade de reduzir o tempo de um ciclo de melhoramento via seleção ultra-precoce, ao predizer os fenótipos em fase de muda. Os ganhos na eficiência de seleção podem chegar a 50% se o tempo do ciclo de melhoramento for reduzido pela metade e pode exceder 100% se a abordagem for mais agressiva, eliminando por completo os testes de progênies e uma etapa de teste clonal. O efeito líquido é o aumento da intensidade de seleção não só para crescimento, mas para todas as características de interesse do melhorista simultaneamente, como por exemplo, as propriedades da madeira, antecipando assim o ciclo de desenvolvimento de novos clones elites em seis a nove anos.

Desta forma, a seleção genômica em espécies florestais e culturas perenes é uma ferramenta revolucionária e particularmente atrativa devido ao seu potencial de ganho crescente por unidade de tempo, além de potencialmente melhorar a seleção para características quantitativas e de baixa herdabilidade (JANNINK et al., 2010; GRATTAPAGLIA; RESENDE, 2011; RESENDE et al., 2012a).

2.3 Valores Genéticos Genômicos (VGGs) e Acurácias de Seleção

Resende (2007) explica que a estimação dos VGGs usa um conjunto de dados de referência que inclui indivíduos com genótipos (marcadores) e fenótipos conhecidos. Os valores genéticos estimados dos marcadores para todos os caracteres quantitativos de interesse do melhorista e efetivamente fenotipados na população de descoberta são usados para predição dos valores genéticos genômicos de indivíduos jovens candidatos à seleção que foram genotipados para os marcadores, mas não possuem informação fenotípica.

A escolha de um método ótimo de estimação/predição de valores genéticos deve basear-se no critério de uma inferência mais precisa e realista possível, a qual deve ser avaliada segundo parâmetros estatísticos adequados. Nesse contexto, os parâmetros mais importantes são a acurácia seletiva e o erro quadrático médio

(RESENDE et al., 2011). A acurácia é conceituada como a correlação entre o valor genético verdadeiro e aquele estimado a partir das informações genóticas (marcadores) e/ou fenotípica dos indivíduos (RESENDE et al., 2011; GRATTAPAGLIA, 2014) e depende da proporção da variação genética explicada pelos marcadores e da acurácia da predição dos efeitos dos haplótipos. Um estimador acurado apresenta menor diferença quadrática entre valores verdadeiros e estimados, ou seja, apresenta mínimo erro quadrático médio (RESENDE et al., 2011).

Em resumo, a acurácia da SGA depende de cinco fatores: (i) da herdabilidade do caráter; (ii) do número de locos controlando o caráter e da distribuição de seus efeitos; (iii) do número de indivíduos na população de descoberta; (iv) do tamanho efetivo populacional; (v) do espaçamento entre marcadores, o qual depende do seu número e do tamanho do genoma. Os dois primeiros fatores não estão sobre o controle do melhorista. Os três últimos podem ser modificados pelo melhorista visando aumentar a acurácia da SGA (RESENDE et al., 2011).

Várias abordagens analíticas em espécies florestais foram propostas para seleção baseada nos valores genéticos genômicos: Resende et al. (2012a) em *Eucalyptus*, Resende et al. (2012b) em *Pinus*, Beaulieu et al. (2014a, 2014b) em *Picea abies*, Gamal El-Dien et al. (2015) em *Picea engelmannii*. Grattapaglia (2014) elaborou uma Tabela com os principais resultados da SGA em espécies florestais com os experimentos realizados até 2013, Tabela esta que atualizada com os trabalhos mais recentes publicados em 2014 e 2015 (Tabela 1).

Tabela 1 – Resumo de resultados experimentais da SGA em espécies florestais apresentados por Grattapaglia (2014) atualizado com os artigos publicados em 2014 e 2015 (continua)

Espécies	Tipo de População	Tamanho da Pop. Treinamento	Tipo de marcador	Tratamentos	Acurácia	Referencia
<i>E. grandis x E. urophylla</i>	Teste de progênies, 43 famílias de irmãos completos, 11 genitores elites	738	3129 DArT	Circunferência	0,74	Resende et al. (2012a)
				Altura	0,79	
				Rendimento Gravimétrico	0,78	
				Rendimento em Celulose	0,88	
<i>E. grandis, E. urophylla, E. globulus, e híbridos</i>	Teste de progênies, 232 famílias de irmãos completos, 51 genitores elites	920	3564 DArT	Circunferência	0,73	Resende et al. (2012a)
				Altura	0,66	
				Rendimento Gravimétrico	0,65	
				Rendimento em Celulose	0,55	
<i>Pinus taeda</i>	Teste de progênies, 61 famílias de irmãos completos, 32 genitores elites	800	4852 SNP	Diâmetro	0,65 - 0,73	Resende et al. (2012b)
				Altura	0,64 - 0,74	
<i>Pinus taeda</i>	Teste de progênies, 61 famílias de irmãos completos, 32 genitores elites	951	4853 SNP	Crescimento	0,38 - 0,49	Resende et al. (2012c)
				Desenvolvimento	0,24 - 0,51	
				Resistência à ferrugem	0,23 - 0,34	
				Liginina	0,17	
<i>Pinus taeda</i>	13 famílias de irmãos completos e 149 parentais comuns	149	3406 SNP	Liginina	0,66 - 0,76	Zapata-Venezuela et al. (2012)
				Celulose	0,61 - 0,83	
				Altura	0,47 - 0,52	
				Volume	0,30 - 0,56	
<i>Picea abies</i>	Teste de progênies, 214 famílias meios-irmãos, 43 genitores	1694	6385 SNP	Qualidade da madeira	0,32 - 0,43	Beaulieu et al. (2014a)
<i>Picea abies</i>	Teste de progênies, 1748 indivíduos, 39 genitores em diferentes locais	1748	6932 SNP	Crescimento	0,52 - 0,69	Beaulieu et al. (2014b)
			Qualidade da madeira	0,71 - 0,79		

Tabela 1 – Resumo de resultados experimentais da SGA em espécies florestais apresentados por Grattapaglia (2014) atualizado com os artigos publicados em 2014 e 2015 (conclusão)

Espécies	Tipo de População	Tamanho da Pop. Treinamento	Tipo de marcador	Tratamentos	Acurácia	Referencia
<i>E. grandis x E. urophylla</i>	Teste de progênies, 58 irmãos completos, 45 genitores	1000	EUchip60K	Crescimento	0,77 - 0,92	Lima (2014)
<i>Picea engelmannii</i>	Teste de Progênies, 25 famílias de meios-irmãos	1126	1,2M SNP	Crescimento Qualidade da Madeira	0,63 – 0,77 0,62 – 0,72	Gamal El-Dien et al. (2015)

Acurácias teóricas da ordem de 85 % para a SGA foram relatadas por Meuwissen et al. (2001) para uma população com 2.200 indivíduos com avaliações fenotípicas utilizando simulações. Tais autores relataram também que equações de predição acuradas (71 %) foram obtidas mesmo para populações de descoberta de tamanho modesto, tal qual com 500 indivíduos com avaliações fenotípicas.

Grattapaglia e Resende (2011) apresentaram um estudo detalhado baseado em simulações determinísticas das perspectivas da SGA no melhoramento florestal avaliando os diferentes aspectos que impactam a capacidade preditiva. Foram estudados o tamanho efetivo da população e a densidade de marcadores que conjuntamente determinam a capacidade de capturar o desequilíbrio de ligação, além da herdabilidade e o número de QTLs envolvidos no controle da característica e o tamanho da população de treinamento. Os resultados indicaram que a SGA tem o potencial de melhorar radicalmente a eficiência do melhoramento genético reduzindo o tempo do ciclo de melhoramento aumentando o ganho por unidade de tempo. A precisão de seleção do REML/BLUP convencional é atingida pela SGA mesmo com uma densidade de marcadores da ordem de ~2 marcadores/cM para tamanhos efetivos populacionais $N_e \leq 30$, enquanto que 20 marcadores/cM são necessários para tamanhos efetivos maiores. Em outro estudo baseado em simulações, Resende et al. (2011) verificaram que, para uma população de eucalipto com tamanho efetivo $N_e = 20$ a acurácia seletiva esperada com a SGA seria de 0.79, para um tamanho amostral de $N = 4000$ indivíduos. Esse valor supera a acurácia máxima (0.70) para a seleção de indivíduos pelo REML/BLUP tradicional na idade adulta.

Resende et al. (2012) relataram acurácias experimentais variando de 55 até 88% para diferentes características de crescimento e qualidade da madeira em um experimento envolvendo duas populações elite de *Eucalyptus* no Brasil genotipadas com cerca de 3.000 marcadores dominantes DArT. Resende et al. (2012) também relataram resultados semelhantes em *Pinus taeda* nos EUA ao utilizar cerca de 3.500 marcadores SNPs. Estes dois estudos foram os primeiros a demonstrar experimentalmente em espécies florestais as perspectivas positivas da utilização da SGA, embora tenham utilizado plataformas de marcadores limitadas em densidade e não tenham avaliado aspectos importantes como o nível de relacionamento genético entre os indivíduos utilizados na validação cruzada.

Lima (2014) examinou o potencial da SGA para quinze características de crescimento e propriedades químicas e físicas da madeira, em uma população de 1000 indivíduos derivados de uma população elite de melhoramento utilizando genotipagem de alta densidade com 29.090 SNPs. Dados fenotípicos e genotípicos foram utilizados na construção de modelos preditivos usando RR-BLUP e LASSO Bayesiano. Capacidades preditivas elevadas foram obtidas para caracteres relacionados à lignina (0,89 a 0,97), densidade (0,86) e crescimento (0,70-0,76), alcançando ou ficando apenas cerca de 10 a 15% abaixo das capacidades preditivas obtidas por REML/BLUP fenotípico. A acurácia preditiva da SGA alcançou 70% em comparação com a seleção fenotípica na seleção das melhores 30 árvores da população e de 60% quando a seleção foi aplicada para crescimento, densidade, relação Siringil:Guaiacil e lignina. Neste estudo não foi analisada a questão do impacto do relacionamento genético entre indivíduos utilizados nas validações cruzadas, aspecto que certamente impactou positivamente as capacidades preditivas alcançadas.

Em um estudo para coníferas (*Picea glauca*), Beaulieu et al. (2014a) utilizando validação cruzada relatou acurácias que variaram de 0,71 a 0,79 para qualidade da madeira e 0,52 a 0,69 para crescimento quando os indivíduos foram analisados dentro do mesmo local. A predição genômica para ambientes não testados permaneceu moderadamente elevada para densidade da madeira ($r \geq 0,61$) e caiu significativamente para crescimento ($r \geq 0,24$), enfatizando a necessidade de desenvolvimento de modelos preditivos nos diferentes ambientes ou a utilização de modelos que incluam a interação genótipo x ambiente para características de crescimento. Neste mesmo estudo foi avaliado o impacto do relacionamento genético entre indivíduos na população de treinamento e de validação na capacidade preditiva. Com forte relacionamento genético entre os indivíduos utilizados na validação cruzada acurácias dentro do mesmo ambiente e população foram elevadas para qualidade da madeira ($r = 0,71-0,79$) e moderadamente altas para crescimento ($r = 0,52-0,69$). Por outro lado ao remover o relacionamento genético entre indivíduos, as acurácias preditivas caíram consideravelmente para todas as populações e características chegando a zero entre grupos de indivíduos sem qualquer ancestralidade em comum. Em vistas da necessidade de elevado relacionamento genético para a obtenção de acurácias satisfatórias, eles recomendam a construção de modelos preditivos dentro da mesma população de

melhoramento. Vale notar, entretanto, que este estudo utilizou apenas 6.932 marcadores SNPs para um genoma de *Picea* da ordem de 2.100 cM, sendo que os SNPs eram localizados exclusivamente dentro de modelos gênicos preditos e não de forma homogênea ao longo do genoma. Em outras palavras a densidade de marcadores além de baixa foi muito mal distribuída o que reduziu consideravelmente a capacidade de captura de desequilíbrio de ligação entre marcadores e polimorfismos envolvidos no controle das características alvo.

Em outro estudo, Beaulieu et al. (2014b) estudaram populações de treinamento de tamanho efetivo elevado compostas por famílias de meios irmãos. Modelos de SGA foram construídos para prever os valores genéticos estimados utilizando toda plataforma SNPs disponível e foram validados utilizando vários esquemas de validação cruzada. A acurácia dos valores genéticos genômicos (VGGs) estimados variou de 0,327-0,435 (BEAULIEU et al., 2014b).

Estes resultados experimentais vêm demonstrando em diferentes espécies florestais o potencial da SGA no melhoramento florestal, destacando, entretanto, os importantes aspectos relativos ao impacto da densidade de marcadores, interação genótipo x ambiente e principalmente do relacionamento genético entre os indivíduos utilizados para o treinamento dos modelos e os indivíduos sobre os quais se deseja praticar a SGA.

2.4 Perspectivas e desafios da Seleção Genômica Ampla em espécies Florestais

Utilizar grandes números de marcadores de DNA para prever o potencial genético (Seleção Genômica Ampla) é uma nova fronteira em programas de melhoramento animal e de plantas. A SGA vem sendo rotineiramente empregada na seleção de touros superiores para criação de gado leiteiro. Em árvores, algumas provas de conceito e estudos empíricos sugerem que a SGA poder ser bem sucedida, no entanto, a aplicação da SGA em florestas ainda está no início (GRATTAPAGLIA; RESENDE 2011; IWATA et al., 2011; ISIK, 2014).

