



BRENO ALVARENGA REZENDE

**SELEÇÃO GENÔMICA AMPLA PARA
VOLUME E QUALIDADE DA MADEIRA EM
EUCALIPTO**

LAVRAS - MG

2015

BRENO ALVARENGA REZENDE

**SELEÇÃO GENÔMICA AMPLA PARA VOLUME E QUALIDADE DA
MADEIRA EM EUCALIPTO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Doutor.

Orientadora

Dra. Flávia Maria Avelar Gonçalves

Coorientadores

Dr. Magno Antonio Patto Ramalho

Dr. Aurélio Mendes Aguiar

LAVRAS - MG

2015

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Rezende, Breno Alvarenga.

Seleção genômica ampla para volume e qualidade da madeira
em eucalipto / Breno Alvarenga Rezende. – Lavras : UFLA, 2015.
74 p.

Tese(doutorado)–Universidade Federal de Lavras, 2015.

Orientadora: Flávia Maria Avelar Gonçalves.

Bibliografia.

1. Melhoramento de plantas. 2. *Eucalyptus spp.* 3. Genome
Wide Selection (GWS). 4. SNP. 5. DArT. I. Universidade Federal
de Lavras. II. Título.

BRENO ALVARENGA REZENDE

**SELEÇÃO GENÔMICA AMPLA PARA VOLUME E QUALIDADE DA
MADEIRA EM EUCALIPTO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Doutor.

APROVADA em 17 de Abril de 2015.

Dr. Magno Antonio Patto Ramalho	UFLA
Dr. Aurélio Mendes Aguiar	FIBRIA
Dr. Alexandre Alves Missiagia	FIBRIA
Dr. Luiz Antônio dos Santos Dias	UFV

Dra. Flávia Maria Avelar Gonçalves
Orientadora

LAVRAS – MG

2015

Aos meus pais Procópio e Maísa, e minha esposa Camila.

DEDICO

AGRADECIMENTOS

A Deus, por me dar incontáveis motivos para agradecer, e poucos para pedir.

Aos meus pais, Procópio e Maísa, por terem me ensinado que a maior herança de uma pessoa é o conhecimento; Pela educação moral e acadêmica que me proporcionaram; Pelo amor com o qual me criaram.

À minha esposa Camila, pelo apoio que me deu para que conseguisse concluir o doutorado, mesmo estando tão longe; pelas incontáveis qualidades que me fazem cada dia mais feliz; por ser a mulher da minha vida.

Aos meus avós João e Nicinha, pelos 9 anos que me acolheram em sua casa, e pelas tantas lições de vida que pude aprender com vocês. Aos tios Altamiro e Elaine, que também me acolheram como a um filho; à Olinda, Lucimara, Tião, Tiana e Ladinho, por terem me ensinado que o querer bem está nos gestos mais simples, vindos das pessoas mais simples, sem querer nada em troca. Ao vô Totonho e vó Telma, por todo o carinho que só os avós têm.

Aos irmãos Renato e Marina, cunhados Fernanda e Júlio, e sobrinha Lara, por todo o companheirismo e momentos bons que passamos em família.

À Universidade Federal de Lavras e seus professores, por todo o conhecimento que me transmitiram, em especial os professores do Departamento de Biologia e os queridos funcionários que ali trabalham.

À orientadora e amiga Flávia, pela paciência e incentivo em todo o doutorado, por sempre estar disponível para ajudar, e por ter me convencido a aceitar um projeto completamente novo até então, mas que contribuiu muito para minha formação profissional.

Ao coorientador Magno, por todas as lições desde o primeiro semestre da graduação.

Ao professor Márcio Balestre, pela ajuda com as análises.

À FIBRIA, na pessoa do Aurélio Mendes Aguiar, pelos dados fornecidos, por promover a parceria entre a empresa e a universidade, mesmo com tanto trabalho, e pela boa vontade em ajudar.

Aos colegas da Du Pont Pioneer, Wila, Sybelli, Paulo e Sheila, pelo grande apoio que me deram para que pudesse terminar a tese.

MUITO OBRIGADO!!!

RESUMO

Objetivou-se, neste trabalho, avaliar a viabilidade da aplicação da seleção genômica ampla (GWS), no melhoramento do eucalipto. Para tanto, foram empregados marcadores SNPs e DArTs na seleção precoce de uma população representativa do programa de melhoramento da empresa Fibria S.A. A população foi composta de híbridos, derivados do cruzamento entre plantas pré-selecionadas de *E. grandis*, *E. urophylla*, *E. globulus* e *E. camaldulensis* e de alguns híbridos entre elas. As combinações híbridas foram feitas ao acaso, gerando 68 combinações híbridas, envolvendo 45 genitores distintos e, ao final, foram obtidos 611 indivíduos. Os dados fenotípicos foram coletados aos 24, 36 e 63 meses de idade, totalizando 20 características relacionadas ao volume e qualidade da madeira, sendo que algumas delas foram mensuradas em mais de uma idade. Estes indivíduos também foram genotipados, utilizando-se marcadores SNPs e DArTs, totalizando 36.812 e 15.742 marcadores polimórficos, respectivamente. As análises da GWS foram feitas pela metodologia do RR-BLUP. Acurácias preditivas acima de 0,7 foram encontradas para a maioria das características, para ambos marcadores. Acurácias de validação da GWS foram 46% e 40% inferiores as da predição, em média, para SNPs e DArTs, respectivamente, sendo elas altamente influenciadas pela herdabilidade da característica e também da marca. SNPs e DArTs foram equivalentes em termos de capacidade preditiva (acurácia). A GWS mostrou-se eficiente para seleção precoce principalmente para as características circunferência à altura do peito e altura total das plantas, para ambos marcadores.

Palavras-chave: Melhoramento de Plantas. *Eucalyptus spp.*. Genome Wide Selection (GWS). SNP. DArT.

ABSTRACT

This project aimed to assess the feasibility of Genome Wide Selection (GWS) on eucalypt breeding. For that, SNP and DArT markers were used for the early selection of a representative population from Fibria S.A. The population was composed from hybrids, derived from the mating of pre-selected *E. grandis*, *E. urophylla*, *E. globules* and *E. camaldulensis* plants, and also some interspecies hybrids. Matings were carried out randomly between 45 distinct genitors, which generated 68 hybrid combinations and 611 individuals. Phenotypic data were collected at 24, 36 and 63 months old, for 20 traits related to volume and wood quality. Some traits were measured at two or more ages. Plants were also genotyped using SNP and DArT markers, totaling 36,812 and 15,752 polymorphic markers, respectively. GWS analysis were carried out using RR-BLUP methodology. Predictive accuracies of 0.7 and higher were found for most characteristics, for both markers. In average, GWS validation accuracies were 46% and 40% inferior than the predictive ones, for SNPs and DArTs, respectively. Validation accuracies were highly influenced by trait heritability and also marker heritability. SNPs and DArTs were very similar for predictive capacity (accuracy). GWS showed to be efficient for early selection, mainly for traits related to circumference at breast height and plant height, for both markers.

Key words: Plant breeding. *Eucalyptus spp.*. Genome Wide Selection (GWS). SNP. DArT.

LISTA DE FIGURAS

Figura 1	Acurácias preditivas e da validação de 20 características de eucalipto, utilizando o RR-BLUP	45
Figura 2	Acurácias da validação e herdabilidades de 20 características de eucalipto, utilizando o RR-BLUP (*Herdabilidades no sentido amplo).....	46
Figura 3	Manhattan plots dos efeitos de SNPs e DArTs para as 20 características avaliadas	51

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO	13
2.1	A Eucaliptocultura e a produção de celulose no Brasil	13
2.2	Características-Alvo do melhoramento do Eucalipto	14
2.3	Melhoramento de plantas com auxílio de marcadores moleculares	17
2.4	Histórico dos marcadores moleculares no melhoramento de plantas	23
2.5	Seleção Genômica Ampla	26
2.5.1	Princípios da seleção genômica ampla	26
2.5.2	Operacionalização da seleção genômica ampla	28
2.5.3	Metodologias Estatísticas para Computação da Seleção Genômica Ampla	31
3	MATERIAIS e MÉTODOS	34
3.1	Local	34
3.2	Materiais	34
3.3	Fenotipagem	34
3.4	Genotipagem	36
3.5	Estimação dos efeitos dos marcadores	37
3.6	População de validação dos efeitos dos marcadores	39
3.7	Viabilidade da seleção precoce	40
3.8	Viabilidade da seleção precoce via dados fenotípicos	40
3.9	Viabilidade da seleção precoce via dados moleculares	42
4	RESULTADOS	44
5	DISCUSSÃO	56
6	CONCLUSÕES	63
	REFERÊNCIAS	64

1 INTRODUÇÃO

O melhoramento de plantas, associado às práticas adequadas de manejo, são os grandes responsáveis pelos ganhos em produtividade na agricultura e no setor florestal brasileiro. O Eucalipto é um ótimo exemplo. Nos últimos 30 anos, a produtividade média de madeira de eucalipto teve um impressionante salto, passou de 24 m³/ha/ano para 44 m³/ha/ano, isto sem mencionar os demais atributos referentes à qualidade da madeira (ASSOCIAÇÃO BRASILEIRA DE CELULOSE E PAPEL- BRACELPA, 2014).

Este incremento em produtividade foi devido, em grande parte ao melhoramento genético, e, com certeza, a vertente do melhoramento “clássico”, como é chamado o melhoramento baseado exclusivamente em dados fenotípicos, teve e continua tendo papel fundamental e insubstituível nesse processo. No entanto, com o advento da biotecnologia, uma enorme expectativa foi criada no sentido de aumentar os ganhos por unidade de tempo, unindo dados fenotípicos e genotípicos, para auxiliar o melhorista no momento da seleção.

A inserção da biotecnologia no melhoramento de plantas – mais precisamente a biologia molecular - foi baseada no desenvolvimento de marcadores moleculares, ou seja, técnicas que permitem a localização específica de uma região do genoma das plantas. Tais marcas seriam úteis especialmente se estivessem bem próximas a um QTL (*Quantitative trait loci*) que controla o caráter de interesse, de modo que a seleção para este QTL poderia ser baseada no próprio genótipo da marca. Deste modo, teríamos uma “seleção assistida por marcadores”.

Apesar das promessas, todavia, até o início dos anos 2000, os marcadores moleculares tiveram aplicação restrita no melhoramento de plantas, principalmente no caso de caracteres quantitativos, controlados por muitos genes. O alto custo da técnica era um fator limitante no número de marcas

utilizadas e no número de indivíduos genotipados, de modo que, na maioria dos casos, um pequeno número de QTLs com grandes efeitos eram encontrados, mas que explicava uma pequena proporção da variação genética.

O desenvolvimento de técnicas de genotipagem em larga escala, a custo reduzido, surgiu como uma promessa para atenuar as falhas proporcionadas pelos marcadores moleculares até então utilizados. Por meio dos marcadores SNP (*Single Nucleotide Polymorphism*) e DArT (*Diversity Arrays Technology*), seria possível saturar o genoma com milhares de marcas, aumentando em muito a probabilidade de que, pelo menos uma delas, estivesse bem próxima dos QTLs que controlam o caráter. Assim sendo, grande parte da variação genética seria capturada, permitindo a aplicação da “Seleção Genômica Ampla”, ou “*Genome Wide Selection – GWS*”.

No caso de plantas perenes, como o Eucalipto, vislumbrou-se uma ampla gama de aplicações para a GWS. A maior delas é a possibilidade de selecionar indivíduos ainda em fase de viveiro, reduzindo, em muito, o tempo para completar um ciclo de melhoramento. Como a GWS é uma teoria muito recente (MEUWISSEN; GODDARD; HAYES, 2001), poucos estudos foram conduzidos com dados reais no intuito de se avaliar a técnica (RESENDE et al., 2012; RESENDE JÚNIOR, 2010). Assim, várias dúvidas e questionamentos persistem no processo, como qual seria a sua acurácia em prever fenótipos baseada apenas no genótipo, a idade ideal da planta para fenotipagem dos caracteres, o melhor marcador a ser utilizado dentre os disponíveis. Desta maneira, apesar de parecer extremamente promissora, estudos básicos devem ser conduzidos para validar a implementação da GWS, na cultura do Eucalipto.

Objetivou-se, neste trabalho, avaliar a viabilidade da aplicação da seleção genômica ampla (GWS) no melhoramento do eucalipto. Para tanto, foram empregados marcadores SNPs e DArTs na seleção precoce de uma população representativa do programa de melhoramento da empresa Fibria S.A.

2 REFERENCIAL TEÓRICO

2.1 A Eucaliptocultura e a produção de celulose no Brasil

O Brasil é responsável por, aproximadamente, 20% da área plantada com Eucalipto em todo o planeta, o que corresponde a, aproximadamente, 4 milhões de hectares. Destes, aproximadamente 2 milhões de hectares (ou 50%) são plantados com florestas clonais. O uso de clones, técnicas adequadas de manejo e o melhoramento de plantas proporcionaram um salto de produtividade de 24 m³/ha/ano para 44 m³/ha/ano de madeira nos últimos 30 anos, o que corresponde a um aumento de 83% em produtividade. Quando se compara a produtividade do Brasil, no período de rotação de 7 anos (o menor do mundo), com outros países, como, por exemplo, a África do Sul, que tem o segundo menor período de rotação (8 a 10 anos) e uma produtividade de 20 m³/ha/ano, esse incremento é ainda mais impressionante (BRACELPA, 2014).

No ano de 2011, o Brasil produziu 195 milhões de m³ de madeira de eucalipto. Dentre os diversos usos, 35,2% desse total foi utilizado para produção de polpa de celulose e papel, 2,7% para produção de compensado, 16,4% para serraria (móveis, pisos, dentre outros), 7,1% para painéis de madeira industrializada e 38,7% para carvão, lenha e outros fins (ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS - ABRAF, 2014).

O consumo médio de papel no mundo é de 57 kg per capita. O país maior consumidor é a Finlândia, com 281 kg per capita. Apesar de o Brasil ser o quarto maior produtor de celulose e o nono maior produtor de papel, é apenas o 13º maior consumidor de papel com 49 kg per capita, abaixo da média mundial (BRACELPA, 2014).

Neste cenário, o setor de eucalipto para produção de celulose tem papel de destaque. Conforme comentado anteriormente, 35,2% do total de eucalipto plantado tem como finalidade a produção de polpa de celulose, para confecção de papel e seus derivados. Acompanhando a tendência da produtividade, entre 1980 e 2013 o crescimento médio anual da produção de celulose no País foi de 7,1%, passando de 3,1 para 15 milhões de toneladas de celulose, neste período. O saldo da balança comercial de celulose, em 2013 foi de 4,8 bilhões de reais, e a indústria empregou mais de 768 mil pessoas, entre empregos diretos e indiretos (BRACELPA, 2014).

2.2 Características-Alvo do melhoramento do Eucalipto

Dentre as características da planta, observadas no melhoramento do Eucalipto, podemos citar algumas como a resistência à doenças, forma do tronco e a capacidade de se propagar vegetativamente. No entanto, dentre as que interferem no maior rendimento de celulose, com certeza aquelas relacionadas ao volume de madeira, ou seja, acircunferência à altura do peito e a altura da planta, receberam a maior atenção dos melhoristas, ao longo do tempo. Vale ressaltar que o progresso alcançado foi enorme (ASSIS; MAFIA, 2007; SILVA; BARRICHELLO, 2006), resultando como já comentado, em um aumento de 83% em produtividade, entre 1980 e 2013 (BRACELPA, 2014). No entanto, projeções indicam que o potencial produtivo da cultura no Brasil é de 70 m³/ha/ano, o que justifica os contínuos esforços dos melhoristas e, também, as constantes melhorias nas técnicas de manejo, corresponsáveis pelo incremento em produtividade, alcançado (BRACELPA, 2014).

