UNIVERSIDADE DE BRASÍLIA

Instituto de Ciências Biológicas

Departamento de Biologia Celular

TESE DE DOUTORADO

CARACTERIZAÇÃO GENÔMICA DE MARCADORES DArT COM BASE EM MAPEAMENTO GENÉTICO E FÍSICO E DETECÇÃO DE QTLs EM *Eucalyptus*

CÉSAR DANIEL PETROLI

BRASILIA – DF

Abril, 2013

UNIVERSIDADE DE BRASÍLIA

Instituto de Ciências Biológicas

Departamento de Biologia Celular

CARACTERIZAÇÃO GENÔMICA DE MARCADORES DArT COM BASE EM MAPEAMENTO GENÉTICO E FÍSICO E DETECÇÃO DE QTLs EM *Eucalyptus*

Orientador: Dr. Dario Grattapaglia

Tese apresentada ao Departamento de Biologia Celular do Instituto de Biologia, da Universidade de Brasília, como requisito parcial para obtenção do grau de Doutor em Ciências Biológicas, Área de Concentração: Biologia Molecular

BRASILIA – DF

Abril, 2013

TERMO DE APROVAÇÃO

Tese apresentada ao Departamento de Biologia Celular da Universidade de Brasília, como requisito parcial para obtenção de grau de Doutor em Ciências Biológicas, área de concentração Biologia Molecular.

Tese defendida e aprovada em 26/04/2013 por:

Dr. Dario Grattapaglia - Orientador
Embrapa Recursos Genéticos e Biotecnologia

Dr. Marcio Elias Ferreira
Embrapa Recursos Genéticos e Biotecnologia

Dr. Georgios Joannis Pappas Júnior
Universidade de Brasília

Dr. Márcio de Carvalho Moretzsohn
Embrapa Recursos Genéticos e Biotecnologia

Dr. Alan Carvalho Andrade
Embrapa Recursos Genéticos e Biotecnologia

A meus grandes amores

Caro e nossa princesa Isabella

DEDICO

AGRADECIMENTOS

A realização dessa tese somente foi possível pelo apoio recebido de diversas pessoas e instituições, mas gostaria de agradecer de maneira especial:

À Universidade de Brasília (UnB), e especialmente ao programa de pós-graduação em Biologia Molecular, representados por todos os professores e funcionários, pela oportunidade de realização deste curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nivel superior (CAPES) pelo apoio financeiro que tornou possível a realização desse trabalho.

À EMBRAPA Recursos Genéticos e Biotecnologia, pela infra-estrutura de trabalho.

A todas as empresas florestais e Instituições que facilitaram as amostras e os recursos necessários para a realização deste trabalho.

Devo muito à orientação do Dr. Dario Grattapaglia, com quem estabeleci uma relação profissional e pessoal, a qual desejo seja duradoura após a finalização deste trabalho. É em pessoas como ele que eu me espelho e busco inspiração para o meu próprio desenvolvimento profissional e pessoal.

Ao Dr. Andrzej Kilian, director da empresa *Diversity Arrays Technology*, por ter acreditado no projeto e aberto as portas do DArT para recebernos com muito carinho em Canberra, Australia. Pelos grandes conhecimentos transmitidos que são os pilares para o meu futuro profissional.

Aos membros da banca de avaliação desta tese, o Dr. Marcio Elias Ferreira, Dr. Georgios J. Pappas Júnior, Dr. Marcio de Carvalho Moretzsohn e o Dr. Alan Carvalho Andrade, pelos comentários, sugestões e questionamentos que foram essenciais para a melhoria e o estabelecimento do formato final.

Aos pesquisadores e técnicos do Laboratorio de Genética Vegetal da Embrapa Recursos Genéticos e Biotecnología, Marcão, Marcio, Vânia, Glaucia, Zilneide, Lorena, Peter e muito especialmente a minha grande amiga e conselheira Marília.

Aos amigos que tive o prazer de conhecer no Laboratorio de Genética Vegetal, Bruna (e Daniel), Thaisa, Marília (Georgios e meu príncipe Nicolas), Dione (Andrê e Ian), Ediene e Rodrigo, Tati (Bruno e Arthur), Natália, Mariana, Marco (Bia e Teo), Tulio (e Cecilia) e Pedro pela amizade de tantos anos. A muitos de vocês considero irmãos do coração que a vida me presenteu e tenho certeza que nem a distância ou o tempo vai destruir essa amizade tão linda e verdadeira.

Aos meus amigos do laboratório DarT na Austrália, Colleen, Michael, Kasia, Grzegorz, Jason, Ling, Eric, Damian, Puthick, Vanessa, Cina, Cleare, Frank, Gosia, Hang e Adriane, por todo o apoio e carinho durante minha estadia em Canberra, onde passei dois anos maravilhosos da minha vida, cehio de experiências inesquecíveis.

Aos meus irmãos brasileiros Eduardo, Genny, Arnon, Vinicius, Leo, Gabriel, William e Reinaldo, você sempre estarão presentes.

A meus pais, Elena e Cacho, os principais responsáveis por minha educação e formação como pessoa. Sou eternamente agradecida pelos ensinamentos, apoio incondicional e incentivo para meu crescimento pessoal e profissional. Tudo o que eu sou eu devo a vocês.

Ao meu irmão David e sua esposa Sonia pelo apoio, carinho e principalmente por ter me presenteado com duas crianças maravilhosas: Sofia e Tisiano, amos muito vocês!

Ao meu irmão Gastón, pelo carinho e alegria de sempre. Apesar da juventude você é um exemplo para mim.

À família que me acolheu como mais um filho: Juan, Norma, Adrian, Sole, Vale, Flor e Vicky, tios e primos, pelo carinho e apoio recebido sempre.

À Carol, meu grande amor e companheira de vida, pelo seu apoio, amizade, compreensão e amor incondicional. Este trabalho é compartilhado 100%. Obrigado por ser a melhor esposa do mundo e por ter cumprido meu grande sonho de formar juntos, uma família maravilhosa.

E finalmente a pessoa que conseguiu mudar completamente o sentido da minha vida, nossa princessa Isabella. Nunca imaginei que meu coração tinha tanto amor para dar. Cada dia com um simples sorriso você me faz o papai mais feliz do mundo. Te amo filha! Também para a pessoa que está chegando, você iluminara nossas vidas tanto como a sua irmazihna.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS

ANEXOS

ANEXO I. Cópia de artigo publicado:

***Petroli CD**, Sansaloni CP, Carling J, Steane DA, Vaillancourt RE, Myburg AA, Bonfim da Silva O, Pappas GJ, Kilian A, Grattapaglia D (2012) Genomic Characterization of DArT Markers Based on High-Density Linkage Analysis and Physical Mapping to the Eucalyptus Genome. PLoS ONE 7(9): e44684*

ANEXO II. Cópia de artigo publicado:

*Hudson CJ, Freeman JS, Kullan ARK, **Petroli CD**, Sansaloni CP, Kilian A, Detering F, Grattapaglia D, Potts BM, Myburg AA, Vaillancourt RE (2012) A reference linkage map for Eucalyptus. BMC Genomics 13:240.*

ANEXO III. Cópia de resumo expandido publicado:

*Sansaloni CP, **Petroli CD**, Pappas GJ, Bonfim da Silva O, Grattapaglia D (2011) How many genes might underlie QTLs for growth and wood quality traits in Eucalyptus? BMC Proceedings 5(Suppl 7):P37.*

ANEXO IV. Cópia de resumo expandido publicado:

*García M, Villalba P, Acuña C, Oberschelp J, Harrand L, Surenciski M, Martínez M, **Petroli CD**, Sansaloni C, Faria D, Grattapaglia D, Poltri S. (2011) A genetic linkage map for a full sib population of Eucalyptus grandis using SSR, DArT, CG-SSR and EST-SSR markers. BMC Proceedings 2011, 5(Suppl 7):P26*

RESUMO

Diversity Arrays Technology (DArT) fornece um sistema robusto, de alto rendimento e baixo custo para a identificação de milhares de polimorfismos na sequência do DNA de indivíduos. Apesar da extensa utilização desta plataforma de genotipagem para diversas espécies de plantas, pouco se conhece sobre os atributos dos marcadores DArT do ponto de vista do conteúdo das sequências e sua distribuição em genomas de plantas. Neste trabalho foram investigadas as propriedades genômicas das 7.680 sondas do microarranjo DArT desenvolvido para *Eucalyptus*, por meio do seu sequenciamento e mapeamento genético e físico no genoma de referência de *Eucalyptus grandis*. Em seguida, o mapa genético construído foi utilizado para o mapeamento de QTLs (Quantitative Trait Loci) para sete características quantitativas de crescimento e qualidade da madeira em *Eucalyptus*. Finalmente, para cada QTL identificado foi realizada uma análise preliminar do número de genes anotados na sequência do genoma de *Eucalyptus* incluídos no intervalo genômico correspondente. Um mapa consenso com 2.274 marcadores DArT ancorado a 210 microssatélites com uma média de distância entre marcadores consecutivos na ordem de sub-centimorgan e um mapa genético *framework* com 1.029 marcadores ordenados com maior suporte estatístico foram construídos. Ambos exibiram uma extensa colinearidade com a sequência do genoma quando mapeados fisicamente. O número de sequências únicas observadas para as sondas DArT foi consistente com o número de sequências DArT alinhadas em uma posição única no genoma de *Euaclyptus*. Com uma cobertura do 97% do genoma físico, estes marcadores demonstraram uma ampla e uniforme distribuição ao longo do genoma. Apenas 1,4 Mpb dos 85,4 Mpb das sequências de *scaffolds* ainda não ancorados no atual genoma de *Eucalyptus* foi capturado por 45 marcadores DArT geneticamente mapeados mas fisicamente não alinhados nos 11 pseudo-cromossomos de *Eucalyptus*, fornecendo evidência sobre a qualidade e a completude da montagem atual do genoma de *Eucalyptus*. A maioria das 89 sondas do microarranjo DArT que não mapearam fisicamente no genoma correspondem a sequências provavelmente ausentes em *E. grandis*, o que sugere um caráter pan-genômico do microarranjo DArT de *Eucalyptus*. Uma sobreposição significativa foi encontrada entre o posicionamento dos marcadores DArT e o espaço gênico, com 69%

das sondas DArT mapeando dentro de modelos gênicos preditos. Uma correlação significativa foi identificada entre o número de modelos gênicos preditos e o número total de sondas DArT encontradas em todo o genoma (índice de Spearman ρ = 0,682, ρ = 3,79e-18). O mapeamento de QTLs detectou 35 QTLs para as características avaliadas, sendo que vários deles sugestivos de sintenia em nível cromossômico com QTLs detectados em estudos anteriores. Um total de 8.036 modelos gênicos preditos foi observado nos intervalos genômicos compreendidos pelos marcadores flanqueantes aos 18 QTLs mapeados no parental materno e 6.678 nos intervalos dos 17 QTLs identificados no parental paterno. Centenas desses genes poderiam ser sugeridos tentativamente como genes candidatos para as características fenotípicas avaliadas, cuja validação demandaria um esforço considerável. Em conclusão, as propriedades genômicas dos marcadores DArT levantadas neste estudo são particularmente interessantes para os investigadores que trabalham com culturas as quais já contam com microarranjos DArT mas para as quais não existe ainda um genoma de referência que permita uma caracterização detalhada. Estas propriedades são potencialmente úteis para a realização de estudos filogenéticos, de genética de populações e predição de fenótipos via seleção genômica, explorando a proximidade destes marcadores a genes.

ABSTRACT

Diversity Arrays Technology (DArT) provides a robust, high throughput, cost-effective method to query thousands of sequence polymorphisms in a single assay. Despite the extensive use of this genotyping platform for numerous plant species, little is known regarding the sequence attributes and genome-wide distribution of DArT markers. We investigated the genomic properties of the 7,680 DArT marker probes of a *Eucalyptus* array, by sequencing them, constructing a high density linkage map and carrying out detailed physical mapping analyses to the *Eucalyptus grandis* reference genome sequence. The genetic map was used for QTL (Quantitative trait loci) mapping for seven complex growth and wood quality traits in *Eucalyptus*. Finally, for each QTL we preliminarily assessed the number of annotated gene models in the *Eucalyptus* genome found in its corresponding genomic interval delimited by flanking markers. A

consensus linkage map with 2,274 DArT markers anchored to 210 microsatellites and a framework map, with 1,029 markers ordered with high support, displayed extensive collinearity with the genome sequence. Only 1.4 Mbp of the 75 Mbp of still unplaced scaffold sequence was captured by 45 linkage mapped but physically unaligned markers to the 11 *Eucalyptus* pseudochromosomes, providing compelling evidence for the quality and completeness of the current *Eucalyptus* genome assembly. A highly significant correspondence was found between the locations of DArT markers and predicted gene models, while most of the 89 DArT probes unaligned to the genome correspond to sequences likely absent in *E. grandis*, consistent with the pan-genomic feature of this multi-*Eucalyptus* species DArT array. DArT markers preferentially target the gene space and display a largely homogeneous distribution across the genome, thereby providing superb coverage for mapping and genome-wide applications in breeding and diversity studies. QTL mapping detected 35 QTLs, several of them syntenic to QTLs in previous studies. A total of 8,036 annotated gene models were found within the genomic interval comprised by the 18 QTLs detected in the maternal parent and 6,678 in the 17 QTLs identified in the paternal parent. Hundreds of such predicted genes could be tentatively suggested as candidates involved in trait variation, whose testing and validation would require a gigantic effort. In conclusion, the comprehensive linkage-to-physical mapping analyses reported herein, provide novel data regarding the genomic attributes of DArT markers in plant genomes in general and for *Eucalyptus* in particular. These results should be valuable to other species for which DArT arrays are available but not yet reference genomes. Based on the genomic characterization of the DArT probe sequences reported, phylogenies, population genetic surveys and phenotype prediction by genomic selection can now be further explored according to the gene proximity or gene content of particular markers sets.

1. INTRODUÇÃO

O mapeamento genético permite a identificação de locos que controlam características quantitativas, chamados QTLs (Quantitative Trait Loci). A informação destes locos tem sido utilizada no melhoramento assistido de algumas espécies, bem como em estudos de sintenia, mapeamento comparativo e clonagem posicional de genes (Carneiro and Vieira 2002; Jones, Ougham et al. 2009). Em espécies de *Eucalyptus*, apesar do grande número de QTLs publicados para características complexas, existem limitações experimentais que dificultam a estimativa precisa da quantidade, posicionamento e magnitude dos efeitos destes locos e consequentemente sua utilização na prática do melhoramento. Entre essas limitações, tem destaque o tamanho pequeno das progênies empregadas, o que tende a superestimar a magnitude dos efeitos, o número reduzido de cruzamentos e ambientes utilizados nos experimentos e as limitações inerentes ao tipo de marcador molecular e, métodos estatísticos de detecção de QTLs (Grattapaglia, Plomion et al. 2009).

Métodos de genotipagem utilizados no mapeamento genético devem ser capazes de gerar milhares de marcadores cobrindo todo o genoma de uma forma rápida, robusta e de baixo custo. Novas tecnologias de sequenciamento e genotipagem vem sendo desenvolvidas para espécies de *Eucalyptus* (Novaes, Drost et al. 2008; Külheim, Yeoh et al. 2009; Grattapaglia, Silva-Junior et al. 2011; Neves, Mamani et al. 2011). Entretanto estas metodologias ainda tem um custo relativamente alto o que dificulta seu uso para a genotipagem em larga escala de grandes números de amostras. Alto rendimento, cobertura genômica alta transferibilidade de marcadores entre espécies são aspectos chave para aplicações na genética de *Eucalyptus*. Neste sentido, a tecnologia de genotipagem DArT (Diversity Arrays Technology) oferece marcadores moleculares com estas características (Jaccoud, Peng et al. 2001; Wenzl, Carling et al. 2004). DArT permite a identificação em paralelo de centenas a milhares de marcadores polimórficos em um experimento simples (Akbari, Wenzl et al. 2006) a custos por "*data point*" menores que os atuais custos de genotipagem via SNPs para um número

similar de marcadores informativos. Esta tecnologia tem demonstrado ser eficiente para a montagem de mapas genéticos densos, assim como a identificação de QTL's, a partir dos quais foram identificadas regiões cromossômicas ou genes associados à expressão de características quantitativas e qualitativas em plantas cultivadas como trigo, sorgo centeio e aveia, entre outras (Crossa, Burgueno et al. 2007; Pozniak, Knox et al. 2007; Lillemo, Asalf et al. 2008; Bariana, Bansal et al. 2010; Sadeque and Turner 2010; Zhang, Dong et al. 2011).

Um microarranjo de alta densidade com 7.680 marcadores DArT foi desenvolvido recentemente (Sansaloni, Petroli et al. 2010) e utilizado com sucesso para estudos filogenéticos (Steane, Nicolle et al. 2011), de mapeamento (Hudson, Freeman et al. 2012) e seleção genômica (Resende, Resende et al. 2012). Neste trabalho é apresentada uma caracterização detalhada dos 7.680 marcadores DArT que compõem este arranjo em relação a sua composição de sequência, distribuição no genoma e vizinhança a modelos gênicos preditos. Um mapa genético de alta densidade com estes marcadores foi construido permitindo assim estabelecer uma relação entre distância de recombinação e distância física em todo o genoma. Em seguida este mapa genético foi utilizado para o mapeamento de QTLs para sete características de crescimento e qualidade da madeira e uma análise preliminar realizada de genes preditos no genoma localizados nos intervalos genômicos compreendidos pelos QTLs.

2. REVISÃO BIBLIOGRÁFICA

2.1. O gênero *Eucalyptus*

Eucalyptus L'Hérit. é um gênero que compreende mais de 700 espécies, sendo de vital importância para a indústria florestal no mundo (Brooker 2000; Poke, Vaillancourt et al. 2005). Este gênero pertence à família *Myrtaceae*, sendo *Symphyomyrtus* seu principal subgênero, o qual está constituído por mais de 300 espécies. Dentre essas, *E.grandis*, *E. globulus*, *E. urophylla*, *E. camaldulensis*, *E. saligna* e *E. tereticornis*, são as mais utilizadas para fins comerciais (Eldridge, Davidson et al. 1994). Originário da

Austrália e regiões próximas, esta árvore foi introduzida no Brasil na segunda metade do Século XIX (Firmino-Winckler, Wilcken et al. 2009).

A importância econômica decorre do seu rápido crescimento, da sua capacidade produtiva, da sua adaptabilidade em diversos ambientes e facilidade de manejo por plantio direto e rebrota. Porem, a diversidade de espécies deste gênero é o que torna possível atender à demanda de grande parte dos segmentos que utilizam produtos florestais (Potts 2004).

Com relação à produção de celulose, entre as espécies silviculturalmente adaptadas às condições edafoclimáticas do Brasil, destacam-se *Eucalyptus grandis, E. saligna, E. globulus* e híbridos de *E. grandis x E. urophylla* (Ferreira and Santos 1997; Vencovsky and Ramalho 2000; Gonçalves, Rezende et al. 2001). *Eucalyptus dunnii* pode ser considerada uma opção adequada para a produção de celulose. A madeira apresenta maior densidade básica, menor conteúdo de lignina e maior conteúdo de pentosanas com relação a *E. saligna* e *E. grandis*, permitindo desta maneira, menor consumo de madeira e reagentes durante a cocção da polpa (Ferreira, Gonzaga et al. 1997). Contudo, é importante a introgressão de alelos de outras espécies visando ampliar a base genética e também possibilitar o melhoramento para diferentes caracteres. Este é o caso de *E. camaldulensis*, com maior densidade de madeira e maior resistência à seca, hibridizada com clones elite interespecíficos, resultando em ganhos de volume e qualidade de madeira (Bison, Ramalho et al. 2009). A espécie *E. nitens* é reconhecida por ser altamente resistente a geadas e, embora suscetível à ferrugem, exibe ampla variabilidade quanto à resistência entre procedências. Isto permite a seleção de genótipos resistentes superiores, e pode ser recomendada para incorporação em programas de hibridização visando a produção de clones superiores.

2.2. Atualidade e impacto econômico do *Eucalyptus* no Brasil

A cultura do *Eucalyptus* é de grande importância econômica, ambiental e social para o Brasil. Segundo a Associação Brasileira de Produtores de Florestas Plantadas (ABRAF 2012), atualmente existem aproximadamente 4,9 milhões de hectares

plantados com *Eucalyptus* no território nacional, com destaque para as regiões sudeste (54%), nordeste (16%) e centro-oeste (12%), sendo Minas Gerais (28,8%) e São Paulo (21%) os estados com maior superfície plantada. Esta área continua em expansão em todo o Brasil, com o intuito de suprir a demanda de madeira para produção de celulose e carvão vegetal para a indústria siderúrgica, além de outros derivados da madeira como materiais para a construção, móveis, papelão, óleos, e outros (Mora and Garcia 2000).

A utilização de florestas plantadas minimiza a pressão extrativista sobre as espécies autóctones, contribuindo assim diretamente para a conservação ambiental e de espécies nativas. Isto e a necessidade imperiosa de abastecer de madeira, tanto o mercado interno como o externo, são apontados como os fatores principais que levaram à busca de espécies de rápido crescimento e ao desenvolvimento de tecnologias apropriadas para responder às inquietudes das indústrias.

No ano 2011, as exportações brasileiras de produtos de florestas plantadas atingiram US$ 8,0 bilhões, isto significa 3,1% do total de produtos exportados pelo Brasil, um crescimento de 5,3% quando comparado com o ano 2010. A celulose é o produto de maior impacto na economia florestal, somente no ano 2011, foram exportados aproximadamente US$ 5,0 bilhões, apresentando um crescimento de 5,0% em relação ao ano anterior. O Instituto Brasileiro de Planejamento Tributário (IBPT), estimou que a contribuição tributária do setor florestal foi de R$ 7,6 bilhões no ano 2011, representando 0,51% do total arrecadado no país. Atualmente, o setor de florestas gera 4,7 milhões de empregos, incluindo empregos diretos (640,4 mil), indiretos (1,45 milhões) e empregos resultantes do efeito‑renda (2,60 milhões), o que demonstra a importância deste mercado como um grande suporte na criação de fontes de trabalho no Brasil (http://www.abraflor.org.br). Todos estes dados evidenciam a importância do *Eucalyptus* para a economia do país. Por tanto, existe uma necessidade clara de criar novos recursos e estratégias para atingir a máxima competitividade possível dentro do mercado mundial.