A era genômica está trazendo novas oportunidades para os melhoristas florestais. Sequenciamento completo do genoma das principais espécies florestais vem sendo recentemente concluída: *Populus trichocarpa* (TUSKAN et al., 2006), *Picea glauca* (BIROL et al., 2013), *Picea abies* (NYSTEDT et al., 2013), *E. grandis* (MYBURG et al., 2014). Estes genomas vem sendo importantes recursos para a

descoberta de genes, detecção de marcadores SNPs, construção de mapas genéticos de alta densidade e desenvolvimento de plataformas de genotipagem de alto rendimento, como o EuchIP 60k proposto por SILVA-JUNIOR et al. (2015). Tais ferramentas serão muito importantes para o melhoramento genético florestal futuro.

Um grande desafio continua a sendo a grande quantidade de dados e convertê-los para conhecimento usando algoritmos estatísticos e poder de computação. A aplicação rotineira de SGA vai exigir novas habilidades de computação para gerenciar e processar dados. Além disso, populações bem estruturadas, com pedigrees conhecidos e alta precisão de fenotipagem também são necessários para implementar a SGA (ISIK, 2014).

Outro desafio são as interações genótipo x ambiente ($G \times E$). Elas são fato em todos os programas de melhoramento genético florestal, embora em níveis diferentes, dependendo das espécies, variabilidade ambiental e tipo de material plantado, seja famílias ou clones, com clones sendo tipicamente mais interativos do que famílias (GRATTAPAGLIA, 2014). Dados experimentais de SGA em duas populações independentes de eucalipto avaliadas em dois ambientes diferentes mostraram que o modelo preditivo desenvolvido para uma população não poderia prever fenótipos na outra. Os resultados indicaram que os modelos preditivos serão em princípio para população específica (RESENDE et al., 2012a).

Enquanto todos estes desafios estão sendo estudados, cabe aos melhoristas explorar estratégias de como utilizar a seleção genômica na tomada de decisões, tais como a estratégia de uso da predição do valor genético de cada população, experimentação de campo, que ainda deverá ser importante nos primeiros ciclos, mas deve diminuir à medida que mais indivíduos são genotipados (ISIK, 2014). Portanto, a era genômica está trazendo novas oportunidades para a aceleração e dinamização do melhoramento florestal, porém ainda existem diversas questões a serem investigadas experimentalmente para o uso operacional da SGA.

3 MATERIAL E MÉTODOS

3.1 População de Treinamento e Modelos Preditivos

Em 2014 o trabalho proposto por Lima desenvolveu modelos preditivos utilizando uma população elite da International Paper do Brasil como população de descoberta. Um total de 1000 indivíduos foram fenotipados para diversas características de relevância para o melhoramento e genotipados com um chip EucHIP60k (SILVA-JUNIOR et al., 2015), que fornece até 60 mil marcadores SNPs em diferentes espécies de *Eucalyptus*.

Estes 1000 indivíduos foram selecionados em um testes de progênies híbridas de irmãos-germanos, com tamanho efetivo de 46, cujos critérios de seleção foram garantir a distribuição normal dos indivíduos, sendo selecionados tanto indivíduos superiores quanto indivíduos inferiores, que representassem todas as famílias contidas no teste de progênies.

Modelos preditivos foram construídos para 15 características: crescimento volumétrico (DAP – diâmetro a 1,30m de altura - altura, IMA – Incremento Médio Anual), teor de celulose, teor de hemicelulose, relação SG, lignina insolúvel, lignina solúvel, lignina total, densidade básica, ângulo de microfibrila, comprimento de fibra, espessura de fibra e coarseness (LIMA, 2014). A construção dos modelos preditivos foi feita com base na estimativa dos valores genéticos genômicos (VGG) para as várias características nos 1000 indivíduos da população de descoberta da empresa, utilizando os dados genotípicos e fenotípicos levantados e diferentes métodos analíticos incluindo regressão de cumieira (RR-BLUP) e abordagens Bayesianas.

A acurácia de seleção das equações de predição foi avaliada por meio do método "Jackknife" no qual para cada característica cada um dos 1000 indivíduos foi predito pelos demais 999 que foram utilizados para construir o modelo e com base nele estimar o VGG do indivíduo. Isso foi repetido 1000 vezes de forma a maximizar o tamanho da população de descoberta e de validação que os dados permitiam. Acurácias preditivas da ordem de 75 até 85% foram obtidas com base em um catálogo com 29090 marcas com seus efeitos relativos. Este catálogo de SNPs com seus efeitos foram utilizados para a avaliação das capacidades preditivas nas populações de avaliação neste trabalho

3.2 Populações utilizadas para avaliação da acurácia dos modelos preditivos

Os modelos preditivos foram testados em duas populações, denominadas a partir de agora “Populações de avaliação” que diferiam quanto à existência ou não de relacionamento genético com a população de descoberta, ambas instaladas em testes de progênies ou testes clonais de *Eucalyptus* da empresa International Paper do Brasil, para um total de 400 indivíduos genotipados. Um resumo das principais características de cada população é apresentado nas Tabelas 2 e 3.

3.2.1 Populações Geneticamente Relacionadas

MHSFC200P: Teste de Progênies Híbridas *E. grandis* x *E. urophylla*

Teste de progênies de polinização controlada de *E. grandis* x *E. urophylla* de onde foi amostrada a População de Treinamento utilizada por Lima (2014) para desenvolver os modelos preditivos. Para este trabalho, dentro do mesmo experimento foram selecionados indivíduos pertencentes às mesmas famílias dos indivíduos utilizados na população de descoberta, porém árvores diferentes, e também indivíduos de famílias não genotipadas na população de descoberta, porém relacionadas geneticamente com a população de descoberta pelo lado materno ou paterno.

O experimento está localizado no município de Brotas-SP, foi instalado em 2006 e é composto por 58 tratamentos e quatro testemunhas plantadas em blocos ao acaso, parcelas lineares de seis plantas com oito repetições, em um total de 3.024 árvores.

Para esta população de avaliação foram genotipados 100 indivíduos geneticamente relacionados com a população de descoberta, sendo considerados os indivíduos fenotipicamente mais representativos da população. Aos cinco anos foi realizada mensuração para obtenção das características fenotípicas de altura e diâmetro.

MHGI0002R: Teste de Progênies de *Eucalyptus urophylla*

Teste de progênies de meios-irmãos (polinização aberta), cujos indivíduos são progênies oriundas das matrizes de *E. urophylla* polinizados por *E. urophylla* e *E. grandis*, utilizadas nos cruzamentos controlados que deram origem à População de Treinamento utilizada por Lima (2014).

O experimento está localizado no município de Aguaí-SP, foi instalado em 2009 e é composto por 69 tratamentos e seis testemunhas plantadas em blocos ao acaso, parcelas lineares de seis plantas com seis repetições, em um total de 2.700 árvores.

Nesta população foram genotipados 100 indivíduos que possuem a princípio relacionamento genético com a população de treinamento, embora por serem de polinização aberta, não controlada, o nível de relacionamento poderá ser menor do que na população MHSFC200P, além da possibilidade real de contaminação de pólen externo o que poderia gerar inclusive indivíduos não relacionados. Aos cinco anos foi realizada mensuração para obtenção das características fenotípicas de altura e diâmetro.

CHNS0048R: Teste Clonal de *E. grandis* x *E. urophylla*

Teste clonal composto por clones selecionados em teste de progênes híbridas das mesmas famílias que deram origem à População de Treinamento. O experimento está localizado no município de Mogi Guaçu-SP, foi instalado em 2012 e é composto por 95 tratamentos e sete testemunhas plantadas em blocos ao acaso, uma planta por parcela e 20 repetições, em um total de 2.184 árvores. Nesta população foram genotipados 100 indivíduos geneticamente relacionados com a população de descoberta, sendo considerados os indivíduos fenotipicamente mais representativos da população. Aos dois anos foi realizada mensuração para obtenção das características fenotípicas de altura e diâmetro.

3.2.2 População Geneticamente Não-Relacionada

MHMA0005P: Teste de Progênes de *Eucalyptus grandis*

Teste de progênes de polinização aberta, cujos indivíduos são progênes de matrizes elites de *E. grandis* polinizados por *E. grandis*, geneticamente não relacionadas com a população de descoberta. O experimento está localizado no município de Altinópolis-SP, foi instalado em 2011 e é composto por 71 tratamentos e cinco testemunhas plantadas em blocos ao acaso, parcelas lineares de seis plantas com sete repetições, em um total de 3.192 árvores. Nesta população foram genotipados 100 indivíduos geneticamente não-relacionados com a população de descoberta, sendo considerados os indivíduos fenotipicamente mais representativos da população. Nota-se também o fato que se trata de material de *E. grandis* puro, ao

contrário das demais populações que são híbridas de *E. grandis* x *E. urophylla*. Aos três anos foi realizada mensuração para obtenção das características fenotípicas de altura e diâmetro.

Tabela 2 – Descrição dos experimentos das populações de avaliação e da população de descoberta

ATRIBUTO	CHNS0048R	MHGI002R	MHSFC200P	MHMA0005P	Pop. Descoberta - MHSFC200P
Tipo de Experimento	Teste Clonal	Teste de Progênes Polinização Aberta	Teste de Progênes Polinização Controlada	Teste de Progênes Polinização Aberta	Teste de Progênes Polinização Controlada
Espécies	híbridos	<i>E. urophylla</i>	híbridos	<i>E. grandis</i>	híbridos
Grau de relação genética com Pop. Descoberta	relacionada	relacionada	relacionada	não-relacionada	-
Grau de parentesco com a Pop. Descoberta	clones	meios irmãos	irmãos germanos	nenhum	-
Numero de Arvores no Experimento	2184	2700	3024	3192	3024
Numero de famílias	95	69	58	71	58
Numero testemunhas	7	6	4	5	4
Delineamento	blocos ao acaso	blocos ao acaso	blocos ao acaso	blocos ao acaso	blocos ao acaso
Tipo de parcela	STP	parcela linear	parcela linear	parcela linear	parcela linear
Numero de plantas por parcela	1	6	6	6	6
Numero de blocos	20	6	8	7	8
Município	Mogi Guacu	Aguai	Brotas	Altinópolis	Brotas
Altitude m	700	650	600	600	600
Temperatura media °C	20,3	21,4	21,4	20,4	21,4
Precipitação media mm	1344	1458	1209	1338	1209
Tipo de solo	Latossolo	Latossolo	Neossolo	Neossolo	Neossolo
Número de indivíduos genotipados	100	100	100	100	1000
Número genitores	91	69	46	71	46
Data de plantio	23/11/2012	12/11/2009	18/07/2006	09/11/2011	18/07/2006
Idade de genotipagem	2,07	5,11	8,43	3,12	6,17
Idade de medição	2,28	4,91	4,53	3,44	4,53

Tabela 3 – Cruzamentos entre espécies nas populações de avaliação e número de indivíduos genotipados por cruzamento

Cruzamentos	CHNS0048R	MHGI0002R	MHMA0005P	MHSFC200P
CAMALDULENSISxGRANDIS	0/0	0/0	0/0	3/48
GRANDIS x ?	2/60	0/0	100/2982	0/48
GRANDISxUROGRANDIS	4/80	0/0	0/0	23/672
GRANDISxUROPHYLLA	47/940	0/0	0/42	56/1536
SALIGNA x ?	8/160	0/36	0/0	0/0
UROGRANDIS x ?	12/317	0/144	0/84	0/96
UROGRANDISxUROGRANDIS	7/157	0/0	0/0	5/96
UROGRANDISxUROPHYLLA	2/40	0/0	0/0	11/432
UROPHYLLA x ?	18/430	100/2520	0/84	0/96

Nota: O símbolo “?” significa que o macho parental é desconhecido (progênie de meios-irmãos); O denominador fração é o total de indivíduos apresentados em cada população, o numerador fração é a porção genotipados.

3.3 Fenotipagem das populações de avaliação e estimativa dos parâmetros genéticos

Foram feitas mensurações de crescimento volumétrico dos indivíduos das populações de avaliação. Foram mensurados diâmetro (DAP) e altura das árvores para determinação do volume individual (Volume) e incremento médio anual (IMA).

O volume individual e IMA foram determinados a partir da equação de predição desenvolvida pela International Paper:

$$Vol = exp((-10,3741 + 1,9801 \cdot \ln(DAP) + 1,0245 \cdot \ln(HEST)))$$

Em que:

DAP = diâmetro à 1,3 metros de altura em cm

HEST = altura em m

$$IMA = vol \cdot \left(\frac{10000}{\text{espaçamento}} \right) \cdot \frac{1}{idade}$$

Em que:

IMA = Incremento Médio Anual em $m^3 \text{ ha}^{-1} \text{ ano}^{-1}$

Vol = volume individual em m^3

Espaçamento = área por planta em m^2

Idade = idade da medição

As análises foram feitas utilizando o software R versão 3.2.2. (2015), com os pacotes lme4 de acordo com os modelos:

Famílias

$$y = X_{\beta} + Z_s + W_f + T_b + P_p + e$$

Clones

$$y = X_{\beta} + Z_s + W_c + T_b + e$$

Em que:

β = efeito de bloco, considerado fixo

s = efeito de espécie, considerado aleatório

f = efeito de família, considerado aleatório

b = efeito de local, considerado aleatório

p = efeito de parcela, considerado aleatório

e = efeito residual

c = efeito de clones, considerado aleatório

Para as estimativas das herdabilidades, foram utilizadas as seguintes equações, segundo Resende (2014).

$$h_i^2 = \frac{\sigma_G^2}{\sigma_y^2}$$

$$\sigma_y^2 = \sigma_B^2 + \sigma_S^2 + \sigma_G^2 + \sigma^2$$

$$h_t^2 = \frac{\sum_i^3 h_i^2}{3}$$

Em que:

$h_i^2 = \text{herdabilidade individual}$

$\sigma_G^2 = \text{variância genotípica}$

$\sigma_y^2 = \text{variância fenotípica}$

$\sigma_B^2 = \text{variância do bloco}$

$\sigma_S^2 = \text{variância da espécie}$

$\sigma_G^2 = \text{variância de família}$

$\sigma^2 = \sigma_e^2$

$h_f^2 = \text{herdabilidade individual média de famílias}$

3.4 Genotipagem das populações de avaliação

Foram coletadas amostras de folhas para extração de DNA de todos os indivíduos das duas populações de avaliação. Esta extração foi realizada pela empresa Heréditas (Brasília, DF) utilizando o método CTAB (GRATTAPAGLIA; SEDEROFF, 1994) e quantificados em equipamento nanodrop para concentrações finais entre 20-40ng.uL⁻¹.