Para uma medição mais acurada do volume da madeira, é comum mensurar também a espessura da casca. Não faz sentido a seleção de progênies com alto volume de madeira, porém com um percentual grande desse volume

ocupado pela casca. A casca da árvore não é aproveitada pela indústria para extração de celulose, sendo inclusive removida e deixada no campo durante o corte. Por conseguinte, os melhoristas almejam selecionar as plantas com menor volume de casca.

Apesar do rendimento bruto (volume) de celulose ser o principal alvo dos melhoristas, uma série de outras propriedades tecnológicas da madeira são levadas em consideração no momento da seleção. Talvez a principal delas seja a densidade básica, que pode ser definida como a relação entre a massa de madeira seca em estufa e o volume obtido acima do ponto de saturação das fibras. Assim sendo, além do volume ela é influenciada também pela composição química da madeira (SHIMOYAMA; BARRICHELO, 1989), afetando o custo de transporte (em função da maior ou menor massa transportada em um mesmo volume) e também o processamento pela indústria, uma vez que, com madeiras mais densas, uma maior quantidade de fibra é processada por unidade de tempo (ASSIS, 1996; BORRALHO; ALMEIDA; POTTS, 2007).

Existe um método indireto de medição da densidade básica, através de um instrumento chamado “Pilodyn”. Este aparelho mede a resistência à penetração no tronco (sem casca) de um pino rígido sob pressão constante, cuja profundidade de penetração é, em princípio, inversamente proporcional à densidade da madeira. Além de ser um método não destrutivo, apresenta ainda as vantagens de rapidez, isenção de erros sistemáticos por parte do operador, a não necessidade de extração de amostras de madeira com sondas e, ainda, custos inferiores a outros métodos (GREAVES et al., 1996; MOURA; SANTIAGO, 1991). Todavia, como esta metodologia faz uma medida indireta da densidade da madeira, não é seguro que ele consiga traduzir com um mínimo de rigor a densidade real da madeira (MUNERI; RAYMOND, 2000). Além do método do Pilodyn, é comum a mensuração do número de fibras por grama, utilizado

também como método indireto para a densidade básica, já que ambas estão bastante correlacionadas (FOELKEL, 1997).

Considerando-se que a celulose é uma fibra da madeira, e que a finalidade do uso depende das propriedades desta fibra, existe uma série de características que são utilizadas para verificar tais qualidades. O número de fibras por grama, por exemplo, além de ser um método indireto para a densidade, também é uma medida importante para a produção de papéis para imprimir e escrever. Quanto maior for este número, melhor será a formação e a estabilidade dimensional do papel. Já o *coarseness* é a relação entre o peso das fibras, por unidade de comprimento das mesmas. Fibras mais pesadas, de paredes mais espessas, elevam o *coarseness*. No caso de papéis para imprimir e escrever, celuloses com baixo *coarseness* são mais indicadas. O comprimento ponderado, por sua vez, é uma medida da relação entre a largura e o comprimento da fibra, característica que também influencia na qualidade do papel. O rendimento depurado, como o próprio nome sugere, é o rendimento da polpa, ou seja, porcentagem da massa seca que passou pelo depurador, em função da massa seca total que entrou no digestor. Por fim, o teor de pentosanas (hemicelulose que compõe a parede celular) é importante para facilitar o refino, promover a ligação entre fibras e melhorar a resistência da folha de papel. Assim sendo, este teor deve ser alto para fabricação de papéis para imprimir e escrever, e baixo para fabricação de papéis sanitários (FOELKEL, 1997).

Apesar de a celulose ser o componente da madeira utilizado pela indústria para fabricação dos mais diversos produtos, outro componente que também atrai a atenção dos melhoristas de eucalipto é a lignina. Ela compõe a parede celular, juntamente com a celulose, tendo como funções, dentre várias, a de conferir rigidez à madeira, impermeabilidade e resistência contra ataques biológicos aos tecidos vegetais. Apesar de possuir estas nobres funções, ela é um empecilho no processo de extração da celulose, pois ambas estão associadas e

devem, portanto, ser separadas quimicamente. Assim sendo, é comum a mensuração do teor de lignina da madeira, e o melhoramento busca a redução deste teor para facilitar o processo de extração da celulose. Uma vez que a lignina é importante para evitar o tombamento das plantas, sendo assim inviável removê-la totalmente, busca-se também o aumento do teor de um tipo de lignina mais facilmente removível da madeira, a Siringil, em detrimento do tipo Guaiacil, mais difícil de ser removida. Assim sendo, uma maior relação de lignina Siringil/Guaiacil também é alvo do melhoramento.

2.3 Melhoramento de plantas com auxílio de marcadores moleculares

O objetivo de qualquer programa de melhoramento de plantas é obter cultivares/híbridos/clones geneticamente superiores. Para este fim, são necessárias três etapas básicas:

- a) geração de variabilidade;
- b) avaliação da variabilidade gerada;
- c) seleção dos genótipos superiores. Para a geração de variabilidade são realizados cruzamentos controlados entre genitores que possuam características florestais de interesse, ou seja, que possuam alelos favoráveis para tais características.

Para a segunda etapa, a avaliação da população segregante gerada, são planejados e instalados os experimentos, principalmente de campo, para então se proceder à avaliação dos genótipos. Devido à necessidade constante de se gerarem genótipos superiores, o planejamento dos experimentos e as avaliações devem ser conduzidos com grande precisão, para então se aumentarem as chances de selecionar genótipos superiores, que é a terceira parte deste ciclo.

No caso do Eucalipto, as avaliações dos genótipos são principalmente fenotípicas, o que envolve grande quantidade de mão de obra, além de um longo período de tempo, uma vez que o gênero *Eucalyptus* é perene. Esta etapa é, portanto, a mais laboriosa, dispendiosa e que consome maior período de tempo, em um programa de melhoramento (RAMALHO; FERREIRA; OLIVEIRA, 2012). Além disso, para caracteres com baixa herdabilidade, ou seja, que são altamente influenciados pelo ambiente, como é o caso da maioria das características quantitativas, muitas vezes o efeito do genótipo fica confundido com o efeito do ambiente, o que leva a uma baixa acurácia no momento da seleção (RAMALHO et al., 2012). Visando aumentar a eficiência deste processo, portanto, vislumbrou-se a necessidade e a possibilidade de se realizarem as avaliações dos indivíduos, com base não somente em seus fenótipos observados, mas também em seus genótipos. Este momento, então, marcou o início da seleção assistida por marcadores moleculares, ou MAS, da expressão em inglês “*Marker Assisted Selection*”.

O princípio da MAS consiste em localizar no genoma da planta, marcas (ou sequências de pares de bases), que estejam próximas ou mesmo dentro de um alelo responsável por um fenótipo de interesse. Quando isso acontece, é dito que existe ligação gênica entre marcador e o loco de interesse. Obviamente, o loco marcado deve apresentar polimorfismo entre os indivíduos da população, ou seja, deve existir variabilidade para o caráter em estudo. Desta maneira, a ideia é de que a marca e o gene estejam intimamente associados, de modo que a presença da marca molecular, em teoria, indica a presença do alelo de interesse.

Na cultura do Eucalipto, marcadores moleculares têm sido utilizados com alguns propósitos, como verificação da identidade de árvores (GAIOTTO; BRAMUCCI; GRATTAPAGLIA, 1997; KEIL; GRIFFIN, 1994; KIRST et al., 2005), distância genética (BARIL et al., 1997; NESBITT et al., 1995), e construção de mapas genéticos (BRONDANI et al., 2006; MARQUES et al.,

2002). No caso de características qualitativas, controladas por um ou poucos genes, o uso de marcadores para seleção assistida tem sido eficaz, em alguns casos, como na seleção de genótipos resistentes à doenças (FREEMAN; POTTS; VAILLANCOURT, 2008; JUNGHANS et al., 2003) e florescimento precoce (MISSIAGIA; PIACEZZI; GRATTAPAGLIA, 2005).

No caso de características quantitativas, que são controladas por muitos genes, a eficiência da MAS está diretamente ligada a alguns pressupostos (RESENDE et al., 2008). O primeiro pressuposto é que deve haver desequilíbrio de ligação (*Linkage Disequilibrium* - LD), entre as marcas e os QTLs. Este é o pressuposto básico para a MAS, e para qualquer análise que envolva marcadores moleculares.

O LD pode ser definido como a dependência entre alelos de dois ou mais locos. Se a frequência observada de dois alelos de locos diferentes é maior do que a esperada (via produto das frequências de cada alelo separadamente), estes alelos estão em desequilíbrio de ligação, e depreende-se, portanto, que os mesmos estão localizados próximos, no cromossomo (FALCONER; MACKAY, 1996). Uma vez que exista LD e, por conseguinte, o marcador e o alelo de interesse estiverem próximos, existirá, portanto, uma correlação genética entre eles. Quanto menor a distância entre o marcador e o loco, menor será a frequência de recombinação entre eles, e mais informativo será o marcador nesse caso. No entanto, caso a frequência de recombinação entre eles seja muito grande, pode-se considerar que o marcador e o loco são independentes e, em um caso como este, o marcador não tem utilidade alguma.

Outro fator, que afeta significativamente a eficiência da MAS, é o percentual da variação genética explicada pelos marcadores. Como visto, para que o marcador seja informativo, é preciso que ele esteja bem próximo ao alelo de interesse. Para caracteres qualitativos, isto é relativamente simples de ser conseguido e tem sido bastante empregado, como visto anteriormente. Nestes

casos, poucos QTLs de grande efeito explicam quase ou toda a variação genética, fazendo com que a herdabilidade do caráter seja um ou próxima da unidade, situação almejada por todos os melhoristas. Assim, poucos marcadores bem próximos aos QTLs são necessários. No entanto, no caso de caracteres quantitativos, a situação é bem mais complexa. Isto porque caracteres como altura da planta, volume de madeira, dentre outros, são controlados por inúmeros genes, de pequeno efeito cada. Fica fácil perceber que conseguir marcadores com forte LD, com cada um desses genes, é uma tarefa extremamente difícil.

Um terceiro pressuposto para a MAS é o tamanho da família fenotipada e genotipada. Segundo Resende et al. (2008), este deve ser superior a 500, para que a maioria dos alelos que representam o caráter estejam presentes na população. Se inferior, a seleção fenotípica deve ser preferida.

Desde os primeiros trabalhos que incentivaram o uso de marcadores moleculares e a MAS em plantas perenes (FERNANDO; GROSSMAN, 1989; GODDARD, 1991, 1992; KENNEDY; QUINTON; ARENDONK, 1992; LANDE; THOMPSON, 1990; MEUWISSEN; ARENDONK, 1992), uma expectativa enorme foi criada em torno dos benefícios que os marcadores e a MAS poderiam agregar ao melhoramento de plantas, chegando ao ponto de se imaginar que a seleção fenotípica logo seria substituída (MAZUR, 1995). Os ganhos reais, no entanto, foram muito aquém dos esperados (BERNARDO, 2008; DEKKERS, 2004). Diversas críticas à metodologia da MAS surgiram então como forma de justificar os limitados ganhos obtidos, os quais serão discutidos à seguir.

Para caracteres quantitativos, que são a grande maioria, e os principais alvos dos melhoristas, o número de genes que controla o caráter pode ser muito grande. Isto implica que primeiro é necessário que cada um desses genes seja detectado e sua localização mapeada, para então se proceder à MAS propriamente dita. Fica fácil perceber, que essa tarefa se torna quase impossível,

à medida que o número de genes envolvidos aumenta, juntamente com a complexidade do caráter (DUDLEY, 1993).

Um exemplo interessante, que representa bem este cenário, foi dado por MacKay (2001), utilizando o gênero *Drosophila*. Ele verificou que o tamanho médio do intervalo, contendo um QTL neste gênero é de 8,9 cM, podendo variar de 0,1 a 44,7 cM. Quando se considera 8,9 cM, isto equivale a uma região do DNA de 4.459 Kb. Ramalho e Lambert (2004), ao analisarem os resultados de MacKay (2001) inferiram que, como o genoma da *Drosophila* possui 120 Mb, é provável que ele tenha 13.600 genes, com tamanho médio de 8,8 Kb. Portanto, em um QTL médio estariam incluídos 507 genes, podendo variar de 11 a 2.191 genes. Como se observa, a incerteza, mesmo em espécies modelo, é enorme.

Talvez essa mesma situação ocorra em plantas, na qual apenas um pequeno número de QTLs de grandes efeitos são detectados, sendo que um número grande de genes passa despercebido no momento da seleção, e junto deles grande parte da variação genética do caráter não é capturada. Isto leva à conclusão que a MAS não é útil para caracteres de baixa herdabilidade (DUDLEY, 1993).

O método mais comum usado para detecção de QTLs é o dos cruzamentos biparentais. Apesar de estes serem importantes para a compreensão da arquitetura genética da população, na grande maioria das situações não refletem bem as populações de melhoramento, em que diversas situações mais complexas podem acontecer. No caso de vários parentais envolvidos em uma população, por exemplo, a acurácia na detecção de QTLs e seus efeitos fica comprometida (DEKKERS; HOSPITAL, 2002; SCHON et al., 2004). Além disso, a diversidade alélica de uma população de melhoramento não é capturada com um simples cruzamento biparental (HEFFNER; SORRELLS; JANNINK, 2009). Agregado a este fato, a MAS requer o estabelecimento de associações marcadores-QTLs, para cada família em avaliação. No caso de espécies

alógamas, como o Eucalipto, por exemplo, estas associações são úteis apenas para a família avaliada. O mesmo procedimento deve ser repetido (ou validado), portanto, nas demais famílias do programa. Considerando que os marcadores anteriores aos SNPs (RFLP, RAPD, AFLP, Microsatélites, etc) são extremamente laboriosos, muitas vezes dispendiosos e demorados, fica claro que a aplicabilidade destes é limitada, tendo em vista o grande número de famílias a serem avaliadas em um programa de melhoramento de grande porte.

Apesar de um dos incentivos à MAS ter sido a substituição das avaliações fenotípicas por moleculares apenas, na prática isto nunca aconteceu por completo. Pelo contrário, uma fenotipagem precisa é fundamental para a MAS. Desta maneira, no caso de caracteres de alta herdabilidade, que propiciam ganhos com a seleção apenas pelo fenótipo, a utilização da MAS é desnecessária. No caso de caracteres de baixa herdabilidade, ainda assim a avaliação fenotípica se faz necessária. Neste caso, porém, a confiabilidade dos dados fica comprometida pelo efeito do ambiente. Mais comprometida ainda fica a eficiência da MAS, pelas razões apresentadas anteriormente. Desta maneira, o melhor uso da MAS se faz sob condições de alta herdabilidade, quando a característica de interesse é de difícil mensuração (DUDLEY, 1993).