Produtividades florestais crescentes e refinamentos na qualidade dos produtos de madeira por meio de melhoramento genético tornar-se-ão cada vez mais estratégicos

para a indústria florestal, independentemente do uso final da madeira ser para energia, fibra, celulose ou produtos estruturais de madeira sólida. Ferramentas moleculares baseadas na identificação de polimorfismos no DNA, envolvidos no controle genético de fenótipos de interesse, prometem fornecer novas oportunidades para a seleção de características de crescimento, adaptabilidade a novas condições climáticas e propriedades da madeira de árvores cultivadas (Grattapaglia, Plomion et al. 2009).

2.3. Mapeamento genético

No ano 1910 e através de cruzamentos controlados de moscas (*Drosophila melanogaster*), Thomas Hunt Morgan e colaboradores observaram proporções fenotípicas não coincidentes com as propostas pela segunda lei de Mendel de "segregação independente dos genes". Sugeriu assim a presença de alguns genes agrupados num mesmo cromossomo, e a ocorrência ocasional de permutações nas quais existiria uma troca de segmentos cromossômicos entre homólogos. Segundo Morgan, essas variações nas proporções de segregantes, de algum modo, refletiam a distância linear entre dois genes em um mapa genético. Três anos depois, A.H. Sturtevant, interpretando dados oriundos da segregação de genes ligados, sugeriu o uso da porcentagem de recombinantes como indicador quantitativo da distância linear entre dois genes na construção de mapas genéticos (Coelho and Silva 2005).

Os mapas genéticos podem ser definidos como uma sequência de elementos genéticos ordenados de acordo com seus padrões de segregação. Os mapas mostravam que a posição dos genes correspondia à sua ordem linear nos cromossomos. Assim, o conceito de localização dos genes em uma ordem linear passou a ser incorporado à "Teoria Cromossômica da Herança", a qual foi criada no começo do século XX por W. S. Sutton e T. Boveri (Gardner and Snustad 1986). A transmissão, dos parentais para a progênie, dos marcadores moleculares que estão ligados no mesmo cromossomo é que determina a ordem dos marcadores ao longo do cromossomo. Em geral, quanto mais próximos os marcadores, menor será a possibilidade de ocorrer uma recombinação, logo, esses marcadores segregarão

juntos; ou quanto mais distantes estiverem os marcadores, maior a chance de ocorrer um *crossing-over* (Paterson 1996). A prova definitiva da associação entre mapas de ligação e cromossomos veio em 1931 com os estudos realizados por Creighton e McClintock em cromossomos de milho, demonstrando que o *crossing-over* é resultado de troca entre segmentos cromossômicos (Creighton and McClintock 1931).

Os primeiros mapas genéticos em plantas foram construídos com base em marcadores morfológicos e citológicos, principalmente, nas culturas de milho, tomate e ervilha. Características de herança discreta, cujas classes são facilmente distinguíveis, como nanismo, deficiência de clorofila, cor das pétalas e morfologia foliar foram usadas como marcadores fenotípicos. Entretanto, a disponibilidade de marcadores morfológicos é restrita.

Com o surgimento de marcadores isoenzimáticos, na década de 60, foi possível a construção de mapas para várias espécies. Sem embargo, o número de marcadores dificilmente passava de 30, o que não permitia uma ampla cobertura genômica. O surgimento de marcadores moleculares de DNA, na década de 80, permitiu a saturação de mapas já existentes e, até mesmo, a construção de mapas para espécies de plantas e animais para os quais os estudos de herança eram restritos (Carneiro and Vieira 2002; Coelho and Silva 2005; Pereira and Pereira 2006). Também, se tornaram possíveis a identificação, localização e medição da magnitude do efeito de genes envolvidos no controle de características monogênicas ou quantitativas (Ferreira and Grattapaglia 1998). O número de indivíduos genotipados estabelece o nível máximo de resolução que pode ser alcançado com um número ilimitado de marcadores no mapa genético. Assim frequências de recombinação abaixo de 10% apenas podem ser obtidas pela avaliação de mais de dez gametas passíveis de recombinação e, portanto, informativos (Coelho 2000).

A construção de mapas genéticos baseia-se na existência de desequilíbrio de ligação (DL), que é definido como desvios das frequências haplotípicas observadas, em relação às frequências esperadas sob a hipótese de independência dos locos considerados (Coelho and Silva 2002). Fatores como migração e seleção podem ser responsáveis pela segregação conjunta de segmentos genômicos, causando o DL em

consequência. Por isso, para a construção de mapas genéticos é necessário o desenvolvimento de populações de mapeamento. Nesse caso, o desequilíbrio de ligação é provocado, basicamente, pela ligação física entre os locos. Os indivíduos utilizados na formação dessas populações devem ser contrastantes para as características de interesse, além de possuírem uma distância genética suficiente para a identificação de marcadores polimórficos, possibilitando a construção do mapa. O tamanho das populações de mapeamento em plantas varia, de maneira geral, entre 50 e 250 indivíduos, mas populações maiores são necessárias para mapeamento de alta resolução (Mohan, Nair et al. 1997; Collard, Jahufer et al. 2005). Diferentes tipos de populações podem ser utilizados para a construção de mapas, sendo as mais comuns as obtidas por retrocruzamentos, populações F_2, Linhagens Puras Recombinantes (RILs – "Recombinant Inbred Lines"), linhagens duplo-haploides e, no caso de espécies de fecundação cruzada, cruzamentos entre indivíduos heterozigotos (Ferreira and Grattapaglia 1998; Coelho 2000; Collard, Jahufer et al. 2005).

O uso de mapeamento para estudos comparativos ou de sintenia colabora com o entendimento sobre a evolução dos genomas. Mapas genéticos são úteis para comparar as estruturas genômicas das diferentes espécies, observando a homologia dos genes, conservação da distância e a ordem de ligação nos cromossomos; também como um meio de obtenção de mapa único de referência (Carneiro and Vieira 2002).

2.4. Análise de QTLs (Quantitative Trait Loci)

Com o desenvolvimento das teorias de genética quantitativa (Fisher 1918), reconheceu-se que os caracteres complexos ou quantitativos são controlados pelos mesmos fatores hereditários descritos por Mendel. Vários desses fatores, cada um contribuindo com uma pequena parcela, estão envolvidos na expressão do caráter. Apesar da dificuldade de identificar os locos responsáveis pelas características quantitativas, foi Karl Sax (Sax 1923), quem encontrou uma associação direta entre o tamanho das sementes (característica complexa) e sua coloração (característica monogênica) em populações F2 de feijão. Este tipo de associação entre marcadores morfológicos e locos que controlam caracteres quantitativos foi reportado para vários

caracteres, mas as análises dependiam, exclusivamente, da suficiente quantidade de marcadores morfológicos que permita uma razoável cobertura do genoma.

Com o advento das diversas classes de marcadores moleculares, tornou-se possível a busca de regiões do genoma responsáveis por parte da variabilidade fenotípica em qualquer espécie de interesse (Tanksley 1993). O cálculo das frequências de recombinação e o ordenamento dos marcadores possibilitou a obtenção de estimativas mais acuradas da localização e do efeito dos QTLs em estratégias de mapeamento por intervalo (Lander and Botstein 1989; Haley and Knott 1992). Neste método estatístico, a informação dos genótipos de ambos os marcadores que flanqueiam um intervalo são tomadas simultaneamente. A análise estatística progrediu com a proposição do mapeamento por intervalo composto (Jansen 1993; Zeng 1993; Zeng 1994), em que marcadores sabidamente relacionados aos QTLs são incluídos como variáveis independentes no modelo de regressão múltipla, diminuindo a variância residual e, com isso, aumentando o poder do teste.

Estas análises permitem, em princípio, obter estimativas do número mínimo de locos envolvidos no controle dos caracteres, o posicionamento ou localização desses locos dentro do genoma, a quantificação das suas contribuições para a variação fenotípica e a avaliação dos seus efeitos em diferentes ambientes e genótipos. Entretanto, apesar de dezenas ou mesmo centenas de QTLs terem sido detectados em diversos estudos, a informação gerada não tem sido imediatamente útil para a seleção assistida no melhoramento de *Eucalyptus*. As razões para isso incluem a proporção limitada da variação explicada pelos QTLs e a magnitude superestimada dos seus efeitos, além do comportamento imprevisível da interação entre QTLs e diferentes *background* genéticos, diferentes locais e diferentes idades (Grattapaglia, Ribeiro et al. 2004; Grattapaglia and Kirst 2008; Grattapaglia, Plomion et al. 2009; Grattapaglia and Resende 2011).

2.5. Melhoramento genético de *Eucalyptus*

O melhoramento genético das espécies sobre as quais a indústria florestal vem centrando seus esforços nas últimas décadas, o qual tem colaborado de uma maneira considerável para o aumento da produtividade e qualidade dos produtos florestais. Várias são as características que o melhoramento florestal pretende atingir, entre elas, incremento no crescimento e na produtividade, alterações das propriedades químicas e físicas da madeira, resistência a doenças, tolerância a estresse abiótico, melhoria da capacidade fotossintética, dos caracteres fisiológicos, uso em biorremediação, produção de compostos farmacêuticos e alterações na arquitetura da árvore. Com esse fim, são utilizadas diversas estratégias de melhoramento florestal, tanto por meio do uso de técnicas clássicas, e, mais recentemente, com o uso da biotecnologia.

A possibilidade de prever ganhos é considerada uma das maiores contribuições para o melhoramento. Os materiais genéticos utilizados no setor florestal apresentam, de modo geral, grande variabilidade genética (Paula, Pires et al. 2002). Isto é de grande importância se consideramos que é nessa variabilidade genética que temos a base para tentar prever ganhos que possam melhorar a produtividade no setor. Com este objetivo é que são realizados diferentes programas de seleção de indivíduos superiores, no intuito de melhorar a produção de matéria prima para as diferentes indústrias. Porém, a seleção individual com altas intensidades é uma estratégia arriscada num programa de melhoramento genético, a redução no tamanho efetivo da população pode levar à endogamia e perda de vigor (Oda, Menck et al. 1989). Por este motivo, é de fundamental interesse utilizar métodos elaborados de melhoramento e critérios de seleção. De forma geral, um bom programa de melhoramento genético deve permitir a manutenção da variabilidade em longo prazo, tão grande quanto possível, sacrificando minimamente os resultados de curto prazo (Matheson 1990).

As espécies perenes possuem características que comprometem o sucesso de programas de melhoramento genético que dependam de técnicas clássicas. Existe uma série de obstáculos que dificultam a implementação de programas de melhoramento em espécies florestais (Strauss, Lande et al. 1992). Fatores como sobreposição de gerações, dificuldades de controle nos processos de polinização e fecundação, e complexidade na análise fenotípica dos descendentes. O grande período necessário para atingir o ciclo reprodutivo, a reprodução sexuada e assexuada, a expressão de

caracteres ao longo de várias idades, a necessidade de grandes áreas de plantio, dentre outros, também dificultam a execução de técnicas clássicas de melhoramento (Tzfira, Zuker et al. 1998; Resende 2001). A depressão endogâmica é também um fator importantíssimo, que torna impraticável a criação de linhagens endogâmicas. Além dos problemas técnico-práticos, é necessário considerar as questões de caráter financeiro. Em programas de melhoramento florestal, independentemente da espécie, os cruzamentos controlados são geralmente caros, consomem tempo e exigem pessoal treinado e área para a realização dos mesmos.

No Brasil, o melhoramento genético do *Eucalyptus* alcançou enorme sucesso e contribuiu para o expressivo aumento da produtividade de celulose e carvão. Entretanto, para se continuar obtendo resultados adicionais no melhoramento genético é preciso utilizar novas estratégias (Gonçalves, Rezende et al. 2001). A biotecnologia pode ser utilizada para obtenção de ganhos em produtividade, qualidade e sustentabilidade. Tais técnicas passarão, cada vez mais, a integrar as rotinas dos programas de melhoramento, acelerando o alcance de resultados (Golle, Reiniger et al. 2009). Entre as várias ferramentas da biotecnologia, metodologias de marcadores moleculares têm sido desenvolvidas para a análise da diversidade genética.

2.6. Desenvolvimento de marcadores moleculares em *Eucalyptus*

Um marcador genético representa a variação num sítio particular do genoma, o qual é herdado de maneira Mendeliana, é fácil de identificar e pode ser acompanhado através das gerações. Existem três tipos de marcadores que podem ser empregados: morfológicos (fenotípicos), bioquímicos (proteínas e isoenzimas) e moleculares (baseados no DNA). A decisão sobre qual sistema de marcadores é o mais apropriado para ser usado depende da espécie, do objetivo do trabalho e dos recursos disponíveis. Marcadores genéticos vêm sendo utilizados em *Eucalyptus* desde os anos 80, quando isoenzimas permitiram realizar os primeiros estudos de sistemas de cruzamento em pomares de sementes e produzir as primeiras versões de mapas genéticos de coníferas (Moran and Bell 1983), além de investigações sobre relacionamentos filogenéticos entre espécies próximas em angiospermas (Burgess and Bell 1983). Desde então, a

análise genética de Eucalyptus progrediu, essencialmente, em razão do desenvolvimento de novas técnicas moleculares, principalmente, pela característica que os marcadores moleculares apresentam de não sofrerem influência ambiental, ao contrário de marcadores morfológicos e citológicos.

Marcadores moleculares foram utilizados para responder questões evolutivas, descrever a estrutura genética de populações naturais e otimizar o gerenciamento e o avanço de populações de melhoramento de espécies de Eucalyptus. No final dos anos 80 e início dos anos 90, começou-se a utilizar marcadores moleculares gerados com base em fragmentos genômicos produzidos a partir do corte com enzimas de restrição denominados RFLP (*Restriction Fragment Length Polymorphism*) (Devey, Jermstad et al. 1991; Steane, West et al. 1992; Sale, Potts et al. 1993; Byrne, Murrell et al. 1995). Estes foram seguidos por métodos que utilizam a amplificação de fragmentos de DNA via PCR (*Polymerase Chain Reaction*). Surgiram assim os marcadores RAPD (*Random Amplified Polymorphism DNA*), produzidos pela amplificação aleatória de fragmentos polimórficos de DNA (Carlson, Tulsieram et al. 1991; Grattapaglia, Wilcox et al. 1991; Grattapaglia and Sederoff 1994; Nesbitt, B.M. Potts et al. 1995), os AFLP (*Amplified Fragment Length Polymorphism*) amplificando seletivamente um subgrupo de fragmentos genômicos gerados com enzimas de restrição (Vos, Hogers et al. 1995; Gaiotto, Bramucci et al. 1997; Marques, Araújo et al. 1998) e os microssatélites (Condit and Hubbell 1991; Byrne, Marquezgarcia et al. 1996; Brondani, Brondani et al. 1998), que são marcadores caracterizados pela amplificação de fragmentos de DNA de um a sete pares de bases repetidos em tandem e que frequentemente se encontram de forma abundante através do genoma de plantas.

No começo do século XXI, o grande destaque no estudo da diversidade genética em espécies florestais foi para o uso de polimorfismos de nucleotídeos únicos ou SNP (*Single Nucleotide Polymorphism*), vindos da análise de sequência (Brown, Gill et al. 2004; Gonzalez-Martinez, Ersoz et al. 2006; Novaes, Drost et al. 2008; Grattapaglia, Silva-Junior et al. 2011). Recentemente, a avaliação de polimorfismos de sequência, incluindo SNPs e indels tem sido feita em *Eucalyptus* através do método DArT (Diversity Arrays Technology) (Sansaloni, Petroli et al. 2010). Esta técnica é baseada na

hibridização de representações genômicas de complexidade reduzida dispostas em microarranjos (Jaccoud, Peng et al. 2001).

2.7. Estudos de genética de populações e evolução em *Eucalyptus*

Um variado número de marcadores moleculares vem sendo utilizado para responder diversas questões evolutivas em diferentes níveis taxonômicos no gênero *Eucalyptus*. Uma revisão recente detalhou as várias aplicações (Grattapaglia, Vaillancourt et al. 2012). Investigações foram conduzidas sobre o relacionamento entre subgêneros (Steane, Nicolle et al. 2002; Whittock 2003) ou espécies (Nesbitt, B.M. Potts et al. 1995; Butcher, Otero et al. 2002; Elliott and Byrne 2004; McKinnon, Vaillancourt et al. 2008), em estudos filogeográficos (Jackson, Steane et al. 1999; Byrne and Hines 2004) e para a descrição da estrutura genética de populações naturais (Steane, Conod et al. 2006; Butcher, McDonald et al. 2009). Marcadores RAPD e microssatélites foram utilizados para discriminar genótipos individuais (Keil and Griffin 1994; Rocha, Abad et al. 2002), estes últimos têm sido também empregados em testes de parentesco (Kirst, Cordeiro et al. 2005; Jones, Shepherd et al. 2008). Análises de variabilidade genética em pomares de sementes foram realizadas através de informações obtidas com a combinação de marcadores AFLP e microssatélites (Marcucci Poltri, Zelener et al. 2003). Métodos de sequenciamento de DNA utilizados para genotipagem de SNPs têm auxiliado na identificação do relacionamento filogenético existente entre *E. nitens* e *E. globulus* (Külheim, Yeoh et al. 2009) e, recentemente, marcadores DArT foram utilizados para diferenciar espécies, identificar híbridos interespecíficos e resolver disjunções biogeográficas entre espécies (Steane, Nicolle et al. 2011). Este último trabalho foi realizado com 94 espécies e os resultados revelaram um poder de resolução muito elevado.

2.8. Desenvolvimento de mapas genéticos em *Eucalyptus*

Mapas de ligação são ferramentas úteis para os melhoristas que desejam incluir o uso de marcadores moleculares em programas de melhoramento (Staub, Serquen et al. 1996). Mapas genéticos foram desenvolvidos para diversas espécies florestais com o objetivo de localizar QTLs e prover as potenciais bases para a Seleção Assistida por Marcadores (*Marker Associated Selected - MAS*), em *Pinus* (Groover, Devey et al. 1994), *Populus* (Bradshaw and Stettler 1995) e *Eucalyptus* (Grattapaglia and Sederoff 1994; Byrne, Murrell et al. 1997).

A principal estratégia utilizada para produzir os primeiros mapas de ligação em *Eucalyptus* foi a de pseudo-cruzamento teste (Grattapaglia and Sederoff 1994), onde os marcadores segregantes são analisados separadamente para cada parental. Este método inclui o cruzamento de indivíduos heterozigotos e a geração de híbridos que, frequentemente, demonstram uma combinação das características favoráveis das espécies parentais. Dentro do subgênero, a hibridização tem mostrado ser um fenômeno relativamente comum entre espécies, embora isto não ocorra entre subgêneros distintos (Griffin, Burgess et al. 1988).

Os primeiros mapas de *Eucalyptus* foram construídos com centenas de marcadores dominantes, como RAPD (Grattapaglia and Sederoff 1994; Verhaegen and Plomion 1996; Bundock, Hayden et al. 2000) e AFLP (Marques, Araújo et al. 1998; Remington, Whetten et al. 1999). Marcadores mais informativos e transferíveis entre as diferentes espécies foram utilizados em seguida. Assim, os RFLP e microssatélites foram utilizados no desenvolvimento de mapas genéticos, com densidade equivalente porém maior conteúdo informativo e transferibilidade entre *pedigrees* (Byrne, Murrell et al. 1995; Bundock, Hayden et al. 2000; Gion, Rech et al. 2000; Brondani, Brondani et al. 2002; Thamarus, Groom et al. 2002; Brondani, Williams et al. 2006). Mais recentemente, um painel de marcadores SNPs combinado com microssatélites foi utilizado na construção de um mapa genético consenso incluindo dois pedigrees não relacionados (Lima, Silva-Junior et al. 2011) e mapas genéticos com alguns milhares de marcadores DArT foram construídos (Hudson, Kullan et al. 2011; Hudson, Freeman et al. 2012), incluindo mapas derivados da presente tese (Petroli, Sansaloni et al. 2012).

2.9. Identificação de QTLs em *Eucalyptus*

O mapeamento de QTLs em *Eucalyptus* tem encontrado, invariavelmente, locos de maior efeito para todas as características consideradas, apesar da precisão experimental limitada, a falta de um pré-desenho dos *pedigrees* que maximize a segregação fenotípica, e as populações de mapeamento limitadas. O sucesso na detecção de QTLs de maior efeito pode ser explicado pela natureza selvagem e a ampla heterogeneidade genética apresentada pelo gênero (Grattapaglia and Kirst 2008). QTLs para características juvenis, como a altura das plântulas, área foliar e tolerância de mudas a geadas foram mapeados em *E. nitens* (Byrne, Murrell et al. 1997; Byrne, Murrell et al. 1997). Também têm sido revelados QTLs que regulam elementos relacionados com habilidade de propagação vegetativa em *E. tereticornis* e *E. globulus* (Marques, Vasques-Kool et al. 1999), assim como em híbridos *E. grandis* x *E. urophylla* (Grattapaglia, Bertolucci et al. 1995) e *E. tereticornis* x *E. globulus* (Marques, Carocha et al. 2005). Foram identificados QTLs para florescimento precoce em híbridos *E. urophylla* x *E. grandis* e em *E. globulus* (Missiaggia 2005; Bundock, Potts et al. 2008). QTLs para resistência a insetos foram mapeados numa população híbrida (Shepherd, Chaparro et al. 1999), e para a produção de óleos essenciais e terpenos em uma população composta unicamente por indivíduos *E. nitens* (Henery, Moran et al. 2007). QTLs de efeito maior para resistência a ferrugem provocado por *Puccinia psidii* foram mapeados e validados em diversos pedigrees de *E. grandis* (Junghans, Alfenas et al. 2003; Mamani, Bueno et al. 2010; Alves, Rosado et al. 2011) e dois QTLs de resistência a *Mycosphaerella* em *E. globulus*, (Freeman, O'Reilly-Wapstra et al. 2008), evidenciando um controle possivelmente mais simples da resistência a estas doenças provocadas por fungos em *Eucalyptus*.

