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

**Construção de mapa genético integrado em uma progênie de
irmãos-completos proveniente do cruzamento entre *Eucalyptus grandis*
e *Eucalyptus urophylla***

Cristiane Hayumi Taniguti

Dissertação apresentada para obtenção do título de
Mestra em Ciências. Área de concentração: Genética
e Melhoramento de Plantas

**Piracicaba
2016**

Cristiane Hayumi Taniguti
Bacharela em Biotecnologia

**Construção de mapa genético integrado em uma progênie de
irmãos-completos proveniente do cruzamento entre *Eucalyptus grandis*
e *Eucalyptus urophylla***

Orientador:

Prof. Dr. **ANTONIO AUGUSTO FRANCO GARCIA**

Dissertação apresentada para obtenção do título de
Mestra em Ciências. Área de concentração: Genética
e Melhoramento de Plantas

Piracicaba
2016

RESUMO

Construção de mapa genético integrado em uma progênie de irmãos-completos proveniente do cruzamento entre *Eucalyptus grandis* e *Eucalyptus urophylla*

O eucalipto é amplamente cultivado em diversos países, dentre os quais, o Brasil se destaca pela sua alta produção. A cultura tem grande importância comercial e atende à uma ampla variedade de setores do mercado, entre eles o de celulose. Apesar disso, a cultura ainda está nas fases iniciais de domesticação devido, principalmente, ao seu longo ciclo reprodutivo e tempo de rotação, uma vez que os cortes são feitos entre 5 e 15 anos. A aplicação de tecnologias de marcadores moleculares é uma proposta promissora para acelerar o melhoramento do eucalipto. Com o desenvolvimento de metodologias modernas de sequenciamento tornou-se acessível a obtenção de grande quantidade de marcadores a baixo custo. Uma das aplicações de tais marcadores é a construção de mapas genéticos de ligação, os quais permitem a caracterização genética de caracteres quantitativos, além de estudos comparativos entre populações e o auxílio na montagem de genomas. No presente trabalho, objetivou-se a construção de um mapa genético integrado em uma progênie F_1 segregante com 200 indivíduos, proveniente do cruzamento entre *Eucalyptus grandis* e *Eucalyptus urophylla*. Para identificação dos marcadores, foi realizado o ressequenciamento do genoma completo (WGS) dos genitores e a genotipagem por sequenciamento (GBS) da progênie. A metodologia de construção de mapa foi adaptada para o conjunto de dados obtido, que apresenta grande quantidade de marcadores do tipo SNP, pouco informativos e com maior probabilidade de erro comparado com os marcadores tradicionais mais utilizados nos últimos anos. Para isso foram propostas duas estratégias: i) utilização da posição dos marcadores no genoma de referência para auxílio na ordenação dos marcadores no mapa; ii) alteração do parâmetro de probabilidade de erro da abordagem implementada no software *Onemap*. O mapa obtido apresentou padrão de taxa de recombinação semelhante a outros mapas construídos para eucalipto. O mapa apresentou tamanho total de 1471.91 cM e 1512 marcadores, com distância média entre eles de 1.85 cM. Os marcadores formaram 11 grupos de ligação, que corresponderam aos cromossomos do genoma de referência. Em média, foram cobertos 96.8% dos cromossomos. Também foram agrupados, junto aos 11 grupos, 61 marcadores localizados em outros *scaffolds* no genoma de referência, os quais podem servir para elucidação na montagem destes. Utilizando as estratégias propostas, foi obtido um mapa integrado adequado para o experimento em questão, considerando o tamanho da população de mapeamento.

Palavras-chave: Mapa de ligação, Recombinação, Eucalipto, Sequenciamento

ABSTRACT

Development of an integrated genetic map for a full-sib progeny from crossing between *Eucalyptus grandis* and *Eucalyptus urophylla*

The eucalyptus is widely cultivated in several countries, among which Brazil is highlighted by its high production. This culture has great commercial importance and supplies a wide variety of markets, including cellulose. However, the culture is still in the early stages of domestication due, mainly, to its long reproductive cycle and rotation time, since cuts are made between 5 and 15 years. The application of molecular marker technologies is a promising proposal to accelerate the improvement of eucalyptus. By the development of modern sequencing methodologies it was possible to obtain a large quantity of markers at low cost. One of the applications of such markers is the construction of genetic linkage maps, which allow the genetic characterization of quantitative traits, as well as comparative studies between populations and the support in the assembly of genomes. In the present work, the aim was to construct an integrated genetic map in a segregating F_1 progeny with 200 individuals, derived from the cross between *Eucalyptus grandis* and *Eucalyptus urophylla*. For the identification of the markers, it was performed a complete genome re-sequencing (WGS) of the parents and genotyping-by-sequencing (GBS) of the progeny. The mapping methodology was adapted to the obtained data set, which presents a large amount of SNP-type markers, with little information and with a greater probability of error compared to the most used traditional markers in the last years. For this, two strategies were proposed: i) use of the position of the markers in the reference genome to aid in the ordering of the markers on the map; ii) change of the error probability parameter in the approach implemented in the software *Onemap*. The obtained map showed recombination rate pattern similar to other maps constructed for eucalyptus. The map presented a total size of 1471.91 cM and 1512 markers, with a mean distance between them of 1.85 cM. The markers formed 11 linkage groups, which corresponded to chromosomes of the reference genome. On average, 96.8 % of chromosomes were covered. 61 markers located in other scaffolds in the reference genome were grouped with the 11 groups. They may serve to elucidate the assembly of these. Using the proposed strategies, a suitable integrated map was obtained for the present experiment, considering the size of the mapping population.