As amostras de DNA dos 400 indivíduos foram genotipadas pela Geneseek (Lincoln, NE), empresa que presta serviço com o chip de SNPs EuchIP60k.br desenvolvido por (SILVA JUNIOR et al., 2015), e que fornece até um total de 59.222 marcadores SNPs polimórficos em diferentes espécies de *Eucalyptus* utilizando a tecnologia Infinium (Illumina).

Os dados genotípicos foram extraídos a partir de dados de intensidade de sinal utilizado algoritmos de clusterização otimizados para *Eucalyptus grandis* e *E. urophylla* desenvolvidos por Silva Junior et al. (2015)

3.5 Acurácias de Seleção e Capacidades Preditivas

As análises foram feitas utilizando o software R, através do pacote RRBLUP. Foram utilizados para as análises somente SNPs com frequência de declaração de genótipo (call rate) ≥ 0.90 e MAF (menor frequência alélica) ≥ 0.01 . Foram adotadas as mesmas marcas utilizadas no trabalho de Lima (2014). Para tal, foi realizada uma verificação do catálogo de marcas derivadas do trabalho de Lima (2014) com o catálogo de marcas das quatro populações genotipadas de forma a utilizar

exatamente os mesmos marcadores SNPs em ambos os catálogos na população de treinamento e nas populações de avaliação.

A primeira análise consistiu em um teste simples no qual os efeitos estimados dos genótipos aos 29090 SNPs no trabalho de Lima (2014) para população elite de *Eucalyptus* da International Paper foram multiplicados pelos genótipos observados nos 400 indivíduos genotipados das populações de avaliação e gerados os VGG de cada um dos 400 indivíduos. Em seguida foi calculada a correlação entre os VGGs e os fenótipos (VG) observados.

A segunda análise consistiu na construção de novos modelos preditivos utilizando os 400 indivíduos genotipados e fenotipados neste estudo como população de treinamento. Foi utilizada uma validação cruzada independente (Resende et al. 2014) em que 300 indivíduos de 3 populações foram utilizados como população de treinamento e 100 indivíduos de uma quarta população como população de validação.

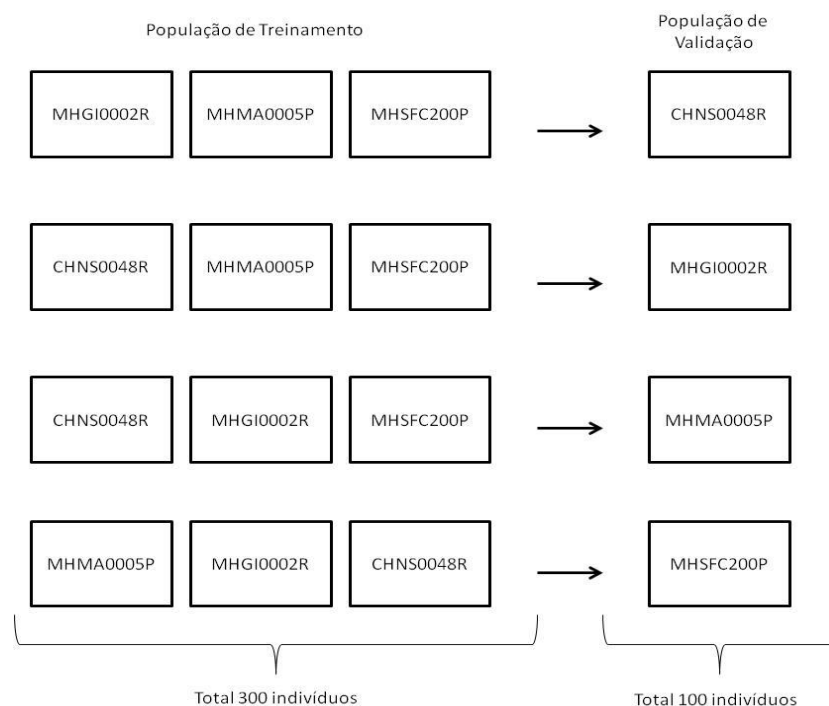


Figura 1 – Esquema de validação independente feitos nas quatro populações de estudo

O modelo misto utilizado para famílias em parcela linear foi:

$$y = X_{\beta} + Z_s + W_f + T_b + P_p + e$$

O modelo misto utilizado para clones em parcelas single tree plot foi:

$$y = X_{\beta} + Z_s + W_c + T_b + e$$

O valor genético foi calculado segundo o modelo:

$$\hat{g} = Z_{\hat{s}} + W_{\hat{f}} + \hat{e}$$

Em que:

β = efeito de bloco, considerado fixo

s = efeito de espécie, considerado aleatório

f = efeito de família, considerado aleatório

b = efeito de local, considerado aleatório

p = efeito de parcela, considerado aleatório

e = efeito residual

c = efeito de clones, considerado aleatório

Na sequência, foram feitos os testes para determinar os conjuntos de marcas de maior efeito dentro das 29090 marcas. Para tal, fez-se uma primeira rodada com RRBLUP onde ordenou-se os efeitos de cada marca da maior para a menor magnitude. A partir disso, foram testados quais marcadores maximizariam as capacidades preditivas utilizando os marcadores do maior para o menor efeito. Para tal, fez-se uso da sequência de Fibonacci, em que foram testados 21 conjuntos de marcas e determinados os marcadores de maior efeito que foram em seguida utilizados nos testes de determinação da máxima acurácia.

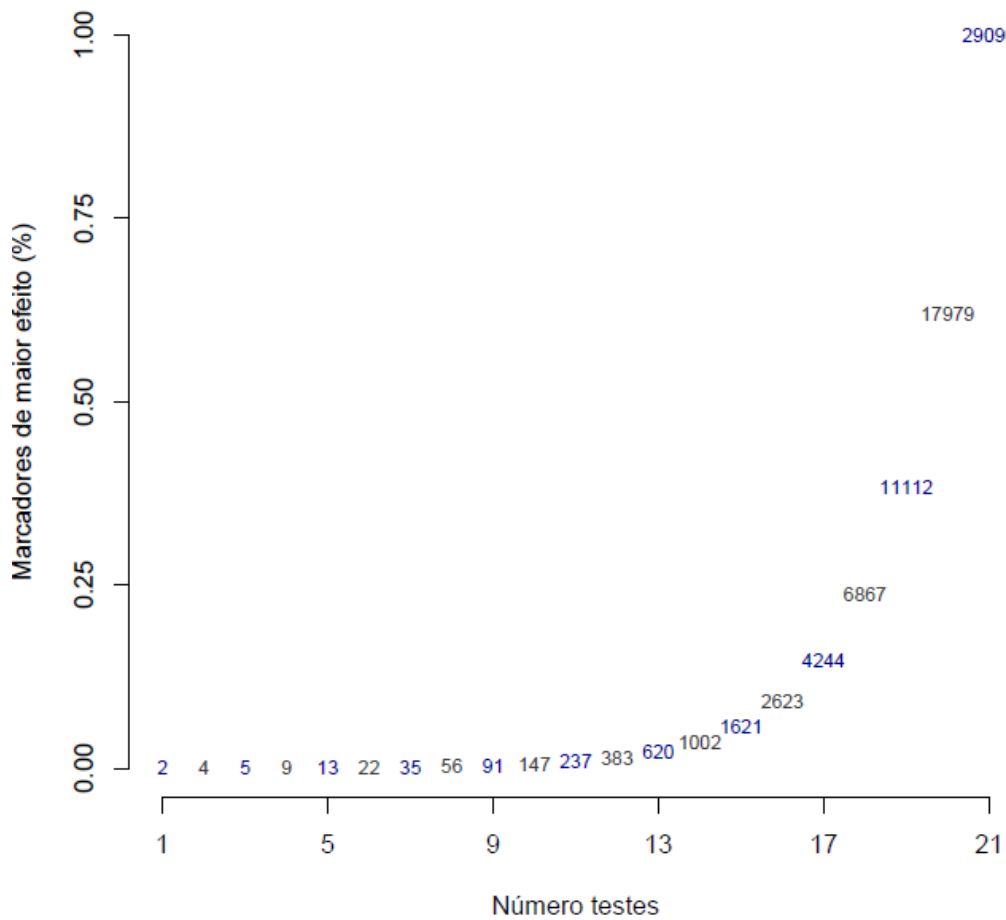


Figura 2 – Teste com os marcadores de maior efeito. Os marcadores destacados em azul foram os testados

Para estimativa das acurácias foram utilizadas as estimativa das herdabilidades moleculares conforme equação proposta por Resende (2014).

$$h_M^2 = \frac{\sum_i^{nM} 2 \cdot p_i \cdot q_i \cdot m_i^2}{\sigma_y^2}$$

Em que:

$h_M^2 = \text{herdabilidade molecular}$

$nM = \text{número de marcas}$

$\sigma_y^2 = \text{variância fenotípica}$

$p_i = \text{frequência de um dos alelos do loco marcador } i$

$q_i = \text{frequência do outro alelo do loco marcador } i$

$m_i = \text{efeito genético de um dos alelos do loco marcador } i$

A capacidade preditiva foi estimada a partir da correlação dos VGGs com os VGs dos 400 indivíduos genotipados e a acurácia de acordo com a equação de Resende (2014).

$$\text{acurácia} = \frac{rg\hat{g}}{\sqrt{h_t^2}} \cdot \sqrt{h_M^2}$$

Em que:

$rg\hat{g} = \text{capacidade preditiva}$

$h_M^2 = \text{herdabilidade molecular}$

$h_t^2 = \text{herdabilidade individual média de famílias}$

3.6 Correlação entre VGGs estimados pelos modelos de SGA e dados fenotípicos mensurados

Os VGG de todos os indivíduos das populações de avaliação foram estimados com base nos dados de marcadores SNPs e os seus respectivos efeitos estimados após validação independente nas quatro populações de avaliação.

Os indivíduos foram classificados (rankeados) com base nos VGGs via RRBLUP. Em paralelo os valores fenotípicos efetivamente mensurados nestes mesmos indivíduos via REML/BLUP foram também utilizados para classifica-los. Em seguida, foi realizada uma comparação entre as classificações obtidas com os dois métodos visando assim avaliar a efetiva capacidade preditiva da SGA em nível de árvore individual e o percentual de coincidência na seleção fenotípica e seleção genômica.

4 RESULTADOS E DISCUSSÃO

4.1 Estimativa dos parâmetros genéticos

A partir dos dados fenotípicos das populações de avaliação foram estimadas as herdabilidades individuais para cada variável e as herdabilidades individuais médias das famílias, conforme apresentado na Tabela 4.

Tabela 4 – Herdabilidades individuais (h^2) para cada população e variável e herdabilidades médias das famílias (h^2_t)

Variável	População	h^2	h^2_t
DAP	CHNS0048R	0,48	-
	MHGI0002R	0,08	
	MHMA0005P	0,08	0,09
	MHSFC200P	0,13	
Altura	CHNS0048R	0,30	-
	MHGI0002R	0,08	
	MHMA0005P	0,08	0,09
	MHSFC200P	0,13	
Volume	CHNS0048R	0,54	-
	MHGI0002R	0,07	
	MHMA0005P	0,11	0,11
	MHSFC200P	0,16	
IMA	CHNS0048R	0,54	-
	MHGI0002R	0,07	
	MHMA0005P	0,11	0,11
	MHSFC200P	0,16	

Conforme esperado, dada a maior precisão experimental de um teste clonal, as maiores herdabilidades individuais ocorreram no teste clonal (CHNS0048R), variando de 0,302 para a variável altura, até 0,542 para volume e IMA. Segundo Resende (2002), resultados médios da herdabilidade no sentido amplo para características de crescimento em *Eucalyptus* apresentam uma magnitude em torno de 0,30, o que está de acordo com os valores encontrados neste trabalho.

As herdabilidades individuais de famílias variaram de 0,074 para a variável volume no teste de progênies MHGI0002R até 0,159 para IMA e volume no teste de progênies MHSFC200P. Macedo et al. (2013) verificaram que o coeficiente de

herdabilidade individual de progênie para os caracteres DAP, altura e volume foram 0,02 e 0,03 e 0,01, respectivamente em uma população de *Eucalyptus*, abaixo dos valores encontrados neste trabalho. Resende (2002), ao compilar resultados de herdabilidades em diferentes estudos determinou valores médios ponderados na ordem de 0,27 para características de crescimento em *Eucalyptus*, acima dos valores encontrados neste trabalho, porém variando de 0,07 a 0,35, o que mostra que os resultados aqui encontrados estão dentro do range esperado, apesar de estarem abaixo da média esperada.