QTLs para características de crescimento e qualidade da madeira foram mapeados em diversos trabalhos em diferentes espécies de *Eucalyptus*. Os primeiros QTLs foram publicados para crescimento em volume e densidade básica da madeira em *E. grandis* (Grattapaglia, Bertolucci et al. 1996) e *E. grandis* x *E. urophylla* (Verhaegen, Plomion et al. 1997). Em seguida QTLs foram descritos para *E. globulus* (Moran, Thamarus et al. 2002; Thamarus, Groom et al. 2004; Freeman, Whittock et al. 2009) e *E. nitens*(Thumma, Baltunis et al. 2010). QTLs e genes candidatos para caracteres de

propriedade da madeira foram mapeados em grupos de ligação (GL) equivalentes. Isto foi observado inclusive em espécies diferentes, com número diferente de QTLs, mas em muitas oportunidades, conservando a posição (Bundock, Potts et al. 2008; Freeman, Whittock et al. 2009; Thumma, Baltunis et al. 2010; Gion, Carouche et al. 2011). O agrupamento de QTLs para diferentes características na mesma posição pode refletir as altas correlações fenotípicas observadas entre as características sugerindo a existência de genes pleiotrópicos. Por exemplo, a concentração de celulose foi positivamente correlacionada com o rendimento de polpa e negativamente com extrativos, da mesma forma como a produção de boa qualidade da polpa está ligada a altos níveis de celulose e baixos de extrativos e lignina (Thumma, Baltunis et al. 2010). Existe também uma relação positiva entre o ângulo da microfibrila e a concentração de lignina (Gindl and Teischinger 2002; Hori, Suzuki et al. 2003; Gindl, Gupta et al. 2004; Thumma, Baltunis et al. 2010), fato acompanhado por um agrupamento dos QTLs que determinam estes caracteres. Em 2005 um estudo de genética de associação demonstrou a existência de polimorfismos (25 SNPs) em um gene envolvido na via de biossíntese da lignina (CCR - *Cinnamoyl CoA reductase*) afetando o desenvolvimento de microfibrilas em uma população de *E. nitens* (Thumma, Nolan et al. 2005). Cinco anos depois, o mesmo gene foi co-localizado também com um QTL que determina o ângulo da microfibrila no grupo de ligação 10 numa população da mesma espécie (Thumma, Baltunis et al. 2010).

Comparações diretas do número e dos efeitos de QTLs entre diferentes estudos são complicados, pois diferentes estudos utilizam diferentes desenhos experimentais, marcadores e técnicas de análise (Freeman, Whittock et al. 2009). Mesmo utilizando marcadores altamente transferíveis entre espécies, devido aos poucos marcadores em comum entre os mapas de ligação publicados para *Eucalyptus*, na maioria dos casos, a resolução é insuficiente para determinar a homologia de QTLs entre diferentes estudos além do nível de grupo de ligação. Diferenças na posição do QTL e do gene candidato, poderiam sugerir a existência de genes ainda não mapeados e com efeitos significativos sobre a característica. Estudos de QTLs são limitados pelo fato de tipicamente envolverem apenas uma ou poucas famílias segregantes e um número restrito de marcadores. Como resultado, somente um pequeno número de QTLs pode

ser detectado, e, além disso, seus efeitos são fortemente superestimados devido ao conhecido efeito "Beavis" (Beavis 1998; Xu 2003).

A análise de QTLs em *Eucalyptus* tem sido conduzida com base em *pedigrees* simples utilizando mapas de ligação. Um dos principais obstáculos para o uso de dados de QTLs no melhoramento é a falta de representação de multiplos *"backgrounds"* genéticos. Um número crescente de experimentos de mapeamento de QTLs em *Eucalyptus* vem sendo publicado utilizando múltiplas famílias. Invariavelmente tem sido encontrada uma ampla interação entre QTLs em diferentes *backgrounds* genéticos com alguns QTLs mais estáveis, ou seja, detectados em diversas famílias, e outros mais específicos (Thamarus, Groom et al. 2004; Freeman, Potts et al. 2013). Embora existam QTLs sintenicos e colineares entre diferentes pedigrees de *Eucalyptus*, a especificidade de muitos deles para famílias e sítios específicos reforça a importância de considerar que os efeitos próprios da configuração genética e do sítio poderiam complicar a aplicação de seleção assistida por marcadores.

QTLs mais estáveis em diferentes *"backgrounds"* genéticos e em uma variedade de ambientes poderão, entretanto, ser os principais alvos de técnicas de mapeamento de alta resolução e potencial clonagem posicional mediante o uso do genoma de *Eucalyptus grandis,* recentemente sequenciado (http://www.phytozome.net/cgi-bin/gbrowse/*Eucalyptus*/). Para isso, entretanto, é necessária uma redução considerável do tamanho dos intervalos de recombinação associados aos QTLs para regiões relativamente pequenas do genoma da ordem de apenas algumas centenas de milhares de pares de bases. Com este objetivo, deverão ser desenvolvidos mapas de ligação de altíssima densidade com milhares de marcadores genotipados em progênies segregantes de milhares de indivíduos, de forma a amostrar um grande número de eventos de recombinação. Além disso, a fenotipagem desses indivíduos deverá ter alta precisão de forma a permitir uma categorização clara nas diferentes classes fenotípicas. Esta abordagem tem sido possível para QTLs de alta penetrância, em geral mapeados para características de alta herdabilidade em espécies modelo como *Arabidopsis* ou em culturas altamente domesticadas como arroz, trigo e tomate para as quais são disponíveis populações de linhagens puras recombinantes e/ou linhagens quase isogênicas (Sugimoto, Takeuchi et al. 2010; Bailey, Cevik et al. 2011). Em

Eucalyptus, entretanto, estas condições não foram satisfeitas, o que torna a clonagem posicional de QTLs um desafio técnico enorme.

2.10. Marcadores DArT (Diversity Arrays Technology)

Apesar de todos os desenvolvimentos recentes, existe uma necessidade de tecnologias que permitam genotipar polimorfismos de sequência em paralelo e em larga escala, ou seja, grandes números de amostras a custos muito reduzidos por genótipo e com elevada robustez analítica. No caso de plantas cultivadas, tecnologias que consolidem todos estes atributos teriam diversas aplicações imediatas e permitiriam a efetiva integração de tecnologias genômicas em procedimentos operacionais de melhoramento genético, estudos de diversidade e distância genética, introgressão de alta precisão de segmentos genômicos, mapeamento genético de alta resolução e genética de associação.

A metodologia DArT (Diversity Arrays Technology) foi desenvolvida para atender várias das limitações de outros métodos. Descrita dez anos atrás (Jaccoud, Peng et al. 2001), apresenta uma série de vantagens que complementam com eficiência as metodologias de análise de polimorfismo ora em utilização, tais como microssatélites, AFLP e SNP. A metodologia DArT envolve tres etapas: 1) *Construção de uma biblioteca* (representação genômica), a qual reune o DNA genômico total de um grupo de diversos indivíduos que representem o germoplasma de interesse. Este DNA é submetido a um processo de redução da complexidade por meio de corte com enzimas de restrição, as quais reconhecem e cortam preferencialmente em regiões hipometiladas do DNA. As representações genômicas obtidas a partir do pool de DNA são clonadas, criando bibliotecas de insertos individuais, os quais são imobilizados sobre um microarranjo geralmente denominado de "descoberta", pois tem por objetivo permitir a seleção de fragmentos que revelem polimorfismo de sequência. 2) G*enotipagem das amostras*, na qual o DNA das amostras individuais de estudo ou *targets* e suas replicas (30%) é cortado com a mesma combinaçao de enzimas de restrição, com o objetivo de gerar a mesma redução de complexidade utilizada anteriormente para o desenvolvimento das bibliotecas, posteriormente os fragmentos

obtidos são amplificados via PCR. As replicadas são geradas para poder estabelecer o nível de reprodutibilidade do experimento por meio da homologia dos dados. Os fragmentos são marcados com fluorescência, hibridizados sobre o arranjo de descoberta e submetidos num processo de lavagem para, finalmente, serem escaneados para a detecção dos sinais fluorescentes (Figura 1). 3) *Análise de dados*, os fragmentos polimórficos detectados mostram sinais de hibridização variáveis entre diferentes indivíduos. Quando a relação de sinal *target*/referência é similar em todos os slides, os fragmentos são considerados monomórficos, enquanto que se dois clusters (alelos) são distinguidos e a variância da intensidade relativa entre eles é de pelo menos 80% da variância total, os clones são considerados polimórficos e genotipados de forma binaria como "0" ou "1". A capacidade do software de posicionar o sinal em um dos clusters (0 ou 1) é medida pela qualidade do marcador (Q) (Figura 1). Quando a diferencia da intensidade do sinal não é suficiente para poder posicionar os clones dentro de um dos dois cluster, estes são considerados como dados faltantes. A porcentagem de genótipos que são chamados de "0" ou "1" para cada clone em todos os *targets* é representado pelo parâmetro Call Rate (Sansaloni et al., 2010). Estas etapas da técnica estão padronizadas, embora possam precisar de pequenas modificações dependendo da espécie alvo.

O método baseia-se na hibridização de DNA com sondas relativamente longas (300-500 pb) derivadas de regiões de baixo número de cópias. A metodologia DArT, normalmente, fornece sinal consistente mesmo entre espécies relacionadas, um recurso especialmente valioso em *Eucalyptus*. Os clones de DNA que compõem o arranjo podem ser sequenciados e as sequências compartilhadas e usadas como marcadores âncoras robustos para mapeamentos de QTL comparativos ou para explorar o genoma referência em projetos de clonagem posicional. DArT fornece uma plataforma de genotipagem padronizada de alto rendimento através de milhares de marcadores, milhares de amostras podem ser facilmente testadas em paralelo.

A tecnologia DArT vem sendo amplamente utilizada nos últimos anos em mais de 60 espécies de plantas, incluindo espécies florestais como *Eucalyptus* (Sansaloni, Petroli et al. 2010) e *Pinus* (Alves-Freitas, Kilian et al. 2010). Devido aos diferentes aperfeiçoamentos na robustez e reprodutibilidade, esta técnica tem se revelado

altamente interessante para diversas aplicações, desde a análise de variabilidade genética em populações, mapeamento genético de alta densidade e em apoio a projetos de montagem de genomas. Entretanto, a maior vantagem desta classe de marcadores é a rapidez de genotipagem e o custo reduzido por genótipo (data point), o que torna esta tecnologia útil para a efetiva aplicação em procedimentos operacionais de seleção assistida por marcadores.

O primeiro microarranjo de genotipagem DArT para *Eucalyptus* com 7.680 marcadores foi desenvolvido, demonstrando seu potencial para a análise da diversidade e estudos de mapeamento de ligação em espécies do gênero (Sansaloni, Petroli et al. 2010). Recentemente, o microarranjo DArT para *Eucalyptus* tem demonstrado ser muito útil na diferenciação de espécies, identificação de híbridos interespecíficos e resolução de disjunções biogeográficas entre espécies (Steane, Nicolle et al. 2011). Este mesmo microarranjo já foi utilizado na construção de alguns mapas genéticos, demonstrando um alto grau de sintenia e colinearidade entre populações de *Eucalyptus* (Hudson, Kullan et al. 2011; Kullan, van Dyk et al. 2012).

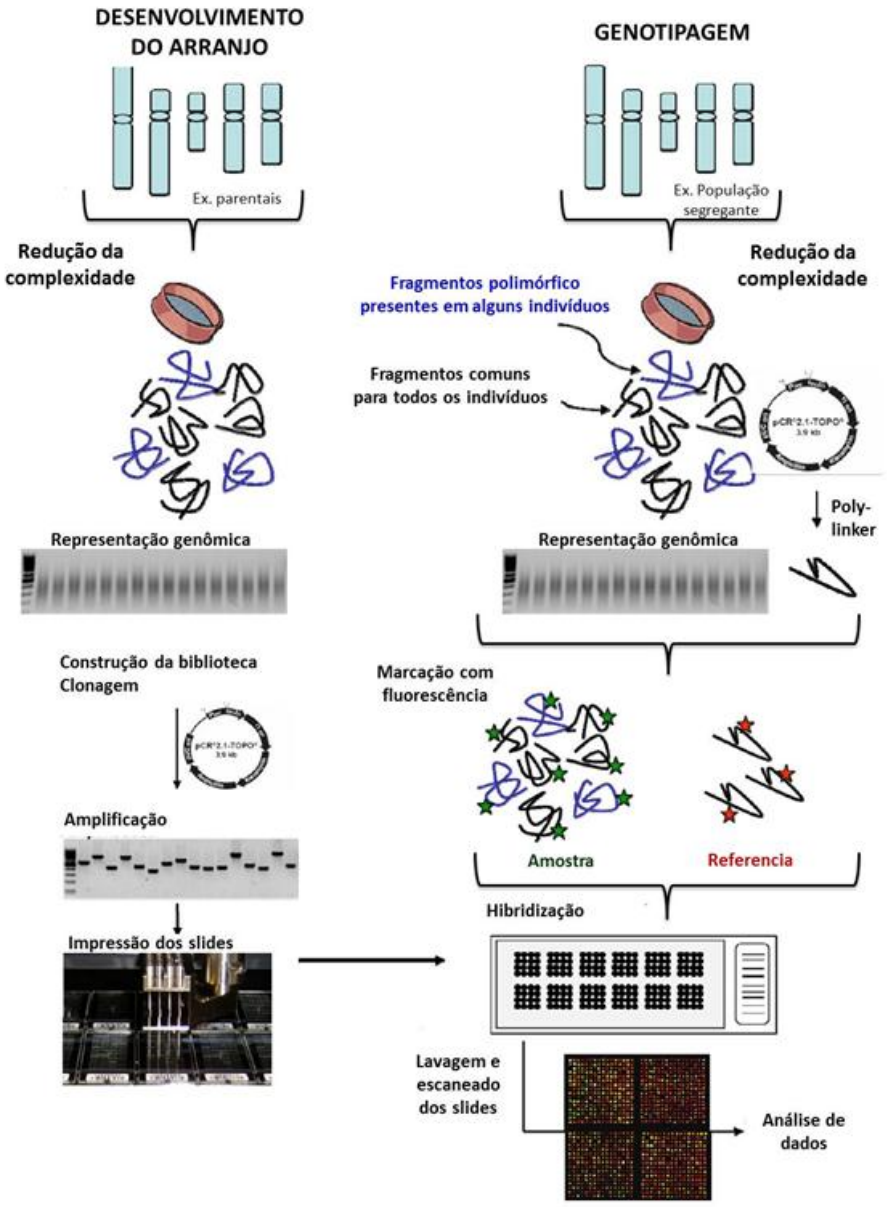


Figura 1: Procedimento de desenvolvimento de marcadores DArT. Na primeira etapa é construída a biblioteca genômica posteriormente impressa no microarranjo. Na segunda etapa, a progênie de uma população segregante ou indivíduos para estudos de diversidade são genotipados por meio da hibridação da representação genômica (i.e. amostra tecnicamente denominada "target" na metodologia DArT) sobre os slides (i.e. microarranjo). Software específico identifica clones polimórficos que são marcados como presente (1) ou ausente (0) na representação genômica.

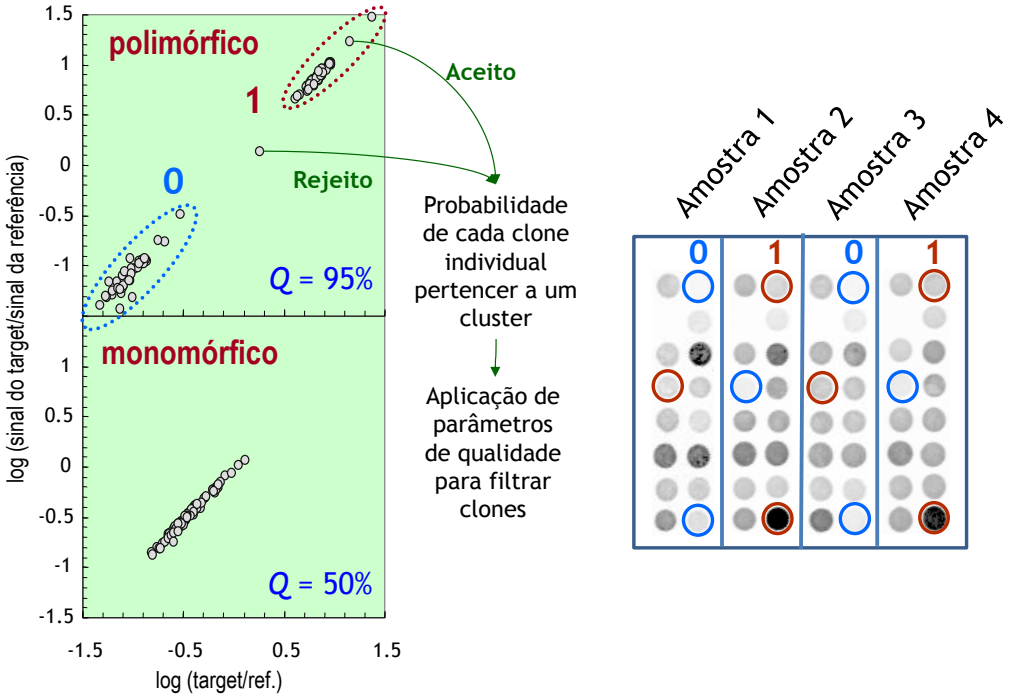


Figura 2. Interpretação de dados calculado pelo programa *DArTSoft* mostrando a diferença de intensidade entre genótipos. À esquerda, os gráficos mostram o log (sinal do *taget*/sinal da referência), quanto maior é a diferença de intensidade entre clusters (Q=95%), maior é o polimorfismo dos marcadores. Se a intensidade é homogênea (Q<50%), os marcadores são monomórficos. À direita está exemplificada a diferença de intensidades entre amostras representadas como ausência (0) e presença (1).

3. CAPÍTULO 1

CARACTERIZAÇÃO GENÔMICA DE MARCADORES DArT BASEADA NA ANÁLISE DE MAPEAMENTO GENÉTICO E FÍSICO NO GENOMA DE *Eucalyptus*

3.1. INTRODUÇÃO

Tecnologias de marcadores de DNA para genotipagem de alto desempenho e custos acessíveis, tornaram-se indispensáveis na caixa de ferramentas do geneticista

de plantas. Uma grande variedade de métodos para detectar polimorfismos de sequência de DNA entre plantas individuais foi desenvolvida e usada amplamente nos últimos 25 anos. Apesar da hibridação baseada em DNA ter inaugurado esta jornada com marcadores RFLP (Tanksley, Young et al. 1989), os métodos baseados em PCR (Williams, Kubelik et al. 1990; Vos, Hogers et al. 1995) foram os responsáveis pela remoção das barreiras que impediam o acesso à análise genômica de plantas para um grande número de espécies, incluindo culturas órfãs e muitas espécies florestais. A maioria dos métodos de marcadores moleculares baseados em PCR, no entanto, possuem baixo desempenho, e, por conseguinte, são muito demorados e economicamente de alto custo, para aplicações que requerem a genotipagem de milhares de amostras, para centenas de marcadores dentro de orçamentos modestos. Apesar de plataformas de SNP terem sido desenvolvidas para um número crescente de espécies vegetais (Ganal, Altmann et al. 2009), elas ainda permanecem em grande parte limitadas para as principais culturas e, seus custos por amostra, são inacessíveis para a maioria dos programas de melhoramento e de conservação de germoplasma de plantas.

Diversity Arrays Technology (DArT), foi descrito há mais de uma década (Jaccoud, Peng et al. 2001) e tem experimentado um interesse crescente nos últimos anos como um método robusto, de alta transferibilidade, capaz de detectar milhares de polimorfismos de presença/ausência que cobrem todo o genoma em um único experimento e a custos muito reduzidos. Embora exista uma propriedade intelectual, esta técnica é livremente licenciada sob um modelo "*open-source*" (Kilian 2009), uma condição que tem estimulado o desenvolvimento de microarranjos de genotipagem para mais de 60 organismos, incluindo muitas culturas menos privilegiadas (Wenzl, Carling et al. 2004; Xia, Peng et al. 2005; James, Schneider et al. 2008; Mantovani, Maccaferri et al. 2008; Bolibok-Bragoszewska, Heller-Uszynska et al. 2009; Tinker, Kilian et al. 2009; Hippolyte, Bakry et al. 2010; Bartos, Sandve et al. 2011; Howard, Whittock et al. 2011; Milczarski, Bolibok-Bragoszewska et al. 2011; Reddy, Rong et al. 2011; Supriya, Senthilvel et al. 2011; Yang, Saxena et al. 2011; Belaj, Dominguez-Garcia et al. 2012; Van Schalkwyk, Wenzl et al. 2012). DArT envolve o isolamento e clonagem de um conjunto aleatório de fragmentos de DNA produzidos a partir da redução de

complexidade genômica de uma, ou um conjunto de amostras de DNA, de vários acessos de germoplasma, de modo que uma coleção representativa de sequências genômicas variáveis de uma ou mais espécies-alvo é capturada. Vários milhares destes clones de DNA são dispostos sobre uma lâmina de vidro e hibridizados com o produto de PCR de uma amostra submetida ao mesmo processo de redução da complexidade genômica que a biblioteca de clones. Sendo um método baseado na hibridação DNA-DNA e utilizando sondas relativamente longas (~300-500 pb), DArT fornece um sinal alto e consistente, mesmo para espécies relacionadas (Steane, Nicolle et al. 2011).