Keywords: Linkage map, Recombination, Eucalipt, Sequencing

1 INTRODUÇÃO

O eucalipto é atualmente a cultura florestal mais amplamente cultivada por todo o mundo, devido ao interesse comercial em sua madeira e em sua capacidade de crescimento rápido (Mora and Garcia, 2000).

De fato, o eucalipto é na verdade um grupo espécies natural das paisagens da Austrália e de algumas ilhas da Indonésia. Inicialmente, houveram muitas controvérsias quanto a classificação dos eucaliptos, ora baseando-se na capacidade de hibridização interespecífica, ora em caracteres morfológicos ou história evolutiva (Johnson, 1976). Recentemente, análises cladísticas de dados morfológicos e moleculares mostraram um alto nível de compatibilidade na filogenia do grupo. Dos sete gêneros agora reconhecidos como pertencentes ao grupo dos eucaliptos, o gênero *Eucalyptus* é o que mais se destaca pelo interesse comercial e distribuição pelo mundo (Ladiges et al., 2003).

Os eucaliptos também se destacam comercialmente pela sua capacidade de adaptação à diversos ambientes. Até quando cultivados como exóticos, superam a produtividade de espécies nativas. Essa atual capacidade de adaptação pode ser consequência de diferentes pressões de seleção que ocorreram na região do continente australiano nos milhões de anos do período Terciário, com as mudanças climáticas, topográficas e do solo. Dos muitos gêneros e espécies disponíveis para a nova colonização, o ancestral do eucalipto teve características adaptativas e competitivas, latentes ou ainda em desenvolvimento, as quais foram necessárias para a dispersão da reduzida floresta tropical para toda parte da Austrália e para as ilhas ao norte (Eldridge et al., 1994). Entre essas características destacam-se: a adaptação aos solos desgastados, com baixo conteúdo de nutrientes, principalmente do fósforo (Mora and Garcia, 2000); mecanismos de resistência ao fogo (Jacobs, 1981); sistema de acasalamento predominantemente cruzado e com facilidade em hibridizações, privilegiando a diversidade genética; a produção abundante de pequenas sementes para eficiente dispersão; e mecanismos para crescimento rápido (Eldridge et al., 1994).

Foi possível maiores elucidações sobre a evolução do genoma e outras características dos eucaliptos com o sequenciamento e montagem de 94% do total de 640 Mb do genoma de *Eucalyptus grandis*, publicado em Myburg et al. (2014). Neste trabalho, foi identificado que 34% dos genes aparecem em repetições em tandem. Tais eventos de duplicação parecem ter um importante papel na adaptação à ambientes com frequentes alterações climáticas. Observa-se também que mesmo dentro do gênero *Eucalyptus* há um dinamismo nessas repetições. Assim, as características adaptativas variam muito entre as espécies do gênero.

Com o sequenciamento do genoma também foi possível identificar regiões específicas com altas taxas de depressão endogâmica. A fecundação preferencialmente cruzada, com taxa de aproximadamente 90%, permite que se mantenham muitos locos heterozigotos por todo genoma, e, com isso, o acúmulo de vários genes recessivos com pequenos efeitos deletérios que podem afetar o desempenho e sobrevivência do indivíduo, quando em homozigose (Myburg et al., 2014).

A disseminação do eucalipto pelo mundo ocorreu a partir do século XIX e se intensificou há aproximadamente 50 anos (Mora and Garcia, 2000). No Brasil, a expansão da cultura de eucaliptos se deu com o pesquisador Navarro de Andrade entre 1904 e 1909. O pesquisador foi responsável pela realização de diversos experimentos no Horto de Jundiá-SP, que tinham o objetivo de comparar o desempenho de várias espécies nativas com o de eucaliptos, os quais se sobressaíram em relação àquelas. Na época, a cultura foi impulsionada pela demanda da também emergente Companhia Paulista de Estradas de Ferro. Assim, foi iniciado o plantio de eucaliptos em escala comercial, especialmente em Rio Claro (SP), com sementes de 144 espécies. O destino da madeira era diverso: desde a lenha para combustíveis das locomotivas até mourões de cercas e postes margeando a ferrovia, fornecendo ainda os dorminetes e o madeiramento para as construções das estações e vilas (Jacobs, 1981; Mora and Garcia, 2000).

Conforme foram aumentando as demandas por seus produtos, a cultura de eucaliptos no Brasil

foi se expandindo e, atualmente, é uma das maiores do mundo. Os plantios são espalhados por todo território nacional, concentrados principalmente no sudeste do país (ABRAF, 2013).

Os eucaliptos possuem tipicamente fibras curtas, característica visada no comércio de celulose, para o qual destina-se aproximadamente 35% de toda produção florestal brasileira, o que coloca o país em terceiro lugar entre países produtores, ficando atrás apenas dos Estados Unidos e Canadá (ABRAF, 2013).

A celulose é destinada a produção principalmente de papéis para escrever, embalagens, papel cartão e papéis sanitários. Além destes, a cultura também abrange outros setores do mercado, tais como geração de energia com sua biomassa, fabricação de móveis, carvão vegetal e construção civil, além de óleos essenciais com os quais são fabricados alimentos, produtos de limpeza, perfumes e remédios, entre outras aplicações (ABRAF, 2013). Portanto, os programas de melhoramento na cultura focam principalmente em características como o volume e densidade da madeira, e rendimento da polpa.