A partir das herdabilidades individuais de famílias foram calculadas as herdabilidades médias de famílias para cada variável e que foram utilizadas no cálculo da herdabilidade molecular. Desta forma, foi desconsiderado nesta média, as herdabilidades do teste clonal, que consideram efeitos aditivos e dominantes, e utilizadas as herdabilidades medias de família, por considerar apenas o efeito aditivo para herdabilidades no sentido restrito. As maiores herdabilidades médias de família foram para as variáveis Volume e IMA (0,11333), seguidas de DAP (0,09667) e Altura (0,09333).

Foram encontrados valores maiores que 0,754 para a correlação entre DAP e Altura, chegando a 1,000 para Volume e IMA, mostrando que se trata de correlações positivas fortes por se tratar de características de crescimento.

Tabela 5 – Correlações Genéticas entre as variáveis de estudo

Variáveis	DAP	Altura	Volume	IMA
DAP	1,00	0,75	0,97	0,97
Altura		1,00	0,81	0,81
Volume			1,00	1,00
IMA				1,00

4.2 Acurácias de Seleção e Capacidades Preditivas

4.2.1 Capacidade preditiva com base no modelo proposto por Lima (2014) para população de Eucalipto da International Paper

Este trabalho foi dividido em duas etapas, a primeira consistiu na utilização do modelo de predição desenvolvido por Lima (2014) para predição das populações de avaliação, ou seja, a utilização dos efeitos estimados naquela população de treinamento para os diferentes genótipos a cada um dos 29090 marcadores SNPs.

Tabela 6 - Capacidade Preditiva utilizando o catálogo de marcas do trabalho de Lima (2014), composto por 29090 marcas

População	DAP	Altura	Volume	IMA
CHNS0048R	0,440	0,219	0,436	0,436
MHGI0002R	0,080	0,008	0,076	0,076
MHMA0005P	-0,156	-0,296	-0,181	-0,181
MHSFC200P	-0,005	-0,102	-0,029	-0,029
Média Geral	0,140	0,009	0,136	0,136

As maiores capacidades preditivas ocorreram para o teste clonal CNHS0048R em todas as variáveis de estudo. Lima (2014) estimou capacidades preditivas da ordem de 0,44 para DAP, 0,35 para altura, 0,40 para volume e 0,43 para IMA, a partir da validação cruzada, valores estes consistentes com os valores estimados neste trabalho para esta população. Para todas as quatro características nas outras três populações avaliadas o modelo preditivo apresentou baixas capacidades preditivas.

Capacidades preditivas satisfatórias eram esperadas para as populações CHNS0048R, MHGI0002R e MHSFC200P, por serem geneticamente relacionadas com a população de treinamento e baixa ou nenhuma capacidade preditiva para a população MHMA0005P por não possuir nenhuma relação genética com a população de treinamento além de ser composta somente por indivíduos de *E. grandis*. Capacidade preditiva satisfatória e consistente com o esperado foi obtida para a população CHNS0048R indicando o bom potencial do modelo desenvolvido por Lima (2014) na seleção de indivíduos em teste clonal. Entretanto, os resultados observados para as outras duas populações relacionadas são conflitantes, principalmente aquele para a população MHSFC200P na qual tecnicamente deveria

se obter uma capacidade preditiva igualmente elevada como na população CHNS0048R. Para a população relacionada MHGI0002R a baixa capacidade preditiva pode ser explicada pelo fato de se tratar de famílias de meios irmãos de polinização aberta o que implica em tamanho efetivo maior, maior diversidade de pólen participante inclusive de pólen estranho o que impacta diretamente o nível de relacionamento genético efetivo e conseqüentemente a capacidade preditiva do modelo desenvolvido previamente por Lima (2014).

Apesar da inconsistência dos resultados obtidos entre as populações CHNS0048R e MHSFC200P os resultados deste trabalho indicam que a SGA permite identificar indivíduos superiores com acurácia satisfatória quando existir relacionamento genético entre a população de treinamento e a população de seleção, mas que na ausência de relacionamento genético, a SGA não permite seleção adequada. Estes resultados estão de acordo com aqueles relatados por Resende et al. (2012) em *Eucalyptus* e Beaulieu et al. (2014a) em *Picea*, corroborando que a capacidade preditiva da SGA depende essencialmente da captura de relacionamento genético e desequilíbrio de ligação a longa distância.

Vale por fim ressaltar que diversos aspectos não considerados na análise realizada podem ainda ter impactado os resultados, seja no que se refere aos métodos utilizados por Lima (2014) nas correções e parametrizações dos dados e seja na metodologia de análise. A forma de correção dos dados fenotípicos e a parametrização dos dados genotípicos na população de treinamento do trabalho de Lima (2014) pode ter sido diferente das correções utilizadas nas quatro populações de avaliação deste estudo, o que pode causar viés pelo fato da modelagem absorver ou não os efeitos de escala no intercepto do modelo. Além disso, o acesso a apenas 400 indivíduos (100 por população), dificilmente torna possível avaliar com rigor a capacidade preditiva dos modelos, e a correlação calculada apenas com 100 indivíduos pode não ter representado adequadamente as populações. Todos estes fatores são importantes e podem ter influenciado significativamente os resultados desta análise. Torna-se assim importante, na continuidade deste trabalho, realizar uma análise conjunta dos dados de Lima (2014) juntamente com os dados gerados neste estudo, de forma a avaliar novamente as capacidades preditivas com base em uma parametrização e correção uniformes.

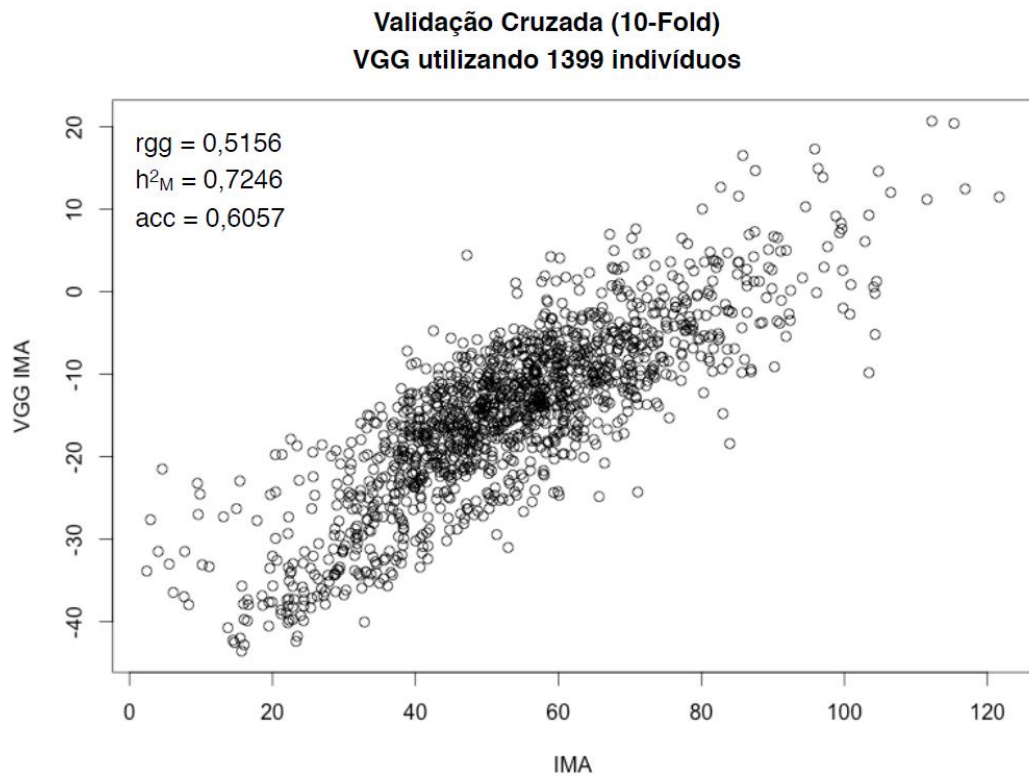


Figura 3 – VGG utilizando 1.399 indivíduos com mesmas correções e parametrizações

A análise utilizando todos os 1399 indivíduos com validação pelo método “Jackknife” mostrou um aumento da capacidade preditiva e acurácia quando comparadas à validação feita no trabalho de Lima (2014) com 999 indivíduos e com apenas 100 indivíduos das populações deste estudo, corroborando a hipótese de que quanto maior o número de indivíduos genotipados, mais robusto o modelo de SGA fica e, por consequência melhores valores de capacidades preditivas e acurácias poderão ser encontrados.

4.2.2 Capacidade preditiva e acurácia de seleção com base no modelo desenvolvido das populações de avaliação

Utilizando os 400 indivíduos das quatro populações de avaliação foi construído novo modelo utilizando desta vez a seleção de marcas. A partir das validações independentes (quatro rodadas, com treinamento de 300 indivíduos e validação em 100 referentes a uma população), foram obtidos valores de herdabilidade molecular para estimativa das acurácias e capacidade preditiva para cada rodada de validação. Embora os valores de herdabilidade molecular tenham

sido baixos, constata-se que os marcadores capturaram uma grande parte da herdabilidade dos caracteres, as quais foram também baixas (em torno de 10%).

Com a sequência de Fibonacci, 21 testes de conjunto de marcas foram realizados para determinação das máximas capacidades preditivas e acurácias.

Tabela 7 – Herdabilidades Moleculares no sentido restrito em nível de indivíduo para cada variável em cada população de estudo com número total de marcas e com as marcas selecionadas para maximizar a capacidade preditiva e acurácia

h^2_M População	DAP		Altura		Volume		IMA	
	Nº SNP	h^2_M	Nº SNP	h^2_M	Nº SNP	h^2_M	Nº SNP	h^2_M
CHNS0048R	237	0,08	237	0,06	237	0,11	237	0,11
MHGI0002R	237	0,08	237	0,05	237	0,09	237	0,09
MHMA0005P	620	0,07	237	0,07	237	0,12	237	0,12
MHSFC200P	237	0,07	237	0,05	237	0,10	237	0,10
Média	-	0,08	-	0,05	-	0,11	-	0,11
CHNS0048R	29090	0,0000	29090	0,0000	29090	0,0000	29090	0,0000
MHGI0002R	29090	0,0002	29090	0,0000	29090	0,0004	29090	0,0004
MHMA0005P	29090	0,0001	29090	0,0000	29090	0,0002	29090	0,0002
MHSFC200P	29090	0,0006	29090	0,0003	29090	0,0009	29090	0,0009
Média	-	0,0002	-	0,0001	-	0,0004	-	0,0004

Tabela 8 – Resultados de RRBLUP para capacidade preditiva e acurácia geral, ou seja, sem seleção de marcas, e a capacidade preditiva e acurácia maximizadas com os SNPs selecionados

Atributos	População	Variáveis			
		DAP	Altura	Volume	IMA
Capacidade Preditiva Geral	CHNS0048R	-0,072	0,018	-0,072	-0,072
	MHGI0002R	-0,028	-0,051	-0,028	-0,028
	MHMA0005P	0,035	0,052	0,035	0,035
	MHSFC200P	-0,058	-0,013	-0,058	-0,058
	Média	-0,031	0,002	-0,031	-0,031
Acurácia Geral	CHNS0048R	-0,003	0,000	-0,002	-0,002
	MHGI0002R	0,001	-0,001	-0,002	-0,002
	MHMA0005P	-0,003	0,104	0,001	0,001
	MHSFC200P	-0,006	-0,001	-0,005	-0,005
	Média	-0,002	0,000	-0,002	-0,002
Número de SNP que maximiza Capacidade Preditiva	CHNS0048R	1621	2623	620	620
	MHGI0002R	620	1621	620	620
	MHMA0005P	2623	2623	2623	2623
	MHSFC200P	1002	1621	1621	1621
	Média	1467	2122	1371	1371
Número de SNP que maximiza a Acurácia	CHNS0048R	237	237	237	237
	MHGI0002R	237	237	237	237
	MHMA0005P	620	237	237	237
	MHSFC200P	237	237	237	237
	Média	333	237	237	237
Capacidade Preditiva Maximizada	CHNS0048R	0,817	0,727	0,821	0,822
	MHGI0002R	0,769	0,756	0,791	0,793
	MHMA0005P	0,852	0,874	0,884	0,884
	MHSFC200P	0,537	0,549	0,509	0,509
	Média	0,744	0,727	0,751	0,752
Acurácia Maximizada	CHNS0048R	0,728	0,492	0,796	0,796
	MHGI0002R	0,621	0,506	0,684	0,684
	MHMA0005P	0,729	0,639	0,802	0,802
	MHSFC200P	0,409	0,353	0,420	0,420
	Média	0,660	0,555	0,743	0,743
Viés da Capacidade Preditiva Maximizada	CHNS0048R	1,368	1,231	1,110	1,113
	MHGI0002R	1,402	1,802	1,651	1,650
	MHMA0005P	1,045	1,023	1,136	1,135
	MHSFC200P	0,785	0,949	0,724	0,724
	Média	1,150	1,251	1,155	1,155
Viés da Acurácia Maximizada	CHNS0048R	1,123	1,081	1,031	1,031
	MHGI0002R	1,309	1,272	1,508	1,509
	MHMA0005P	0,910	0,886	0,864	0,864
	MHSFC200P	0,690	0,902	0,541	0,541
	Média	1,008	1,035	0,986	0,986

As capacidades preditivas e acurácias gerais, ou seja, considerando as 29090 marcas, foram muito baixas, próximas à zero, para todas as populações e variáveis de estudo. Da mesma forma, a herdabilidade molecular considerando os 29090 marcadores também foi baixa, próxima à zero. Isto pode ser explicado pelo fato do

tamanho amostral utilizado ser pequeno e a variância de amostragem da h^2_M estimada ser muito grande.