Apesar da ampla utilização desta plataforma de genotipagem para muitas espécies de plantas, pouco se sabe sobre os atributos das sondas genômicas do microarranjo DArT que geram os vários milhares de marcadores genotipados. Com exceção de um estudo em aveia, e as recentes pesquisas em pequena escala de algumas centenas de sequências das sondas DArT em tomate (Van Schalkwyk, Wenzl et al. 2012) e maçã (Schouten, Weg et al. 2012), microarranjos DArT completos ainda não foram examinados em nível de sequência para uma melhor compreensão da redundância, cobertura do genoma e conteúdo gênico. Além disso, não há informação disponível sobre a distribuição de marcadores DArT ao longo do genoma, principalmente pela ausência de um genoma de referência disponível para a maioria das espécies para as quais esta tecnologia tem sido utilizada.

Recentemente foi desenvolvido um microarranjo de genotipagem DArT de alta densidade com 7.680 sondas selecionadas a partir de uma ampla representação de 64 espécies de *Eucalyptus* (Sansaloni, Petroli et al. 2010). Uma característica particularmente notável desta ferramenta de genotipagem baseada em hibridização tem sido a sua transferibilidade entre as diferentes espécies do gênero, um atributo dificilmente oferecido por microssatélites ou SNPs (Grattapaglia, Silva-Junior et al. 2011). DArT tem proporcionado uma plataforma padronizada de genotipagem de alto desempenho, em que milhares de marcadores podem ser facilmente testados em paralelo para milhares de amostras em todas as espécies de *Eucalyptus*. Este microarranjo DArT tem demonstrado um excelente desempenho para complexas análises filogenéticas e de diversidade (Steane, Nicolle et al. 2011), seleção genômica (Resende, Resende et al. 2012) e mapeamento genético (Sansaloni, Petroli et al. 2010;

Hudson, Kullan et al. 2011; Kullan, van Dyk et al. 2012). Uma descrição detalhada sobre o conteúdo de sequência e a distribuição no genoma das sondas de DNA que compõem este microarranjo DArT deverá incrementar o seu valor para estudos comparativos de mapeamento de QTLs, navegar a partir dos mapas de ligação para o genoma de referência, em projetos de clonagem posicional e para extrair informação genômica adicional a partir de marcadores informativos identificados em estudos filogenéticos, de genética populacional e seleção genômica.

Mapas genéticos têm sido ferramentas fundamentais para a análise da herança de características qualitativas e quantitativas, para o mapeamento comparativo, para montagem de genomas inteiros e para aplicações de melhoramento molecular, incluindo análises de germoplasma, seleção assistida por marcadores e clonagem baseada em mapa (Jones, Ougham et al. 2009). Mapas de ligação genética para espécies de *Eucalyptus* foram relatados a partir de diferentes *pedigrees*, tanto intra como interespecíficos, utilizando diferentes tecnologias de marcadores moleculares (Grattapaglia and Kirst 2008; Grattapaglia, Vaillancourt et al. 2012). Um bom número de mapas de ligação genética tem sido produzido em *Eucalyptus* com tecnologias de marcadores dominantes RAPD e AFLP (Grattapaglia and Sederoff 1994; Verhaegen and Plomion 1996; Marques, Araújo et al. 1998; Bundock, Hayden et al. 2000; Myburg, Griffin et al. 2003; Freeman, Potts et al. 2006), microssatélites (Brondani, Brondani et al. 2002; Brondani, Williams et al. 2006) e SNPs (Lima, Silva-Junior et al. 2011), enquanto RFLPs (Byrne, Murrell et al. 1995; Thamarus, Groom et al. 2002) e um recente microarranjo de genotipagem SFP (Single Feature Polymorphisms) (Neves, Mamani et al. 2011) permitiram posicionar centenas de genes em mapas existentes. Apesar de todos esses avanços, essas tecnologias de marcadores moleculares não foram eficientes em fornecer uma ferramenta amplamente aplicável que possa ser usada para ligar os genótipos aos fenótipos de uma forma extensa, que inclua o mapeamento comparativo, a descoberta de genes e que auxiliasse aos programas de melhoramento. Recentemente, marcadores DArT têm fornecido um mapeamento de ampla cobertura e alta densidade, necessário para caminhar nessa direção (Hudson, Kullan et al. 2011; Kullan, van Dyk et al. 2012), embora não se tenha realizado uma caracterização mais profunda de seu conteúdo genômico.

Neste estudo, investigamos as propriedades genômicas das 7.680 sondas DArT que integram o microarranjo de *Eucalyptus* por meio de seu sequenciamento, construindo um mapa de ligação de alta densidade e realizando uma análise detalhada do mapeamento físico destes marcadores utilizando-se a sequência do genoma referência de *Eucalyptus grandis* (www.phytozome.net). Neste capítulo, estávamos particularmente interessados em verificar o desempenho dos marcadores DArT para mapeamento de ligação em uma família interespecífica de *Eucalyptus*. Também caracterizamos a composição de sequência das sondas do microarranjo DarT, para avaliar a sua distribuição física em termos de cobertura global do genoma e distância em relação aos modelos gênicos preditos. Por último, examinamos a consistência entre o ordenamento de locos físico *versus* o baseado em recombinação.

3.2. OBJETIVOS

1) Construir um mapa genético para um população de referência *Eucalyptus grandis x Eucalyptus urophylla* e verificar o desempenho dos marcadores DArT para mapeamento de ligação;

2) Caracterizar a composição de sequência das sondas do microarranjo DArT em relação à redundância e o seu possível conteúdo gênico;

3) Avaliar a distribuição física dos marcadores, em termos de cobertura do genoma e distância em relação a modelos gênicos preditos;

4) Alinhar o mapa genético aos pseudocromossomos correspondentes para examinar a consistência entre o ordenamento dos locos em termos físicos *versus* genético com base em recombinação;

5) Fornecer estimativas para cada cromossomo, e para o genoma como um todo, da relação entre distância física e distância de recombinação no genoma de *Eucalyptus*.

3.3. MATERIAL E MÉTODOS

3.3.1. Material Vegetal

Uma população de mapeamento F1 de 177 indivíduos foi derivada de um cruzamento interespecífico entre duas árvores elite, *E. grandis* (clone G38) e *E. urophylla* (clone U15). Ambas as espécies são amplamente plantadas nos trópicos e pertencem ao mesmo subgênero *Symphyomyrtus*. Esta população de mapeamento, chamada GxU-IP foi selecionada como um *pedigree* de referência para fins de mapeamento no projeto Genolyptus (Grattapaglia 2004), imortalizada por propagação via mini-estaquia e plantada em ensaio repetido em cinco locais em Julho de 2003 em blocos casualizados com parcelas de árvore única e cinco repetições por local. O DNA foi extraído de ambos os pais e todos os indivíduos F1, utilizando-se 150mg de tecido foliar armazenado a -20°C como descrito anteriormente (Grattapaglia and Sederoff 1994). As amostras de DNA resultantes foram de qualidade consistente e adequada para genotipagem via DArT e microssatélites.

3.3.2. Genotipagem de microssatélites

Uma triagem de 300 marcadores microssatélites EMBRA (Brondani, Williams et al. 2006; Faria, Mamani et al. 2010; Faria, Mamani et al. 2011) foi utilizada na identificação de polimorfismos entre os parentais, com uma análise adicional de seis indivíduos da progênie F1 para verificar a segregação, resultando na seleção de 222 microssatélites informativos. A genotipagem dos microssatélites foi realizada em sistemas multiplex, com detecção de multi-fluorescência em um sequenciador automático ABI 3100XL, como descrito anteriormente (Brondani and Grattapaglia 2001; Faria, Mamani et al. 2011).

3.3.3. Genotipagem de marcadores DArT (Diversity Arrays Technology)

Uma descrição detalhada do método usado para preparar o arranjo de alta densidade DarT, para *Eucalyptus*, foi descrita recentemente (Sansaloni, Petroli et al. 2010). Resumidamente, foram construídas 18 bibliotecas de representações genômicas, com um total de 23.808 sondas de DNA provenientes de 64 espécies diferentes de *Eucalyptus*, utilizando-se o método de redução da complexidade genômica *Pst*I/*Taq*I, as quais foram testadas em um painel de 96 indivíduos. Um conjunto de 7.680 sondas que revelaram polimorfismos robustos foi selecionado e usado para construir o microarranjo de genotipagem operacional DArT. Este procedimento foi otimizado por meio de: (1) amostragem de uma grande coleção de variantes de sequência para aumentar a recuperação de clones polimórficos, e (2) transferibilidade interespecífica dos marcadores registrados. Com o mesmo método de redução de complexidade genômica utilizado na construção da biblioteca, representações genômicas dos parentais e dos 177 indivíduos F1 da população de mapeamento foram geradas para produzir "*targets*", os quais posteriormente foram hibridizados no microarranjo. Um grupo de amostras foi marcado com fluorescência Cy-3 (verde) e outro grupo com Cy-5 (vermelho), para assim otimizar o rendimento de cada lâmina, já que podem ser hibridizadas duas amostras com diferentes cores na mesma lâmina.

Após a hibridação, as lâminas do microarranjo foram lavadas e escaneadas, usando um equipamento TECAN LS300 confocal a laser, com uma resolução de 20 mm por pixel com aquisição sequencial de três imagens para cada lâmina. O sinal emitido a partir do sitio de clonagem do vetor *poli-linker*, marcado com fluorescência FAM, proporcionou um valor de referência para a quantidade de fragmentos de DNA amplificados e, presentes em cada "*spot*" do microarranjo. As imagens resultantes foram analisadas usando *DArTSoft* versão 7.44, um *software* desenvolvido por *Diversity Arrays Technology Pty. Ltd*. para a extração de dados das imagens do microarranjo, a detecção de polimorfismos e a definição do marcador. O valor relativo de intensidade de hibridação foi então calculado para todos os "*spots*" aceitos como log[*target*/referência], ou seja, log[sinal de Cy-3 / sinal de FAM] para os *targets* marcados com Cy-3, e log[sinal de Cy-5 / sinal de FAM] para os *targets* marcados com Cy-5. Posteriormente, *DArTSoft* comparou os valores de intensidade relativa obtidos

para cada um dos clones. Se a relação target/referência entre dois *clusters* possuía uma variância de intensidades relativas, de pelo menos 80% da variância total, os clones eram considerados polimórficos e genotipados de forma binaria como “1” ou “0”. *Targets* com valores de intensidade relativa que não puderam ser atribuídos a um ou outro *cluster* (0 ou 1) foram registados como dados perdidos. Métodos padrão de descoberta de marcadores foram implementados através de uma combinação de parâmetros extraídos automaticamente, a partir dos dados do arranjo usando *DArTsoft*. Os parâmetros utilizados neste estudo foram: (1) reprodutibilidade ≥ 95%, como medida pela concordância do chamado genótipo entre repetições da técnica (*targets* replicados e processados para um mínimo de 30% das amostras de DNA genotipadas); (2) qualidade do marcador Q ≥ 65, o qual representa a variância entre *clusters* como uma porcentagem da variância total na distribuição do sinal de fluorescência entre as amostras testadas; e (3) *Call Rate* do marcador ≥ 75% (porcentagem de *targets* que puderam ser registrados como '0' ou '1').

3.3.4. Construção do mapa genético

Um mapa de ligação genética único e integrado foi construído utilizando ambos os dados de marcadores, seja os codominantes microssatélites e os dominantes DArT, usando JoinMap v3.0 (Van Ooijen and Voorrips 2001). Marcadores microssatélites segregaram a partir de cada um dos parentais individualmente em uma configuração 1:1, ou em uma proporção 1:2:1 em configuração F2 de fase de ligação desconhecida, com ambos os parentais igualmente heterozigotos para o mesmo genótipo, ou ainda em configuração totalmente informativa 1:1:1:1 com três ou quatro alelos diferentes no total, segregando a partir dos dois parentais. Marcadores dominantes DArT, por outro lado, segregaram em uma configuração de pseudo-cruzamento teste 1:1, a partir de cada um dos parentais individualmente, ou em razão 3:1 quando ambos os parentais eram heterozigotos. Para ambos os dados, microssatélites e DArT, os marcadores que mostraram *Call Rate* ≥ 75% e apresentaram uma das proporções de segregação esperadas com $\alpha \leq 0{,}01$ foram usados para a análise de ligação. O agrupamento e ordenamento dos marcadores foram estabelecidos inicialmente pela

aplicação do algoritmo de máxima verossimilhança, usando JoinMap, com população tipo CP; agrupamento a LOD > 15; fração de recombinação ≤ 0,4, com valor de Ripple = 1; *jump in goodness-of-fit threshold* = 5 (correspondendo à diferença normalizada de Qui-quadrado antes e após a adição de um loco); utilizando a função de mapeamento Kosambi. O ordenamento dos marcadores com JoinMap foi realizado por anelamento simulado, excluindo aqueles marcadores que contribuíam para uma ordem instável nas duas primeiras tentativas de ordenamento para produzir um mapa *framework*, com um suporte estatístico para ordenamento de alta verossimilhança. Marcadores segregantes adicionais foram então incorporados ao mapa de ligação com menor rigor em uma terceira e última tentativa do JoinMap, para proporcionar uma posição de mapa a um maior número possível de marcadores DArT segregantes.

3.3.5. Análise comparativa entre o mapa de ligação e a montagem da sequência do genoma

Foi realizada uma avaliação genômica da consistência na ordem dos marcadores, estimada no mapa de ligação *framework* com a posição física dos marcadores no genoma, alinhando estes sobre os 11 *scaffolds* correspondentes aos 11 pseudo-cromossomos da atual versão do genoma de referência de *Eucalyptus* (versão 1.0 disponível em Phytozome 6.0, http://www.phytozome.net/*Eucalyptus*.php), produzido a partir da árvore autofecundada BRASUZ1 (Brasil Suzano S1). Este alinhamento foi também utilizado para proporcionar estimativas para cada cromossomo específico e do genoma completo, com base na correspondência entre distância física e fração de recombinação no genoma de *Eucalyptus*, bem como uma estimativa da eficiência da cobertura genômica fornecida pelo mapa *framework*.

3.3.6. Caracterização genômica dos marcadores DArT

Os clones de *E. coli* contendo as 7.680 sondas DArT de *Eucalyptus* (Sansaloni, Petroli et al. 2010) foram rearranjados em vinte placas de 384 poços e submetidos a sequenciamento Sanger, bidirecional, no laboratório de genômica da Universidade de Purdue (www.genomics.purdue.edu). Após a filtragem de qualidade e a clivagem das regiões do vetor e sitios *Pst*I, as sequências obtidas foram depositadas no GenBank (números de acesso HR865291-HR872186). A redundância das sondas DArT em nível de sequência foi analisada através do *software* Geneious Pro 5.1.7 (Drummond, Ashton et al. 2011), usando um mínimo de sobreposição de 50 pb, para uma sequência ser agrupada dentro de um *contig* e uma identidade de sobreposição de 98% (a "identidade de sobreposição" é a porcentagem mínima de bases que deve ser idêntica dentro da região de sobreposição, para a sequência ser agrupada). Os números de sondas DArT únicas e redundantes foram avaliados através da aplicação de quatro diferentes parâmetros de agrupamento, desde o mais rigoroso (A1) ao mais permissivo (A4). Estes parâmetros foram: (a) tamanho da palavra (word length), ou seja, o número mínimo de bases consecutivas que deveriam coincidir perfeitamente a fim de encontrar uma correspondência entre duas sequências; (b) número máximo de bases únicas incompatíveis permitidas por *read* como uma porcentagem do tamanho da sobreposição entre dois *reads*; (c) número máximo de bases ambíguas permitidas nas palavras coincidentes; (d) o número máximo de *gaps* que podem estar inseridos dentro de cada *read* como uma porcentagem do tamanho da sobreposição entre dois *reads*; (e) tamanho máximo de cada *gap* que pode estar inserido dentro dos *reads*.

Após o pré-processamento para remover os contaminantes, artefatos de sequenciamento e sequências de baixa qualidade, todas as sondas DArT para as quais foram obtidas sequências foram mapeadas sobre o genoma referência de *Eucalyptus grandis* (versão 1.0 disponível em Phytozome 6.0). O mapeamento foi realizado utilizando o *software* BWA-SW (versão 0.5.8) a partir da ferramenta de alinhamento Burrows-Wheeler (Li and Durbin 2010) para produzir um arquivo BAM (Li, Handsaker et al. 2009). O limiar para um *hit* da sequência da sonda ser retido foi definido para um valor fixo (T = 70). Foram considerados até dois *hits* como sendo um "mapeamento de sucesso" e utilizados os valores sub-ótimos de cada um dos *hits* para classificar a "confiabilidade do mapeamento" e a "taxa esperada de erro do mapeamento'' para os

resultados. Finalmente, para examinar as características genômicas dos marcadores DArT em relação aos modelos gênicos preditos na versão 1.0 do genoma de *Eucalyptus grandis*, todos os marcadores pertencentes ao arranjo DarT, incluindo aqueles agrupados e ordenados no mapa genético, foram mapeados fisicamente no genoma de referência. Posteriormente os 11 pseudo-cromossomos foram divididos em intervalos de 5Mbp para realizar uma análise detalhada da localização dos marcadores DArT em relação com os modelos gênicos preditos. Uma correlação de Spearman entre o número de marcadores DArT e o número de modelos gênicos anotados em cada intervalo foi estimado para cada pseudo-cromossomo. Além disso, foi avaliada a distância física em pares de bases a partir de cada sonda DArT sequenciada, com o modelo gênico predito mais próximo, para fornecer uma visão genômica ampla da cobertura do espaço gênico no genoma de *Eucalyptus* propiciada pelo arranjo DArT.

3.4. RESULTADOS

3.4.1. Genotipagem dos marcadores DArT

A distribuição dos marcadores nos diferentes níveis de reprodutibilidade tendeu para as classes de maior qualidade. Por exemplo, a partir dos 3.933 marcadores que tinham reprodutibilidade ≥ 95%, 70% apresentaram reprodutibilidade igual ou maior do que 99%. Para os 4.884 marcadores que passaram a o valor limiar de qualidade, 61% tinham Q ≥ 70, enquanto que dos 5.415 marcadores com um *Call Rate* ≥ 75%, 36% apresentaram um *Call Rate* ≥ 90% (Figura 2). Enquanto os parâmetros de reprodutibilidade e Q são medidas que diretamente avaliam a qualidade da genotipagem, o *Call Rate* reflete essencialmente a porcentagem tolerada de dados perdidos. Um *Call Rate* menos rigoroso de marcadores com valores ≥ 75% foi adotado para maximizar o número de marcadores posicionados no mapa de ligação, uma vez que tal limiar produziria ainda dados de marcadores com boa qualidade para ~128 gametas recombinantes informativos que permitem uma análise satisfatória da ligação e ordenamento dos marcadores durante a construção do mapa. As 7.680 sondas do microarranjo de *Eucalyptus* resultaram em 3.191 marcadores que simultaneamente

passaram todos os parâmetros de filtragem de qualidade de marcador além do filtro de *Call Rate* (Figura 2).

A partir dos 3.191 marcadores testados para observação do comportamento mendeliano, apenas 215 não se encaixaram em qualquer uma das proporções, 1:1 ou 3:1, e foram excluídos das análises posteriores. Assim, os 2.976 marcadores DArT que mostraram comportamento mendeliano foram utilizados na análise de ligação, e destes, 1.777 segregaram numa configuração 1:1 de pseudo-cruzamento e 1.199 numa razão de 3:1, ou seja, foram locos heterozigotos segregando simultaneamente a partir de ambos os parentais. Em relação ao conjunto de dados de microssatélites, 166 locos foram totalmente informativos com três ou quatro alelos segregando em quatro classes genotípicas distintas proporcionando locos âncora para a construção de um mapa de ligação integrado. Quarenta e dois microssatélites segregaram a partir de um dos parentais apenas, 25 a partir de *E. grandis* e 17 de *E. urophylla*, enquanto 14 segregaram em uma configuração F2 1:2:1 de fase desconhecida.

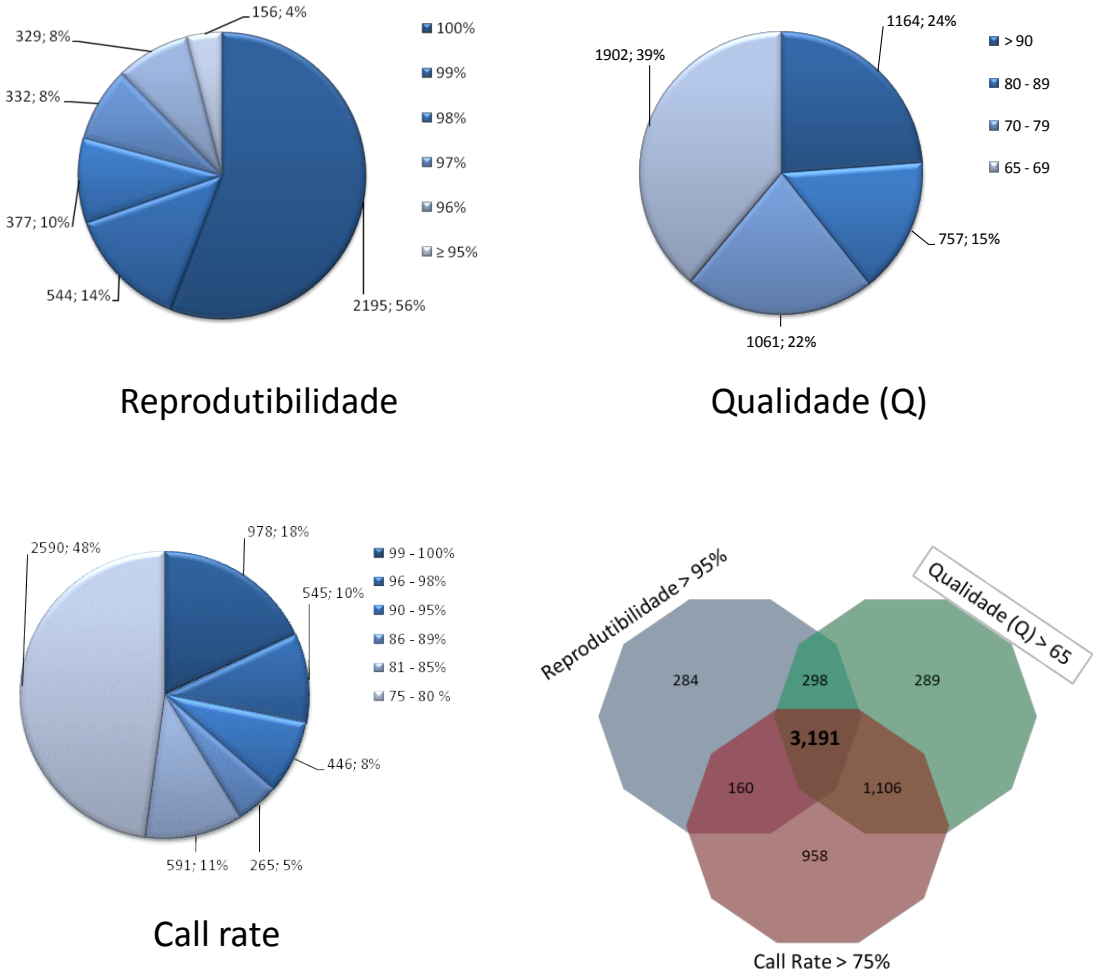


Figura 3. Distribuição do número e porcentagens de marcadores DArT que passaram o limiar de filtragem adotado para reprodutibilidade (≥95%), qualidade (Q≥65%) e *call rate* (≥75%). Um diagrama de Venn consolida a informação mostrando todas as classificações possíveis dos marcadores DArT de acordo com os três critérios adotados. Somente marcadores que satisfizeram simultaneamente a todos os três critérios foram usados para o mapeamento de ligação.