Em resumo, a detecção e estimação do efeito dos vários QTLs de uma característica quantitativa é uma tarefa bastante complicada com o uso de um número reduzido de marcadores, pois fatores como a baixa herdabilidade do caráter, tamanho populacional insuficiente, estruturas populacionais complexas e estabelecimento de limiares de significância arbitrários, levam à detecção de alguns poucos QTLs, com efeitos superestimados na maioria dos casos.

2.4 Histórico dos marcadores moleculares no melhoramento de plantas

A utilização dos marcadores moleculares no melhoramento de plantas teve início em meados de 1980, quando marcadores isoenzimáticos (alozimas) eram utilizados para agilizar a introgressão de genes de parentais exóticos, na cultura do tomate (TANKSLEY; RICK, 1980). Este método, no entanto, era muito precário, pois o número de isoenzimas era muito pequeno, e estas não eram capazes de detectar variação no DNA (SCHLÖTTERER, 2004). A partir daí, com o aperfeiçoamento de técnicas de manipulação de DNA, surgiu o primeiro marcador molecular de DNA propriamente dito, o RFLP – Restriction Fragment Length Polymorphism (BOTSTEIN et al., 1980). Após a descoberta do RFLP, uma série de outros marcadores foram propostos, como os Minissatélites (JEFFREYS; WILSON; THEIN, 1985) e aqueles baseados em técnicas de PCR, a exemplo dos Microsatélites ou SSR – Simple Sequence Repeat (TAUTZ, 1989), RAPD – Randomly Amplified Polymorphic DNAs (WILLIAMS et al., 1990) e AFLP – Amplified Fragment Length Polymorphisms (VOS et al., 1995). Estes marcadores analisavam três maneiras distintas de variação: variação a nível de proteínas (alozimas), variação na sequência de DNA (RFLP, RAPD e AFLP) e variação no tamanho de sequências repetidas de DNA (Minissatélites e Microsatélites). Estes marcadores, no entanto, apesar de muito promissores à primeira vista, contribuíram basicamente no caso de caracteres qualitativos, mas não foram de grande valia para a MAS de maneira geral (XU; CROUCH, 2008). As razões que levaram a isso já foram discutidas anteriormente, mas podem ser resumidas em três: pequeno número de marcadores empregados, a laboriosidade para execução das técnicas, e o tempo e custo despendidos para execução.

Neste contexto, surgiu um novo tipo de marcador, os chamados SNPs – *Single Nucleotide Polymorphisms*. Como o próprio nome sugere, os SNPs são os polimorfismos de um único nucleotídeo, encontrados em todo o genoma. Desta

maneira, o que diferencia os SNPs dos demais marcadores não é a origem do polimorfismo em si, pois todos os demais marcadores já existentes, de certa forma, são também baseados no polimorfismo de uma ou mais bases. A grande diferença dos SNPs para os demais marcadores é o alto rendimento do processo, e o enorme número de marcadores que pode ser obtido. Isto porque os SNPs são a fonte mais abundante de variação no genoma (CHEE et al., 1996). A identificação dos mesmos é possível simplesmente comparando-se a sequência de bases dos indivíduos de uma população (VARSHNEY et al., 2009). Se há polimorfismo há um SNP, mas para que uma variação possa ser considerada SNP, ela deve ocorrer em, pelo menos, 1% da população. Ele é um marcador codominante (ou bialélico), ou seja, consegue distinguir os homocigotos (AA ou aa) dos heterocigotos (Aa) (BERNARDO, 2014; BROOKES, 1999).

Outro marcador que tem recebido atenção, atualmente, é o DArT – *Diversity Arrays Technology* (JACCOUD et al., 2001; WENZL et al., 2004). Como o próprio nome sugere, a base para os marcadores DArT são fundamentadas em técnicas de microarranjo. O resumo da técnica consiste nos seguintes passos: a amostra de DNA de cada indivíduo é digerida com enzimas de restrição, e então são ligados adaptadores aos fragmentos digeridos. Posteriormente, é feita uma PCR onde são utilizados primers complementares à sequência dos adaptadores, para reduzir a complexidade do genoma. Os fragmentos amplificados são clonados e novamente amplificados, purificados e arranjados em um suporte sólido (microarranjo), resultando em um “*discovery array*”. Representações genômicas preparadas, a partir de genomas individuais alvos de estudo são hibridizados a esse “*discovery array*” para identificação de polimorfismos. Os clones polimórficos (marcadores DArT) mostram intensidade de sinais de hibridização variáveis para diferentes indivíduos, conforme os diferentes genótipos (JACCOUD et al., 2001). De acordo com este autor, a grande vantagem do DArT, em relação ao SNP, consiste no fato da tecnologia

dos DArTs não necessitar do sequenciamento do genoma, enquanto que a tecnologia dos SNPs depende de um sequenciamento genômico intensivo e laborioso que, juntamente com a dificuldade de análise dos dados, torna o processo caro. Segundo Killian et al. (2005), a desvantagem dos marcadores DArT é que eles são dominantes. Outra desvantagem é que a quantidade de DNA necessária para a técnica é quase o dobro da de SNPs (dados não publicados).

Recentemente, uma nova técnica de genotipagem para marcadores DArT foi proposta: DArT-seq. Este novo marcador foi desenvolvido a partir da técnica de genotipagem por sequenciamento, ou *GBS (Genotyping-by-Sequencing)* (ELSHIRE et al., 2011). A nova técnica permite um rendimento operacional superior ao DArT convencional, e utiliza as tecnologias de sequenciamento da próxima geração (*Next Generation Sequencing- NGS*). De acordo com Elshire et al. (2011), a técnica do DArT-seq é simples, rápida, extremamente específica, altamente reproduzível e pode alcançar regiões importantes do genoma que antes eram inacessíveis ao sequenciamento. A técnica consiste de três etapas básicas: a redução da complexidade por enzimas de restrição, o preparo das amostras para o sequenciamento e a genotipagem via sequenciamento (SANSALONI et al., 2011). O DArT-seq é capaz de detectar polimorfismos de um único nucleotídeo (SNPs).

Na cultura do Eucalipto, alguns trabalhos mostram a utilização de marcadores DArT, auxiliando o melhoramento de plantas. Resende et al. (2012) utilizaram mais de 3.000 marcadores DArT para características de crescimento e qualidade da madeira, visando avaliar a metodologia da GWS. Até o momento, poucos foram os trabalhos que utilizaram dados reais de SNPs para GWS (LIMA, 2014).

2.5 Seleção Genômica Ampla

Os princípios, a operacionalização e as metodologias para análise serão discutidos a seguir.

2.5.1 Princípios da seleção genômica ampla

O atrativo da genética molecular em benefício do melhoramento de plantas é a utilização direta das informações do DNA na seleção, permitindo assim, alta eficiência seletiva, maior rapidez na obtenção de ganhos genéticos e baixo custo, isto comparado à seleção fenotípica tradicional (RESENDE et al., 2008). Infelizmente, no entanto, isto não foi plenamente alcançado via MAS, pelas razões citadas anteriormente. Visando esses objetivos, Meuwissen et al. (2001) propuseram um novo método de seleção via marcadores moleculares, denominado “Seleção Genômica” ou “Seleção Genômica Ampla – *Genome Wide Selection* (GWS)”.

A Seleção Genômica Ampla (GWS) proposta por Meuwissen et al. (2001) tem como diferencial o grande número de marcadores utilizados (na ordem de milhares), que cobrem de maneira densa o genoma. À medida que o número de marcadores utilizados aumenta, a distância física entre eles no genoma diminui, e a chance de que todos os genes de um caráter quantitativo estejam em desequilíbrio de ligação (LD) com, pelo menos, alguns marcadores aumenta muito. Como somente marcadores em LD são úteis para explicar a variação genética, é esperado que grande parte ou quase a totalidade desta variação genética seja explicada pelos marcadores. Este é um dos pressupostos que muitas vezes não é atendido pela MAS.

Os tipos de marcadores preferencialmente utilizados na GWS são os SNPs e os DArTs. No início da divulgação da técnica, os SNPs eram

considerados muito promissores para o melhoramento de plantas de maneira geral, mas não foram vastamente empregados devido à complexidade de identificação e seu alto custo inicial (GANAL; ALTMANN; RÖDER, 2009). Isto fez com o que o trabalho de Meuwissen et al. (2001) permanecesse discreto por alguns anos. No entanto, o aperfeiçoamento de técnicas e custos decrescentes de genotipagem, atualmente permitem seu uso em diversas culturas (AKBARI et al., 2006; BERNARDO; YU, 2007; OLIVEIRA et al., 2012; RESENDE et al., 2012; WENZL et al., 2004), de modo que a evolução da GWS coincide com a evolução dos SNPs.

As técnicas atuais de genotipagem são de alto rendimento (“*high-throughput*”), o que permitiu a redução dos custos por genótipo. Isto permitiu, conseqüentemente, que um grande número de genótipos pudesse ser genotipado a um preço acessível. Desta maneira, populações inteiras, oriundas de diferentes famílias, podem ser genotipadas simultaneamente, fazendo com que o impacto de famílias específicas, (com padrões específicos de LD), na estimação dos efeitos dos marcadores seja diluído (RESENDE, 2008).

O fato de um grande número de indivíduos de diferentes famílias, de uma população, serem genotipados, com um grande número de marcadores (geralmente distanciados de 1cM ou menos), permite duas vantagens teóricas em relação à MAS: primeiramente, como o efeito dos marcadores são estimados com muitas repetições, (cada indivíduo é uma repetição), a confiabilidade desta estimativa é bem alta, e o efeito destes mesmos marcadores não precisaria (em teoria) ser reestimado para famílias aparentadas. A segunda vantagem consiste no fato de que, com marcadores muito próximos e fortemente ligados aos alelos de interesse ou QTLs, a frequência de recombinação entre eles é bem baixa. Assim sendo, mesmo após uma geração de inter cruzamento, marcador e alelo continuariam próximos e o marcador continuaria informativo. Apesar destas

vantagens teóricas, mais estudos práticos neste sentido são necessários para comprovação.

De forma semelhante à MAS, na GWS também é realizada a associação dos marcadores com o fenótipo. No entanto, a escolha dos marcadores não é realizada previamente com base na significância de suas associações com o fenótipo, ao contrário da MAS. Isto porque na MAS os QTLs de pequeno efeito (ou não significativos) são desprezados, e, pela GWS, os pequenos efeitos dos QTLs são levados em conta e somados àqueles de grande efeito, capturando assim uma parte maior da variância genética. Neste caso, o que realmente importa é o efeito dos marcadores, não sendo necessárias as suas localizações no genoma. Apesar de a GWS levar em conta todas as marcas no momento da análise, algumas metodologias discutem a opção de se considerar apenas uma parte delas, partindo do princípio que, quando um número muito grande de marcas é utilizado, muito provavelmente uma grande porcentagem delas não terá associação alguma com o caráter em estudo e, portanto, deveriam ser eliminadas (Bayes B) ou testadas quanto à sua significância (teste de Wald). Estas metodologias foram utilizadas em algumas oportunidades (RESENDE et al., 2012; RESENDE JÚNIOR, 2010).

2.5.2 Operacionalização da seleção genômica ampla

Para implementar a GWS, primeiramente é necessário estimar os efeitos dos marcadores. O próximo passo é validar estes efeitos, ou seja, verificar a acurácia na estimação dos mesmos. Estes dois passos requerem dados fenotípicos. Por fim, com efeitos de marcadores estimados e validados pode-se proceder à Seleção Genômica propriamente dita, que seria a seleção de indivíduos com base apenas em seus genótipos, ou em seus valores genéticos

genômicos (VGG). Para isto, três populações são necessárias (GODDARD; HAYES, 2007):

População de Descoberta. Nesta população, um grande número de marcadores SNP ou DArT são utilizados para cobrir de maneira densa o genoma dos indivíduos, os quais também têm seus fenótipos avaliados para as características de interesse. O método de estimação dos efeitos dos marcadores é variável, e será discutido posteriormente. É importante ressaltar que a precisão na medição dos fenótipos é tão ou mais importante que o número de marcadores utilizados, uma vez que, o efeito destes marcadores será estimado com base nestes fenótipos. Resende et al. (2008) comentam que esta população deve ter tamanho maior ou igual a 500 indivíduos.

Uma vez estimados os efeitos dos marcadores, calcula-se então o valor genético genômico (VGG) de cada indivíduo desta amostra, da seguinte maneira: o efeito de cada marcador (estimado na população de descoberta) é multiplicado pela codificação dada ao número de alelos favoráveis que o indivíduo apresenta para este loco. No caso de SNPs, este número pode ser -1, 0 ou 1 (homozigoto recessivo, heterozigoto e homozigoto dominante), e no caso de DArTs pode ser 0 ou 1, uma vez que é marcador dominante. Este processo é repetido para todos os locos. Assim, o VGG do indivíduo k , pode ser computado da seguinte maneira: $VGG_k = [(\text{genótipo loco 1}) \times (\text{efeito marca 1})] + [(\text{genótipo loco 2}) \times (\text{efeito marca 2})] + \dots + [(\text{genótipo loco } n) \times (\text{efeito marca } n)]$. Por exemplo: $VGG_k = [0 \times (0,25)] + [1 \times (-0,10)] + \dots + [-1 \times (0,11)]$, em que, o valor entre parêntesis exemplifica o efeito do marcador estimado na população de descoberta. A acurácia da estimação, portanto, nada mais é do que a correlação de Pearson entre o VGG dos indivíduos e seus fenótipos.

População de Validação. Uma vez estimados os efeitos dos marcadores, é necessário validá-los, ou seja, verificar a acurácia na estimação

dos mesmos. Para isso, repete-se o processo descrito na população de descoberta, desta vez porém estimando o efeito das marcas apenas com uma parte dos indivíduos. Os demais indivíduos que não fizeram parte desta população de descoberta são utilizados então para compor a população de validação, em um processo conhecido como “Validação Cruzada”. Essa amostra geralmente é menor que a população de descoberta e contempla indivíduos também fenotipados e genotipados, com os mesmos marcadores SNP ou DArT, utilizados na população de descoberta. Uma vez que este processo pode ser repetido inúmeras vezes, a acurácia da validação nada mais é do que a média de todas as acurácias obtidas pela correlação entre o VGG das plantas que compuseram a população de validação e seus fenótipos.

Como a amostra de validação não foi envolvida na predição dos efeitos dos haplótipos marcadores, os erros dos valores genéticos genômicos preditos e dos valores fenotípicos são independentes e toda correlação entre esses valores é de natureza genética e equivale à própria acurácia (RESENDE et al., 2008).

Os valores de acurácia para validação encontrados na literatura são dependentes da espécie e do caráter em estudo. Meuwissen et al. (2001) encontraram acurácias de 85% e 71% para populações de descoberta de 2.200 e 500 indivíduos, respectivamente, para dados simulados. Já Resende et al. (2012), avaliando os caracteres circunferência à altura do peito, crescimento em altura, peso específico da madeira e produtividade de polpa de celulose encontraram acurácias de 0,54, 0,51, 0,60 e 0,54, respectivamente, para híbridos entre *E. grandis* e *E. urophylla*, pertencentes à empresa CENIBRA. As acurácias também foram computadas para híbridos entre *E. grandis*, *E. urophylla* e *E. globulus*, pertencentes à empresa FIBRIA, e as estimativas para circunferência à altura do peito, crescimento em altura, peso específico da madeira e produtividade de polpa de celulose foram de 0,55, 0,46, 0,42 e 0,38, respectivamente.