As maiores capacidades preditivas foram obtidas por um número de SNPs médio que variou de 1371 para volume e IMA, 1467 para DAP e o máximo de 2122 para Altura, havendo populações com 2623 marcadores selecionados para algumas das variáveis. Desta forma, a média de capacidade preditiva maximizada para DAP foi de 0,744, para altura 0,727, para volume 0,751 e para IMA 0,752. Resende et al. (2012a) encontraram para duas populações de eucalipto capacidades preditivas para características de crescimento que variaram de 0,46 para altura até 0,55 para circunferência à altura do peito. Resende et al. (2012b) obtiveram para uma população de *Pinus taeda* capacidades preditivas para características de crescimento que variaram de 0,39 a 0,49.

O uso de 237 a 5000 marcas proporcionam praticamente o mesmo nível de capacidade preditiva. Isto pode revelar que a acurácia é também dada principalmente pelo grau de parentesco, situação essa em que poucos marcadores são suficientes para capturar adequadamente o parentesco entre os indivíduos. Resultados teóricos e de simulação (Azevedo et al., 2015, no prelo) corroboram os resultados apresentados neste trabalho.

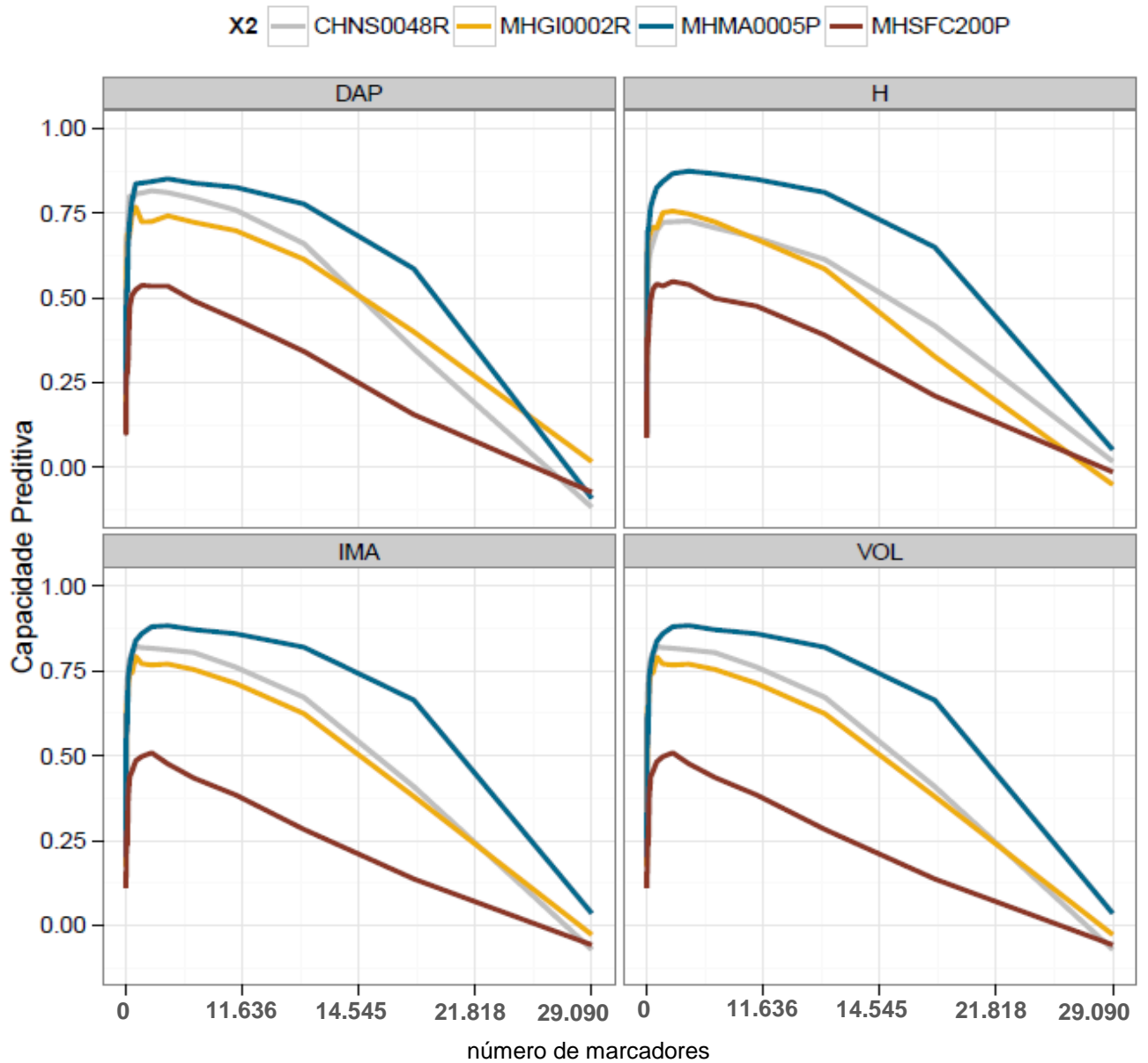


Figura 4 – Capacidades preditivas maximizadas considerando todos os 29090 marcadores

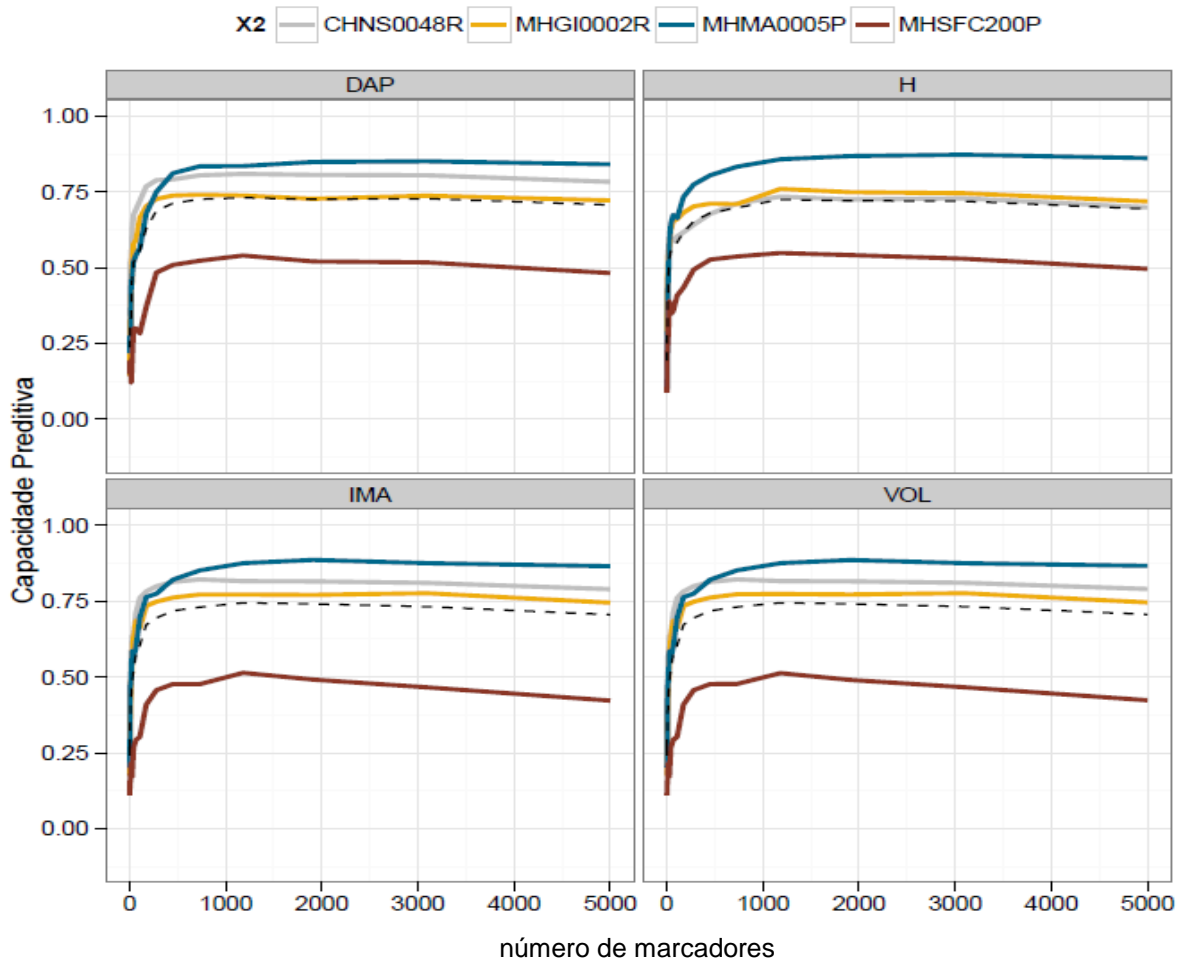


Figura 5 – Capacidades preditivas considerando os 5000 marcadores de maior efeito. As linhas tracejadas correspondem às médias de capacidade preditiva para cada variável

Já para as acurácias maximizadas, o número de SNPs selecionados ficou ao redor de 237. Este número de marcas levou à acurácias na ordem de 0,660 para DAP, 0,555 para altura, 0,743 para volume e 0,743 para IMA. Acurácias de 0,74 e 0,79 foram encontrados por Resende et al. (2012a) para CAP e Altura respectivamente na população CENIBRA e 0,73 e 0,66 para as mesmas variáveis respectivamente na população da FIBRIA. Lima (2014) apresentou acurácias de 0,55 para DAP, 0,50 para altura, 0,50 para volume e 0,56 para IMA.

As acurácias caíram mais rapidamente quando comparadas as capacidades preditivas. Entretanto, fórmulas alternativas para cálculo da acurácia genômica ainda estão sendo pesquisadas.

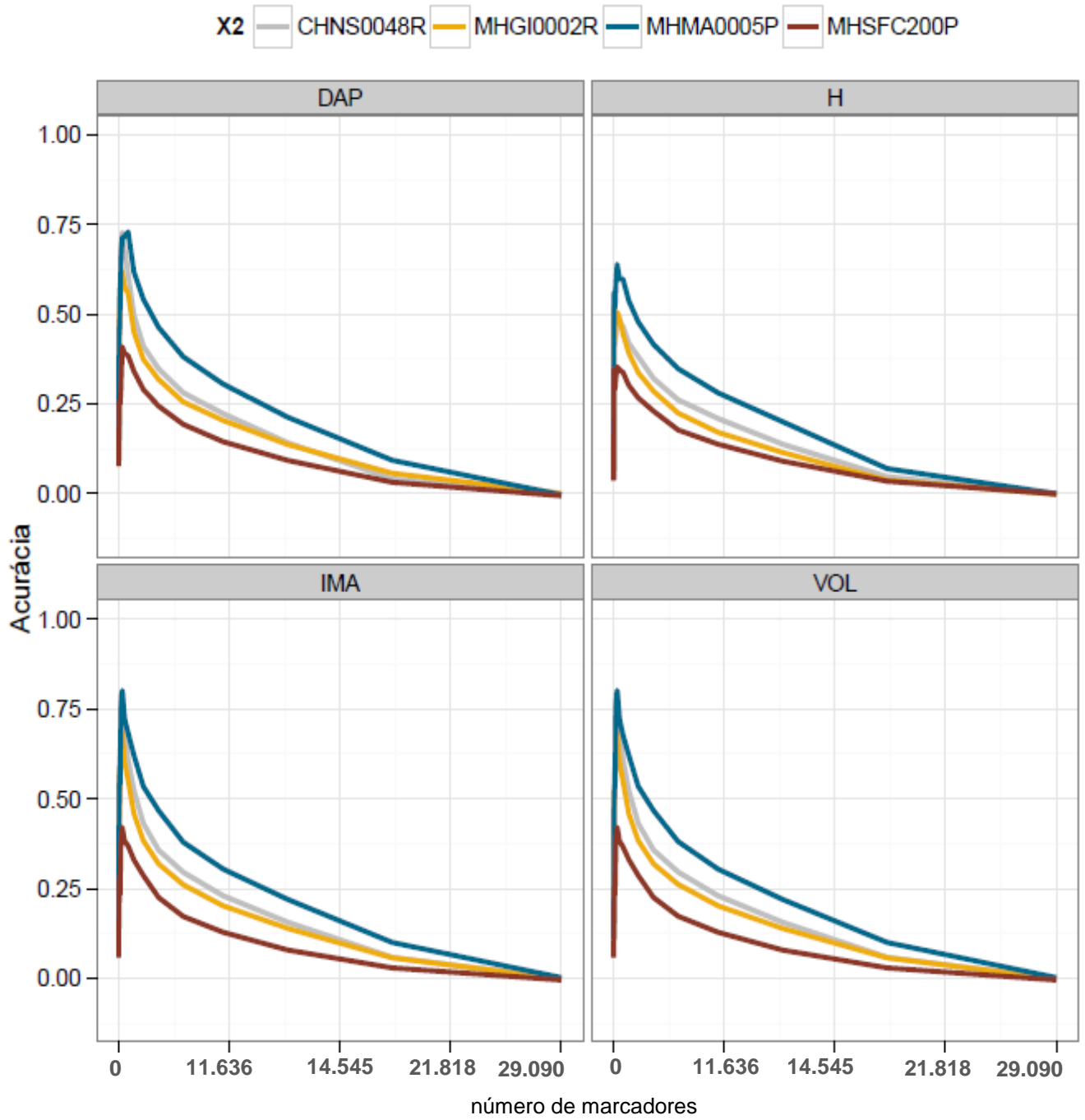


Figura 6 – Acurácias de seleção maximizadas considerando todos os 29090 marcadores