3.4.2. Mapeamento de ligação

Um conjunto de 3.198 marcadores (2.976 DArTs e 222 microssatélites) foi sujeito a uma análise de mapeamento. A análise de agrupamento a LOD>15,0 resultou em 2.980 marcadores reunidos em 11 grupos de ligação numerados como estabelecido anteriormente (Grattapaglia and Sederoff 1994; Brondani, Williams et al. 2006) e

avaliados pela presença de marcadores microssatélites âncoras. O mapa de ligação construído com suporte de alta verossimilhança para o ordenamento dos marcadores após a segunda rodada de JoinMap é apresentado como um "Mapa *Framework*" (Figura 3). O mapa de ligação obtido após a terceira rodada de ordenamento é chamado a seguir de "Mapa *full*" ou mapa completo, e é apresentado como uma maneira de fornecer uma posição preliminar de todos os marcadores informativos para a subsequente caracterização genômica (Tabela 1 e Figura 4). O mapa *framework*, construído seguindo a segunda rodada do JoinMap resultou em 1.029 marcadores posicionados com alta verossimilhança, 861 DArTs e 168 microssatélites. Quando um ordenamento menos rigoroso foi permitido, um total de 2.484 marcadores foi mapeado (2.274 DArTs e 210 microssatélites). Os 496 marcadores remanescentes não puderam ser posicionados, mesmo utilizando um critério mais permissivo, possivelmente como resultado da redundância de marcadores DArT no nível de sequência (ver abaixo) ou devido a uma ligação muito próxima, de modo que não puderam ser amostrados recombinantes suficientes para resolver o ordenamento relativo ao longo do mapa.

Uma proporção maior de microssatélites (80%) foi posicionada no mapa *framework* do que de marcadores DArT (45%), provavelmente devido ao maior conteúdo informativo dos microssatélites multialélicos que fornecem um maior poder para categorizar haplótipos recombinantes contra haplótipos parentais e, assim, determinar a ordem dos marcadores. No entanto, o tamanho final do mapa *framework* foi apenas 9,5% menor do que o mapa *full*, 1.176,7cM e 1.303,9 cM respectivamente (Tabela 1 e Figura 4). As ordens dos marcadores de ambos os mapas (framework e full) foram geralmente consistentes, apesar de que alguns conjuntos de marcadores invertidos puderam ser observados, principalmente nos grupos de ligação (GL) GL1, GL2, GL7 e GL9. Além disso, embora a expectativa fosse de que todos os marcadores do mapa *framework* estivessem contidos no mapa *full* este não foi sempre o caso. O mapa *framework* continha 99 marcadores que foram excluídos quando um limiar de ordenamento mais permissivo foi utilizado para a construção do segundo mapa. Estes se concentraram nos grupos GL2 (42 marcadores), GL3 (26 marcadores), GL1 (19 marcadores) e GL11 (8 marcadores) (Figura 4).

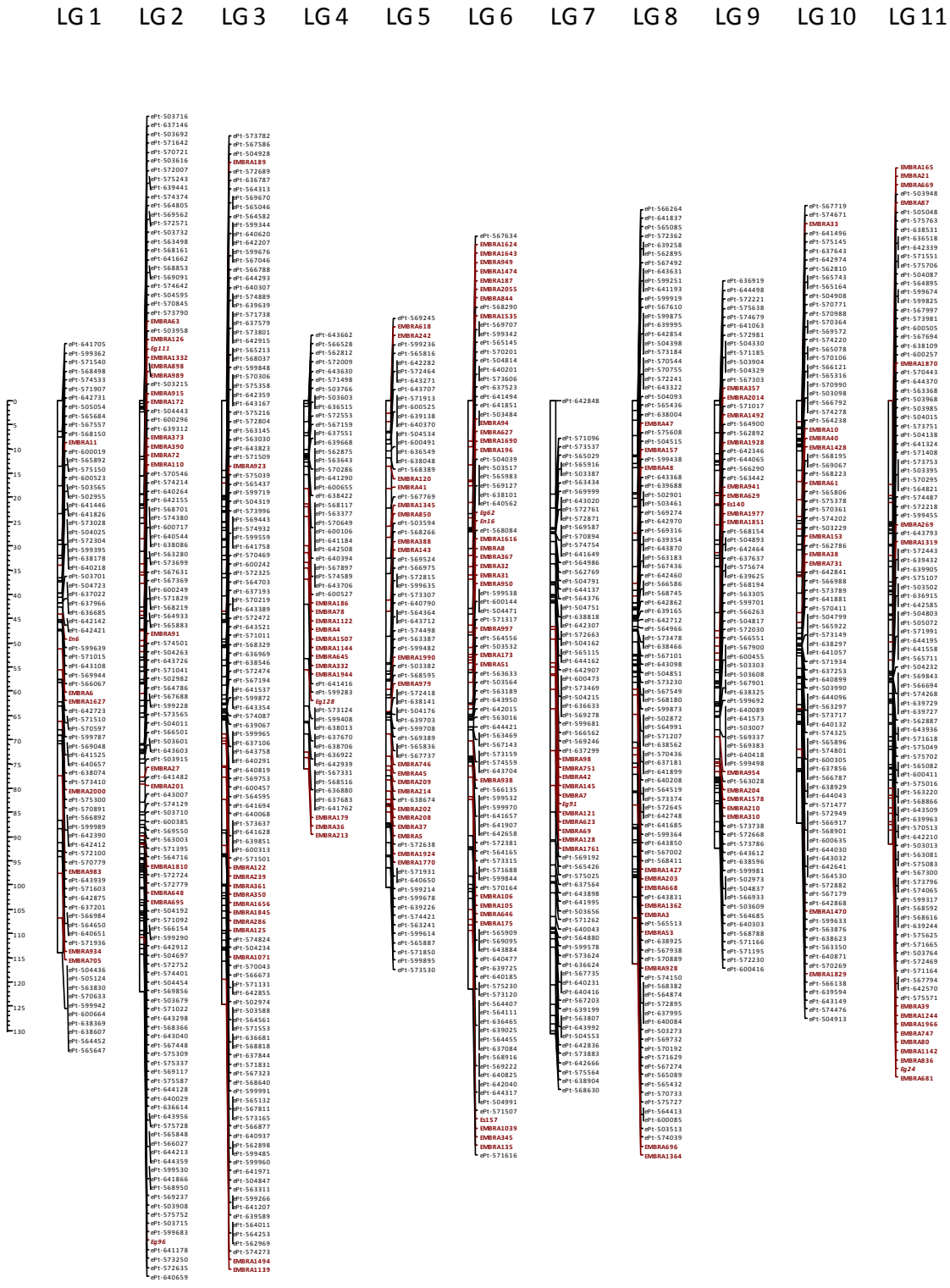


Figura 4. Mapa de ligação *framework* DArT/microssatélites para *Eucalyptus*. O mapa inclui 1.029 marcadores posicionados com alto suporte de verossimilhança para o ordenamento de locos, envolvendo 861 DArTs (em preto) e 168 microssatélites (em vermelho) com uma escala em centiMorgan à esquerda.

Tabela 1. Estatísticas dos mapas consenso de DArT/microssatélites para a população segregante de *E. grandis x E. urophylla*.

Grupo de Ligação/ Pseudo-cromossomo	1	2	3	4	5	6	7	8	9	10	11	Total	Média	Desvio padrão
Mapa *Full*[a]														
N° Total de Marcadores	207	244	270	106	189	271	224	275	220	263	215	2.848	225,8	49,34
N° de Marcadores DarT	191	219	256	93	166	231	210	262	204	246	196	2.274	206,7	47,68
N° de Microssatélites	16	25	14	13	23	40	14	13	16	17	19	210	19,1	7,98
Tamanho Total (cM)	167,8	129	102,4	86,6	130,7	116,5	117,3	118,6	118,3	117,4	99,5	1.303,9	118,5	20,8
Distância Média entre Marcadores (cM)	0,8	0,5	0,4	0,8	0,7	0,4	0,5	0,4	0,5	0,4	0,5	0,5		
Mapa *Framework*[b]														
N° Total de Marcadores	80	131	128	57	74	104	75	107	78	92	103	1.029	93,5	23,4
N° de Marcadores DarT	72	112	115	44	54	72	64	95	64	82	87	864	78,3	22,6
N° de Microssatélites	8	19	13	13	20	32	11	12	14	10	16	168	15,3	6,6
Tamanho Total (cM)	114,0	122,1	124,6	76,1	100,3	121,6	130,5	116,2	92,5	87,4	91,5	1.176,8	107	18,1
Distância Média entre Marcadores (cM)	1,4	0,9	1,0	1,3	1,4	1,2	1,7	1,1	1,2	1,0	0,9	1,1	-	0,3
***Mapa Framework* no Genoma[c]**														
N° Total de Marcadores *Framework*	62	98	102	52	68	88	68	94	66	82	89	869	79	16,5
Distância Física Coberta (Mpb)	40,7	63,8	79,7	41,1	73,8	50,3	51,9	68,4	38,4	38,6	40,8	587,5	-	-
Proporção de kpb/cM	357,3	522,1	639,2	539,9	736,1	413,7	548,4	588,9	415,1	441,4	445,4	-	513,4	112,7

[a] Mapa *Full*: todos os marcadores mapeados com um suporte probabilístico permissivo para ordenamento.
[b] Mapa *Framework*: marcadores ordenados com o maior suporte estatístico.
[c] Mapa *Framework* no genoma: marcadores do mapa *Framework* foram posicionados na sequência do genoma de *Eucalyptus grandis* para fornecer uma correspondência entre distância física e fração de recombinação para cada pseudo-cromossomo e ao nível do genoma total.

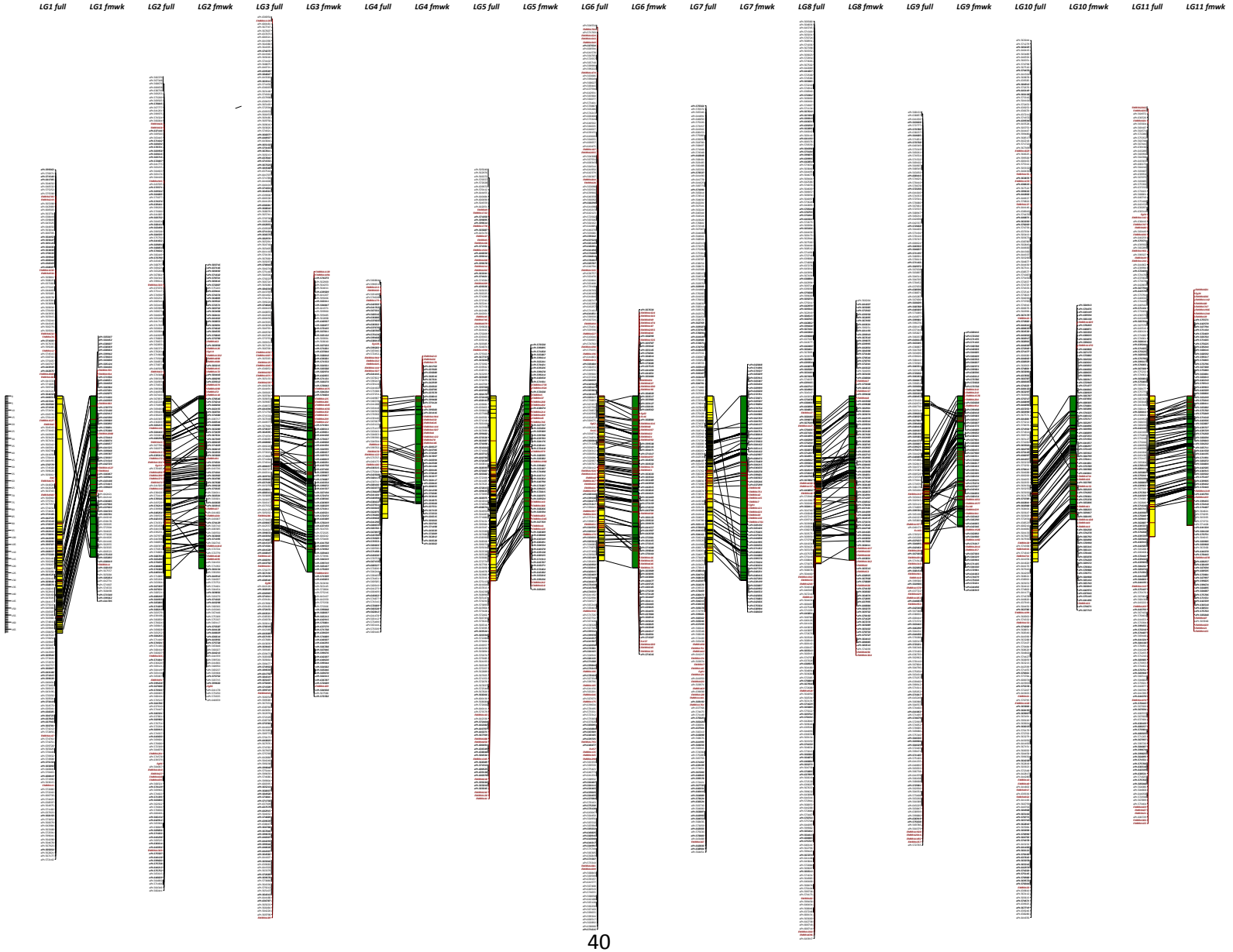

LG1 full
LG1 fmwk
LG2 full
LG2 fmwk
LG3 full
LG3 fmwk
LG4 full
LG4 fmwk
LG5 full
LG5 fmwk
LG6 full
LG6 fmwk
LG7 full
LG7 fmwk
LG8 full
LG8 fmwk
LG9 full
LG9 fmwk
LG10 full
LG10 fmwk
LG11 full
LG11 fmwk

Figura 5. Alinhamento do mapa *full* (barras amarelas) com o mapa *framework* (Fmwk) (barras verdes) para os 11 pseudo-cromossomos do *Eucalyptus* mostrando as conexões entre os mesmos locos em ambos os mapas. O mapa *full* inclui um total de 2.484 marcadores, 2.274 DArTs e 210 microssatélites enquanto que o mapa *framework* possui 1.029 marcadores posicionados com alta verossimilhança para a ordem dos locos, sendo 861 DArTs e 168 microssatélites. Marcadores DArT em preto e microssatélites em vermelho; na esquerda encontra-se a escala em centiMorgan.

O mapa *full* de marcadores DArT continha em média 226 marcadores por GL posicionados a uma distância média entre marcadores consecutivos de 0,5 cM, enquanto que o mapa *framework* tinha em média 93,5 marcadores por GL e uma distância média entre marcadores de 1,1 cM (Tabela 1). A distribuição de distâncias de mapa entre marcadores consecutivos no mapa *framework* foi significativamente diferente daquela no mapa *full* (p = 0,021 em teste não paramétrico de Komolgorov-Smirnov) (Figura 5). Este resultado demonstra que (i) o mapa *framework* espalha marcadores com alto suporte estatístico para proporcionar um ordenamento robusto de locos e (ii) reduz a proporção de distâncias curtas entre marcadores (< 1 cM) em relação ao mapa *full* (87% no mapa full e 65% no mapa framework).

Uma análise da origem dos 861 marcadores DArT mapeados no mapa *framework* mostrou que 197 (23%) que segregaram em uma proporção 1:1 tinham origem a partir do genitor *E. urophylla* e 298 (35%) a partir do genitor *E. grandis*, enquanto que 366 (42%) eram heterozigotos em ambos os pais, segregando 3:1. Proporções muito semelhantes foram observadas quando todos os 2.274 marcadores DArT foram examinados. Estes resultados sugerem uma maior heterozigosidade no genitor *E. grandis* do que no genitor *E. urophylla*. A partir das 7.680 sondas do microarranjo, 2.274 (isto é, aproximadamente 30%), foram por fim mapeados nesta família segregante. No entanto, se os 3.191 marcadores DArT que passaram os filtros de qualidade de genotipagem para este experimento fossem considerados, 71% dos marcadores poderiam ser mapeados. Embora a proporção de marcadores que pode ser mapeada dependa em grande parte da heterozigosidade de sequência dos

parentais e da sua divergência genética, este resultado corrobora o excelente desempenho do microarranjo DArT para fins de mapeamento de ligação em *Eucalyptus*.

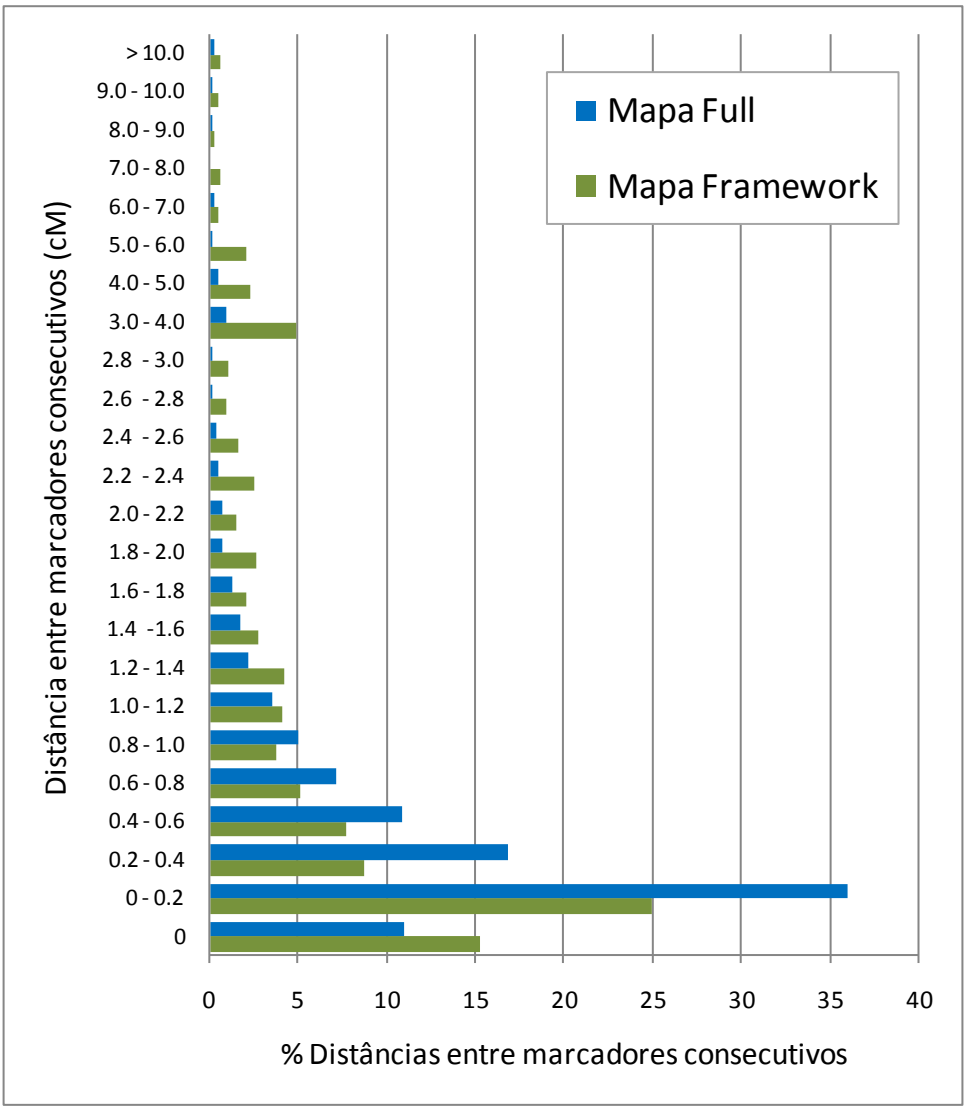
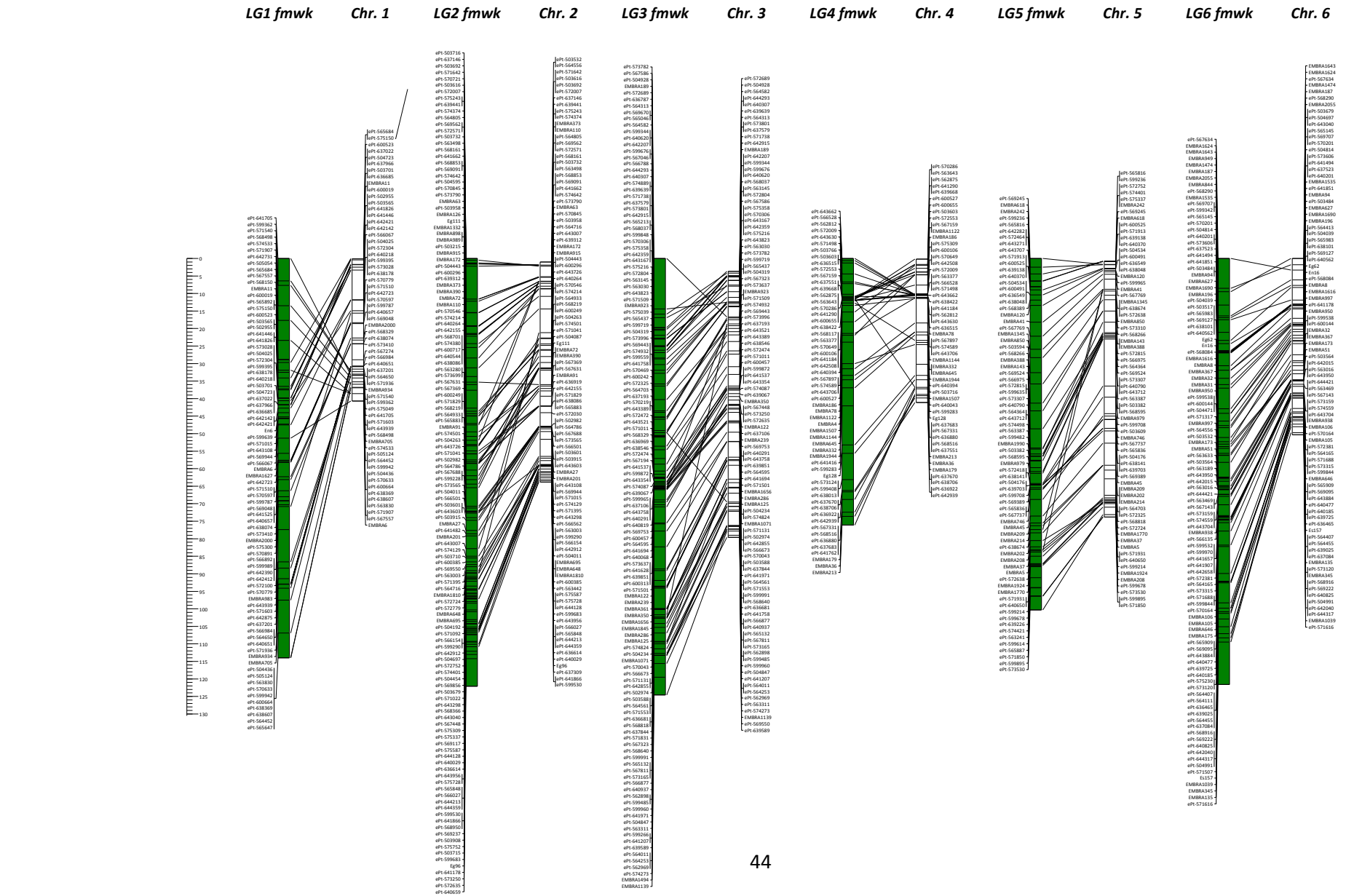