Além da diversidade que existe no grupo, outra característica muito vantajosa para o melhoramento da cultura é seu alto potencial de hibridização interespecífica, pois a utilização de híbridos aumenta ainda mais a diversidade genética. A hibridização combinada com uma seleção intensiva entre as famílias e a propagação clonal já resultaram em grandes ganhos genéticos à cultura, como é o caso do *Eucalyptus urograndis* (*Eucalyptus grandis* x *Eucalyptus urophylla*) (Grattapaglia, 2008). Este híbrido é muito utilizado no Brasil por combinar as melhores características das duas espécies, ou seja, o crescimento e qualidade da madeira de *E. grandis* e a adaptação e resistência a doenças, particularmente ao fungo causador do cancro do eucalipto, do *E. urophylla* (Mora and Garcia, 2000; Eldridge et al., 1994).

Entretanto, comparados às outras culturas, os eucaliptos ainda estão nas fases iniciais da domesticação considerando que a maioria das árvores nos programas de melhoramento estão há apenas algumas gerações retiradas da selva australiana (Grattapaglia, 2008). Um dos principais fatores que dificultam o processo de domesticação são os longos ciclos reprodutivos e os tempos de rotação, uma vez que os cortes em eucaliptos são feitos entre 5 e 15 anos.

Com o desenvolvimento de marcadores moleculares foi possível obter informações sobre a variação alélica de determinados locos no genoma. As aplicações desses conhecimentos são propostas muito promissoras para acelerar o processo de domesticação nos eucaliptos (Harfouche et al., 2012).

A ideia de marcadores genéticos se iniciou junto com o conceito de ligação entre genes, com os estudos de Thomas Morgan e Sturtevant, em que foram observadas variações em características morfológicas acompanhadas por variações em outras características, relação que se repetia conforme certo padrão nas gerações seguintes, diferente dos padrões esperados conforme as Leis de Mendel. Esses primeiros marcadores, denominados de morfológicos, tiveram grande utilidade na construção dos primeiros mapas genéticos e na seleção de características de interesse, mas eles sofrem grande influência do ambiente, o que os tornam imprecisos. Além disso, têm número limitado para estudos mais complexos (Winter and Kahl, 1995). Mais tarde, com o desenvolvimento da eletroforese, iniciou-se o uso de isoenzimas (Tanksley, 1993), as quais têm pouca influência do meio, mas também apresentam número limitado. Então, com a PCR e as enzimas de restrição, iniciou-se o uso de marcadores conforme variações nas sequências de DNA, que estão presentes por todo o genoma e não sofrem influência do meio a ponto de alterar sua sequência (Collard et al., 2005).

Hoje existem diversos tipos de marcadores moleculares. Eles se diferenciam basicamente pelo modo de transmissão (herança biparental nuclear, herança materna nuclear, herança materna por organelas, herança paterna por organelas), sua forma de representação (dominante ou codominante) e no método de análise (baseado em hibridação, PCR ou sequenciamento) (Semagn et al., 2006).

O *single nucleotide polymorphism* (SNP) é o tipo de marcador molecular mais abundante por se tratar de uma variação em um único par de base. Külheim et al. (2009) encontraram a taxa de um SNP a cada 16 a 33 pb entre quatro espécies de eucaliptos, uma das taxas mais altas encontradas em

árvores e em plantas em geral. A alta diversidade em nucleotídeos nos eucaliptos faz com que seja possível encontrar grande número de SNPs, mas também dificulta a identificação desses, uma vez que existe grande variação na sua vizinhança (Grattapaglia et al., 2012). Também são frequentes no genoma marcadores denominados *indels*, que consistem em variações devido a inserções ou deleções de uma ou mais bases em determinadas regiões do genoma. Os avanços nas tecnologias de sequenciamento de nova geração têm facilitado a identificação de SNPs, tornando-os um dos marcadores mais utilizados atualmente (figura 1.1).

Figura 1.1. Número total de citações (escala a direita) envolvendo diferentes tipos de marcadores moleculares entre os anos de 1973 e 2013 (figura retirada de Grover and Sharma (2014)).

Com a técnica de genotipagem por sequenciamento (GBS), cujo termo foi primeiramente proposto por Elshire et al. (2011), é possível a identificação de grande número de polimorfismos por todo o genoma e avaliar diversas amostras simultaneamente. Na técnica são utilizadas uma ou mais enzimas de restrição para redução da complexidade do genoma, destacando apenas parte representativa dele. As sequências são então amplificadas por PCR e sequenciadas, de forma a gerar bibliotecas de representação reduzida.

Diferente de técnicas anteriores, que buscavam polimorfismos no sítio de restrição das enzimas, pelo GBS são gerados dados de sequências adjacentes aos sítios de restrição, o que possibilita a identificação de maior número de polimorfismos. A metodologia GBS apresenta menor custo por amostra comparado com outras metodologias de sequenciamento, possibilitando o sequenciamento de maior número delas em uma única análise (Andrews et al., 2016; Kim et al., 2015).