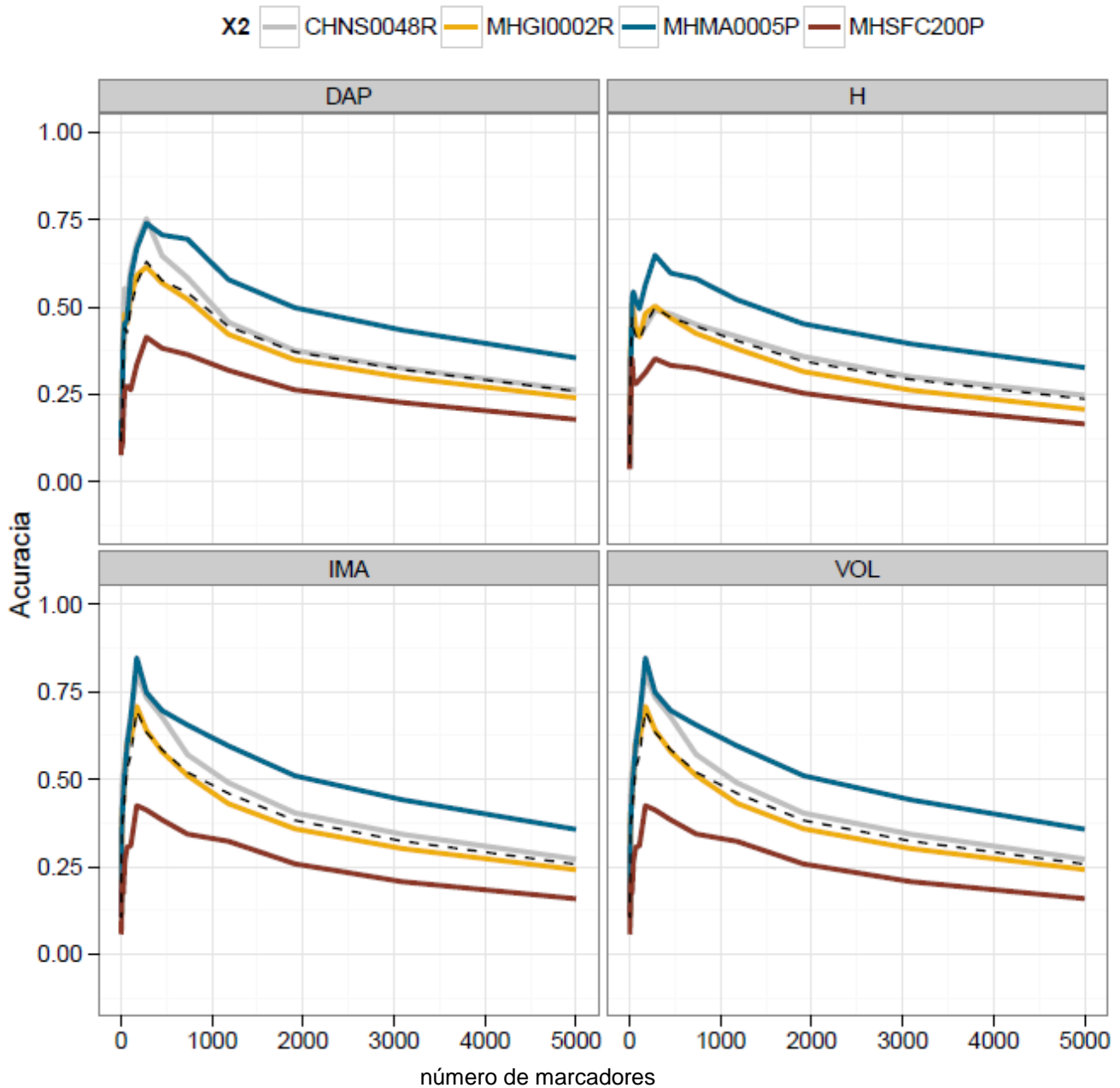


Figura 7 – Acurácia de seleção considerando os 5000 marcadores de maior efeito. As linhas tracejadas correspondem às médias de acurácia para cada variável

Abdollahi-Arpanahi et al. (2014), em um estudo de Seleção Genômica em Frangos concluiu que a capacidade preditiva foi maior quando verificados os SNPs de maior efeito e que aumentando o número de marcadores isto não melhorou a capacidade preditiva. Resende et al. (2015 no prelo) em simulações de dados de seleção genômica evidenciaram que o pico de acurácia foi atingido com cerca de 500 marcadores de maior efeito. No entanto, Weigel et al. (2009) e Vazquez et al.

(2010), estudando a seleção de marcas mostraram que ela pode atingir predições equivalentes àquelas obtidas com todos os marcadores mas nunca melhores do que estes. Desta forma, apesar de ainda haver dúvidas quanto a este assunto, importantes estudos vêm sendo realizados, onde a seleção de marcas se mostra como uma abordagem importante, pois permite avaliar mais diretamente cada caráter de acordo com a sua arquitetura genética, em detrimento do uso de uma mesma matriz de parentesco genômico para todas as variáveis.

As maiores capacidades preditivas (DAP = 0,852; Altura = 0,874; Volume = 0,884; IMA = 0,884) e acurácias (DAP = 0,729; Altura = 0,639; Volume = 0,802; IMA = 0,802) foram obtidas na população MHMA0005P, enquanto que os valores mais baixos para estes dois atributos ocorreram na população MHSFC200P (Capacidade Preditiva: DAP = 0,537; Altura = 0,549; Volume = 0,509; IMA = 0,509 – Acurácia: DAP = 0,409; Altura = 0,353; Volume = 0,420; IMA = 0,420). Estas diferenças estão relacionadas com a estrutura de cada população, sendo a população MHSFC200P formada por híbridos de diferentes espécies, enquanto que MHMA0005P é composta por espécie pura, a qual provavelmente apresenta menor número de alelos em cada loco gênico do que uma população híbrida que congrega alelos de diferentes espécies. Sendo os locos marcadores bialélicos, a situação é mais favorável à espécie pura, pois, haverá uma melhor captura de alelos gênicos na espécie pura.

Foi realizada análise de discriminantes que mostrou o grau de similaridade das populações. Nota-se que há dois grupos distintos, sendo um composto pela população MHMA0005P, composta por *E. grandis* puro, e o outro composto pelas outras três populações (MHSFC200P, MHGI0002R e CHNS0048R), com alto grau de similaridade entre elas, não sendo possível discriminá-las. O elevado grau de similaridade entre estas três populações pode ser explicado pelo fato de possuírem grau de relação genética entre elas.

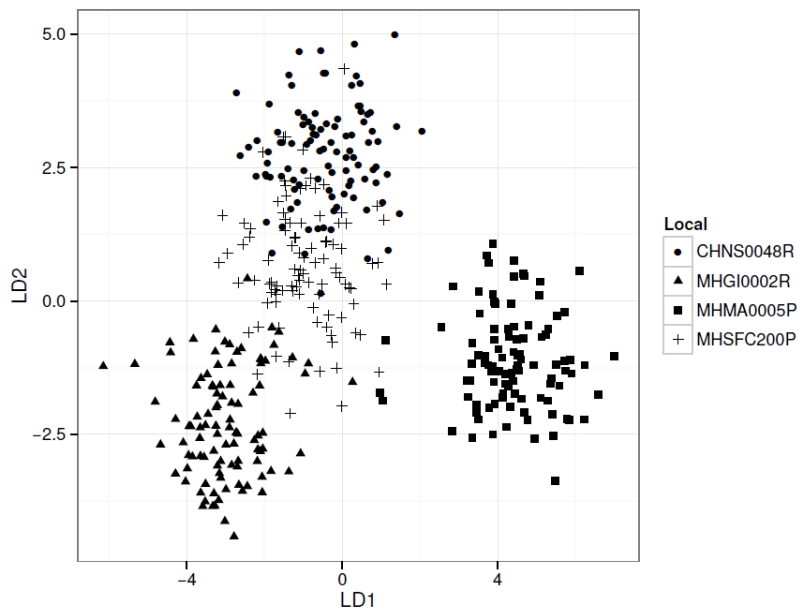


Figura 8 – Análise de discriminantes entre locais para verificar similaridades utilizando 237 marcadores

Por fim, um resumo das capacidades preditivas encontradas utilizando o modelo preditivo proposto por Lima (2014) e as capacidades preditivas e acurácias maximizadas por população.

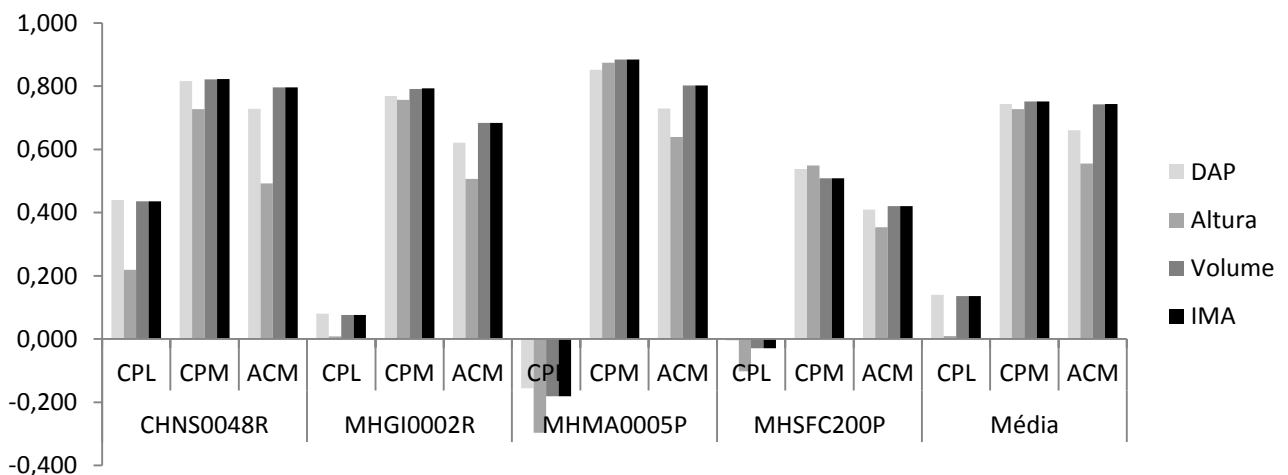


Figura 9 – Capacidade Preditiva obtida a partir do modelo de Lima, 2014 (CPL), Capacidade Preditiva Maximizada pelo método RRBLUP (CPM) e Acurácia Maximizada pelo método RRBLUP (ACM) para as quatro populações

4.3 Comparação do ranking individual utilizando a Seleção Genômica e BLUP para seleção fenotípica

Com base nos resultados das máximas capacidades preditivas, foi gerada a classificação de seleção para a variável IMA dos 20 melhores indivíduos para a característica IMA tanto com o BLUP Fenotípico quanto para a seleção genômica via RRBLUP, conforme Tabela 9.

Os resultados mostram que o percentual de indivíduos coincidentes em cada uma das classificações foi de 50% na população MHSFC200P, 55% em MHGI0002R e 70% para CHNS0048R e MHMA0005P. Lima (2014) observou um percentual de coincidência para análise via RRBLUP de 60%.

As populações de maior percentual de coincidência são clones e espécies puras, sendo populações mais homogêneas.

Tabela 9 – Classificações, geradas via BLUP (BLUP Fen) e genotípico obtido via RRBLUP para os 20 melhores indivíduos o percentual de coincidência (%COI) em cada população para seleção fenotípica e genômica

População	CHNS0048R		MHGI0002R		MHMA0005P		MHSFC200P	
Rank IMA	BLUP Fen	RRBLUP	BLUP Fen	RRBLUP	BLUP Fen	RRBLUP	BLUP Fen	RRBLUP
1	HNS_036	HNS_086	GI_005	GI_023	MA_054	MA_024	BR_030	BR_068
2	HNS_086	HNS_002	GI_072	GI_013	MA_065	MA_065	BR_080	BR_058
3	HNS_090	HNS_038	GI_047	GI_079	MA_024	MA_010	BR_065	BR_017
4	HNS_002	HNS_051	GI_023	GI_031	MA_058	MA_054	BR_063	BR_065
5	HNS_051	HNS_090	GI_032	GI_021	MA_010	MA_096	BR_096	BR_006
6	HNS_029	HNS_036	GI_078	GI_043	MA_088	MA_023	BR_050	BR_027
7	HNS_027	HNS_085	GI_013	GI_063	MA_081	MA_094	BR_070	BR_050
8	HNS_028	HNS_005	GI_083	GI_015	MA_011	MA_088	BR_081	BR_063
9	HNS_057	HNS_018	GI_043	GI_018	MA_009	MA_011	BR_013	BR_013
10	HNS_050	HNS_053	GI_021	GI_032	MA_029	MA_035	BR_041	BR_069
11	HNS_052	HNS_034	GI_015	GI_098	MA_018	MA_018	BR_001	BR_034
12	HNS_038	HNS_027	GI_066	GI_005	MA_037	MA_092	BR_084	BR_084
13	HNS_073	HNS_076	GI_022	GI_016	MA_035	MA_056	BR_073	BR_041
14	HNS_085	HNS_099	GI_016	GI_089	MA_094	MA_029	BR_058	BR_005
15	HNS_005	HNS_073	GI_031	GI_087	MA_056	MA_066	BR_015	BR_061
16	HNS_044	HNS_010	GI_018	GI_083	MA_066	MA_070	BR_039	BR_002
17	HNS_006	HNS_052	GI_001	GI_068	MA_055	MA_069	BR_054	BR_062
18	HNS_054	HNS_029	GI_053	GI_060	MA_046	MA_046	BR_062	BR_096
19	HNS_099	HNS_007	GI_058	GI_025	MA_100	MA_020	BR_042	BR_073
20	HNS_082	HNS_057	GI_094	GI_088	MA_020	MA_087	BR_045	BR_011
%COI	-	70%	-	55%	-	70%	-	50%

5 CONCLUSÕES

Com base no estudo do uso da SGA em duas populações distintas de seleção, uma geneticamente relacionada com a população de treinamento, e outra geneticamente não-relacionada com a população de treinamento, podemos concluir que:

- Os resultados deste trabalho indicam que a SGA permite identificar indivíduos superiores com acurácia satisfatória quando há relacionamento genético entre a população de treinamento e a população de seleção, mas que na ausência de relacionamento genético, a SGA não permite seleção adequada.