Figura 6. Distribuição de frequência das distâncias de recombinação de Kosambi entre marcadores consecutivos ao longo das duas versões do mapa de ligação. A distribuição de distâncias de mapa no mapa *framework* foi significativamente diferente daquela do mapa *full* (p = 0,021 de um teste não paramétrico de Komolgorov-Smirnov), confirmando o fato de que o mapa *framework* espalha os marcadores retidos com elevado suporte de ordenamento e reduz a proporção de distâncias entre marcadores menores do que um centiMorgan, de um total de 87% no mapa *full* para 65% no mapa *framework*.

3.4.3. Distâncias de recombinação e física no genoma de *Eucalyptus*

O alinhamento do mapa de ligação *framework* com a sequência do genoma de *Eucalyptus* indicou que a ordem relativa dos marcadores mapeados concorda amplamente com suas posições físicas (Figura 6). Apenas alguns marcadores esparsos ou pequenos blocos de marcadores (por exemplo, nos GL1 e GL4) mostraram uma ordem localmente inconsistente com a estimada na sequência do genoma. Após a inspeção ainda mais aprofundada dos dados de segregação dos poucos marcadores que mostraram discrepância entre suas posições físicas e aquelas obtidas com base em recombinação, constatou-se que vários deles estavam no limite em termos de parâmetros de qualidade e de *Call Rate*, o que poderia explicar as inconsistências observadas. A partir dos 1.029 marcadores mapeados no *framework*, 869 puderam ser posicionados sobre a sequência do genoma, enquanto que para os 160 remanescentes não foi obtida sequência ou mapearam em algum dos 4.941 *scaffolds* menores adicionais não ancorados na montagem atual do genoma de *Eucalyptus*. Os 869 marcadores mapeados genética e fisicamente cobriram um total de 587,5 Mpb da sequência (Tabela 1), proporcionando assim uma cobertura de 97% dos 605,8Mbp atualmente montados nos 11 *scaffolds* principais do genoma de *Eucalyptus*. As estimativas pseudo-cromossomo-específicas sobre a relação entre distância física em kpb e a fração de recombinação em cM variou entre 357,3 kpb/cM para o pseudo-cromossomo 1 e 736,1 kpb/cM para o pseudo-cromossomo 5, com uma média de 513,4 kpb/cM para o genoma todo (Tabela 1). Quando o mapa *full* foi alinhado com a sequência do genoma (dados não mostrados), dos 2.274 marcadores DArT segregantes e geneticamente mapeados, 1.986 alinharam nos 11 pseudo-cromossomos, enquanto que 45 marcadores mapearam em 31 *scaffolds* não ancorados e para 243 marcadores DArT não foi obtida sequência ou não mapearam na atual vesão do genoma. Com base nos marcadores DArT mapeados, os 31 *scaffolds* não ancorados, adicionando 1,4 Mpb de sequências, puderam ser alocados aos 11 pseudo-cromossomos (Tabela 2).


LG1 fmwk
Chr. 1
LG2 fmwk
Chr. 2
LG3 fmwk
Chr. 3
LG4 fmwk
Chr. 4
LG5 fmwk
Chr. 5
LG6 fmwk
Chr. 6

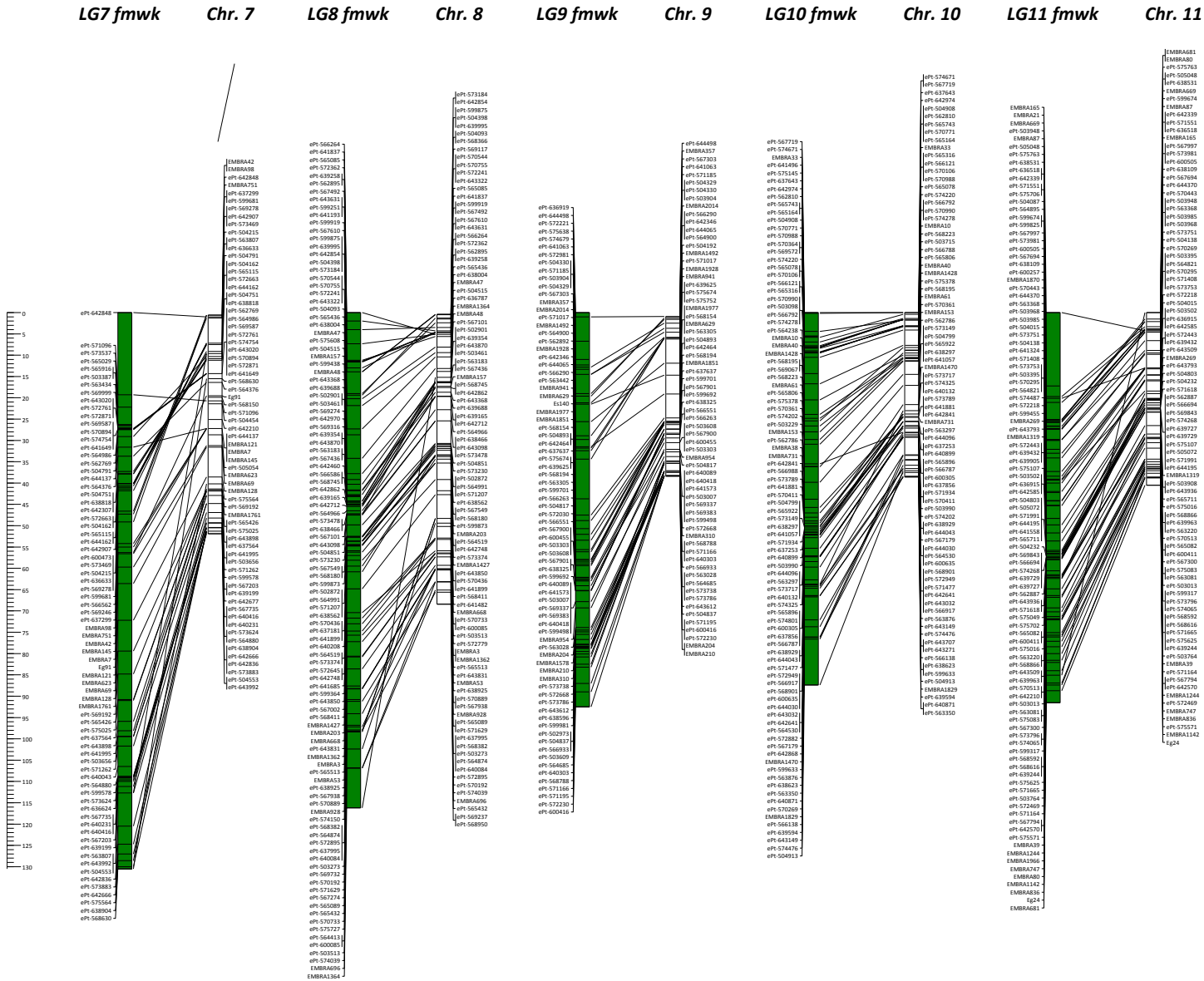

LG7 fmwk
Chr. 7
LG8 fmwk
Chr. 8
LG9 fmwk
Chr. 9
LG10 fmwk
Chr. 10
LG11 fmwk
Chr. 11

Figura 7. Alinhamento do mapa *framework* com o genoma de referência de *Eucalyptus*. Correspondência do posicionamento dos marcadores DArT e microssatélites no mapa de ligação *framework* (barras verdes) e nos 11 *scaffolds* peudo-cromossômicos de *Eucalyptus grandis* (barras brancas). A escala à esquerda corresponde simultâneamente às distâncias em centiMorgan para o mapa de ligação e em Mpb para as sequências nos pseudo-cromossômos.

Tabela 2. Lista dos 45 marcadores DArT mapeados nos 11 grupos de ligação e posicionados em pequenos *scaffolds* não ancorados na atual montagem do genoma de *Eucalyptus grandis* (versão 1.0 no Phytozome 6.0). Estes marcadores DArT mapeados geneticamente permitiram a alocação de 31 pequenos *scaffolds* (1.4 Mpb de sequência total) aos 11 pseudo-cromossomos principais.

Marcadores DArT	Posições de mapa dos marcadores (cM)	Grupos de Ligação/ Pseudo-cromossomo	Tamanho de sequências DArT	N° do *Sacffold* não ancorado	Tamanho do *Scaffold* (pb)
ePt-571990	39.98	7	342	24	625.428
ePt-599305	45.43	3	226	50	200.725
ePt-639198	60.54	2	688	118	109.742
ePt-569673	36.45	2	326	134	73.68
ePt-574447	77.42	2	497	134	
ePt-641713	36.62	2	305	134	
ePt-568315	30.79	7	317	207	54.558
ePt-503680	85.13	8	310	207	
ePt-572981	92.18	9	337	460	27.208
ePt-599970	69.7	6	452	491	26.388
ePt-641657	69.7	6	453	491	
ePt-641907	69.7	6	452	491	
ePt-642658	69.7	6	453	491	
ePt-503696	47.35	6	465	499	25.365
ePt-641193	26.65	8	304	556	23.569
ePt-571092	93.5	2	295	578	23.216
ePt-565718	7	6	610	686	22.065
ePt-641578	6.55	6	438	686	
ePt-503229	11.76	10	241	746	18.491

ePt-570410	34.11	9	522	847	16.12
ePt-642696	42.73	9	536	847	
ePt-643222	44.55	11	554	949	14.791
ePt-599337	148.04	1	377	958	16.434
ePt-643196	148.46	1	376	958	
ePt-567002	64.73	8	445	1033	
ePt-640230	44.41	3	292	1165	12.751
ePt-572676	35.65	10	382	1221	12.304
ePt-575708	98.48	8	549	1340	11.269
ePt-644097	98.48	8	558	1340	
ePt-599556	56.67	2	546	1519	10.167
ePt-599364	63.45	8	245	1533	9.961
ePt-570223	30.13	3	399	1585	10.093
ePt-640659	122.09	2	342	1685	9.193
ePt-504097	31.65	8	601	1742	8.856
ePt-565887	51.24	5	580	1843	8.383
ePt-599614	50.38	5	578	1843	
ePt-643170	51.28	5	578	1843	
ePt-570110	81.04	9	229	1891	8.187
ePt-568521	59.84	2	381	1915	8.269
ePt-564907	29.1	9	356	2966	7.743
ePt-573755	34.37	3	383	3233	4.288
ePt-599351	34.43	3	384	3233	
ePt-568568	85.77	5	383	3329	4.17
ePt-572802	47.53	3	263	4064	3.562
ePt-573923	47.51	3	263	4064	

3.4.4. Análises de redundância das sequências das sondas DArT

Foram obtidas sequências Sanger para 6.918 das 7.680 sondas DArT (90%), com um tamanho médio de 534 pb. Sob os parâmetros mais rigorosos de montagem (ver Material e Métodos), de um total de 6.918 sequências, 3.709 agruparam em *clusters* multi-sequência com duas ou mais sequências por *cluster*. Estas foram agrupadas dentro de 1.374 *clusters* únicos de sequências não redundantes, enquanto que 3.209 tiveram uma única sequência representada, ou seja, *singletons* exclusivos. No total, as

6.918 sondas para as quais foram obtidas sequências representaram efetivamente 4.583 locos únicos, ou seja, uma estimativa da taxa de redundância de 33,75%. Sob parâmetros de montagem mais liberais, o número de locos únicos foi reduzido para um total de 3.864, proporcionando uma estimativa da taxa de sequências redundantes de 44,14% (Tabela 3). Se uma taxa equivalente de redundância é assumida para as 762 sondas DArT para as quais não puderam ser obtidas sequências, as 7.680 sondas no arranjo DArT amostrariam efetivamente entre 4.289 e 5.087 locos únicos no genoma de *Eucalyptus*. Todas as 6.918 sequências foram submetidas ao GenBank e 6.896 foram finalmente aceitas e depositadas (22 foram cortadas por terem tamanhos menores do que o mínimo aceito pelo NCBI), recebendo números de acesso HR865291-HR872186, com identificadores de clones correspondentes à nomenclatura convencionada dos marcadores DArT usada neste trabalho. Estes clones foram apresentados contra o banco de dados completo do NCBI EST, dos quais 3.703 (53,6%) retornaram com *hits* BLASTn positivos.

Tabela 3. Resultados da análise de redundância das 6.918 sequências das sondas DArT sob quatro grupos diferentes de parâmetros de montagem, desde o mais rigoroso (A1) ao mais permissivo (A4) (ver Material e Métodos para detalhes).

Parâmetro	**A1**	**A2**	**A3**	**A4**
Comprimento da palavra (*word length*)	18	14	12	10
Index do compr. da palavra	13	12	11	10
Mismatches	10%	15%	20%	20%
Ambiguidades por leitura	4	4	16	16
% máximo de *gaps* por leitura	10%	15%	20%	20%
Tamanho do *gap*	1bp	2bp	5bp	5bp
Resultados da análise de redundância				
N° de *singletons* exclusivos[a]	3.209	2.607	2.381	2.276
N° de sequências redundantes[b]	3.709	4.311	4.537	4.642
N° de sequências únicas não redundantes[c]	1.374	1.537	1.587	1.588
Total de sequências selecionadas[d]	4.583	4.144	3.968	3.864
Taxa estimada de sequências redundantes	33,75%	40,0%	42,64%	44,14%

[a]Sequências únicas sem correspondência com nenhuma outra leitura.

[b]Sequências que agruparam dentro de *clusters* multi-sequência com mais de duas sequências por *cluster*.
[c]Sequências únicas provenientes de *clusters* redundantes.
[d]Soma de *singletons* (exclusivos) sem correspondência e sequências não redundantes.

3.4.5. Alinhamento das sondas DArT no genoma de *Eucalyptus*

A partir das 6.896 sondas DArT para as quais sequências de qualidade foram obtidas, 6.631 (96%) puderam ser alinhadas com sucesso sobre a sequência do genoma de *Eucalyptus grandis* (versão 1.0 em Phytozome 6.0), enquanto 265 não puderam ser mapeadas usando parâmetros de elevado rigor. Assim, das sondas mapeadas 6.390 foram alinhadas nos 11 principais pseudo-cromossomos e as 241 remanescentes se posicionaram nos 4.941 pequenos *scaffolds* adicionais não ancorados. Quando estes resultados de mapeamento foram utilizados para avaliar a qualidade dos parâmetros de alinhamento de sequências adotados (ver Material e Métodos) uma taxa de erro de mapeamento de 0,002 foi estimada mediante a observação de 12 sequências não mapeadas em 6.631, isto é, uma confiabilidade $\geq$ 99.8%. Curiosamente, este número corresponde precisamente com a média de reprodutibilidade estimada por *DArTsoft* seguindo os valores limiares padrão utilizados para a seleção de marcadores. Usanso-se o valor padrão (default) de BWA (T = 37) como o limiar para o *hit* da sequência ser retido, 166 das 265 sondas não mapeadas puderam ser alinhadas adicionalmente com os 11 pseudo-cromossomos principais, sendo que 91 destas 166 sondas também foram mapeadas no mapa de ligação *full*. Além disso, das 89 sondas que permaneceram fisicamente não mapeadas no genoma de *E. grandis*, 36 foram mapeadas com sucesso também no mapa de ligação.

Adicionalmente, foi realizado um exame mais aprofundado do alinhamento das 6.631 sondas DArT no genoma montado de *Eucalyptus*, incluindo todos os 4.952 *scaffolds* pequenos. Um total de 4.189 sondas foi alinhado com alta confiança em uma simples e única posição do genoma, assim como o seu mapeamento produziu um simples *hit* sem sub-alinhamentos. Para 2.252 das 2.442 sondas remanescentes, um segundo sub-alinhamento foi retido, o qual sobrepôs o mesmo loco que o do primeiro melhor alinhamento. Portanto, no total, 6.441 sondas DArT das 6.631 avaliadas (97,1%) foram consideradas como alinhadas para um único loco no genoma. Para as

190 sondas nas quais um segundo *hit* foi reportado por BWA, realizamos uma análise usando ferramentas quiméricas de detecção fornecidas pelos *softwares* de montagem de fragmentos CD HIT (Li and Godzik 2006) e EULER DNA (Pevzner, Tang et al. 2001), os quais foram executados com parâmetros pré-definidos, resultando na ausência de *reads* quiméricos. Para 135 sondas, o sub-alinhamento retido foi localizado em uma posição diferente no mesmo *scaffold*, e em 95 destes casos, a distância entre alinhamentos foi menor do que 1 kb, o que sugere uma duplicação contígua em tandem. Para 40 sondas DArT as distâncias entre o alinhamento do primeiro e o segundo *hit* eram maiores do que 1 kb, com 13 deles maiores do que 10 kb. Finalmente, apenas 55 sondas foram alinhadas para posições em diferentes pseudo-cromossomos. Em 37 destes casos, ambos os locos foram encontrados nos 11 pseudo-cromossomos principais e em 18 casos um dos locos foi encontrado em um *scaffold* não ancorado. A frequência de sondas DArT multiloco em cromossomos diferentes observada em *Eucalyptus* (55 em 6.631, ou seja, 0,83%) é consistente com a frequência de 1,4% observada em um estudo de mapeamento de ligacão em cevada (Wenzl, Li et al. 2006).

3.4.6. Cobertura do espaço gênico de *Eucalyptus* pelos marcadores DArT

Para caracterizar o espaço gênico de *Eucalyptus* coberto pelo microarranjo DArT, foram consideradas apenas as sondas alinhadas nos 11 pseudo-cromossomos anotados. Estas totalizaram 6.390 sondas, as quais alinharam em um total de 6.571 posições, uma vez que 181 sondas também alinharam em uma segunda posição, de acordo com o limiar de BWA adotado. A distribuição das 6.571 sondas DArT alinhadas e dos 1.986 marcadores DArT mapeados geneticamente, foi plotada em conjunto com a distribuição dos 41.204 modelos gênicos preditos no genoma do *Eucalyptus* (versão 1.0). O genoma foi particionado em 122 intervalos de 5 Mpb cada, os quais, em média, correspondem a ~10 cM de distância no intervalo de mapa, assumindo um total de 1.200 cM de distância de recombinação (Figura 7). O histograma indica que o microarranjo DArT proporciona uma cobertura consideravelmente homogênea de marcadores ao longo do genoma e sugere uma relação constante entre o número de

modelos gênicos e o número de marcadores DArT. Esta relação foi evidenciada através de uma correlação de Spearman relativamente forte e altamente significativa entre o número de modelos gênicos preditos e o número total de sondas DArT encontradas em cada intervalo do genoma (ρ = 0,682, ρ = 3,79e-18), e do mesmo modo, com o número de marcadores DArT mapeados (ρ = 0,467, ρ = 5,19e-8) (Figura 8). Estes resultados indicam que o microarranjo DArT tende a fornecer marcadores segregantes em essencialmente todos os intervalos genômicos de 5 Mpb com o número de marcadores DArT aumentando com o número de genes dentro do intervalo. Em média, cada intervalo contém 16±9,0 marcadores geneticamente mapeados, 53 ±21,6 sondas DArT e 334 ±100,9 modelos gênicos preditos. Apenas quatro intervalos tinham menos de 20 sondas DArT mapeadas e somente 11 dos 122 intervalos tinham menos de 5 marcadores mapeados geneticamente. Além disso, pouco menos de 70% das sequências DArT foram mapeadas a zero pb a partir do modelo gênico predito mais próximo, e menos de 10% localizadas a uma distância maior do que 10 kpb dos modelos gênicos preditos (Figura 9).