População de Seleção. O grande diferencial da GWS no melhoramento do Eucalipto é a possibilidade de seleção, enquanto as plantas ainda são jovens, o que promoveria uma economia enorme de tempo, uma vez que não seria necessário aguardar entre 3 a 7 anos para proceder à seleção. Imaginemos uma situação em que os passos descritos na população de descoberta e de validação foram realizados para uma população, e que os melhores indivíduos selecionados foram inter cruzados. A GWS poderia ser aplicada nos filhos destes cruzamentos ainda na fase de mudas, sem a necessidade de avaliação dos fenótipos, uma vez que o efeito dos marcadores já teria sido validado. Estes indivíduos filhos, portanto, comporiam a população de seleção, com a possibilidade inclusive de serem avaliados diretamente em testes clonais, sem passarem previamente pelo teste de progênes (GODDARD; HAYES, 2007).

2.5.3 Metodologias Estatísticas para Computação da Seleção Genômica

Ampla

O desafio estatístico da GWS consiste em estimar o efeito de cada parâmetro – ou cada marca – da maneira mais acurada possível. Por conseguinte, desde que foi proposta por Meuwissen et al. (2001), diversas metodologias estatísticas frequentistas e bayesianas foram sugeridas para computação de tais efeitos, afim de verificar qual promoveria maior acurácia.

Dentre as metodologias frequentistas propostas, pode-se citar de quadrados mínimos, e os modelos mistos G-BLUP (*Genomic BLUP*) e RR-BLUP (*Ridge Regression BLUP*). O grande problema da metodologia dos quadrados mínimos na GWS é que, uma vez que o número de marcas cujos efeitos precisam ser estimados é dezenas ou centenas de vezes maior que o número de observações fenotípicas, não existem graus de liberdade suficientes

para ajustar todos estes efeitos simultaneamente, sendo necessário estimar um efeito de cada vez e verificar sua significância. Isto conduz a superestimativas dos efeitos significativos, baixa acurácia na estimação, detecção apenas de QTLs de grandes efeitos e uma baixa captura da variância genética (RESENDE et al., 2008). A inferioridade desta metodologia foi demonstrada por Meuwissen et al. (2001), e desde então ela tem sido pouco empregada na GWS.

As metodologias de modelos mistos aplicadas à GWS permitem ajustar todos os efeitos alélicos simultaneamente, mesmo quando existem mais efeitos a serem preditos do que o número total de observações fenotípicas. Tanto o G-BLUP, quanto o RR-BLUP são similares ao BLUP tradicional, porém, na predição dos efeitos aleatórios das marcas, não há necessidade de uso da matriz de parentesco (SCHAEFFER, 2006). A matriz de parentesco baseada em *pedigree*, usada no BLUP tradicional, é substituída por uma matriz de parentesco estimada pelos marcadores. Essa matriz de parentesco é a própria matriz $Z'Z$ presente nas equações de modelo misto, em que Z é a matriz de incidência para os efeitos de marcadores. De acordo com Resende et al. (2008), esse procedimento é superior ao uso do *pedigree*, pois efetivamente captura a matriz de parentesco realizada e não uma matriz de parentesco médio associada ao *pedigree*. Essa estratégia foi adotada por Bernardo e Yu (2007), Kolbehdari, Shaeffer e Robinson(2007), Meuwissen et al. (2001) e Muir (2007).

A maior crítica ao G-BLUP e RR-BLUP é o fato de estas metodologias assumirem o modelo infinitesimal de muitos locos com pequenos efeitos. Os marcadores têm variância comum e, portanto, explicam a mesma proporção da variância genética (MEUWISSEN et al., 2001). Assim sendo, tais metodologias não teriam um embasamento genético, uma vez que é conhecido que nem todas as marcas, distribuídas pelo genoma, têm alguma relação com o caráter e, portanto, deveriam ter efeito nulo. Todavia, Bernardo e Yu (2007) relatam que

essa suposição de iguais variâncias, por loco, não conduz a perdas significativas na acurácia da GWS.

Além das metodologias de modelos mistos, diversas outras metodologias bayesianas foram propostas para a GWS. Por meio de simulações, Meuwissen et al. (2001) propuseram o uso do Bayes A e Bayes B para a GWS. Estas metodologias assumem que cada marca tem sua própria variância, e assim, contribuem diferentemente para a variância genética total. O que as diferencia é que o Bayes B assume *a priori* que muitos marcadores têm efeito zero, enquanto que o Bayes A não faz nenhuma pressuposição *a priori* dos efeitos das marcas.

Em complemento ao Bayes A e Bayes B, também foram propostas diversas outras metodologias bayesianas, como o LASSO (DELOS CAMPOS et al., 2009), Bayes $C\pi$ (HABIER et al., 2011), *machine learning* (LONG et al., 2007) e o “*Bayesian Shrinkage Analysis*” (*Shrinkage*) (XU, 2003). Essas abordagens diferem na suposição sobre o modelo genético associado ao caráter quantitativo.

3 MATERIAIS E MÉTODOS

3.1 Local

O experimento foi instalado na cidade de Aracruz-ES (19°50'10" S e 40°12'43" W), em uma área representativa dos plantios realizados pela empresa Fibria S.A.

3.2 Materiais

Para realização deste trabalho foi utilizada uma população de híbridos, da empresa Fibria S.A., derivados do cruzamento entre plantas pré-selecionadas de *E. grandis*, *E. urophylla*, *E. globulus* e *E. camaldulensis*, e de alguns híbridos entre elas. As combinações híbridas foram feitas ao acaso, gerando 68 combinações híbridas envolvendo 45 genitores distintos. O tamanho efetivo desta população foi 27. Foram avaliados 12 indivíduos por combinação híbrida e devido à perdas, foram obtidos no final 611 indivíduos.

Estas progênies foram plantadas, no ano de 2006, no delineamento de blocos ao acaso, com parcelas de uma planta (*single treeplot*) e 12 repetições. O espaçamento foi de 2 metros entre plantas e 3 metros entre-linhas.

3.3 Fenotipagem

A fenotipagem das plantas ocorreu em três épocas distintas. Aos 24 meses de idade (dois anos), aos 36 meses de idade (três anos) e aos 63 meses de idade (cinco anos e três meses). A descrição das características mensuradas em cada idade da planta está apresentada na Tabela 1.

A circunferência das árvores foram medidas, à altura do peito, à 10 cm da base e 400 cm da base, utilizando uma fita métrica, a qual abraçou o tronco sem a remoção da casca. A altura total e altura comercial foram medidas, utilizando-se um hipsômetro vertex. A espessura da casca foi medida por meio de um paquímetro, em um segmento de casca removido à altura do peito. Uma vez removida a casca à altura do peito, o Pilodyn foi aplicado diretamente ao tronco, e então verificado o deslocamento da agulha. As demais características referentes à qualidade da madeira, tais como densidade básica, rendimento depurado, *coarseness*, comprimento ponderado da fibra, teor de lignina, número de fibras por grama, teor de pentosanas siringil/guaiacil foram mensuradas, utilizando-se o aparelho NIR (*Near Infrared Spectroscopy*).

Tabela 1 Relação das características mensuradas em cada idade da planta

Idade (meses)	Característica Mensurada	Unidade
24	Circunferência à altura do peito (CAP 24)	cm
	Densidade básica(DB)*	Kg/m ³
	Coarseness (COA)*	mg/100m
	Comprimento ponderado da fibra (CPF)*	mm
	Teor de lignina (LIG)*	%
	Número de fibras por grama (NFG)*	Milhões/g
	Teor de pentosanas (PEN)*	%
	Rendimento depurado (RD) *	%
Siringil/Guaiacil (SG) *	-	
36	Circunferência à altura do peito (CAP 36)	cm
	Espessura da casca (EC 36)	mm
	Pilodyn (PIL 36)	mm
	Altura total (AT 36)	m

“Tabela 1, conclusão”

Idade (meses)	Característica Mensurada	Unidade
63	Circunferência à altura do peito (CAP 63)	cm
	Espessura da casca (EC 63)	mm
	Pilodyn (PIL 63)	mm
	Altura comercial (AC) #	m
	Altura total (AT 63)	m
	Circunferência a 10 cm da base (C10)	cm
	Circunferência a 400 cm da base (C400)	cm

* Fenótipos mensurados com NIR; # Altura até o ponto em que o diâmetro do tronco mede 3 cm.

3.4 Genotipagem

Concomitantemente à tomada dos dados fenotípicos, aos dois anos de idade foi extraído o DNA de todas as 611 plantas dessa população, para genotipagem via GBS, método proposto por Elshire et al. (2011). Primeiramente, a genotipagem foi feita com 15.753 marcadores DArT-seq (os quais serão denominados como DArT neste trabalho), e posteriormente com 49.201 marcadores SNP. Após um processamento mínimo dos DArTs e SNPs, observou-se que um subconjunto de 15.742 e 36.812 marcas, respectivamente, apresentavam polimorfismo e também tinham MAF (*minimum allele frequency*) superior a 1%. Assim sendo, as marcas não-polimórficas com MAF menor que 1% foram removidas das análises. Nenhum outro critério foi utilizado para exclusão de marcas. Considerando o tamanho médio do genoma do eucalipto, estima-se uma saturação com, em média 13 e 30 marcas por cM, respectivamente. As genotipagens de DArTs e SNPs foram realizadas pelas empresas *Diversity Arrays Technology PtyLt* e *Gene SeekInc*, localizadas na Austrália e Estados Unidos, respectivamente.

3.5 Estimação dos efeitos dos marcadores

A partir do volume de dados fenotípicos de cada caráter (611 indivíduos) e do número de marcadores moleculares disponíveis (DArT = 15.742; SNP = 36.812), é possível concluir que o número de marcas, cujos efeitos precisam ser estimados, é muito superior ao número de dados fenotípicos para cada caráter. Desta maneira, existe um número insuficiente de graus de liberdade para estimar todos os efeitos simultaneamente pelo método dos quadrados mínimos. Assim sendo, optou-se pelo uso de uma metodologia de modelos mistos – RR-BLUP (MEUWISSEN et al., 2001), que permite a estimação de todos os efeitos, simultaneamente.

Para estimação dos efeitos das marcas via metodologia RR-BLUP, o seguinte modelo linear misto foi utilizado (RESENDE, 2008):

$$y = XB + Za + e,$$

em que:

y é o vetor de observações fenotípicas,

B é o vetor de efeitos fixos,

a é o vetor dos efeitos aleatórios das marcas,

e é o vetor de resíduos aleatórios,

X e Z são as matrizes de incidência para B e a .

A estrutura de médias e variâncias no modelo em questão é definida como:

$$a \sim N(0, G)$$

$$e \sim N(0, R = I\sigma_e^2) \quad E(y) = XB$$

$$G \sim \sum_1^m I \sigma_a^2 \text{Var}(y) = V = ZGZ' + R$$

As equações de modelo misto genômicas para a predição de a via o método RR-BLUP equivalem a (RESENDE, 2008):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I \frac{\sigma_e^2}{(\sigma_g^2/n)} \end{bmatrix} \begin{bmatrix} \hat{B} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix},$$

em que:

σ_g^2 é a variância genética total do caráter

σ_e^2 é a variância residual.

n é o número de marcas.

As análises do RR-BLUP foram feitas no software livre R (R CORE TEAM, 2013), através do pacote rrBLUP (ENDELMAN, 2011).

Uma vez estimado o efeito de cada marca (a_i) para os diferentes caracteres, foi possível estimar o valor genético genômico (VGG) dos indivíduos, pela expressão:

$$V\hat{G}G = \sum_i Z_i \hat{a}_i,$$

em que:

Z_i equivale a 0 ou 1 no caso de marcadores DArTs para os genótipos homocigoto recessivo ou heterocigoto/homocigoto dominante, respectivamente,

e -1, 0 ou 1 no caso de SNPs para os genótipos homozigoto recessivo, heterozigoto ou homozigoto dominante, respectivamente.

De posse dos efeitos de cada marca DArT ou SNP, plotou-se também os Manhattan Plots, utilizando o pacote “gap” (ZAO, 2007) do software R. As herdabilidades de SNPs e DArTs foram estimadas conforme descrito em Resende et al. (2012).

Ao final do processo de estimação das marcas e dos VGG, é necessário estimar a acurácia da estimação, ou seja, com qual precisão o genótipo do indivíduo, representado pelas marcas e seus efeitos estimados, é capaz de prever o fenótipo real observado. A acurácia da estimação foi verificada por meio da correlação de Pearson entre o fenótipo predito pela GWS (VGG) e o fenótipo observado nas plantas.

3.6 População de validação dos efeitos dos marcadores

Para validação dos efeitos dos marcadores, utilizando-se o RR-BLUP, a população total de 611 indivíduos foi particionada de forma aleatória em uma subpopulação de descoberta, contendo 500 indivíduos, e outra subpopulação de validação, contendo os 111 indivíduos restantes. Uma vez estimados os efeitos das marcas na população de descoberta, via RR-BLUP (vide item 3.3.3), estes foram aplicados à população de validação para cômputo do VGG dos indivíduos da mesma. Posteriormente, foi obtida a acurácia da validação, por meio da correlação de Pearson entre os VGG estimados e os fenótipos desta população de validação. Este processo foi repetido 500 vezes, de modo que a acurácia da validação foi a média de todas as 500 acurácias realizadas neste processo. Este método de validação cruzada é conhecido como *Jackknife*.

3.7 Viabilidade da seleção precoce

Além das análises já descritas anteriormente, foi verificada a viabilidade da seleção precoce para os caracteres CAP, AT, EC e PIL. Foram utilizadas apenas estas características, pois elas foram as únicas mensuradas, em mais de uma idade da planta, sendo que AT, EC e PIL foram medidas aos 36 e 63 meses e o CAP, medida aos 24, 36 e 63 meses de idade.

3.8 Viabilidade da seleção precoce via dados fenotípicos

Com a finalidade de se verificar a eficiência da seleção precoce, primeiramente, foi realizada a análise fenotípica nas três idades da planta. Para isto, foi adotado o modelo misto de irmãos germanos que incorpora o efeito do pai e da mãe, conforme Resende (2007b):

$$y = Xb + Zm + Wf + Tc + \varepsilon,$$

em que:

y é o vetor das observações;

b é o vetor dos efeitos de blocos (fixos);

m é o vetor dos efeitos de genitores da população de machos (aleatórios);

f é o vetor dos efeitos de genitores da população de fêmeas (aleatórios);

c é o vetor dos efeitos aleatórios da capacidade específica de combinação dos genitores da população de machos com os genitores da população de fêmeas;

X , Z , W e T são as matrizes de incidência para os referidos efeitos;

ε é o vetor dos erros aleatórios.

As herdabilidades no sentido amplo (\hat{h}_g^2) foram estimadas de acordo com Resende (2007a):

$$\hat{h}_g^2 = \frac{\sigma_a^2 + \sigma_d^2}{\sigma_y^2},$$

onde $\sigma_a^2 = 2(\sigma_{gm}^2 + \sigma_{gf}^2)$, $\sigma_d^2 = \frac{4\sigma_s^2}{\sigma_y^2}$, $\sigma_y^2 = \sigma_{gm}^2 + \sigma_{gf}^2 + \sigma_s^2 + \sigma_e^2$,

em que:

σ_a^2 é a variância genética aditiva interpopulacional;

σ_d^2 é a variância genética de dominância interpopulacional;

σ_y^2 é a variância fenotípica interpopulacional;

σ_{gm}^2 é a variância genética da população utilizada como machos;

σ_{gf}^2 é a variância genética da população utilizada como fêmeas;

σ_s^2 é a variância da capacidade específica de combinação interpopulacional entre dois genitores;

σ_e^2 é a variância residual.