- A capacidade preditiva da SGA depende essencialmente da captura de relacionamento genético e desequilíbrio de ligação a longa distância.

- Capacidades preditivas e acurácias de seleção foram maximizadas quando utilizado métodos de seleção de marcadores de maiores efeitos;

- Populações de espécies pura apresentou maiores capacidades preditivas e acurácias de seleção em virtude do menor número de alelos envolvidos;

- A seleção genômica propicia alta coincidência com a seleção fenotípica, sendo que sua vantagem potencial é principalmente em ganho genético por unidade de tempo;

REFERÊNCIAS

- ABDOLLAHI-ARPAHAHI, R.; NEJATI-JAVAREMI, A.; PAKDEL, A.; MORADI-SHAHRBABA, M.; MOROTA, G.; VALENTE, B.D.; KRANIS, A.; ROSA GJM.; GIANOLA, D. Effect of allele frequencies, effect sizes and number of markers on prediction of quantitative traits in chickens. **Journal of Animal Breeding Genetics**, Nova Jersey, EUA , v. 131, p. 123–133, 2014. doi:10.1111/jbg.12075
- ASSIS, T.F. Melhoria genética do eucalipto. **Informe Agropecuário**, Belo Horizonte, v. 1, n. 4, p. 56-66, 1986. Disponível em: <<http://www.celsofoelkel.com.br/artigos/outros/Arquivo%2014.%20melhoramento%20eucalipto.doc>>. Acesso em: 01 ago. 2015.
- ASSOCIAÇÃO BRASILEIRA DE CELULOSE E PAPEL. **Dados do setor**. São Paulo, 2011. Disponível em: <<http://bracelpa.org.br/>>. Acesso em: 05 maio 2014.
- AZEVEDO, C.F.; RESENDE, M.D.V.; SILVA, F.F.; VIANA, J.M.S.; VALENTE, M.S.F.; RESENDE JUNIOR, M.F.R. New estimator for the experimental accuracy of genome selection. **Crop Science**, Madison 2015. In press.
- BEAULIEU, J.; DOERKSEN, T.K.; CLEMENT, S.; MACKAY, J.; BOUSQUET, J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. **Journal of Heredity**, Oxford, v. 113, p. 343–352, Apr. 2014a. doi:10.1038/hdy.2014.36.
- BEAULIEU, J.; DOERKSEN, T.K.; MACKAY, J.; RAINVILLE, A.; BOUSQUET, J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. **BMC Genomics**, London, v. 15, p. 1048, 2014b. Disponível em: <<http://www.biomedcentral.com/1471-2164/15/1048>>. Acesso em: 12 ago. 2015.
- BIROL, I.; RAYMOND, A.; JACKMAN, S.D.; PLEASANCE, S.; COOPE, R.; TAYLOR, G.A.; SAINT YUEN, M.M. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Oxford Journals - **Bioinformatics**, Oxford, v. 31, n. 22, p. 1-6, 2013. doi:10.1093/bioinformatics/btt178.
- ELDRIDGE, K.; DAVIDSON, J.; HARWOOD, C.; VAN WYK, G. **Eucalyptus domestication and breeding**. New York: Oxford University Press 1993. 312 p.
- FERREIRA, M.; SANTOS, P.E.T. Melhoria genética florestal dos *Eucalyptus* no Brasil: breve histórico e perspectivas. In: IUFRO CONFERENCE ON SILVICULTURE AND IMPROVEMENT OF EUCALYPTS, 1., 1997, Salvador. **Proceedings...** Colombo: EMBRAPA, CNPF, 1997. v. 1, p. 14-34.
- FONSECA, S.M.; RESENDE, M.D.V.; ALFENAS, A.C.; GUIMARÃES, L.M.S.; ASSIS, T.F.; GRATTAPAGLIA, D. **Manual prático de melhoria genética do eucalipto**. Viçosa: UFV, 2010. 200 p.

GAMAL EL-DIEN, O.; RATCLIFFE, B.; KLÁPŠTĚ, J.; CHEN, C.; PORTH, I.; EL-KASSABY, A.Y. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. **BMC Genomics**, London, v. 16, p. 16-37, 2015. doi:10.1186/s12864-015-1597-y. Disponível em: <<http://www.biomedcentral.com/1471-2164/16/370>>. Acesso em: 15 set. 2015.

GODDARD, M. Genomic selection: prediction of accuracy and maximization of long term response. **Genetica**, Bethesda, v. 136, p. 245–257, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18704696>>. Acesso em 20 de setembro de 2015

GODDARD, M.E.; HAYES, B.J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews Genetics**, London, v. 10, p. 381–391, 2009.

GRATTAPAGLIA, D. Breeding forest trees by genomic selection: current progress and the way forward. In: TUBEROSA, R.; GRANER, A.; FRISON, E. **Genomics of plant genetic resources**. Berlin: Springer, 2014. chap. 26, p. 652-678. DOI 10.1007/978-94-007-7572-5_26.

GRATTAPAGLIA, D.; RESENDE, M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, Berlin, v. 7, p. 241-255, 2011.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and Rapp markers. **Genetics**, Bethesda, v. 137, n. 4, p. 1121-1137, 1994.

INDÚSTRIA BRASILEIRA DE FLORESTAS. **A Indústria Brasileira de Árvores (Ibá) em números**. Disponível em: <<http://www.iba.org/pt/dados-e-estatisticas>>. Acesso em: 10 ago. 2015.

ISIK, F. Genomic selection in forest tree breeding: the concept and an outlook to the future. **New Forests**, Raleigh, v. 45, p. 379-401, May 2014.

IWATA, H.; HAYASHI, T.; TSUMURA, Y. Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. **Tree Genetics & Genomes**, Berlin, v. 7, p. 747–758, 2011. DOI 10.1007/s11295-011-0371-9.

JANNINK, J.L.; LORENZ, A.J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Brief Functional Genomics**, Oxford, v. 9, p. 166–177, 2010.

LEGARRA, A.; ROBERT-GRANIE, C.; MANFREDI, E.; ELSEN, J.M. Performance of genomic selection in mice. **Genetics**, Bethesda v. 180, p. 611-618, 2008.

LIMA, B.M. **Bridging genomics and quantitative genetics of *Eucalyptus*: genome-wise prediction and genetic parameter estimation for growth and wood properties using high-density SNP data**. 2014. 93 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2014.

MACEDO, H.R.; FREITAS, M.L.M.; MORAES, M.L.T.; ZANATA, M.; SEBEN, A.M. Variação, herdabilidade e ganhos genéticos em progênies de *Eucalyptus tereticornis* aos 25 anos de idade em Batatais-SP. **Scientia Forestalis**, Piracicaba, v. 41, n. 100, p. 533-540, dez. 2013.

MEUWISSEN, T.H.E.; GODDARD, M.E.; HAYES, B.J. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, p. 1819-1829, 2001.

MYBURG AA, GRATTAPAGLIA D, TUSKAN GA, HELLSTEN U, HAYES RD, GRIMWOOD J, JENKINS J, LINDQUIST E, TICE H, BAUER D, GOODSTEIN DM, DUBCHAK I, POLIAKOV A, MIZRACHI E, KULLAN AR, HUSSEY SG, PINARD D, VAN DER MERWE K, SINGH P, VAN JAARVELD I, SILVA-JUNIOR OB, TOGAWA RC, PAPPAS MR, FARIA DA, SANSALONI CP, PETROLI CD, YANG X, RANJAN P, TSCHAPLINSKI TJ, YE CY, LI T, STERCK L, VANNESTE K, MURAT F, SOLER M, CLEMENTE HS, SAIDI N, CASSAN-WANG H, DUNAND C, HEFER CA, BORNBERG-BAUER E, KERSTING AR, VINING K, AMARASINGHE V, RANIK M, NAITHANI S, ELSEER J, BOYD AE, LISTON A, SPATAFORA JW, DHARMWARDHANA P, RAJA R, SULLIVAN C, ROMANEL E, ALVES-FERREIRA M, KÜLHEIM C, FOLEY W, CAROCHA V, PAIVA J, KUDRNA D, BROMMONSCHENKEL SH, PASQUALI G, BYRNE M, RIGAUULT P, TIBBITS J, SPOKEVICIUS A, JONES RC, STEANE DA, VAILLANCOURT RE, POTTS BM, JOUBERT F34, BARRY K, PAPPAS GJ, STRAUSS SH, JAISWAL P, GRIMAPETTENATI J, SALSE J, VAN DE PEER Y, ROKHSAR DS, SCHMUTZ J. The genome of *Eucalyptus grandis*. **Nature**, London, v. 510, p. 356–362, June 2014. doi:10.1038/nature13308.

NYSTEDT B, STREET NR, WETTERBOM A, ZUCCOLO A, LIN YC, SCOFIELD DG, VEZZI F, DELHOMME N, GIACOMELLO S, ALEXEYENKO A, VICEDOMINI R, SAHLIN K, SHERWOOD E, ELFSTRAND M, GRAMZOW L, HOLMBERG K, HÄLLMAN J, KEECH O, KLASSON L, KORIABINE M, KUCUKOGLU M, KÄLLER M, LUTHMANN J, LYSHOLM F, NIITTYLÄ T, OLSON A, RILAKOVIC N, RITLAND C, ROSSELLÓ JA, SENA J, SVENSSON T, TALAVERA-LÓPEZ C, THEIßEN G, TUOMINEN H, VANNESTE K, WU ZQ, ZHANG B, ZERBE P, ARVESTAD L, BHALERAO R, BOHLMANN J, BOUSQUET J, GARCIA GIL R, HVIDSTEN TR, DE JONG P, MACKAY J, MORGANTE M, RITLAND K, SUNDBERG B, THOMPSON SL, VAN DE PEER Y, ANDERSSON B, NILSSON O, INGVARSSON PK, LUNDEBERG J, JANSSON S. The Norway spruce genome sequence and conifer genome evolution. **Nature**, London, v. 497, n. 7451, p. 579–584, 2013. doi:10.1038/nature12211.

PIRES, I.E.; RESENDE, M.D.V.; SILVA, R.L.; RESENDE JR., M.F.R. **Genética florestal**. Viçosa: UFV, Departamento de Engenharia Florestal; Ed. Arka, 2011. 318 p.

THE R project for statistical computing: R version 3.2.2 (Fire Safety) - RRBLUP package. Disponível em: <<https://www.r-project.org/>> Acesso em: 14 ago. 2015.

RESENDE, M.D.V. Correlação intraclasse entre valores genéticos e implicações na comparação de métodos de seleção. **Floresta**, Curitiba, v. 21, n. ½, p. 37-48, jun./dez. 1994.

_____. Delineamento de experimentos de seleção para a maximização da acurácia seletiva e progresso genético. **Revista Árvore**, Viçosa, v. 19, n. 4, p. 479-500, out./dez. 1995.

_____. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975 p.

_____. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007. 561p.

RESENDE, M.D.V.; SILVA, F.F.E.; AZEVEDO, C.F. **Estatística matemática, biométrica e computacional**: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência. Visconde do Rio Branco: Suprema, 2014. v. 1, 881 p.

RESENDE, M.D.V.; SILVA, F.F.; LOPES, P.S. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: UFV, Departamento de Estatística. 2011. 180 p. Disponível em: <http://www.det.ufv.br/ppestbio/corpo_docente.php>. Acesso em: 11 jul. 2015.

RESENDE, M.D.V.; LOPES, P.S.; SILVA, R.L.S.; PIRES, I.E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, Colombo, n. 56, p. 63-77, jan./jun. 2008.

RESENDE, M.D.V.; RESENDE, M.F.R.; SANSALONI, C.P.; PETROLI, C.D.; MISSIAGGIA, A.A.; AGUIAR, A.M.; ABAD, J.M.; TAKAHASHI, E.K.; ROSADO, A.M.; FARIA, D.A.; PAPPAS, G.J.; KILIAN, A.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, London, v. 194, p. 116–128, 2012a.

RESENDE, M.D.V.; RESENDE, M.F.R.; SANSALONI, C.P.; PETROLI, C.D.; MISSIAGGIA, A.A.; AGUIAR, A.M.; ABAD, J.M.; TAKAHASHI, E.K.; ROSADO, A.M.; FARIA, D.A.; PAPPAS, G.J.; KILIAN, A.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, London, v.194, p. 116-128, 2012b.

RESENDE, M.F.R. **Seleção genômica no melhoramento vegetal**. 2010. 78 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Viçosa, Viçosa, 2010.