A técnica exige planejamento prévio do número e do tipo de enzimas de restrição que serão utilizadas, e da quantidade de amostras agrupadas numa mesma *lane* de sequenciamento. Enzimas de cortes mais frequentes pelo genoma e maior número de amostras numa mesma *lane* reduzem o número de leituras por sítio de restrição, ou seja, a profundidade dos dados. Tais fatores devem ser determinados conforme os objetivos do estudo e de características da espécie estudada: tamanho do genoma, complexidade, a ploidia e a heterozigosidade esperada.

A técnica de GBS tem ganhado popularidade nos últimos anos, com aplicações em estimativas de diversidade genética, estudos evolutivos, mapeamento de QTLs e desenvolvimento de mapas genéticos de alta densidade de marcadores (Grover and Sharma, 2014). Conforme sua popularidade cresce, são adaptadas as metodologias de análise e interpretação para os dados gerados.

Apesar dos recentes avanços nas tecnologias de marcadores moleculares e das metodologias de aplicação, até então o seu uso mais amplo na cultura do eucalipto tem sido na identificação de clones, como um método de controle de qualidade para plantações clonais de larga escala (Grattapaglia, 2008).

As populações experimentais disponíveis de eucaliptos geralmente envolvem famílias com genitores, heterozigotos para maior parte dos locos, e sua progênie de irmãos-completos ou meio-irmãos maternos (Grattapaglia and Sederoff, 1994).

Em caso de genitores heterozigotos, é possível nas populações de irmãos-completos ou meio-irmãos a avaliação da segregação, no entanto, há menor número de recombinações comparado a populações F_2 ou RILs, pelos haplótipos dos genitores terem passado por apenas um evento de meiose. Além disso a segregação pode apresentar 18 diferentes padrões, já que existe a possibilidade de existir até quatro alelos diferentes em cada loco: aqueles locos com três ou quatro diferentes alelos nos parentais, podem gerar quatro diferentes classes de genótipos nos filhos, outros locos podem seguir o padrão 1:2:1, ou seja, os dois parentais heterozigotos com apenas dois alelos diferentes, ou ainda o padrão 1:1 com um parental homozigoto e outro heterozigoto com dois alelos diferentes. A maior dificuldade em analisar o conjunto desses marcadores é a incerteza sobre as fases de ligação desses marcadores nos parentais. As fases de ligação descrevem a configuração dos alelos em um par de locos heterozigotos em cromossomos homólogos em um único pai. Diferente do observado em linhagens endogâmicas em que os parentais possuem apenas uma fase ligação possível, nas espécies de fecundação cruzada os parentais podem ter quatro diferentes fases o que dificulta a identificação de eventos de recombinação na progênie. Portanto, para esses casos, também devem ser estimadas as fases de ligação (Wu et al., 2002a).

A partir dessas dificuldades tem-se desenvolvido métodos alternativos aos usados em outras populações de melhoramento para construção de mapas de ligação a partir de populações de irmãos-completos.

No primeiro mapa genético construído para eucaliptos, foram utilizados marcadores RAPD (*Random amplified polymorphic DNA*) para identificar segregação e ligação em híbridos de *E. grandis* e *E. urophylla* utilizando a estratégia de duplo pseudo-testcross, descrita em Grattapaglia and Sederoff (1994), tal abordagem considera que a configuração dos alelos parentais não é conhecida *a priori*, mas é inferida *a posteriori* pela análise de segregação na progênie. Para isso, utiliza-se apenas marcadores que segregam na proporção 1:1, os quais se apresentam de forma heterozigota em um pai e homozigota no outro. Cada marcador detém informação de segregação individual de cada genitor o que gera mapas de ligação separados em função de qual é o genitor informativo.

Esta estratégia foi implementada em softwares como o MAPMAKER/EXP (Lander et al., 1987) e PGRI (Liu, 1997) e foi amplamente utilizada na construção de mapas para eucaliptos com o uso de diversos marcadores como RAPD (*Random Amplified Polymorphic DNA*) (Verhaegen and Plomion, 1996; Bundock et al., 2000), AFLP (*Amplified fragment length polymorphism*) (Marques et al., 1998; Myburg et al., 2003), microssatélites (Brondani et al., 1998, 2002, 2006; Marques et al., 2002; Bundock et al., 2000) e SNPs (Bartholomé et al., 2015). Atualmente, são acessíveis marcadores mais informativos que AFLP e RAPD, e existem abordagens melhores para construção de mapas. Portanto, é surpreendente o recente uso de duplo pseudo-testcross com dados de SNPs.

No intuito de aprimorar o método de construção de mapa genético, Ritter et al. (1990) propuseram um método de estimação de dois-pontos de frequência de recombinação por máxima verossimilhança para diferentes configurações de alelos baseados em marcadores RFLP considerando genitores heterozigotos. Arús et al. (1994) e Ritter and Salamini (1996) acrescentaram fórmulas para estimação da fração de recombinação para outras configurações alélicas. Ridout et al. (1998) propuseram aprimoramento do método aplicando análises de ligação de três pontos também por máxima verossimilhança. Maliepaard et al. (1997) apresentam uma visão geral de todas as configurações possíveis em famílias de irmãos-completos, mostrando as fórmulas de estimadores de máxima verossimilhança para todos os casos.

Em Jansen and Nap (2001) também é proposto um modelo de estimador de verossimilhança multiponto para construção de mapas em cruzamentos entre linhagens endogâmicas. Ooijen (2011) expandem esse modelo para famílias de irmãos-completos. O algoritmo também gera dois mapas para

cada genitor por análise multiponto e os une em um único mapa utilizando a média dos comprimentos entre os marcadores informativos para ambos genitores, e interpolando ou extrapolando para marcadores segregando em apenas um genitor. A metodologia em Ooijen (2011) foi implementada no software JoinMap, versão 4.1.