De posse dos dados dos valores genotípicos (via BLUP) em cada idade, os indivíduos foram ranqueados pelo BLUP ideal (maior ou menor, em função do caráter) em cada uma delas, e foi estimada a correlação de Spearman entre as idades.

Além disso, foi verificado o índice de coincidência (IC) entre os 30 melhores indivíduos em cada idade (intensidade de seleção = 5%). A

coincidência foi estimada, utilizando-se a expressão de Hamblin e Zimmermann (1986): $IC = \frac{A-C}{B-C}$, em que: A é o número de plantas comuns às duas seleções; B é o número de plantas selecionadas em uma dada época (30); C é o número de plantas em comum nas duas seleções, unicamente devido ao acaso ($C = i \times B$), em que i é a proporção selecionada, no caso 5%.

3.9 Viabilidade da seleção precoce via dados moleculares

Conforme comentado anteriormente, os indivíduos da população foram fenotipados em três idades distintas, e também genotipados, utilizando os marcadores DArT e SNP. Ao longo das idades, é esperado então que os fenótipos das plantas para CAP, AT, EC e PIL sofram alterações. No entanto, se desconsiderarmos a ocorrência de mutações, o genótipo das plantas não sofre variação alguma neste período. Assim sendo, um questionamento que pode ser levantado, é de que maneira efeitos de marcas ou regiões do genoma seriam afetados pela alteração do fenótipo, ou se marcas com pequeno efeito em idades precoces passam ou não a ter efeitos expressivos em idades mais avançadas, considerando o ambiente em que foram avaliadas.

Neste sentido, de modo análogo à análise descrita utilizando apenas dados fenotípicos, é importante verificar a eficácia da seleção precoce também via GWS. Para tanto, foi estimada a correlação entre os efeitos da marca de ambos os marcadores, aos 24 e 36 meses, com o efeito aos 63 meses para CAP, e também entre 36 e 63 meses para EC, PIL e AT. Vale ressaltar que, quanto mais próximo da idade de corte, mais confiável é a fenotipagem para seleção. Além da estimativa da correlação, também foi verificado o índice de coincidência entre os 30 (5%) melhores indivíduos selecionados, em cada idade,

com base em seus valores genéticos genômicos (VGG), também utilizando a expressão de Hamblin e Zimmermann (1986).

Por fim, foi realizada uma variação da metodologia de validação já descrita, sendo a estimação do modelo feita utilizando fenótipos da idade precoce, e a sua validação, utilizando-se os fenótipos da idade subsequente, mais tardia. Este processo também foi repetido 500 vezes.

4 RESULTADOS

A variação fenotípica encontrada para as 20 características, pode ser observada na Tabela 2. Como pode se observar, todas as características apresentaram uma ampla variação fenotípica, ou seja, um bom indicativo da existência de variabilidade genética e possibilidade de ganho com a seleção. O coeficiente de variação (CV) foi baixo para a maioria das características, indicando boa precisão experimental.

Tabela 2 Média, valor mínimo (MÍN), máximo (MÁX) e coeficiente de variação (CV) das 20 características fenotipadas para volume e qualidade da madeira

	DB	COA	CPF	LIG	NFG	PEN	RD	SG	PIL36	PIL63
MÉDIA	415,5	6,7	0,7	30,1	25,2	16,7	52,7	0,7	14,5	11,7
MÍN	324,9	3,9	0,6	22,5	12,8	13,1	47,8	0,3	5,0	6,5
MÁX	582,1	9,9	0,8	33,2	39,5	19,5	60,3	1,1	28,0	21,0
CV	6,4	11,7	5,5	3,7	11,6	5,0	2,3	16,9	15,5	17,4
	CAP24	CAP36	CAP63	AT36	AT63	AC	EC36	EC63	C10	C400
MÉDIA	30,3	37,9	47,0	17,7	21,9	19,0	5,4	6,3	55,9	41,3
MÍN	20,2	21,2	23,0	9,8	11,7	8,2	1,9	2,3	26,8	19,1
MÁX	44,4	60,5	84,6	24,1	29,3	27,3	9,5	10,9	101,0	71,8
CV	13,3	18,8	23,1	13,0	17,8	22,1	17,5	21,4	23,1	24,3

As 20 características relacionadas ao volume e qualidade da madeira, em eucalipto, foram analisadas por meio do RR-BLUP para cômputo da GWS (Figura 1). As menores e maiores acurácias preditivas foram coincidentes entre SNPs e DArTs, e variaram de 0,44 e 0,51 para COA, até 0,83 e 0,89 para PEN, respectivamente. Em todos os casos, os marcadores DArT apresentaram

acurácias superiores aos SNPs, sendo em média 8% superiores. A maior diferença foi observada para CAP63, 14%.

A validação teve acurácias bastante reduzidas em relação à predição, em média 46% e 40% menores para SNPs e DArTs, respectivamente (Figura 1). Mais uma vez, o COA teve a menor acurácia, tanto para SNP, quanto DArT, enquanto que PIL36 teve os melhores valores da validação, para os dois marcadores. Na validação, no entanto, SNP e DArT tiveram performance bastante semelhante, com algumas poucas exceções.

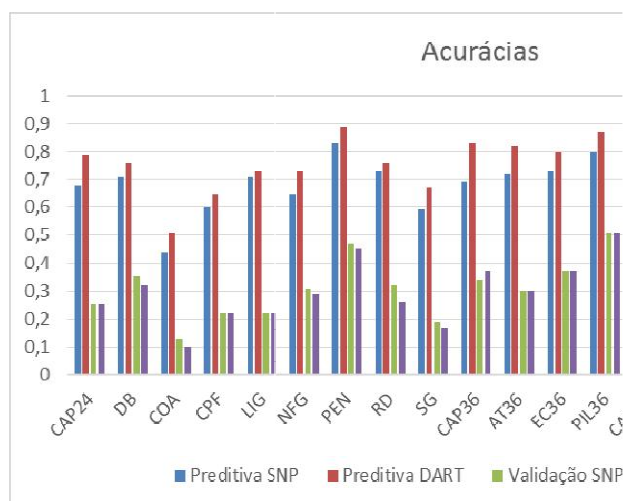


Figura 1 Acurácias preditivas e da validação de 20 características de eucalipto, utilizando o RR-BLUP

As herdabilidades, no sentido amplo, dos caracteres via BLUP foram altamente variáveis, porém é notável que as menores foram obtidas para algumas características ligadas à qualidade da madeira, sendo elas: SG (0,07), COA (0,11) e LIG (0,14) (Figura 2). Outras características importantes ligadas à qualidade da madeira tiveram herdabilidades mais satisfatórias, como a DB, PIL 36 (que é uma medida indireta da DB), PEN e NFG. Herdabilidades medianas foram observadas para características relacionadas ao volume da madeira (AC,

EC e circunferências) em todas as idades, com destaque para C400 e AT63 cujas herdabilidades foram 0,72 e 0,76, respectivamente.

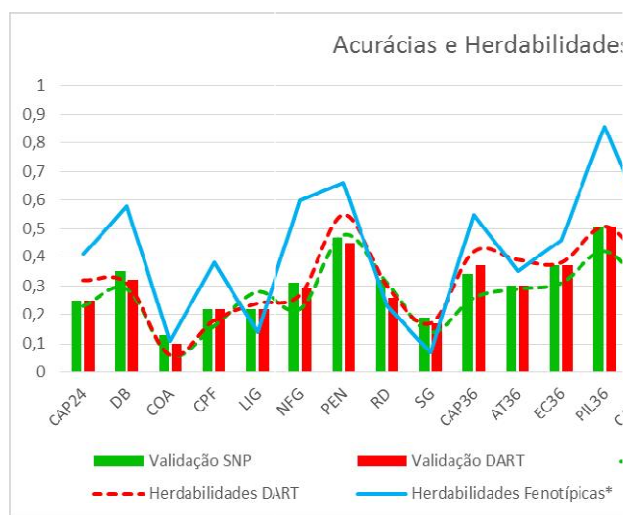


Figura 2 Acurácias da validação e herdabilidades de 20 características de eucalipto, utilizando o RR-BLUP (*Herdabilidades no sentido amplo)

As herdabilidades das marcas evidenciaram que os DArTs capturaram uma variação genética superior aos SNPs. As maiores herdabilidades das marcas foram encontradas para PEN, seguidas por AT63, PIL36, C10 e C400. Mais uma vez, os valores mais baixos foram encontrados para COA e SG (Figura 2).

De maneira geral, é possível notar claramente, por meio da Figura 2, que as acurácias da validação são altamente influenciadas pelas herdabilidades, tanto fenotípica quanto do marcador. Isto se torna ainda mais claro, quando observamos as correlações entre estes dois parâmetros (Tabela 3). Foram encontradas correlações acima de 0,85 entre acurácias da validação e herdabilidades em todos os casos, indicando claramente a interdependência entre eles.

Tabela 3 Correlações entre herdabilidades de SNPs, DArTs e fenotípicas (BLUP), com acurácias da validação

Acurácias Validação	Herdabilidades		
	SNP	DART	BLUP
SNP	0,89	-	0,85
DART	-	0,90	0,85

Como o próprio nome da metodologia sugere, o maior objetivo da GWS é a seleção de genótipos superiores. Sem a utilização de marcadores, no entanto, grande parte dos melhoristas de eucalipto obtiveram ganhos expressivos por meio de fenótipos bem mensurados e analisados pelo BLUP, e até então não foram encontrados relatos, na literatura, de clones lançados utilizando a GWS, até mesmo por ser uma técnica bastante recente. Assim, uma boa maneira de comparar as metodologias seria por meio do índice de coincidência entre as melhores plantas selecionadas pelo BLUP e pela GWS (Tabela 4).

Pôde-se observar primeiramente que, de maneira geral, o índice de coincidência entre o BLUP e a GWS, utilizando DArT foi ligeiramente superior àquele utilizando SNPs. Favoravelmente à utilização da GWS, algumas características importantes no melhoramento, visando celulose, tiveram um índice de coincidência elevado com o BLUP, como a DB, PIL36, PIL63, CAP63, EC36 e NFG. A maioria das características ligadas à qualidade da madeira tiveram menor coincidência entre as plantas selecionadas, como foi o caso mais uma vez de COA, CPF, SG e RD. De forma análoga, no entanto, algumas medições de altura tiveram índices menores que 30% (Tabela 4).

Ao longo da vida da planta, diversos alelos atuam com maior ou menor expressão na construção dos fenótipos. No entanto, não sabemos se os alelos com maior efeito, para um determinado caráter em uma idade precoce, continuam tendo efeito pronunciado em idades mais avançadas da planta. Uma vez que a GWS necessita de fenótipos para seu cômputo, uma alta correlação

entre efeitos de marcas - que representam os alelos - em idades distintas da planta permitem que o fenótipo seja mensurado mais precocemente, já para o cálculo da GWS. Consequentemente, isto viabilizaria a seleção precoce de genótipos superiores via GWS.

Tabela 4 Índices de coincidência (%) entre as 30 melhores plantas selecionadas pelo BLUP e pela GWS (SNPs e DArTs)

Caráter	SNP	DART
DB	44	47
COA	26	16
CPF	05	09
LIG	37	37
NFG	54	61
PEN	40	23
RD	30	23
SG	23	19
PIL36	51	58
PIL63	51	54
CAP24	19	33
CAP36	12	19
CAP63	30	40
AT36	23	37
AT63	12	26
AC	09	19
EC36	61	61
EC63	33	30
C10	23	30
C400	23	33

Como é possível observar na Tabela 5, a correlação de Spearman entre os efeitos das marcas foi maior que 0,75 para os caracteres altura e circunferência nas diferentes idades. Correlações abaixo de 0,5, no entanto, foram encontradas para EC e PIL. Resultados muito semelhantes foram encontrados para SNPs e DArTs. Por meio destes resultados infere-se que, em se tratando de alturas e circunferências, marcas importantes em idades precoces, ou seja, de grande efeito, se mantiveram em idades mais avançadas. As maiores correlações de Spearman entre BLUPs de plantas foram coincidentes com as maiores correlações entre efeitos de marcas. Destaque para correlação entre BLUPs de CAP36 X CAP63 que foi de 0,96 (Tabela 5).

Altas correlações refletem normalmente boa coincidência na seleção de plantas. No caso do caráter circunferência, a coincidência entre as 30 melhores plantas selecionadas pelo BLUP variou entre 51 e 72% (Tabela 5). Isto não foi observado, no entanto, para as alturas e espessura da casca. Apesar da alta correlação entre BLUPs (0,71), apenas 33% das plantas selecionadas foram coincidentes para as alturas e 26% para EC, enquanto que 72% de coincidência foi observada para o PIL, mesmo com correlação mediana entre BLUPs. Com exceção das alturas e EC, todos os IC das plantas selecionadas pelo BLUP foram maiores que os das plantas selecionadas pela GWS, utilizando SNPs e DArTs. Os IC dos marcadores, para as alturas e EC, foram bastante superiores ao do BLUP. De maneira geral, a GWS utilizando SNPs teve maior IC entre idades do que DArTs, com destaque positivo para IC-SNP das alturas (72%) e negativo para IC-SNP do PIL (37%).

Tabela 5 Correlação de Spearman entre efeitos das marcas para SNPs e DArTs, entre BLUPs das 611 plantas avaliadas, índice de coincidência (IC - %) entre as 30 melhores plantas selecionadas pelos SNPs, DArTs e BLUP em idades distintas e acurácia da validação na idade mais tardia

Caracteres	Correlações			IC (%)			Validação*	
	SNP	DART	BLUP	SNP	DART	BLUP	SNP	DART
CAP24 x CAP36	0,86	0,86	0,77	51	47	65	0,30	0,30
CAP24 x CAP63	0,75	0,76	0,69	47	44	51	0,33	0,31
CAP36 x CAP63	0,91	0,91	0,96	61	54	72	0,39	0,41
AT36 x AT63	0,83	0,83	0,71	72	65	33	0,31	0,30
EC36 x EC63	0,51	0,47	0,58	44	54	26	0,29	0,29
PIL36 x PIL63	0,47	0,44	0,52	37	40	72	0,36	0,35

* Estimação dos efeitos com fenótipos da idade precoce e validação com fenótipos da idade tardia.

Além do IC entre as melhores plantas selecionadas pelo BLUP e pela GWS, foi estimada também a acurácia da GWS em predizer os fenótipos das idades mais tardias, como maneira de verificar a aplicabilidade da GWS precoce (Tabela 5). As acurácias encontradas variaram entre 0,29 para EC36 x EC63 (SNPs e DArTs), até 0,39 e 0,41 para CAP36 x CAP63, para SNPs e DArTs, respectivamente. Mais uma vez, houve grande coincidência entre os dois marcadores.

Em adição à utilização de marcadores que capturem maior parte da variação, é importante observar se as marcas utilizadas estão associadas com o caráter em estudo e, mais precisamente, a algum cromossomo em específico. Para ilustrar isto, foram construídos Manhattan Plots dos efeitos de SNPs e DArTs (Figura 3).