RESENDE, M.F.R.; MUNOZ, P.; ACOSTA, J.J.; PETER, G.F.; DAVIS, J.M.; GRATTAPAGLIA, D.; RESENDE, M.D.V.; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytologist**, London, v. 193, p. 617-624, 2012a.

RESENDE, M.F.R.; MUNOZ, P.; RESENDE, M.D.V.; GARRICK, D.J.; FERNANDO, L.; DAVIS, J.M.; JOKELA, E.J.; MARTIN, T.A.; PETER, G.F.; KIRST, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). **Genetics**, Bhaseda, v. 190, p. 1503–1510, Apr. 2012b.

SANSALONI, C.P.; GRATTAPAGLIA, D.; RESENDE, M.D.V.; RESENDE, M.F.R.; PETROLI, C.D.; MISSIAGGIA, A.A.; TAKAHASHI, E.K.; ZAMPROGNO, K.C.; KILIAN, A. Genomic selection for growth traits in *Eucalyptus*: accuracy, within and across breeding populations. **IUFRO Tree Biotechnology**, Porto Seguro, 113-114, 2011.

SANTOS, G.A.S.; XAVIER, A.; LEITE, H.G. Desempenho silvicultural de clones de *Eucalyptus grandis* em relação às árvores matrizes. **Revista Árvore**, Viçosa, v. 30, n. 5, p. 737-747, 2006.

SCHAEFFER, L.R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, Nova Jersey, v. 123, p. 218-223, 2006.

SILVA, R.L. **Influência do tamanho da parcela experimental em testes clonais de eucalipto**. 2001. 67 p. Dissertação (Mestrado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa, 2001.

SILVA-JUNIOR, O.B.; GRATTAPAGLIA, D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. **New Phytologist**, London, v. 208, n. 3, p. 830-845, 2015.

SILVA-JUNIOR, O.B.; FARIA, D.A.; TOGAWA, R.C.; GRATTAPAGLIA, D. *Eucalyptus* genotyping taken to the next level: development of the "EucHIP60k.br" based on large scale multi-species SNP discovery and ascertainment. In: IUFRO TREE BIOTECHNOLOGY CONFERENCE 2013, 2013, Asheville. **Proceedings...** Asheville, 2013. p. 101-102.

SOLBERG, T.R.; SONESSON, A.K.; WOOLLIAMS, J.A.; MEUWISSEN, T.H.E. Genomic selection using different marker types and densities. **Journal of Animal Science**, Champaign, v. 86, p. 2447-2454, 2008.

TUSKAN GA, DIFAZIO S, JANSSON S, BOHLMANN J, GRIGORIEV I, HELLSTEN U, PUTNAM N, RALPH S, ROMBAUTS S, SALAMOV A, SCHEIN J, STERCK L, AERTS A, BHALERAO RR, BHALERAO RP, BLAUDEZ D, BOERJAN W, BRUN A, BRUNNER A, BUSOV V, CAMPBELL M, CARLSON J, CHALOT M, CHAPMAN J, CHEN GL, COOPER D, COUTINHO PM, COUTURIER J, COVERT S, CRONK Q, CUNNINGHAM R, DAVIS J, DEGROEVE S, DÉJARDIN A, DEPAMPHILIS C, DETTER J, DIRKS B, DUBCHAK I, DUPLESSIS S, EHLTING J, ELLIS B, GENDLER K, GOODSTEIN D, GRIBSKOV M, GRIMWOOD J, GROOVER A, GUNTER L, HAMBERGER B, HEINZE B, HELARIUTTA Y, HENRISSAT B, HOLLIGAN D, HOLT R, HUANG W, ISLAM-FARIDI N, JONES S, JONES-RHOADES M, JORGENSEN R, JOSHI C, KANGASJÄRVI J, KARLSSON J, KELLEHER C, KIRKPATRICK R, KIRST M, KOHLER A, KALLURI U, LARIMER F, LEEBENS-MACK J, LEPLÉ JC, LOCASCIO P, LOU Y, LUCAS S, MARTIN F, MONTANINI B, NAPOLI C, NELSON DR, NELSON C, NIEMINEN K, NILSSON O, PEREDA V, PETER G, PHILIPPE R, PILATE G, POLIAKOV A, RAZUMOVSKAYA J, RICHARDSON P, RINALDI C, RITLAND K, ROUZÉ P, RYABOY D, SCHMUTZ J, SCHRADER J, SEGERMAN B, SHIN H, SIDDIQUI A, STERKY F, TERRY A, TSAI CJ, UBERBACHER E, UNNEBERG P, VAHALA J, WALL K, WESSLER S, YANG G, YIN T, DOUGLAS C, MARRA M, SANDBERG G, VAN DE PEER Y, ROKHSAR D.; The genome of black cottonwood, *Populus trichocarpa*. (Torr. & Gray). **Science**, Washington, v. 313, n. 5793, p. 1596–1604, 2006. doi:10.1126/science.1128691.

VAZQUEZ, A.I.; ROSA, G.J.M.; WEIGEL, K.A.; DE LOS CAMPOS, G.; GIANOLA, D.; ALLISON, D.B. Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. **Journal of Dairy Science**, Champaign, v. 93, p. 5942-5949, 2010.

WEIGEL, K.A.; DE LOS CAMPOS, G.; GONZÁLEZ-RECIO, O.; NAYA, H.; WU, X.L.; LONG, N.; ROSA, G.J.M.; GIANOLA, D. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. **Journal of Dairy Science**, Champaign, v. 92, p. 5248–5257, 2009.

ZAPATA-VALENZUELA, J.; ISIK, F.; MALTECCA, C.; WEGRZYN, J.; NEALE, D; MCKEAND; WHETTEN, R. SNP markers trace familial linkages in a cloned population of *Pinus taeda*: prospects for genomic selection. **Tree Genetic Genomes**, Berlin, v. 8, p. 1307–1318, 2012.

APÊNDICE

Apêndice A -

(continua)

ATRIBUTOS	N SNP	CAPACIDADE PREDITIVA				ACURACIA			
		CHNS0048R	MHGI0002R	MHMA0005P	MHSFC200P	CHNS0048R	MHGI0002R	MHMA0005P	MHSFC200P
DAP	2	0,342	0,329	0,218	0,150	0,123	0,143	0,122	0,078
	4	0,417	0,187	0,273	0,191	0,167	0,130	0,176	0,120
	5	0,442	0,212	0,274	0,189	0,165	0,141	0,169	0,115
	9	0,512	0,266	0,221	0,190	0,167	0,174	0,132	0,104
	13	0,557	0,392	0,407	0,101	0,348	0,303	0,287	0,093
	22	0,605	0,456	0,476	0,171	0,434	0,360	0,375	0,171
	35	0,677	0,568	0,504	0,266	0,572	0,473	0,465	0,253
	56	0,666	0,603	0,530	0,279	0,524	0,491	0,443	0,257
	91	0,694	0,689	0,518	0,276	0,555	0,513	0,513	0,250
	147	0,751	0,674	0,657	0,318	0,684	0,564	0,655	0,290
	237	0,801	0,689	0,724	0,471	0,728	0,621	0,714	0,409
	383	0,788	0,742	0,785	0,508	0,684	0,578	0,717	0,392
	620	0,807	0,769	0,837	0,524	0,612	0,558	0,729	0,385
	1002	0,810	0,725	0,839	0,537	0,499	0,451	0,618	0,341
	1621	0,817	0,726	0,844	0,535	0,412	0,375	0,542	0,291
	2623	0,811	0,743	0,852	0,535	0,348	0,319	0,464	0,244
	4244	0,794	0,724	0,839	0,492	0,281	0,256	0,382	0,193
	6867	0,759	0,699	0,827	0,437	0,223	0,204	0,306	0,145
11112	0,661	0,614	0,778	0,343	0,142	0,137	0,213	0,093	
17979	0,350	0,400	0,586	0,156	0,041	0,057	0,093	0,032	
29090	-0,117	0,017	-0,091	-0,073	-0,003	0,001	-0,003	-0,006	
Altura	2	0,298	0,286	0,100	0,088	0,064	0,063	4,02E-02	0,036
	4	0,383	0,363	0,209	0,153	0,176	0,181	1,28E-01	0,094
	5	0,282	0,423	0,218	0,221	0,154	0,204	1,34E-01	0,138
	9	0,458	0,421	0,409	0,227	0,302	0,322	3,02E-01	0,195
	13	0,559	0,515	0,431	0,255	0,392	0,410	3,54E-01	0,233
	22	0,569	0,566	0,581	0,371	0,484	0,520	5,26E-01	0,346
	35	0,595	0,597	0,643	0,383	0,458	0,490	5,62E-01	0,332
	56	0,602	0,640	0,698	0,357	0,480	0,465	5,50E-01	0,297
	91	0,584	0,665	0,699	0,380	0,414	0,432	5,19E-01	0,291
	147	0,618	0,680	0,722	0,430	0,448	0,489	5,83E-01	0,322
	237	0,641	0,695	0,767	0,491	0,492	0,506	6,39E-01	0,353
	383	0,664	0,709	0,792	0,526	0,479	0,491	5,99E-01	0,345

	620	0,699	0,706	0,826	0,541	0,466	0,445	5,98E-01	0,338
	1002	0,724	0,752	0,844	0,535	0,419	0,392	5,37E-01	0,302
	1621	0,725	0,756	0,867	0,549	0,382	0,338	4,79E-01	0,267
	2623	0,727	0,748	0,874	0,540	0,321	0,286	4,16E-01	0,230
	4244	0,707	0,725	0,867	0,499	0,262	0,225	3,48E-01	0,178
	6867	0,677	0,673	0,850	0,476	0,209	0,171	2,81E-01	0,138
	11112	0,614	0,586	0,812	0,390	0,138	0,115	2,01E-01	0,091
	17979	0,418	0,327	0,650	0,211	0,047	0,038	7,00E-02	0,035
	29090	0,018	-0,051	0,052	-0,013	0,000	-0,001	1,04E-07	-0,001
	2	0,340	0,314	0,201	0,110	0,114	0,136	0,107	0,057
	4	0,455	0,177	0,235	0,158	0,124	0,111	0,132	0,087
	5	0,475	0,188	0,234	0,146	0,118	0,112	0,126	0,076
	9	0,450	0,383	0,446	0,139	0,377	0,324	0,327	0,135
	13	0,542	0,446	0,483	0,156	0,482	0,432	0,401	0,163
	22	0,597	0,484	0,542	0,153	0,478	0,386	0,417	0,152
	35	0,631	0,578	0,623	0,207	0,500	0,440	0,482	0,207
	56	0,684	0,642	0,534	0,254	0,571	0,523	0,518	0,255
	91	0,718	0,648	0,593	0,236	0,589	0,525	0,540	0,232
	147	0,760	0,707	0,730	0,373	0,790	0,655	0,750	0,369
Volume	237	0,785	0,743	0,762	0,437	0,796	0,684	0,802	0,420
	383	0,800	0,743	0,801	0,453	0,712	0,603	0,724	0,383
	620	0,821	0,791	0,837	0,482	0,609	0,547	0,677	0,366
	1002	0,819	0,771	0,860	0,497	0,523	0,458	0,621	0,330
	1621	0,817	0,768	0,881	0,509	0,432	0,383	0,534	0,286
	2623	0,813	0,770	0,884	0,476	0,357	0,319	0,467	0,225
	4244	0,804	0,755	0,872	0,436	0,296	0,261	0,381	0,173
	6867	0,761	0,713	0,860	0,384	0,230	0,202	0,304	0,128
	11112	0,673	0,624	0,820	0,283	0,156	0,139	0,219	0,078
	17979	0,409	0,380	0,664	0,136	0,058	0,057	0,099	0,029
	29090	-0,072	-0,028	0,035	-0,058	-0,002	-0,002	0,001	-0,005
	2	0,340	0,314	0,200	0,110	0,114	0,136	0,107	0,057
	4	0,454	0,177	0,235	0,158	0,124	0,111	0,132	0,087
	5	0,475	0,188	0,234	0,146	0,118	0,112	0,126	0,076
IMA	9	0,449	0,383	0,446	0,139	0,377	0,324	0,327	0,135
	13	0,541	0,446	0,483	0,156	0,482	0,432	0,401	0,163
	22	0,596	0,484	0,542	0,152	0,478	0,386	0,417	0,152
	35	0,630	0,578	0,623	0,207	0,500	0,440	0,482	0,207

56	0,693	0,652	0,564	0,278	0,615	0,576	0,564	0,283
91	0,718	0,648	0,593	0,236	0,589	0,525	0,540	0,233
147	0,760	0,706	0,730	0,373	0,790	0,655	0,750	0,369
237	0,785	0,743	0,762	0,437	0,796	0,684	0,802	0,420
383	0,800	0,743	0,801	0,453	0,712	0,603	0,724	0,383
620	0,822	0,793	0,840	0,485	0,609	0,551	0,684	0,369
1002	0,819	0,771	0,860	0,498	0,524	0,459	0,621	0,330
1621	0,817	0,768	0,881	0,509	0,432	0,383	0,534	0,286
2623	0,813	0,771	0,884	0,476	0,357	0,319	0,467	0,225
4244	0,804	0,754	0,872	0,434	0,295	0,261	0,379	0,172
6867	0,761	0,713	0,860	0,384	0,230	0,202	0,304	0,128
11112	0,673	0,624	0,820	0,283	0,156	0,139	0,220	0,078
17979	0,409	0,380	0,664	0,136	0,058	0,057	0,099	0,029
29090	-0,072	-0,028	0,035	-0,058	-0,002	-0,002	0,001	-0,005

(conclusão)