Em Wu et al. (2002a) é apresentado um algoritmo baseado nos estimadores de máxima verossimilhança já apresentados em Maliepaard et al. (1997), mas estimando simultaneamente a ligação e fase de ligação para conjuntos de marcadores de diferentes tipos, o que inclui os marcadores totalmente informativos (segregação 1:1:1:1) e os parcialmente informativos (1:2:1, 3:1 e 1:1), em famílias de irmãos-completos (também conhecido como F_1 segregante. Esta metodologia permite gerar mapas integrados, ou seja, não mais separados para cada genitor. Em Wu et al. (2002b) aprimorou-se o método com um algoritmo multiponto para estimação da máxima verossimilhança com o uso da cadeia de Markov oculta (HMM). A metodologia de Wu et al. (2002a) e Wu et al. (2002b) foi implementada no software *Onemap* (Margarido et al., 2007).

Os mapas de ligação podem ser utilizados para caracterização genética de variações fenotípicas quantitativas, com os denominados QTLs (*Quantitative trait loci*) (Sax, 1923). Assim como os métodos de construção de mapa de ligação, as metodologias de mapeamento de QTLs para espécies de cruzamento preferencialmente cruzado foram sendo aprimoradas com o passar do tempo (Lander and Botstein, 1989; Jansen, 1993; Jansen and Stam, 1994; Zeng, 1994; Wu et al., 2002a; Lin et al., 2003; Gazaffi et al., 2014). No modelo desenvolvido em Gazaffi et al. (2014), considera-se um mapa genético integrado construído com a abordagem multiponto e com marcadores com diversos tipos de segregação para identificar QTLs por mapeamento por intervalo composto.

Para o eucalipto, o mapeamento de QTLs tem sido particularmente produtivo em híbridos interespecíficos onde alelos segregantes de grande efeito representam potencialmente maiores efeitos gênicos na diferença entre espécies Grattapaglia et al. (2012).

Mapas de ligação também são utilizados em estudos comparativos entre taxa (Paterson et al., 2000). Nesses estudos, a posição de marcadores mapeados em mapas genéticos de múltiplas taxa são comparados em função da localização (sintenia) e ordem (colinearidade) correspondentes dos locos nos grupos homólogos. Ao destacar as diferenças e similaridades entre espécies é possível a compreensão da evolução de cromossomos e genomas (Paterson et al., 2000). Também possibilita a transferência segura de informações (como de marcadores moleculares, QTLs e posição de genes candidatos) entre espécies (Hudson et al., 2012b).

Como exemplo, Hudson et al. (2012b) utilizou de estudos comparativos entre mapas, construídos com marcadores DArT, para verificar a possibilidade de transferência de informações do genoma de *E. grandis* para as espécies *E. urophylla* e *E. globulus*. Foi concluído então que a alta similaridade entre essas espécies possibilita o uso das informações disponíveis do genoma de *E. grandis* para todas elas.

Com as modernas tecnologias de sequenciamento, os mapas genéticos tem realizado também uma importante função ao elucidar montagens de genomas, resolvendo agrupamentos duvidosos ou adicionando novas sequências aos cromossomos montados (Fierst, 2015).

Mapas importantes no auxílio à montagem da primeira versão do genoma de *E. grandis* (Myburg et al., 2014) foram construídos utilizando marcadores DArT, com a metodologia implementada no JoinMap (Petroli et al., 2012) e também combinada com a estratégia de duplo pseudotestcross (Kullan et al., 2012).

Mais recentemente, para atualização do genoma, foram construídos mapas com a estratégia de duplo pseudotestcross utilizando SNPs identificados a partir de tecnologias de sequenciamento com *arrays* e genotipagem de progênie de irmãos-completos (Bartholomé et al., 2015).

Por combinar diferentes informações de todos os tipos de marcadores, os mapas integrados possuem diversas vantagens em relação aqueles construídos individualmente para cada parental. Entre

elas estão a maior saturação do mapa e extensão para todo o genoma da caracterização de polimorfismos. Além disso, o uso de mapas integrados facilitam na posterior localização de QTLs (Garcia et al., 2006). Apesar da acessibilidade das tecnologias de marcadores e de softwares capazes, até o nosso conhecimento, ainda não há registros de mapas genéticos integrados para eucaliptos. O presente trabalho apresenta um mapa integrado da população de irmãos-completos progênie de *Eucalyptus grandis* e *Eucalyptus urophylla*, construído a partir de SNPs e *indels* identificados pela metodologia GBS e re-sequenciamento do genoma dos genitores.

2 CONSIDERAÇÕES FINAIS

Foram utilizadas estratégias para construção do mapa genético adaptadas à realidade dos conjuntos de dados atuais, provindos de tecnologias de sequenciamento de baixo custo e de eficiência na obtenção de marcadores moleculares. Tais estratégias consistiram na adaptação do parâmetro de probabilidade de erro do teste multiponto para construção do mapa, e no uso da posição conforme o genoma de referência para auxílio na ordenação dos marcadores.