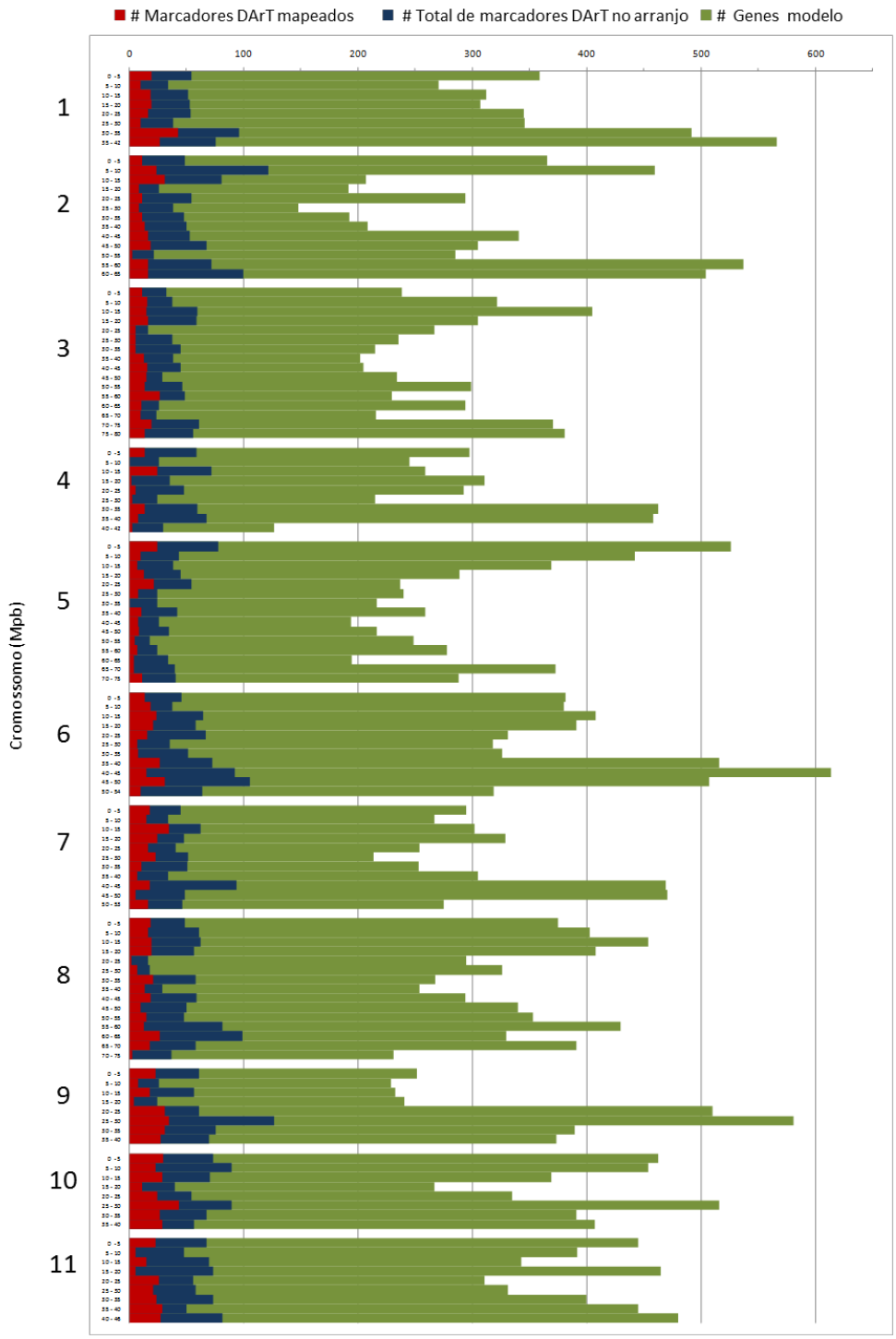
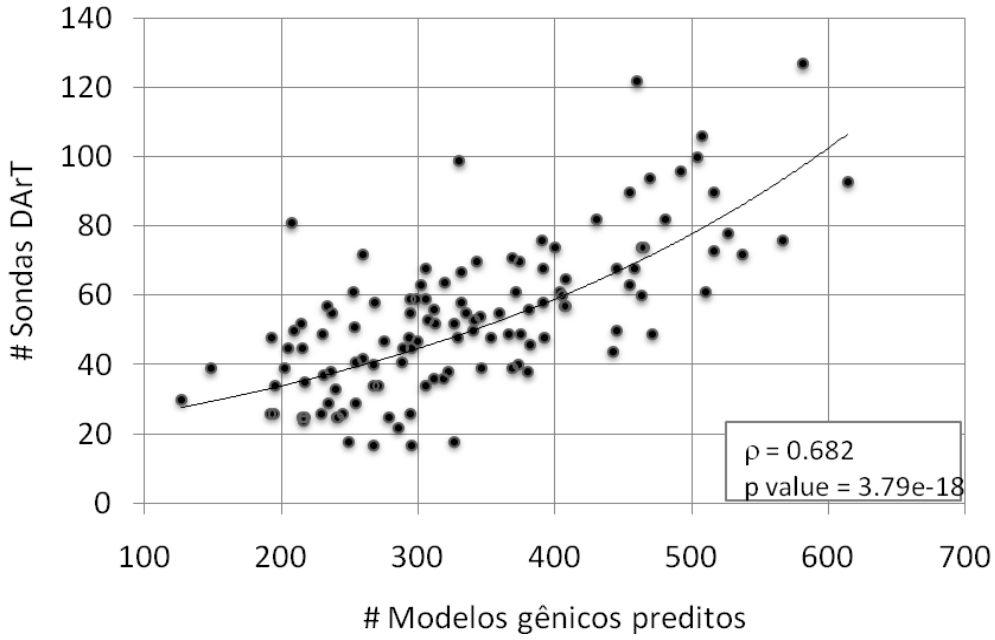


Figura 8. Correspondência posicional entre marcadores DArT e modelos gênicos preditos no genoma de *Eucalyptus grandis*. Os 11 pseudo-cromossomos do genoma de *Eucalyptus grandis* (Versão 1.0 no Phytozome 6.0), foram divididos em 122 intervalos de 5 Mpb. Para cada intervalo foram plotados o número de sondas DArT (barras azuis), o número de marcadores DArT mapeados geneticamente (barras vermelhas) e o número de modelos gênicos preditos (barras verdes).

A



B

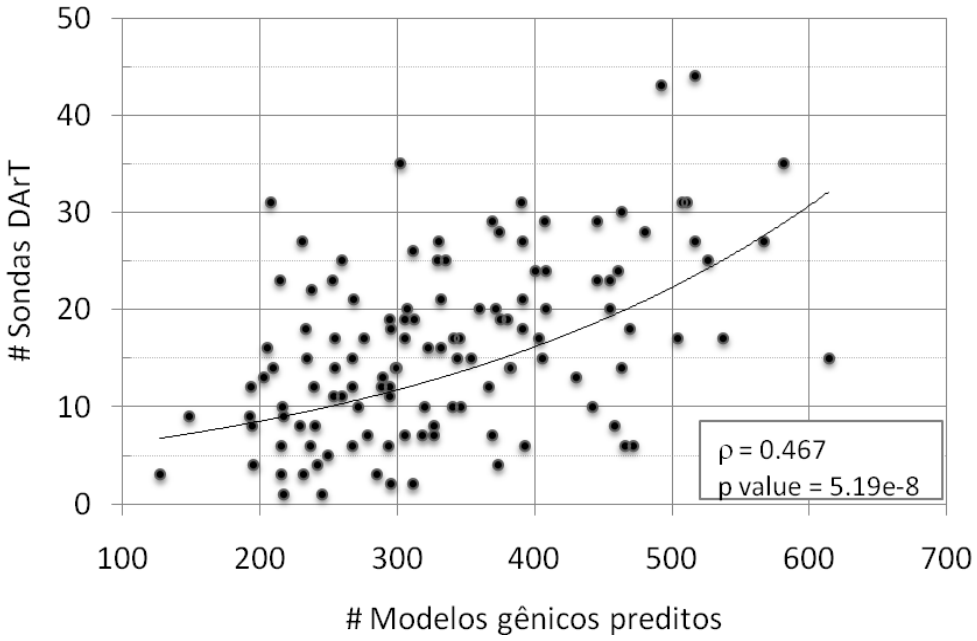


Figura 9. Correlações entre o número de sondas DArT, marcadores DArT mapeados e modelos gênicos no genoma de *Eucalyptus*. Correlações de Spearman (ρ) foram estimadas entre: (A) o número de sondas DArT e o número de modelos gênicos, e (B) o

número de marcadores DArT mapeados e o número de modelos gênicos, para intervalos genômicos de 5 Mpb.

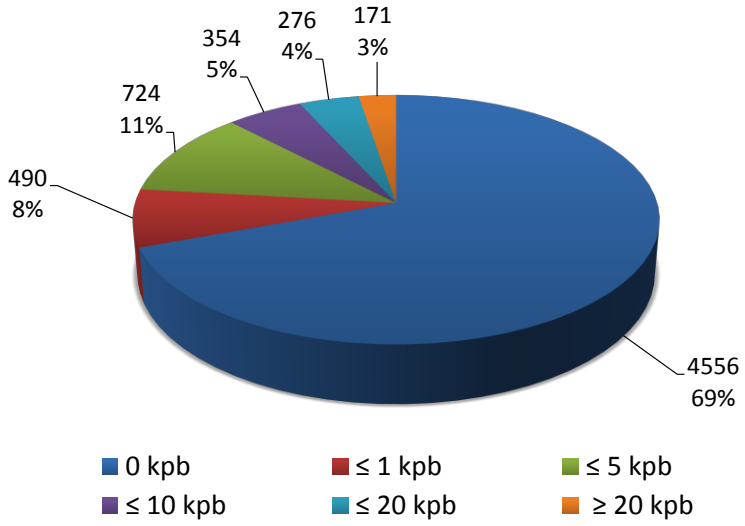


Figura 10. Distribuição do número e porcentagens de marcadores DArT de acordo com intervalos crescentes de distância física entre o marcador DArT e o modelo gênico mais próximo no genoma do *Eucalyptus*.

3.5. DISCUSSÃO

Este estudo fornece dados inéditos e detalhados sobre os atributos genômicos dos marcadores DArT no genoma de uma planta. Após o desenvolvimento de um microarranjo DArT de alto desempenho para *Eucalyptus* (Sansaloni, Petroli et al. 2010) examinamos as propriedades genômicas das sondas DArT que integram este microarranjo por meio do seu sequenciamento, a construção de um mapa de ligação de alta densidade e a comparação física com a sequência genômica de *Eucalyptus grandis* BRASUZ1 recentemente disponibilizada. Demonstramos que as sondas DArT têm preferencialmente o espaço gênico como alvo e exibem uma distribuição uniforme no genoma todo, proporcionando uma excelente cobertura genômica para aplicações em estudos de melhoramento e diversidade. Tais propriedades gerais de marcadores DArT não tinham sido ainda descritas, apesar dos vários estudos já

publicados com base na aplicação destes marcadores em um grande número de espécies de plantas.

3.5.1. Eficiência da genotipagem de marcadores DArT para análise genética em *Eucalyptus*

Um conjunto de parâmetros relativamente rigorosos de filtragem foi aplicado às intensidades de sinal obtidas a partir das 7.680 sondas DArT no microarranjo de *Eucalyptus*. Um total de 3.191 marcadores (41,5%) passou todos os limiares de qualidade do marcador e de *call rate* (Figura 2), uma proporção consistente com as estimativas iniciais relatadas durante a validação do microarranjo (Sansaloni, Petroli et al. 2010) e estudos de mapeamento recentes com *pedigrees* interespecíficos similares (Hudson, Kullan et al. 2011; Kullan, van Dyk et al. 2012). Além da filtragem pela qualidade do marcador, foi aplicada uma triagem estrita para examinar o comportamento mendeliano esperado. Uma distorção da segregação foi relatada em alguns estudos anteriores de mapeamento em *Eucalyptus*, embora, na maioria das vezes, em uma proporção não diferente da esperada pelo simples acaso (Grattapaglia and Sederoff 1994; Brondani, Williams et al. 2006). Isto se tornou um tema de interesse como uma forma de avaliar as interações heteroespecíficas nos híbridos F1 que afetam as taxas de introgressão entre espécies distantes (Myburg, Vogl et al. 2004). Neste *pedigree* particular, no entanto, assim como no primeiro estudo de mapa de ligação em *Eucalyptus* (Grattapaglia and Sederoff 1994), não seria de se esperar distorção de segregação, uma vez que a segregação do marcador foi observada a partir de cada uma das espécies puras parentais e não nos gametas derivados de um híbrido F1 onde uma distorção poderia ser esperada. Assim, apenas 215 marcadores (6,7%) foram excluídos devido a desvios de segregação esperados, uma proporção próxima dos 5% esperados pelo acaso. Além do erro de amostragem, estes marcadores distorcidos podem incluir casos de locos duplicados, tal como as 55 sondas DArT alinhadas com alta confiabilidade em posições de pseudo-cromossomos diferentes, e as 13 sondas alinhadas a distâncias maiores do que 10 kpb. Em ambos os casos, estas ocorrências podem dar origem a uma mistura dos sinais de hibridação e taxas de

segregação distorcidas. Os 97% de cobertura do genoma físico total fornecido pelo mapa de ligação construído neste estudo indica que a manutenção de marcadores distorcidos na análise de ligação poderia não melhorar a cobertura, mas, em vez disso, poderia complicar o ordenamento do marcador se as distorções derivadas de dados perdidos em excesso ou de erros de genotipagem fossem incluídas (Hackett and Broadfoot 2003).

Dentre os 2.976 marcadores DArT incluídos na análise de ligação, 2.274 foram por fim mapeados e ordenados no mapa consenso *full* e 864 na versão *framework* (Tabela 1). A proporção de marcadores DArT mapeados está dentro do número previsto de marcadores úteis para o mapeamento, entre 1.818 e 2.553, tal como foi inicialmente relatado quando o microarranjo DArT para *Eucalyptus* foi desenvolvido (Sansaloni, Petroli et al. 2010). Este número também é comparável aos 2.229 marcadores DArT mapeados no mapa consenso de duas famílias de retrocruzamento relacionadas *E. grandis* x *E. urophylla* (Kullan, van Dyk et al. 2012) e para os 1.845 ESTs baseados em marcadores *Single Feature Polymorphism* (SFP) relatado anteriormente para o mesmo *pedigree* utilizado neste estudo (Neves, Mamani et al. 2011). A proporção de marcadores informativos em tais *pedigrees* interespecíficos de mapeamento foi consideravelmente maior do que o número observado em *pedigrees* intraespecíficos. Hudson *et al*. (Hudson, Kullan et al. 2011) puderam mapear somente 1.060 marcadores DArT em uma família F2 derivada de dois indivíduos F1 da mesma procedência e apenas 569 em um cruzamento da mesma origem em *E. globulus*. A divergência genética entre as espécies e os níveis correspondentes de heterozigosidade diferencial de sequência nos locos DArT determinam, em grande parte, a proporção de marcadores informativos capturados. A plataforma de genotipagem DArT tem proporcionado uma ordem de magnitude maior no número de marcadores para o mapeamento em *Eucalyptus* do que as tecnologias anteriores, como RAPD, AFLP e microssatélites (Grattapaglia and Kirst 2008). A natureza não domesticada do *Eucalyptus* resultou em um maior número de marcadores mapeados do que a maioria dos mapas de ligação construídos com microarranjos DArT extensivamente otimizados tais como aqueles para trigo, aveia, (Tinker, Kilian et al. 2009), sorgo (Mace, Rami et al. 2009), e cevada (Wenzl, Li et al. 2006).

Recentemente, entre 3.100 e 3.500 marcadores polimórficos DArT de alta qualidade foram detectados em populações de melhoramento compostas de várias centenas de indivíduos de *E. grandis* e *E. urophylla*, no contexto de experimentos de seleção genômica (Resende, Resende et al. 2012). Como esperado, quando comparado com uma média de 2.200 a 2.300 marcadores capturados e mapeados em *pedigrees* biparentais, o microarranjo DArT fornece entre 40 e 50% de marcadores a mais quando em nível de população. Se proporções semelhantes são mantidas, pode-se antecipar que, em populações de melhoramento de *E. globulus*, o microarranjo DArT irá proporcionar entre 750 e 1.500 marcadores informativos, dependendo da variabilidade geral da população e da procedência.

3.5.2. Estimativas de redundância das sondas DArT

Estimativas reportadas da redundância de marcadores DArT obtidas comparando o padrão de segregação em população de mapeamento ou estimando as distâncias de *Hamming* entre os marcadores, têm variado de 38% em cevada (Wenzl, Li et al. 2006), a 43% em *Arabidopsis* (Wittenberg, Lee et al. 2005). Depois de sequenciar todos as sondas do microarranjo DArT para *Eucalyptus*, foi calculada uma redundância com base na comparação direta de sequências entre 33,75 e 44,14% (Tabela 3). Isto é consistente com a taxa de redundância de 46% para milhares de sondas DArT sequenciadas a partir do microarranjo de genotipagem para aveia (Tinker, Kilian et al. 2009). Sob o mesmo protocolo de redução da complexidade genômica, a redundância irá variar de acordo com a estrutura do genoma específico da espécie-alvo, a diversidade das amostras utilizadas para construir representações genômicas e, em grande parte, com o critério de triagem das sondas e o número final de sondas selecionadas. Quanto mais sondsa são analisadsa, uma maior redundância irá resultar. A redundância pelo sequenciamento de apenas algumas centenas de sondas DArT, previamente selecionadas para o polimorfismo, foi estimada em 15% para um microarranjo de maçã, apesar de que uma redundância potencial de 50% foi reconhecida para todos os clones do microarranjo sequenciado (Schouten, Weg et al. 2012). Considerando que uma distribuição física e de mapeamento genético

razoavelmente uniforme das sondas DArT foi alcançada através do genoma de *Eucalyptus* (Figuras 6 e 7), um certo nível de redundância das sondas no microarranjo DArT pode ser uma característica desejável. As sondas DArT de *Eucalyptus* variam em tamanho (534 ±215 pb) e, embora compartilhem porções de sequência de DNA, terão capacidade variável para detectar polimorfismos de sequência entre indivíduos e populações, proporcionando assim uma melhora no potencial e na flexibilidade de genotipagem no genoma completo.

Um aspecto interessante da análise de redundância de sequência das sondas DArT surgiu, entretanto, quando se comparou o alinhamento destas com o genoma de referência. Entre 3.864 e 4.583 sequências únicas foram observadas para as sondas DArT (Tabela 3), uma estimativa consistente com as 4.189 sondas DArT alinhadas com confiabilidade em uma posição única dentro do genoma. No entanto, o alinhamento gerado pelo programa BWA para o genoma revelou que 2.252 sondas adicionais mapearam em posições exclusivas, de modo que, no total, 6.441 locos únicos não redundantes no genoma foram obtidos através do microarranjo DArT indicando que a redundância real é muito menor do que aquela estimada pela simples comparação de sequências. Esta é a primeira vez que uma análise deste tipo foi efetuada numa espécie para a qual um microarranjo DArT e um genoma de referência estão disponíveis. Esta avaliação comprova que as estimativas de redundância baseadas em uma simples montagem de sequências de sondas DArT tendem a ser altamente conservadoras e que os marcadores DArT de Eucalyptus tem, na verdade muito baixa redundância quando avaliados em termos de posições únicas no genoma.

3.5.3. O mapa genético *framework* permite estimativas mais confiáveis da relação kpb/cM no genoma de *Eucalyptus*

Duas versões do mapa de ligação foram construídas neste estudo, cada uma com objetivos específicos. Um mapa *framework* foi construído como a "hipótese" que melhor explicou os dados de segregação observados, para fornecer informações mais precisas referentes ao ordenamento dos marcadores (Keats, Sherman et al. 1991; Vision, Brown et al. 2000) e para ser usado na estimativa da relação entre a distância

física e a fração de recombinação (Figura 3). Por outro lado, o mapa *full*, que incluía todos os marcadores DArT segregantes, foi construído para proporcionar uma posição preliminar para todos os marcadores DArT possíveis e, assim, permitir uma avaliação mais extensiva da cobertura do genoma e da distribuição de marcadores DArT relativa aos genes. Além disso, através da inclusão de um maior número de marcadores (mesmo com ordenamento mais permissivo), este mapa ofereceu uma melhor probabilidade de alocação dos *scaffolds* não ancorados aos pseudo-cromossomos do atual genoma referência de *Eucalyptus*. O ordenamento dos marcadores de ambos os mapas, *framework* e *full*, foi geralmente comparável e os tamanhos totais dos mapas foram também próximos (Figura 4 e Tabela 1). No entanto, conjuntos de marcadores invertidos foram observados entre estas duas versões do mapa, bem como marcadores que não mapearam quando se passou de um mapa para o outro e *vice versa*. Estes resultados fundamentam o fato bem conhecido de que a resolução da ordem do marcador não é uma questão trivial ainda mais quando um grande número de marcadores é mapeado com um tamanho de progênie limitado (Cheema and Dicks 2009). Esta observação também suporta o fato de que inversões semelhantes e não colinearidades relatadas em estudos anteriores de mapeamento comparativos entre espécies de *Eucalyptus* (Hudson, Kullan et al. 2011) são, em geral, reflexos de ordenamentos inconsistentes de marcadores devido a várias fontes de erro experimental (Hackett and Broadfoot 2003) e raramente podem ser tomadas como evidências de qualquer ocorrência biológica relevante, a menos que dados de validação independente estejam disponíveis. Com a disponibilidade de um genoma de referência para *Eucalyptus*, aliada a tecnologias de sequenciamento de alto rendimento e procedimentos de montagem poderosos, tais validações podem agora tornar-se possíveis.