Nota-se que o cromossomo 6 foi o que apresentou maior número de marcas com efeito pronunciado, em relação aos demais, principalmente nos

caracteres mais relacionados ao volume de madeira, como circunferências e alturas. Isto foi mais notável, no entanto, para SNPs. Um pico pronunciado no cromossomo 6 também foi observado para PIL36 e PIL 63 (SNPs). Em se tratando de qualidade da madeira, observa-se um pico nos cromossomos 2 e 3, os quais são semelhantes para LIG e RD, tanto utilizando SNPs quanto DArTs. Também observou-se um pico nos cromossomos 3 e 10 para SG (SNPs).

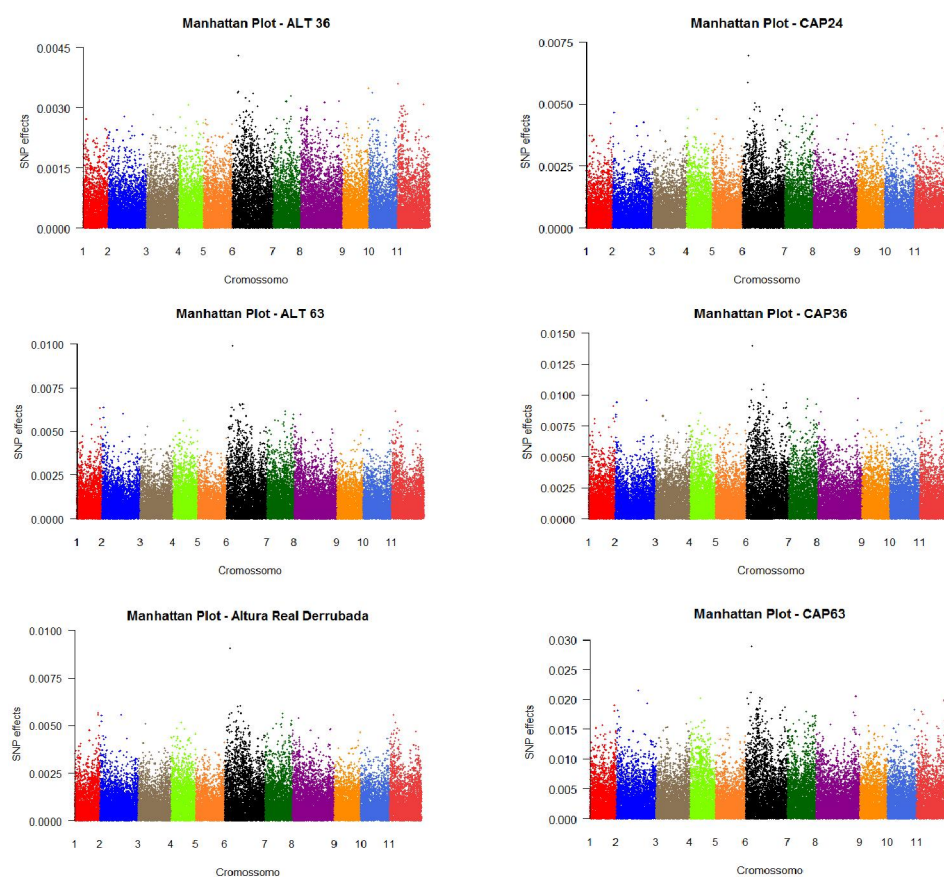
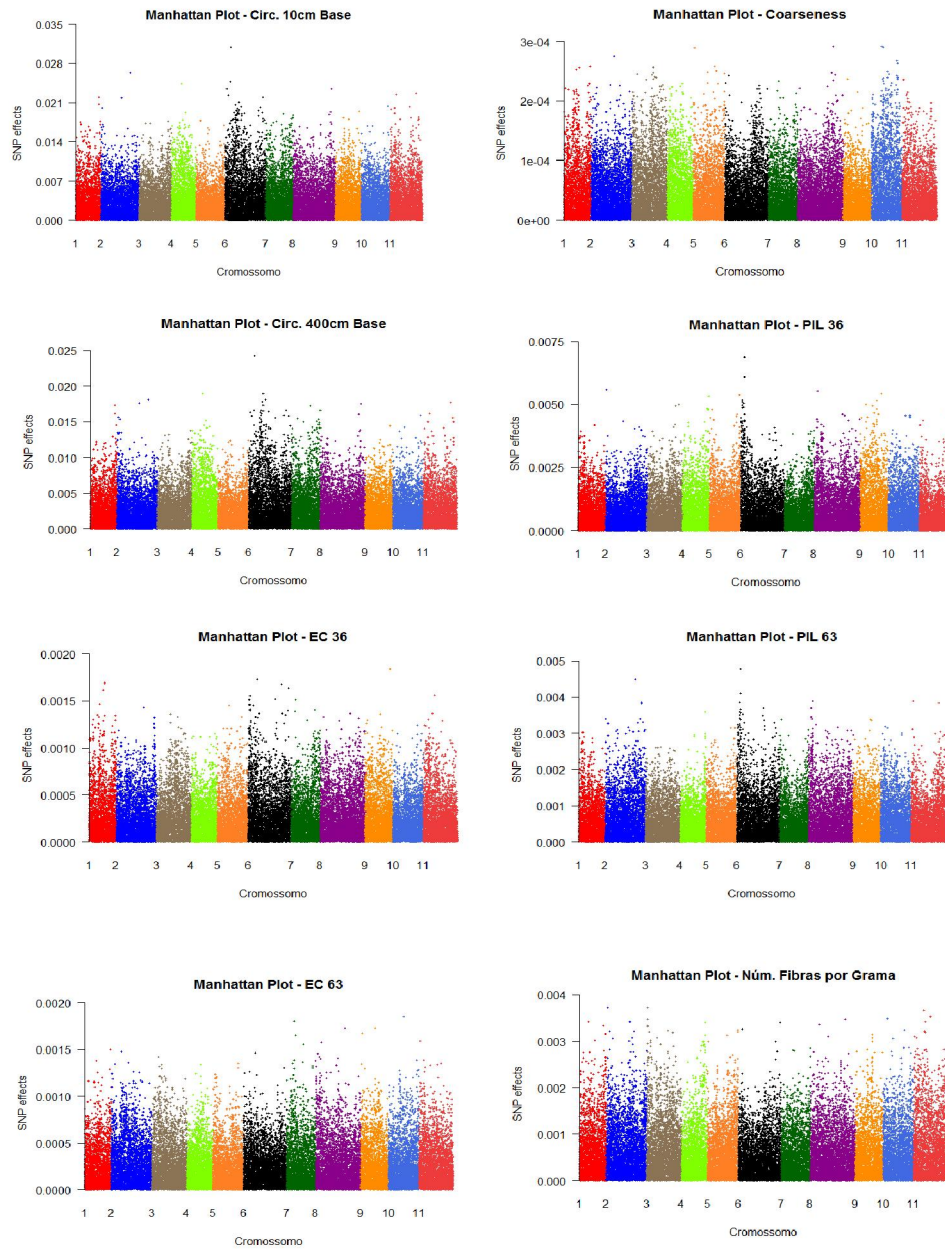


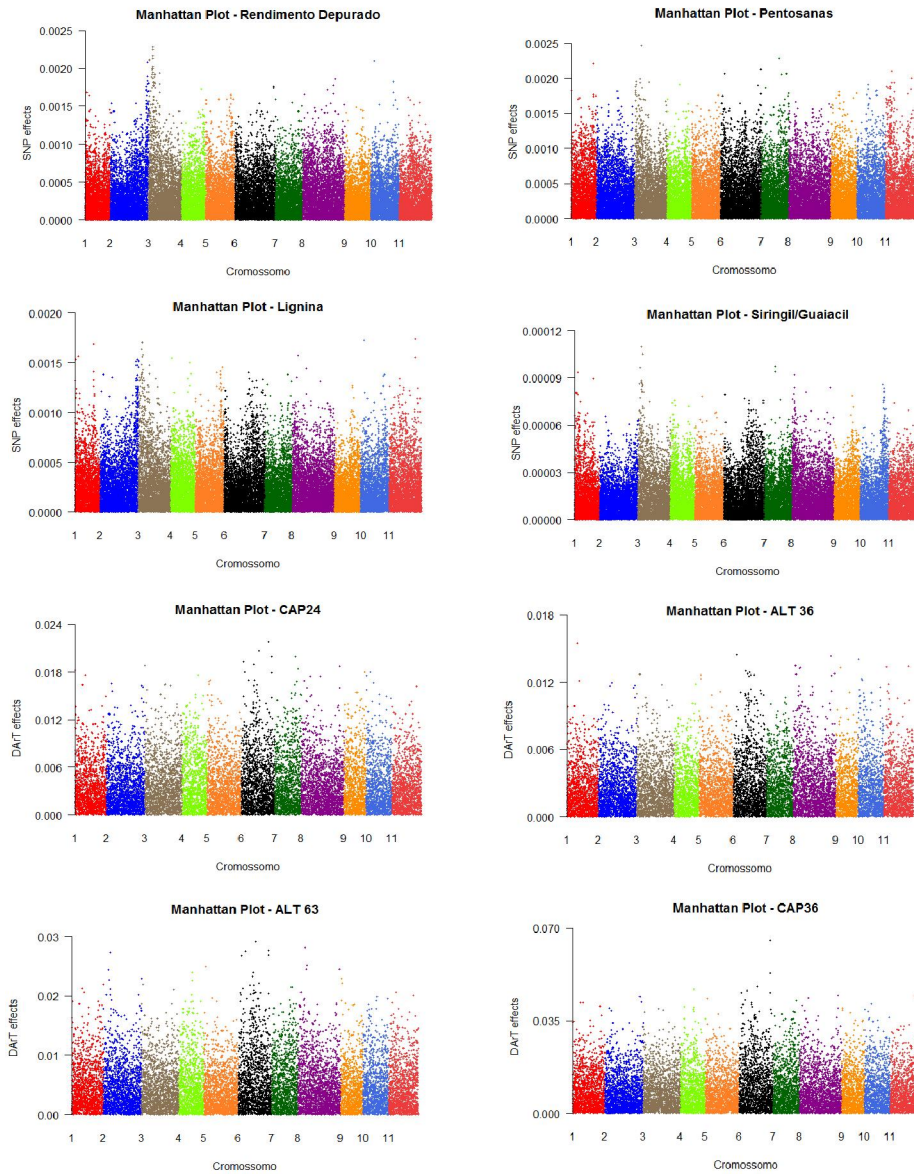
Figura 3 Manhattan plots dos efeitos de SNPs e DArTs para as 20 características avaliadas

(...continua...)

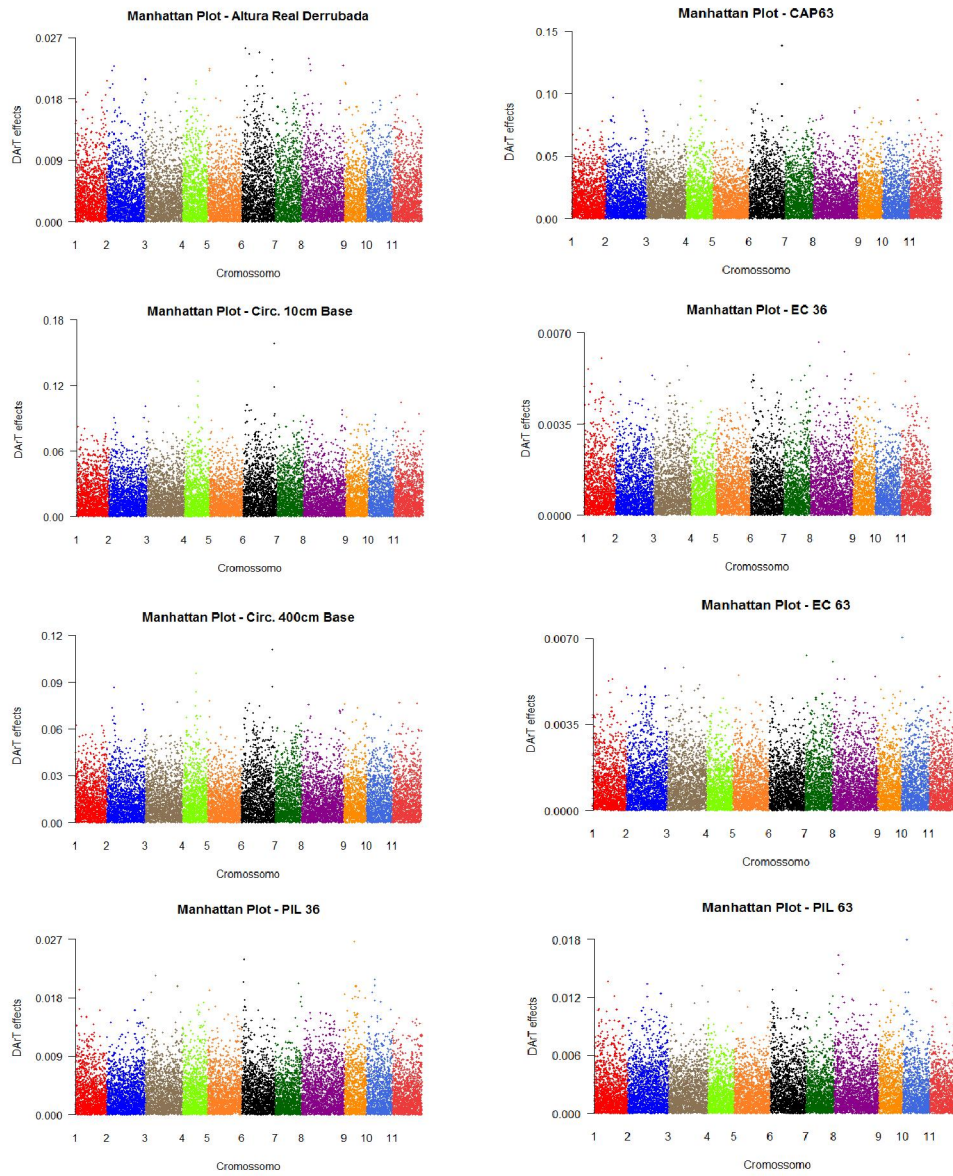
“Figura 3, continua”