Parte dos marcadores do mapa, construído com probabilidade de erro 0.001%, se tornaram redundantes quando a probabilidade foi alterada para 5%. Nos grupos referentes aos cromossomos 3 e 7, o número de marcadores redundantes superou o número dos marcadores efetivos. Na sua maioria, o agrupamento desses marcadores com o aumento da probabilidade de erro é concordante com o observado no genoma de referência.

O maior intervalo médio entre marcadores foi obtido no cromossomo 5, com valor de 2.53 cM, e o menor no cromossomo 1, com valor de 1.24 cM. Foram mapeados 780 marcadores efetivos (não-redundantes) e 1512 marcadores no total. O mapa totalizou 1471.91 cM e todos marcadores mapeados estão bem distribuídos pelo genoma, com cobertura acima de 93% para todos os cromossomos. Os maiores *gaps* observados no mapa coincidem com regiões altamente repetitivas ou com baixa densidade de genes e porcentagem CG do genoma de referência, nas quais são esperadas baixas frequências de recombinação ou dificuldades na identificação de marcadores.

Com o mapa, foi possível agrupar junto aos 11 cromossomos, 61 marcadores localizados em outros *scaffolds* na versão 2.0 do genoma de *E. grandis*, embora com pouca precisão quanto ao ordenamento em relação aos outros marcadores. Avaliação mais detalhada da relação entre esses *scaffolds* e a região dos cromossomos em que os marcadores foram inseridos no mapa de ligação, poderá contribuir para melhorias no genoma de referência, inserindo sequências cuja localização ainda não foi elucidada.

O tamanho populacional influencia diretamente na identificação de recombinações menos frequentes, ou seja, das menores distâncias no mapa. Experimentos com maior número de indivíduos permitem obtenção de menores distâncias médias entre os marcadores do mapa e, quando há grande quantidade de marcadores, menor número de marcadores redundantes. Portanto, com 200 indivíduos, espera-se uma distância média entre 1 e 2 cM, uma vez que são necessárias duas recombinações para ser identificado 1 cM. Em outras palavras, a mínima fração de recombinação possível de ser identificada entre dois marcadores seria 0.5 cM, quando apenas um indivíduo apresentar recombinação. Portanto, com as estratégias utilizadas, foi possível a construção de um mapa genético integrado adequado para o experimento realizado.

REFERÊNCIAS

- ABRAF ABDFP. 2013. Anuário estatístico da ABRAF 2013 ano base 2012. Technical report, Brasília.
- Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T, and Stern DL. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* **21**: 610–617.
- Andrews KR, Good JM, Miller MR, Luikart G, and Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**: 81–92.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Arús P, Olarte C, Romero M, and Vargas F. 1994. Linkage analysis of ten isozyme genes in F1 segregating almond progenies. *Journal of the American Society for Horticultural Sciences* **119**: 339–344.
- Bali S, Mamgain A, Raina SN, Yadava SK, Bhat V, Das S, Pradhan AK, and Goel S. 2015. Construction of a genetic linkage map and mapping of drought tolerance trait in Indian beverage tea. *Molecular Breeding* **35**: 112.
- Bartholomé J, Mandrou E, Mabilia A, Jenkins J, Nabihoudine I, Klopp C, Schmutz J, Plomion C, and Gion JM. 2015. High-resolution genetic maps of Eucalyptus improve Eucalyptus grandis genome assembly. *New Phytologist* **206**: 1283–1296.
- Bentley DR, Gormley NA, Balasubramanian S, Baybayan PA, Benoit VA, Swerdlow HP, Smith GP, Bridgham JA, Brown AA, Milton J, et al.. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–59.
- Broman KW and Sen S. 2009. *A Guide to QTL mapping with R/qtl*, volume 66 of *Statistics for Biology and Health*. Springer New York, New York, NY.
- Brondani RPV, Brondani C, and Grattapaglia D. 2002. Towards a genus-wide reference linkage map for Eucalyptus based exclusively on highly informative microsatellite markers. *Molecular genetics and genomics : MGG* **267**: 338–47.
- Brondani RPV, Brondani C, Tarchini R, and Grattapaglia D. 1998. Development, characterization and mapping of microsatellite markers in Eucalyptus grandis and E. urophylla. *TAG Theoretical and Applied Genetics* **97**: 816–827.
- Brondani RPV, Williams ER, Brondani C, and Grattapaglia D. 2006. A microsatellite-based consensus linkage map for species of Eucalyptus and a novel set of 230 microsatellite markers for the genus. *BMC plant biology* **6**: 20.
- Bundock PC, Hayden M, and Vaillancourt RE. 2000. Linkage maps of Eucalyptus globulus using RAPD and microsatellite markers. *Silvae Genetica* **49**: 223–232.
- Casbon JA, Osborne RJ, Brenner S, and Lichtenstein CP. 2011. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Research* **39**: 1–8.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, and Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–40.

- Collard BCY, Jahufer MZZ, Brouwer JB, and Pang ECK. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**: 169–196.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al.. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.
- Dempster AP, Laid NM, and Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**: 1–38.
- Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, Guigó R, and Ribeca P. 2012. Fast computation and applications of genome mappability. *PLOS ONE* **7**: 1–16.
- Doerge RW. 1996. Constructing genetic maps by rapid chain delineation. *Journal of Quantitative Trait Loci* **2**: 1–14.
- Doyle J. 1991. DNA protocols for plants. In *Molecular Techniques in Taxonomy*, pp. 283–293. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Durbin RM, Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, De La Vega FM, Donnelly P, et al.. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–1073.
- Eldridge K, John Davidson J, Hardwood C, and Gerrit Van Wyk G. 1994. *Eucalypt domestication and breeding*. Oxford University Press Inc.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, and Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE* **6**: 1–10.
- Fierst JL. 2015. Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Frontiers in Genetics* **6**: 1–8.
- Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, da Silva JAG, Ulian EC, Figueira A, et al.. 2006. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *TAG. Theoretical and applied genetics* **112**: 298–314.
- Garrison E and Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *ArXiv e-prints* p. 9.
- Gazaffi R, Margarido GRa, Pastina MM, Mollinari M, and Garcia AAF. 2014. A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny. *Tree Genetics & Genomes* **10**: 791–801.
- Gion JM, Hudson CJ, Lesur I, Vaillancourt RE, Potts BM, and Freeman JS. 2016. Genome-wide variation in recombination rate in *Eucalyptus*. *BMC Genomics* **17**: 590.
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, and Buckler ES. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLOS ONE* **9**: 1–11.
- Grattapaglia D. 2008. Perspectives on genome mapping and marker-assisted breeding of eucalypts. *Southern Forests: a Journal of Forest Science* **70**: 69–75.
- Grattapaglia D and Bradshaw Jr H. 1994. Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Canadian Journal of Forest Research* **24**: 1074–1078.

- Grattapaglia D, Mamani EMC, Silva-Junior OB, and Faria DA. 2015. A novel genome-wide microsatellite resource for species of *Eucalyptus* with linkage-to-physical correspondence on the reference genome sequence. *Molecular Ecology Resources* **15**: 437–448.
- Grattapaglia D and Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**: 1121–37.
- Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Külheim C, Potts BM, and Myburg Aa. 2012. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genetics and Genomes* **8**: 463–508.
- Grover A and Sharma PC. 2014. Development and use of molecular markers: past and present. *Critical reviews in biotechnology* **8551**: 1–13.
- Hackett Ca and Broadfoot LB. 2003. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90**: 33–38.
- Harfouche A, Meilan R, Kirst M, Morgante M, Boerjan W, Sabatti M, and Scarascia Mugnozza G. 2012. Accelerating the domestication of forest trees in a changing world. *Trends in Plant Science* **17**: 64–72.
- Hudson CJ, Freeman JS, Kullán AR, Petroli CD, Sansaloni CP, Kilian A, Detering F, Grattapaglia D, Potts BM, Myburg AA, et al.. 2012a. A reference linkage map for *Eucalyptus*. *BMC Genomics* **13**: 240.
- Hudson CJ, Kullán ARK, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Myburg AA, Potts BM, and Vaillancourt RE. 2012b. High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. *Tree Genetics & Genomes* **8**: 339–352.
- Ihaka R and Gentleman R. 1996. R: A Language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**: 299–314.
- Jacobs MR. 1981. *Eucalypts for planting*, volume 11.
- Jansen RC. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* **135**: 205–211.
- Jansen RC and Nap JP. 2001. Genetical genomics: the added value from segregation. *Trends in genetics* **17**: 388–391.
- Jansen RC and Stam P. 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- Jiang C and Zeng ZB. 1997. Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* **101**: 47–58.
- Johnson LAS. 1976. Problems of species and genera in *Eucalyptus* (Myrtaceae). *Plant Systematics and Evolution* **125**: 155–167.
- Kim C, Guo H, Kong W, Chandnani R, Shuang LS, and Paterson AH. 2015. Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science* **242**: 1–9.
- Kosambi D. 1944. The estimation of map distance from recombination values. *Annals of Eugenics* **12**: 172–175.
- Külheim C, Hui Yeoh S, Maintz J, Foley WJ, and Moran GF. 2009. Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* **10**: 452.

- Kullan ARK, van Dyk MM, Jones N, Kanzler A, Bayley A, and Myburg Aa. 2012. High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis* $\times$ *E. urophylla*. *Tree Genetics & Genomes* **8**: 163–175.
- Ladiges PY, Udovicic F, and Nelson G. 2003. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography* **30**: 989–998.
- Lander ES and Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, and Newberg La. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.
- Le SQ and Durbin R. 2011. SNP detection and genotyping from low coverage sequencing data on multiple diploid samples. *Genome research* **21**: 952–960.
- Li H and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R. 2009a. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Li H, Ruan J, Durbin R, Li H, Ruan J, and Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* pp. 1851–1858.
- Li R, Li Y, Fang X, Res G, Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, et al.. 2009b. SNP detection for massively parallel whole-genome resequencing SNP detection for massively parallel whole-genome resequencing pp. 1124–1132.
- Li Y, Sidore C, Kang HM, Boehnke M, and Abecasis GR. 2011. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Research* **21**: 940–951.
- Lin M, Lou X, Chang M, and Wu R. 2003. A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. *Genetics* **165**: 901–913.
- Liu BH. 1997. *Statistical genomics: linkage, mapping, and QTL analysis*. Press, CRC.
- Maliepaard C, Jansen J, and Van Ooijen JW. 1997. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genetical Research* **70**: 237–250.
- Margarido GRA, Souza AP, and Garcia AAF. 2007. OneMap: software for genetic mapping in outcrossing species. *Hereditas* **144**: 78–9.
- Marques CM, Araújo Ja, Ferreira JG, Whetten R, O'Malley DM, Liu BH, and Sederoff R. 1998. AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *TAG Theoretical and Applied Genetics* **96**: 727–737.
- Marques M, Brondani V, Grattapaglia D, and Sederoff R. 2002. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **105**: 474–478.

- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al.. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.
- Mézard C. 2006. Meiotic recombination hotspots in plants. *Biochemical Society transactions* **34**: 531–534.
- Mora AL and Garcia CH. 2000. *A cultura do Eucalipto no Brasil*. Sociedade Brasileira de Silvicultura, São Paulo, Brasil.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, et al.. 2014. The genome of *Eucalyptus grandis*. *Nature* **510**: 356–62.
- Myburg AA, Griffin AR, Sederoff RR, and Whetten RW. 2003. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. *TAG Theoretical and Applied Genetics* **107**: 1028–1042.
- Oliveira EJ, Vieira MLC, Garcia AAF, Munhoz CF, Margarido GRA, Consoli L, Matta FP, Moraes MC, Zucchi MI, and Fungaro MHP. 2008. An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. *Journal of the American Society for Horticultural Sciences* **133**: 35–41.
- Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, Figueira AV, Ulian EC, Garcia AAF, and Souza AP. 2007. Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. *Molecular Breeding* **20**: 189–208.
- Ooijen JWV. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetics research* **93**: 343–349.
- Paterson AH, Bowers JE, Burow MD, Draye X, Elisk CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R, et al.. 2000. Comparative genomics of plant chromosomes. *The Plant cell* **12**: 1523–40.
- Petroli CD, Sansaloni CP, Carling J, Steane DA, Vaillancourt RE, Myburg AA, da Silva Jr OB, Pappas Jr GJ, Kilian A, and Grattapaglia D. 2012. Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. *PLOS ONE* **7**: 1–14.
- Pompanon F, Bonin A, Bellemain E, and Taberlet P. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* **6**: 847–846.
- Quezada M, Pastina MM, Ravest G, Silva P, Vignale B, Cabrera D, Hinrichsen P, Garcia AAF, and Pritsch C. 2014. A first genetic map of *Acca sellowiana* based on ISSR, AFLP and SSR markers. *Scientia Horticulturae* **169**: 138–146.
- Ridout MS, Tong S, J VC, and R T. 1998. Three-point linkage analysis in crosses of allogamous plant species. *Genet Res* **72**: 111–121.
- Ritter E, Gebhardt C, and Salamini F. 1990. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* **125**: 645–654.
- Ritter E and Salamini F. 1996. The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. *Genetical Research* **67**: 55.
- Rosado CCG, da Silva Guimarães LM, Faria DA, de Resende MDV, Cruz CD, Grattapaglia D, and Alfenas AC. 2016. QTL mapping for resistance to *Ceratocystis* wilt in *Eucalyptus*. *Tree Genetics & Genomes* **12**: 72.

- Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**: 552–60.
- Schlötterer C. 2004. Opinion: The evolution of molecular markers — just a matter of fashion? *Nature Reviews Genetics* **5**: 63–69.
- Semagn K, Bjornstad a, and Ndjiondjop MN. 2006. An overview of molecular marker methods for plants. *African Journal of Biotechnology* **5**: 2540–2568.
- Serres-Giardi L, Belkhir K, David J, and Glemin S. 2012. Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell* **24**: 1379–1397.
- Silva-junior OB, Faria DA, and Grattapaglia D. 2015. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species pp. 1527–1540.
- Smukowski CS and Noor MAF. 2011. Recombination rate variation in closely related species. *Heredity* **107**: 496–508.
- Souza L, Mantello C, Suzuki F, Gazaffi R, Garcia D, Le Guen V, Garcia AA, and Souza A. 2011. Development of a genetic linkage map of rubber tree (*Hevea brasiliensis*) based on microsatellite markers. *BMC Proceedings* **5**: P39.
- Steane DA, Nicolle D, Sansaloni CP, Petroli CD, Carling J, Kilian A, Myburg AA, Grattapaglia D, and Vaillancourt RE. 2011. Population genetic analysis and phylogeny reconstruction in Eucalyptus (Myrtaceae) using high-throughput, genome-wide genotyping. *Molecular Phylogenetics and Evolution* **59**: 206–224.
- Tanksley SD. 1993. Mapping polygenes. *Annual Review of Genetics* **27**: 205–233.
- Verhaegen D and Plomion C. 1996. Genetic mapping in Eucalyptus urophylla and Eucalyptus grandis using RAPD markers. *Genome* **39**: 1051–1061.
- Voorrips RE. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**: 77–78.
- Winter P and Kahl G. 1995. Molecular marker technologies for plant improvement. *World Journal of Microbiology & Biotechnology* **11**: 438–448.
- Wu R, Ma CX, Painter I, and Zeng ZB. 2002a. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theoretical population biology* **61**: 349–363.
- Wu R, Ma CX, Wu SS, and Zeng ZB. 2002b. Linkage mapping of sex-specific differences. *Genetical research* **79**: 85–96.
- Zarpelon TG, da Silva Guimarães LM, Faria DA, Coutinho MM, Cápua Neto B, Teixeira RU, Grattapaglia D, and Alfenas AC. 2015. Genetic mapping and validation of QTLs associated with resistance to Calonectria leaf blight caused by Calonectria pteridis in Eucalyptus. *Tree Genetics & Genomes* **11**: 803.
- Zeng ZB. 1994. Precision mapping of quantitative trait loci. *Genetics* **1136**: 1457–1468.