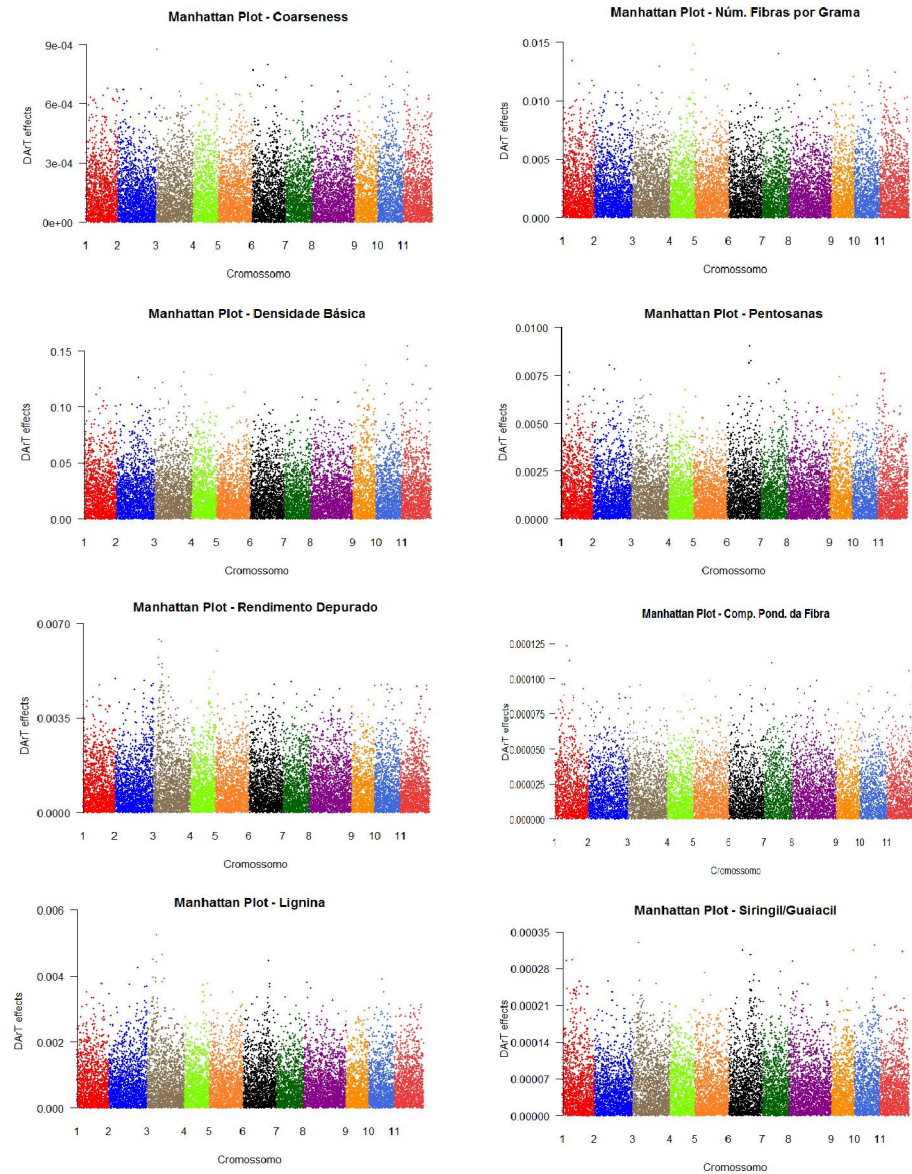
“Figura 3, continua”



“Figura 3, continua”



“Figura 3, conclusão”



5 DISCUSSÃO

Até o presente momento não foram encontrados relatos na literatura de estudos que aplicassem a GWS em uma população, visando à seleção recorrente entre várias espécies de Eucalipto (*E. grandis*, *E. urophylla*, *E. globulus* e *E. camaldulensis*) e de híbridos entre elas, e não apenas um cruzamento biparental ou entre duas espécies apenas. Uma vez que a maioria das empresas de melhoramento utilizam várias espécies em diversos cruzamentos no âmbito de obter o maior número possível de características favoráveis, acreditamos que esta seja uma situação rotineira para melhoristas e que, portanto, fornece parâmetros práticos e factíveis.

Dentre as diversas metodologias envolvendo marcadores moleculares no melhoramento de árvores e especialmente o eucalipto, a GWS provavelmente tem gerado a maior expectativa, em termos de possibilidade de seleção para caracteres quantitativos (RESENDE et al., 2008; RESENDE, M. D. et al., 2012; RESENDE, M. F. et al., 2012), uma vez que utiliza uma grande quantidade de marcas. Utilizando-se um maior número de marcas em desequilíbrio de ligação, a probabilidade de capturar maior proporção da variação genética é maior (MEUWISSEN et al., 2001; RESENDE, M. D. et al., 2012; WIENTJES et al., 2013), facilitando a seleção de genótipos superiores.

Em termos de acurácia preditiva, os resultados da Figura 1 são bastante favoráveis à GWS, uma vez que altas acurácias preditivas ($>0,7$) foram encontradas para a maioria das características, tanto utilizando SNPs quanto DArTs. Não parece haver uma tendência de maiores ou menores acurácias preditivas para características ligadas à qualidade da madeira ou volume, já que os menores e maiores valores absolutos foram encontrados para duas características ligadas à qualidade da madeira, COA e PEN, respectivamente.

Ao se analisarem os resultados da acurácia da validação, no entanto, fica claro que, mesmo a validação sendo feita em indivíduos não amostrados, porém oriundos da mesma população, as acurácias são bastante reduzidas em relação à preditiva (Figura 1). Esta acurácia da validação está intimamente relacionada às herdabilidades, tanto dos marcadores quanto fenotípica, conforme demonstrado pela Figura 3 e Tabela 3. De fato, uma vez que, para estimação dos efeitos das marcas, o efeito fenotípico faz parte do modelo, isto gera uma reação em cadeia. Quanto menor a herdabilidade do caráter, menor a precisão dos dados fenotípicos e, portanto, menor a herdabilidade dos efeitos dos marcadores. Consequentemente, menor será a capacidade em prever acertadamente fenótipos de indivíduos não amostrados para cômputo do modelo. Isto foi demonstrado por meio de simulações por Grattapaglia e Resende (2011), os quais verificaram que o aumento da herdabilidade do caráter provoca o aumento da acurácia da GWS.

Como se sabe, os marcadores SNP são codominantes, enquanto os marcadores DArT são dominantes. Assim sendo, a estimação de efeitos aditivos por meio de DArTs fica confundida. Era esperado, portanto, que SNPs fossem mais informativos que DArTs, capturando maior parte da variação genética aditiva. Levando-se em consideração as acurácias preditivas, no entanto, isto não foi observado para a população do presente estudo, sendo DArTs 8% superiores em média quanto ao poder de predição de fenótipos (Figura 1). Ao analisarmos as acurácias da validação, porém, esta superioridade desaparece, mas mesmo assim DArTs e SNPs são praticamente equivalentes para todas as características (Figuras 1 e 2). Vale mencionar que, o número de SNPs polimórficos utilizados neste trabalho foi maior que o dobro do número de DArTs, indicando que o maior número de marcas, ocupando todo o genoma aumenta a probabilidade de desequilíbrio de ligação com os alelos que controlam o caráter

(GRATTAPAGLIA; RESENDE, 2011), mas não garantem que a maior parte delas esteja intimamente associadas aos mesmos (RESENDE JÚNIOR, 2010).

Mesmo SNPs sendo codominantes, eles não são multialélicos, a exemplo dos microssatélites. Assim sendo, utilizando a tecnologia empregada neste trabalho, SNPs conseguem distinguir genótipos $A^1A^1 \times A^1a \times aa$, e também $A^2A^2 \times A^2a \times aa$, mas por meio da genotipagem não é possível verificar que os alelos A^1 e A^2 são diferentes. O mesmo raciocínio é válido para DArTs, porém para estes não é possível diferenciar os homozigotos dominantes dos heterozigotos. Considerando que a população de árvores em estudo é composta por quatro espécies e também por híbridos entre elas, é bastante plausível a hipótese de ocorrência de alelismo múltiplo para os genes que controlam as características, o que reduz a acurácia, tanto de DArTs, quanto de SNPs, já que o efeito de marcas fica confundido. Vale ressaltar, no entanto, que este tipo de população envolvendo várias espécies é bastante comum em programas de melhoramento de eucalipto e que, portanto, reflete a realidade.

Conforme comentado anteriormente, a GWS é a união da genotipagem com grande número de marcadores, com a fenotipagem das plantas (MEUWISSEN et al., 2001). Assim sendo, ao contrário do que se possa imaginar, a fenotipagem e todos os fatores que afetam a precisão experimental, como o delineamento, têm o mesmo peso da genotipagem na condução da GWS e, portanto, é e sempre será indispensável no melhoramento de plantas. Além disso, o modelo de predição somente é eficiente, quando a seleção dos indivíduos para a formação da população é muito bem feita nos programas de melhoramento.

Uma vez que eucalipto é uma planta perene, com ciclo médio de 7 anos, no Brasil (BRACELPA, 2014), já no melhoramento clássico vislumbrou-se a possibilidade de fenotipar e selecionar plantas precocemente, aumentando assim os ganhos com a seleção. Diversos trabalhos mostram que, para várias

características, isto é plenamente possível (BELTRAME et al., 2012; LIMA et al., 2011; MASSARO et al., 2010; PAVAN et al., 2014; PEREIRA et al., 1997; REZENDE; BERTOLUCCI; RAMALHO, 1994; TOLFO, 2003). No entanto, até o presente momento não foram encontrados relatos, na literatura, de trabalhos com eucalipto que testassem o efeito da fenotipagem precoce no cômputo da GWS, considerando sempre que a fenotipagem mais próxima à idade de corte é a ideal.

Analisando-se a Tabela 5, é possível notar que, dentre as características fenotipadas, em mais de uma idade, a seleção via melhoramento clássico pelo BLUP teria maior aplicabilidade para CAP, que teve as maiores correlações entre o BLUP das plantas, ao longo das idades, e também os maiores IC das 5% melhores plantas selecionadas, bem como para PIL, que apesar da baixa correlação entre BLUPs, teve um alto IC das melhores plantas selecionadas (72%). Considerando que o CAP tem grande peso no cálculo de volume, e que o PIL é uma medida indireta da DB, que também é um ponto de gargalo no melhoramento do eucalipto, estes resultados são bastante favoráveis à seleção precoce fenotípica, por meio destas características. Para ALT, apesar de o IC ter sido inferior ao de CAP e PIL, 33%, esta porcentagem pode estar sendo influenciada pelo índice de seleção, de apenas 5%. Assim sendo, a seleção um pouco mais branda poderia ser uma alternativa. Os resultados de Massaro et al. (2010) corroboram com estes aqui encontrados, pelos quais constataram que a seleção precoce pode ser efetuada aos 25 meses de idade para o caráter DAP (medida indireta de CAP). Lima et al. (2011) também comentam que a seleção para CAP, tanto de genitores quanto de árvores para clonagem, pode ser realizada precocemente aos 30 meses de idade. Pavan et al. (2014) enfatizam que a seleção precoce para volume, aos 36 meses, é viável para seleção, tanto das melhores plantas, quanto das melhores famílias.

A eficácia na seleção fenotípica precoce, por si só, não garante que a GWS também possa ser feita precocemente, uma vez que deve haver uma coincidência entre as plantas selecionadas em idades precoces e idades tardias. Neste sentido, o IC entre as melhores plantas selecionadas pelos marcadores determinam os parâmetros que auxiliam nesta decisão. Estes parâmetros foram estimados e, além destes, foi estimada também a acurácia da GWS em prever os fenótipos das idades mais tardias, utilizando um processo parecido com a validação já descrita anteriormente (Tabela 5).

Correlações entre efeitos de marcas acima de 0,75 foram encontradas para CAP entre as três idades e para AT, tanto utilizando SNPs, quanto DArTs. Estas altas correlações refletiram-se em um alto IC entre plantas selecionadas pelos marcadores, sendo os maiores IC encontrados para AT36 x AT63 (72% e 65%) e para CAP36 x CAP63 (61% e 54%), utilizando SNPs e DArTs, respectivamente. Estes são ótimos indícios que o cômputo da GWS para AT e CAP pode ser feito precocemente aos 36 meses de idade. Somando a estes fatos, a acurácia da validação na idade mais tardia foi satisfatória principalmente entre CAP36 x CAP63, para SNPs e DArTs, respectivamente, mas também para as demais características. Estes resultados reforçam a possibilidade de GWS precoce, principalmente por se tratar de uma validação cruzada com 500 repetições.

Analisando-se os Manhattan Plots de SNPs e DArTs, para estas características avaliadas em mais de uma idade, esta possibilidade faz-se ainda mais factível (Figura 3). Principalmente para SNPs, picos nos mesmos cromossomos, especialmente no cromossomo 6, foram encontrados para as circunferências, alturas e PIL ao longo das idades. Estes picos pronunciados nos mesmos cromossomos, ao longo das idades, sugerem que as mesmas regiões do genoma continuam atuando fortemente no controle do caráter, fato este necessário para seleção precoce. Para DArTs, de maneira geral, um número

menor picos foi encontrado, apesar de as correlações entre efeitos de marcas, ao longo das idades, também terem sido elevadas.

Para qualidade da madeira, chama a atenção a semelhança entre os picos encontrados entre RD e LIG nos cromossomos 2 e 3, tanto utilizando DArTs como SNPs. Seria precipitado, no entanto, inferir que as mesmas marcas teriam maior influência para controle destes caracteres, ou que este seria um caso de pleiotropia. Estudos mais aprofundados seriam necessários para isto, como a utilização de populações em desequilíbrio de ligação para o caráter em que se deseja avaliar a associação, e também a utilização de modelos estatísticos mais robustos para o estudo da arquitetura genética do caráter, entre outros. No entanto, é importante ressaltar que, em relação à característica SG, apesar das baixas acurácias e herdabilidades encontradas, o pico observado no cromossomo 10, utilizando SNPs foi coincidente com o encontrado por Lima (2014), sugerindo assim uma provável existência de marcas de maior efeito controlando o caráter nesta região do genoma.

De acordo com os resultados encontrados neste trabalho, pode-se verificar que a GWS tem sim boa utilidade no melhoramento do eucalipto, porém alguns aspectos precisam ser observados. Primeiramente, a precisão experimental na tomada de dados fenotípicos deve ser a mais rigorosa possível. Esta fenotipagem terá impacto direto em todas as etapas da GWS, desde a calibração do modelo preditivo até na seleção de plantas geneticamente superiores. A herdabilidade fenotípica reduzida gera uma reação em cadeia, já que influencia diretamente a acurácia da validação e que, por sua vez, influencia na estratégia do melhorista para seleção de plantas. Acurácias medianas não inviabilizam a utilização da GWS, mas afetam a pressão de seleção exercida pelo melhorista, no sentido de reduzi-la, fato este que também segue o mesmo raciocínio do melhoramento clássico. Quanto aos marcadores, SNPs e DArTs, a tendência é que sejam cada vez mais numerosos e bem posicionados em regiões

sabidamente importantes para o controle dos caracteres, o que, agregados a metodologias de análises robustas, deverão contribuir para maior aplicabilidade da GWS (GRATTAPAGLIA;RESENDE, 2011).

De qualquer modo, a utilização da GWS, não somente para seleção acurada das melhores progênies apenas, mas pelo menos para descarte das piores, já seria de grande utilidade e promoveria economia de tempo e de recursos em programas de melhoramento de eucalipto, especialmente por se tratar de uma planta perene, com ciclo médio entre seis e sete anos no Brasil. A GWS precoce, de acordo com os resultados deste trabalho, revelou-se uma realidade a ser explorada, principalmente para aqueles caracteres em que já se pratica a seleção fenotípica precocemente. Isto torna a aplicabilidade da GWS ainda maior, já que a fenotipagem tardia reduziria o ganho por unidade de tempo, ou seja, o ganho com a seleção já alcançado simplesmente pelo melhoramento clássico.

6 CONCLUSÕES

- a) Acurácias de validação 46% e 40% inferiores que as da estimação, em média, foram encontradas para SNPs e DArTs, respectivamente;
- b) SNPs e DArTs são equivalentes em termos de capacidade preditiva (acurácia);
- c) A GWS mostrou-se eficiente para seleção precoce principalmente para circunferência à altura do peito e altura total das plantas.

REFERÊNCIAS

AKBARI, M. et al. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. **Theoretical and Applied Genetics**, Berlin, v. 113, n. 8, p. 1409-1420, Nov. 2006.

ASSIS, T. F. Melhoria genética do eucalipto. **Informe Agropecuário**, Belo Horizonte, v. 18, n. 185, p. 32-41, 1996.

ASSIS, T. F.; MAFIA, R. G. Hibridação e clonagem. In: BORÉM, A. (Ed.). **Biociência florestal**. Viçosa, MG: UFV, 2007. p. 93-121.

ASSOCIAÇÃO BRASILEIRA DE CELULOSE E PAPEL. **Dados do setor**. Disponível em: <<http://bracelpa.org.br/bra2/sites/default/files/estatisticas/booklet.pdf>>. Acesso em: 6 out. 2014.

ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS. **Anuário estatístico da ABRAF 2013, ano base 2012**. Brasília, 2014. 146 p.

BARIL, C. P. et al. Structure of the specific combining ability between two species of *Eucalyptus* I. RAPD data. **Theoretical and Applied Genetics**, Berlin, v.94, n. 6/7, p. 796-803, June 1997.

BELTRAME, R. et al. Silvicultural performance and early selection of clones from eucalyptus hybrids. **Pesquisa Agropecuária Brasileira**, Brasília, v. 47, n. 6, p. 791-796, jun. 2012.

BERNARDO, R. Genomewide selection of parental inbreds: classes of loci and virtual biparental populations. **Crop Science**, Madison, v. 54, n. 6, p. 2586-2595, Nov./Dec. 2014.

BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, Madison, v. 48, n. 5, p. 1649-1664, Sept. 2008.

BERNARDO, R.; YU, J. Prospects for Genome Wide Selection for quantitative traits in maize. **Crop Science**, Madison, v. 47, n. 3, p. 1082-1090, May 2007.

BORRALHO, N. M. G.; ALMEIDA, M. H.; POTTS, B. M. O melhoramento do eucalipto em Portugal. In: _____. **Eucalipto em Portugal: impactos ambientais e investigação científica**. Lisboa: ISA, 2007. p. 61-110.

BOTSTEIN, D. et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **The American Journal of Human Genetics**, Chicago, v.32, p. 314-331, 1980.

BRONDANI, R. P. V. et al. A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. **BMC Plant Biology**, London, v. 6, n. 20, p. 1-16, 2006.

BROOKES, A. J. The Essence of SNPs. **Gene**, Amsterdam, v. 234, n. 8, p. 177-186, July 1999.

CHEE, M. et al. Accessing genetic information with high-density DNA arrays. **Science**, New York, v. 274, n. 25, p. 610-614, Oct. 1996.

DE LOS CAMPOS, G. et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, Austin, v. 182, n. 1, p. 375-385, May 2009.

DEKKERS, J. C. M. Commercial application of marker and gene assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, Champaign, v. 82, p. 313-328, Feb. 2004. Supplement.

DEKKERS, J. C. M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, London, v. 3, n. 1, p. 22-32, Jan. 2002.

DUDLEY, J. W. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. **Crop Science**, Madison, v. 33, n. 4, p. 660-667, July/Aug. 1993.

ELSHIRE, R. J. et al. A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. **PLoS ONE**, San Francisco, v. 6, n. 5, 2011. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019379>>. Acesso em: 10 nov. 2014.

ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome**, Madison, v. 4, n. 3, p. 250-255, Nov. 2011.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th ed. Harlow: Longman, 1996. 463 p.

FERNANDO, R. L.; GROSSMAN, M. Marker-assisted selection using best linear unbiased prediction. **Genetics Selection Evolution**, London, v. 21, p. 467-477, 1989.

FOELKEL, C. E. B. Qualidade da madeira de eucalipto para atendimento das exigências do mercado de celulose e papel. In: CONFERÊNCIA IUFRO SOBRE SILVICULTURA E MELHORAMENTO DE EUCALIPTOS, 3., 1997, Salvador. **Anais...** Salvador: EMBRAPA/SIF/IPEF, 1997. p. 15-22.

FREEMAN, J. S.; POTTS, B. M.; VAILLANCOURT, R. E. Few Mendelian genes underlie the quantitative response of a forest tree, *Eucalyptus globulus*, to a natural fungal epidemic. **Genetics**, Austin, v.178, n. 1, p. 563-571, Jan. 2008.

GAIOTTO, F. A.; BRAMUCCI, M.; GRATTAPAGLIA, D. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophyllas*.t. Blake with dominant RAPD and AFLP markers. **Theoretical and Applied Genetics**, Berlin, v. 95, n. 5/6, p. 842-849, Oct. 1997.

GANAL, M. W.; ALTMANN, T.; RÖDER, M. S. SNP identification in crop plants. **Current Opinion in Plant Biology**, Oxford, v. 12, n. 2, p. 211-217, Apr. 2009.

GODDARD, M. E. Mapping genes for quantitative traits using linkage disequilibrium. **Genetics Selection Evolution**, London, v. 23, n. 1, p. 131-134, 1991. Supplement.

GODDARD, M. E. A mixed model for analysis of data on multiple genetic markers. **Theoretical and Applied Genetics**, Berlin, v. 83, n. 6/7, p. 878-886, Apr. 1992.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, n. 6, p. 323-330, Dec. 2007.

GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, Heidelberg, v. 7, n. 2, p. 241-255, Apr. 2011.

GREAVES, B. L. et al. Use of a Pilodyn for the indirect selection of basic density in *Eucalyptus nitens*. **Canadian Journal of Forest Research**, Ottawa, v. 26, n. 9, p. 1643-1650, 1996.

HABIER, D. et al. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, London, v. 12, n. 1, p. 186, May 2011.

HAMBLIN, J.; ZIMMERMANN, M. J. de O. Breeding common bean for yield in mixtures. **Plant Breeding Reviews**, New York, v. 4, p. 245-272, 1986.

HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection for crop improvement. **Crop Science**, Madison, v. 49, n. 1, p. 1-12, Jan. 2009.

JACCOUD, D. et al. Diversity arrays: a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, Oxford, v. 29, n. 4, p. 1-7, Feb. 2001.

JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. Hypervariable 'minisatellite' regions in human DNA. **Nature**, London, v. 314, p. 67-73, 1985.

JUNGHANS, D. T. et al. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetics**, Berlin, v. 108, n. 1, p. 175-180, Dec. 2003.

KEIL, M.; GRIFFIN, A. R. Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*. **Theoretical and Applied Genetics**, Berlin, v. 89, n. 4, p. 442-450, Oct. 1994.

KENNEDY, B. W.; QUINTON, M.; ARENDONK, J. A. M. van. Estimation of effects of single genes on quantitative traits. **Journal of Animal Science**, Champaign, v. 70, n. 7, p. 2000-2012, July 1992.

KILIAN, A. et al. The fast and the cheap: SNP and DArT-based whole genome proWling for crop improvement. In: INTERNATIONAL CONGRESS "IN THE WAKE OF THE DOUBLE HELIX: FROM THE GREEN REVOLUTION TO THE GENE REVOLUTION", 2003, Bologna. **Proceedings...** Oxford: Oxford University, 2005. p. 443-461.

KIRST, M. et al. Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. **Journal of Heredity**, Cary, v. 96, n. 2, p. 161-166, Dec. 2005.

KOLBEHDARI, D.; SHAEFFER, L. R.; ROBINSON, J. A. B. Estimation of genome-wide haplotype effects in half-sib designs. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, n. 6, p. 356-361, Dec. 2007.

LANDE, R.; THOMPSON, R. Efficiency of marker assisted selection in the improvement of quantitative traits. **Genetics**, Austin, v. 124, n. 3, p. 743-756, Mar. 1990.

LIMA, B. M. de. **Bridging genomics and quantitative genetics of Eucalyptus:** genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data. 2014. 92 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2014.

LIMA, J. L. et al. Early selection of parents and tres in Eucalyptus full-sib progeny tests. **Crop Breeding and Applied Biotechnology**, Londrina, v. 11, n. 1, p. 10-16, Mar. 2011.

LONG, N. et al. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, n. 6, p. 377-389, Dec. 2007.

MACKAY, T. F. C. The genetic architecture of quantitative traits. **Annual Review of Genetics**, Palo Alto, v. 35, n. 1, p. 303-339, Dec. 2001.

MARQUES, C. M. et al. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. **Theoretical and Applied Genetics**, Berlin, v. 105, n. 2/3, p. 474-478, Aug. 2002.

MASSARO, R. A. M. et al. Viabilidade de aplicação da seleção precoce em testes clonais de *Eucalyptus* spp. **Ciência Florestal**, Santa Maria, v. 20, n. 4, p. 597-609, out./dez. 2010.

MAZUR, B. J. Commercializing the products of plant biotechnology. **Trends in Biotechnology**, Amsterdam, v. 13, n. 9, p. 319-323, Sept. 1995.

MEUWISSEN, T. H. E.; ARENDONK, J. A. M. van. Potential improvements in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. **Journal of Dairy Science**, Champaign, v. 75, n. 6, p. 1651-1659, June 1992.

MEUWISSEN, T. H. E.; GODDARD, M. E.; HAYES, B. J. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Austin, v. 157, n. 4, p. 1819-1829, Apr. 2001.

MISSIAGGIA, A. A.; PIACEZZI, A. L.; GRATTAPAGLIA, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. **Tree Genetics and Genomes**, Heidelberg, v. 1, n. 1, p. 79-84, Oct. 2005.

MOURA, V. P. G.; SANTIAGO, J. **Densidade básica da madeira de espécies de Pinus tropicais determinada através de métodos não-destrutivos**. Planaltina: EMBRAPA-CPAC, 1991. 14 p. (Boletim de Pesquisa, 33).

MUIR, W. M. Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, n. 6, p. 342-355, Dec. 2007.

MUNERI, A.; RAYMOND, C. A. Genetic parameters and genotype-by-environment interactions for basic density, pilodyn penetration and stem diameter in *Eucalyptus globules*. **Forest Genetics**, Zvolen, v. 7, n. 4, p. 317-328, 2000.

NESBITT, K. A. et al. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globules* (Myrtaceae). **Heredity**, Cary, v. 74, n. 6, p. 628-637, June 1995.

OLIVEIRA, E. J. et al. Genome-wide selection in cassava. **Euphytica**, Wageningen, v. 187, n. 2, p. 263-276, Sept. 2012.

PAVAN, B. E. et al. Early selection in open-pollinated *Eucalyptus* families based on competition covariates. **Pesquisa Agropecuária Brasileira**, Brasília, v. 49, n. 6, p. 483-492, jun. 2014.

PEREIRA, A. B. et al. Eficiência da seleção precoce em famílias de meios-irmãos de *Eucalyptus camaldulensis* Dehnh., avaliadas na região noroeste do Estado de Minas Gerais. **Cerne**, Lavras, v. 3, n. 1, p. 67-81, jan./dez. 1997.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 fev. 2014.

RAMALHO, M. A. P. et al. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Lavras: UFLA, 2012. 522 p.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, 2012. 300 p.

RAMALHO, M.A.P.; LAMBERT, E. de S. Biometria e o melhoramento de plantas na era da genômica. **Revista Brasileira de Milho e Sorgo**, Sete Lagoas, v.3, n. 2, p.228-249, jan./dez. 2004.

RESENDE, M. D. V. de. **Genômica quantitativa e seleção no melhoramento de plantas perenes e animais**. Colombo: EMBRAPA Florestas, 2008. 330 p.

RESENDE, M. D. V. de. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: EMBRAPA Florestas, 2007a. 560 p.

RESENDE, M. D. V. de. **Software Selegen - REML/ BLUP: sistema estatístico e seleção genética computadorizada via modelos mistos**. Colombo: EMBRAPA Florestas, 2007b. 359 p.

RESENDE, M. D. V. de et al. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, Cambridge, v. 194, n. 1, p. 116-128, Dec. 2012.

RESENDE, M. D. V. de et al. 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. de et al. **Seleção Genômica Ampla (GWS) via modelos mistos (REML/BLUP), inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e estatística espacial**. Viçosa, MG: UFV, 2012. 291 p.

RESENDE, M. F. R. et al. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinustaeda* L.). **Genetics**, Austin, v. 190, n. 4, p. 1503-1510, Jan. 2012.

RESENDE JÚNIOR, M. F. R. **Seleção genômica ampla no melhoramento vegetal**. 2010. 67 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Viçosa, Viçosa, MG, 2010.

REZENDE, G. D. S. P.; BERTOLUCCI, F. de L. G.; RAMALHO, M. A. P. Eficiência da seleção precoce na recomendação de clones de eucalipto avaliados no norte do Espírito Santo e sul da Bahia. **Cerne**, Lavras, v. 1, n. 1, p. 45-50, jan./dez. 1994.

SANSALONI, C. et al. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. **BMC Proceedings**, London, v. 5, n. 7, p. 1-2, 2011.

SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, Berlin, v. 123, n. 4, p. 218-223, Aug. 2006.

SCHLOTTERER, C. The evolution of molecular markers: just a matter of fashion? **Nature Reviews Genetics**, London, v. 5, n. 1, p. 63-69, Jan. 2004.

SCHÖN, C. S. et al. QTL mapping based on resampling in a vast maize testcross experiment confirms the infinitesimal model of quantitative genetics for complex traits. **Genetics**, Austin, v. 167, n. 1, p. 485-498, May 2004.

SHIMOYAMA, V. R. de S.; BARRICHELLO, L. E. G. **Densidade básica da madeira, melhoramento e manejo florestal**. Piracicaba: IPEF, 1989. 22 p. (Série Técnica IPEF, 20).

SILVA, P. H. M.; BARRICHELLO, L. E. G. Progressos recentes na área florestal. In: PATERNIANI, E. (Ed.). **Ciência, agricultura e sociedade**. Brasília: EMBRAPA Informação Tecnológica, 2006. p. 439-456.

TANKSLEY, S. D.; RICK, C. M. Isozyme gene linkage map of the tomato: applications in genetics and breeding. **Theoretical Applied Genetics**, Berlin, v. 57, p. 161-170, 1980.

TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, Oxford, v. 17, p. 6463-6471, 1989.

TOLFO, A. L. T. **Estudos da viabilidade de aplicação da seleção precoce em testes clonais de *Eucalyptus* spp. e qualidade da madeira para polpa celulósica**. 2003. 50 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Estadual Paulista, Jaboticabal, 2003.

VARSHNEY, R. K. et al. Next-generation sequencing technologies and their implications for crop genetics and breeding. **Trends in Biotechnology**, Amsterdam, v. 27, n. 9, p. 522-530, Aug. 2009.

VOS, P. et al. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, Oxford, v. 23, n. 21, p. 4407-4414, Nov. 1995.

WENZL, P. et al. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v.101, n. 26, p. 9915-9920, June 2004.

WIENTJES, Y. C. J. et al. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. **Genetics**, Austin, v. 193, n. 2, p. 621-631, Feb. 2013.

WILLIAMS, J. G. K. et al. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research**, Oxford, v. 18, n. 25, p. 6531-6535, Nov. 1990.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, Baltimore, v. 163, n. 2, p. 789-801, Feb. 2003.

XU, Y. B.; CROUCH, J. H. Marker-assisted selection in plant breeding: from publications to practice. **Crop Science**, Madison, v. 48, n. 2, p. 391-407, Mar. 2008.

ZHAO, J. H. Gap: genetic analysis package. **Journal of Statistical Software**, Los Angeles, v. 23, n. 8, p. 1-18, Dec. 2007.