Uma simples inspeção visual dos mapas, *framework* e *full*, alinhados (Figura 4) e a diferença significativa encontrada entre as distribuições das distâncias de mapa entre marcadores consecutivos nas duas versões de mapa ($p = 0,021$) (Figura 5), sustentam a conclusão de que a construção do mapa *framework* remove, em grande parte, conjuntos altamente clusterizados de marcadores DArT, deixando um mapa esparso com uma cobertura genômica equivalente e suporte estatístico melhorado para o

ordenamento relativo. Além disso, quando o mapa de ligação de microssatélites e DArTs foi comparado com um mapa constituído unicamente pelo marcador co-dominante, a distância de recombinação total não foi alterada (dados não apresentados). Este resultado adicional sugere que, enquanto os microssatélites fornecem uma cobertura adequada do genoma, os marcadores DArT efetivamente cobrem o genoma em regiões não amostrados anteriormente, proporcionando assim a densidade necessária para mapeamento de alta resolução e estudos genômicos mais detalhados. O mapa *framework* deveria, portanto, ser tomado como o mapa mais confiável quando se trata de análise comparativa com o genoma de referência ou estudos com base em mapas (tais como procurar a co-localização de genes com potenciais QTLs de grande efeito).

O alinhamento e mapeamento físico de 869 marcadores *framework*, DArTs e microssatélites, nos 11 *scaffolds* principais da sequência genômica de *Eucalyptus* permitiu uma estimativa da relação entre distância física e fração de recombinação em cada pseudo-cromossomo e para todo o genoma. Esta estimativa variou consideravelmente (357-736 Kpb/cM), com uma média genômica de 513 kpb/cM (Tabela 1). Interessante observar que esta estimativa coincide com as estimativas preliminares de 395-559 kpb/cM relatadas com base nos primeiros mapas de ligação disponíveis (Grattapaglia and Sederoff 1994) e as primeiras estimativas do tamanho do genoma de *Eucalyptus* (Grattapaglia and Bradshaw 1994). Desta vez, contudo, utilizando a sequência do genoma no qual os marcadores *framework* foram mapeados fisicamente, uma estimativa melhorada foi possível. Usando uma abordagem diferente, baseada exclusivamente em um conjunto selecionado de 153 pares de marcadores flanqueadores mapeados em aproximadamente 1 cM de distância, uma estimativa de 633 kpb/cM foi proposta (Kullan, van Dyk et al. 2012). Além do potencial viés introduzido pelos pares de marcadores especificamente selecionados e, como consequência, impedindo a variação intrínseca entre distância de recombinação e distância física ao longo do genoma, esta estimativa foi baseada em marcadores ordenados com uma probabilidade permissiva. Consideramos, portanto, as estimativas apresentadas neste trabalho, tanto em nível pseudo-cromossômico como de média genômica as melhores aproximações para *Eucalyptus* até o momento (Tabela 1),

apesar de reconhecer que a taxa de recombinação ao longo do genoma pode variar em ordens de magnitude (Rafalski and Morgante 2004).

3.5.4. O alinhamento do mapa genético à sequência física sugere uma característica pan-genômica do microarranjo DArT e a completude da montagem do genoma de *Eucalyptus*

A consistência entre a ordem dos marcadores DArT *framework* mapeados e sua ordem física na sequência do genoma demonstra a qualidade da montagem dos *scaffolds* do genoma atual de *Eucalyptus* (Figura 6). Enquanto o mapa de ligação relatado por Kullan *et al*. (Kullan, van Dyk et al. 2012) foi utilizado de forma eficaz para assistir no ordenamento de *scaffolds* durante a montagem do genoma (J. Schmutz, comunicação pessoal) o mapa de ligação aqui apresentado não foi, e, portanto, constitui uma validação independente do atual genoma de *Eucalyptus*. Além disso, do ponto de vista de completude do genoma, apenas 45 marcadores dos 2.274 mapeados não puderam ser alinhados nos 11 *scaffolds* principais. Por outro lado, apenas 89 sondas permaneceram fisicamente não mapeadas no genoma. Embora estes marcadores não mapeados possam corresponder a seções faltantes da montagem do genoma, também podem corresponder a sequências que não existem no genoma de *E. grandis*, lembrando que as 7.680 sondas do microarranjo foram desenvolvidas a partir de 18 representações genômicas envolvendo 64 espécies diferentes de *Eucalyptus* com uma ampla diversidade filogenética. Uma análise minuciosa da fonte original dessas 89 sondas DArT não mapeadas revelou que 65 delas vieram de bibliotecas de representação genômicas construídas com DNA de espécies diferentes de *E. grandis*, enquanto que 24 sondas eram provenientes desta espécie (Sansaloni, Petroli et al. 2010). No entanto, foi observado um número significativamente maior de sondas DArT com origem diferente de *E. grandis*, não mapeadas no genoma, do que o que seria esperado pelo simples acaso (Pearson Qui-quadrado 5,03; valor de p = 0,0249). Este resultado, em conjunto com o excelente desempenho do microarranjo em investigações filogenéticas do gênero (Steane, Nicolle et al. 2011), sugere um atributo pan-genômico único deste microarranjo DArT para *Eucalyptus*. Enquanto a maioria das

sondas corresponde ao componente núcleo ("core") que envolve características genômicas comuns entre indivíduos e espécies de *Eucalyptus*, algumas sondas podem ser derivadas a partir do "genoma dispensável" composto de elementos de sequências de DNA parcialmente compartilhadas e/ou não compartilhadas entre as espécies (Morgante, De Paoli et al. 2007; Cao, K. Schneeberger et al. 2011).

A completude da montagem atual do genoma também foi sustentada pela observação de que, dos 85,4 Mpb de sequência não ancorada dos 4.941 *scaffolds* pequenos, apenas 1,4 Mpb foram capturados por 45 marcadores mapeados em 31 *scaffolds* e a maioria foi localizada em posições intermediárias ao longo dos grupos de ligação e não nos extremos (Tabela 2). Se a montagem do genoma fosse incompleta, seria de se esperar a captura de uma proporção consideravelmente maior de *scaffolds* e sequências não ancoradas. De fato, durante a montagem do genoma de *Populus*, Drost *et al*. (Drost, Novaes et al. 2009), valendo-se de um mapa de densidade média com 608 marcadores, foram capazes de ancorar 116 sequências *scaffolds* em posições genéticas únicas dentro dos grupos de ligação, o que estendeu o genoma em aproximadamente 35,7 Mpb de sequência dos 75 Mpb ainda não ancorados até aquele momento. Estes resultados, junto com o fato de que 86% dos 4.941 *scaffolds* não ancorados do genoma de *Eucalyptus* são menores do que 20 kpb, sugerem que a grande maioria dos *scaffolds* não ancorados corresponde a fragmentos de haplótipos alternativos de pseudo-cromossomos já montados, possivelmente derivados de regiões de elevada heterozigosidade no genoma do *Eucalyptus* e não a porções faltantes na montagem do genoma.

3.5.5. O microarranjo DArT oferece cobertura uniforme do genoma e amostra preferencialmente regiões ricas em genes

Os resultados a partir da análise de BLAST das sequências de sondas DArT no genoma de *Eucalyptus* (Figura 7) corroboram estudos anteriores em outras espécies vegetais relatando que marcadores DArT obtidos com base na endonuclease *Pst*I estão predominantemente localizados em regiões do genoma de baixa cópia e ricas em genes (Akbari, Wenzl et al. 2006; Wenzl, Li et al. 2006; Tinker, Kilian et al. 2009). No

entanto, a possibilidade de mapear as sondas DArT em um genoma de referência anotado além de uma análise BLAST simples contra ESTs, revelou uma relação significativa entre o número de marcadores DArT e os modelos gênicos preditos (Figura 8), com uma pequena proporção de sondas DArT localizadas a mais de 10 kpb do gene mais próximo (Figura 9). Este resultado é significativo, pois pode ajudar a explicar o excelente nível de resolução que o microarranjo DArT tem proporcionado para estudos de genética populacional e melhoramento em uma ampla representação de espécies de *Eucalyptus* (Steane, Nicolle et al. 2011; Resende, Resende et al. 2012). Baseadas na caracterização genômica das sondas DArT relatadas neste estudo, pesquisas sobre filogenia ou genética de populações com base nestes marcadores podem agora ser realizadas explorando a proximidade dos marcadores a genes ou o conteúdo gênico de conjuntos específicos de marcadores. Como alternativa, os marcadores DArT provenientes de genes específicos podem ser selecionados *a priori* para reconstruções filogenéticas diferenciais. Além disso, a combinação entre a cobertura genômica fornecida e a associação predominante com o espaço gênico pode também ter contribuído para o bom desempenho do microarranjo DArT no fornecimento de marcadores para modelos preditivos precisos em estudos recentes de seleção genômica (SG) (Resende, Resende et al. 2012). Torna-se possível agora correlacionar os atributos genômicos dos marcadores DArT com suas contribuições específicas para a capacidade preditiva de modelos de SG ou para a resolução de filogenias específicas e, portanto, apontar para marcadores específicos ou segmentos genômicos de particular interesse em estudos posteriores.

3.6. CONCLUSÃO

Os resultados deste trabalho, seguindo os estudos genéticos em *Eucalyptus* recentemente publicados e baseados em marcadores DArT (Hudson, Kullan et al. 2011; Steane, Nicolle et al. 2011; Kullan, van Dyk et al. 2012; Resende, Resende et al. 2012), destacam o valor desta plataforma de genotipagem para pesquisas genéticas, melhoramento e evolução do genoma em espécies deste gênero. Dada a uniformidade dos métodos utilizados no desenvolvimento de microarranjos DArT, as propriedades

genômicas dos marcadores descritos neste estudo são possivelmente comuns à maioria, se não todos, os genomas de angiospermas. A tecnologia DArT tem evoluído atualmente aproveitando o sequenciamento de alto desempenho de *reads* curtos (Sansaloni, Petroli et al. 2011). Ao combinar o método estabelecido de redução de complexidade do genoma, também adotado pelos protocolos recentemente descritos de genotipagem por sequenciamento (Genotyping-by-Sequencing - GbS) (Elshire, Glaubitz et al. 2011; Poland, Brown et al. 2012), um salto considerável tem ocorrido na capacidade de detecção de polimorfismos. No entanto, os atributos genômicos gerais de marcadores obtidos via GbS, assim como a cobertura do genoma e a ocorrência preferencial em regiões ricas em genes, deverão ser essencialmente os mesmos que os descritos no presente estudo. GbS, entretanto, fornece a vantagem potencial adicional de que um número muito maior de marcadores com base em *counts* de sequência digitais em vez do sinal analógico do microarranjo é obtido, além da possibilidade de genotipar marcadores co-dominantes SNPs. Este avanço poderá incentivar uma diminuição nos atuais custos de genotipagem em grande escala para plantas, além do que as plataformas de DArT e SNP têm feito nos últimos anos. Porém, a infraestrutura de informática necessária para manejar, armazenar e analisar os arquivos enormes de sequência gerados por GbS para milhares de amostras não será imediatamente disponível na esfera da maior parte dos programas de recursos genéticos e melhoramento de plantas. A genotipagem baseada no microarranjo DArT com seus protocolos de processamento e análise padronizados deverão, portanto, continuar a ser uma ferramenta útil para uma série de aplicações na análise genética de plantas, especialmente aquelas que não necessariamente requerem uma genotipagem de muito alta densidade.

4. CAPÍTULO 2

MAPEAMENTO DE QTLs PARA CARACTERÍSTICAS DE IMPORTÂNCIA ECONÔMICA E ANÁLISE DO CONTEUDO GÊNICO NOS INTERVALOS GENÔMICOS CORRESPONDENTES

4.1. INTRODUÇÃO

O mapeamento de QTLs (Quantitative Trait Loci) é uma abordagem não enviesada na qual a variação fenotípica observada é analisada frente à segregação de marcadores discretos, os quais revelam a localização de regiões genômicas que afetam a característica fenotípica mensurada. O princípio do mapeamento de QTLs é associar estatisticamente as regiões genômicas responsáveis pelo controle genético da característica com a segregação de marcadores moleculares (Jones, Ougham et al. 2009). A precisão da localização de QTLs é limitada pela informação, em particular o número de recombinantes, que é obtido a partir da observação dos genótipos dos marcadores. Estes recombinantes observados podem ser limitados pelo tamanho reduzido da amostragem e a falta de dados genotípicos (Mackay 2001; Doerge 2002). As estimativas dos efeitos de QTLs são geralmente confusas por causa da distância entre o marcador e o QTL. Este problema é minimizado quando é utilizado um mapa de alta densidade, e o QTL é localizado em intervalos mais estreitos entre marcadores adjacentes ligados (Thoday 1961; Lander and Botstein 1989). Porém, apesar dos avanços nas plataformas de genotipagem, os de QTLs detectados têm capturado proporções limitadas da variação genética. O desenvolvimento de marcadores moleculares transferíveis e o aumento do uso de *pedigrees* múltiplos para mapeamento de QTLs permitirão realizar análises comparativas de QTLs detectados em estudos independentes, fornecendo assim, dados de validação. A informação do posicionamento do QTL, juntamente com a disponibilidade de sequências genômicas anotadas, abre a perspectiva de identificar genes candidatos ao QTL para características complexas (Price 2006).

Em *Eucalyptus*, vários estudos têm identificado QTLs de maior efeito responsáveis pelo controle de características quantitativas de importância econômica. Esses locos controladores estão envolvidos no desenvolvimento de componentes de produtividade (crescimento volumétrico e forma), qualidade da madeira (densidade básica, teor de lignina, rendimento em celulose), resistência a estresses abióticos (tolerância ao frio e à seca) e resistência a patógenos, principalmente fungos (Junghans, Alfenas et al. 2003; Freeman, O'Reilly-Wapstra et al. 2008; Grattapaglia, Plomion et al. 2009; Mamani, Bueno et al. 2010; Gion, Carouche et al. 2011). Entretanto, apesar de QTLs de maior efeito terem sido mapeados, a informação gerada tem sido pouco útil para a seleção assistida por marcadores (SAM). As principais razões pelas quais estes locos não têm auxiliado de maneira eficaz os programas de melhoramento florestal foram amplamente discutidas (Grattapaglia and Kirst 2008; Grattapaglia, Plomion et al. 2009; Grattapaglia and Resende 2011). Dentre esses motivos, podemos considerar a detecção limitada de variação alélica, uma vez que apenas a variação alélica dos parentais da população é amostrada; um efeito dos QTLs superestimado devido ao pequeno tamanho das populações utilizadas; e o comportamento imprevisível da interação entre alelos favoráveis aos QTLs em diferentes *backgrounds* genéticos, diferentes locais e idades das populações envolvidas (Grattapaglia, Plomion et al. 2009; Grattapaglia and Resende 2011). Outro fator limitante na identificação e utilização efetiva de QTLs é a restrita resolução e cobertura genômica fornecida pelos marcadores moleculares. Estudos de QTL identificam regiões genômicas amplas que incluem várias centenas de genes ou elementos reguladores e, por isso, representam apenas um passo muito preliminar na identificação dos polimorfismos causantes (Grattapaglia and Kirst 2008).

Embora muitas tecnologias de marcadores moleculares tenham sido desenvolvidas com grande sucesso para *Eucalyptus* nas últimas décadas (Grattapaglia and Sederoff 1994; Gaiotto, Bramucci et al. 1997; Brondani, Brondani et al. 1998; Brondani, Williams et al. 2006; Grattapaglia, Silva-Junior et al. 2011; Neves, Mamani et al. 2011), todas elas apresentam limitações. Por exemplo, no caso de microssatélites, a descoberta e validação de um grande número de marcadores que cobrem o genoma inteiro é lenta porque envolve várias etapas. microssatélites e SNPs, têm um custo alto por basear-se

em informação de sequência, embora esta tenha se tornado bem mais accessível hoje. O desenvolvimento de técnicas robustas que permitam a genotipagem de milhares de marcadores em milhares de amostras em um simples experimento resolveria grande parte das limitações mencionadas. Assim, para aplicações que demandam uma análise ampla do genoma, a tecnologia ideal deve oferecer não somente milhares de marcadores moleculares cobrindo todo o genoma, mas também estes devem ser obtidos preferencialmente em um experimento único, simples e de baixo custo. Neste sentido, a metodologia DArT (Diversity Arrays Technology) foi desenvolvida para atender várias das limitações expostas por outros métodos. Descrita no ano 2001 (Jaccoud, Peng et al. 2001), esta tecnologia de genotipagem apresenta uma série de vantagens que complementam com eficiência as metodologias de análise de polimorfismo atualmente utilizadas, tais como microssatélites, AFLP e SNP. A metodologia DArT normalmente fornece sinal consistente mesmo entre espécies relacionadas, um recurso especialmente valioso em *Eucalyptus*. Os clones de DNA que compõem o arranjo podem ser sequenciados e as sequências compartilhadas e usadas como marcadores âncoras robustos para mapeamentos de QTL comparativos ou para explorar o genoma referência em projetos de clonagem posicional. Esta tecnologia tem mostrado resultados extremamente positivos em estudos que variam desde a obtenção de perfis genéticos para a identificação individual até a análise de diversidade (Wenzl, Carling et al. 2004; White, Law et al. 2008; Steane, Nicolle et al. 2011; Zhang, Liu et al. 2011; He and Bjørnstad 2012). Além disso, possibilita a construção rápida de mapas de ligação de alta densidade (Wenzl, Li et al. 2006; Hippolyte, Bakry et al. 2010; Milczarski, Bolibok-Bragoszewska et al. 2011; Oliver, Jellen et al. 2011; Thudi, Bohra et al. 2011), mapeamento físico, em projetos que envolvem sequenciamento dos marcadores (Paux, Sourdille et al. 2008; Rodríguez-Suárez, Giménez et al. 2012), seleção assistida por marcadores (McCartney, Stonehouse et al. 2011) e seleção genômica ampla (Crossa, Campos et al. 2010; Grattapaglia, Sansaloni et al. 2010; Resende, Resende et al. 2012). Especificamente no que refere à identificação de QTLs, este marcador dominante tem demonstrado eficiência na captura de locos ou regiões genômicas controladoras da expressão de algumas características em cevada e trigo (Bedo, Wenzl et al. 2008; Huynh, Wallwork et al. 2008; Sadeque and Turner 2010; Zhang, Dong et al. 2011).

Como grandes *pedigrees* tornaram-se disponíveis e o mapeamento de alta resolução com tecnologias tais como SNPs, DArTs e Genotipagem por Sequenciamento se tornaram rotina em árvores, a informação posicional de QTLs poderia ser uma alternativa às atuais abordagens que dependem de genes candidatos tentativos para estudos de associação genética. Entre as várias características para as quais QTLs foram mapeados em espécies florestais aquelas que apresentaram maior herdabilidade, tais como composição química da madeira, são mais propensos a envolver genes candidatos de maior efeito, embora recentes estudos de associação mostrem que mesmo tais genes explicam uma proporção muito pequena da variação (Wegrzyn, Eckert et al. 2010).

Neste segundo capítulo foi utilizado um mapa genético construído com mais de 2.000 marcadores DArT e microssatélites para realizar um experimento de mapeamento de QTLs. Com base nos QTLs detectados, foi realizada uma avaliação do número de genes anotados na sequência do genoma de referência de *Eucalyptus* que estariam incluídos no intervalo genômico correspondente a cada QTL identificado.

4.2. OBJETIVOS

1) Construção de um mapa genético com alto suporte estatístico para ordenamento de marcadores para cada genitor da família IP (*E. grandis* x *E. urophylla*).

2) Mapeamento de QTL's (Quantitative Trait Loci) para características de crescimento e qualidade da madeira, utilizando o mapa genético construído e os dados fenotípicos disponíveis.

3) Análise da co-localização de QTLs com genes anotados no genoma referência de *Eucalyptus.*

4.3. MATERIAL E MÉTODOS

4.3.1. Material vegetal

Um total de 177 indivíduos F1 derivados do cruzamento *E. grandis* x *E. urophylla*, assim como os seus parentais foram genotipados com o microarranjo DArT para *Eucalyptus* como descrito recentemente (Sansaloni, Petroli et al. 2010). Esta população de mapeamento é a mesma família IP utilizada na análise do capítulo anterior, porém, o mapeamento de QTLs foi realizado a partir de 171 indivíduos pertencentes à progênie F1 desta população. Os seis indivíduos faltantes da população original genotipada não foram incluídos na avaliação devido à ausência de mensurações fenotípicas. As medições fenotípicas de sete características foram utilizadas na análise: Crescimento em Altura (CA); Diâmetro à Altura do Peito (DAP); Densidade Básica da madeira (DB); Rendimento Depurado da Celulose (RDC); Lignina Total (LT); Relação Siringil/Guaiacil (S/G) e Penetração do Pilodyn (PP), uma medida indireta de densidade da madeira. Estes dados foram levantados ao longo do projeto Genolyptus envolvendo equipes de campo e de laboratórios de diversas instituições e fornecidos para este estudo. Notadamente os dados laboratoriais de qualidade da madeira foram levantados pela equipe do Laboratório de Celulose e Papel da Universidade Federal de Viçosa e os dados de crescimento em campo pela equipe de melhoramento da Fibria S.A. no experimento montado em Guaíba, RS aos 3 anos de idade.

4.3.2. Construção dos mapas e detecção de QTLs

Marcadores DArT foram combinados com microssatélites e um novo mapa de ligação foi construído para cada parental da família IP usando o *software* JoinMap v3.0 (Van Ooijen and Voorrips 2001). Ambos os mapas parentais foram montados com os mesmos critérios e valores limiares para cada parâmetro usados pelo *software* na produção do primeiro mapa (ver capítulo 1). Um mapeamento de QTLs para cada característica foi conduzido nos mapas parentais de maneira individual através do

método de mapeamento por Marca Simples, Intervalo Simples e Intervalo Composto, recursos disponíveis no *software* para Windows QTL Cartographer versão 2.5 (Wang, Basten et al. 2007). Para cada característica foi adotado um nível de significância de 5% com base em um teste envolvendo 1.000 permutações de acordo com o procedimento descrito por Churchill e Doege (1994) (Churchill and Doerge 1994) e implementado no QTL Cartographer. Isto permitiu determinar o LOD mínimo para detectar significância nas análises.

O mapeamento de intervalos utiliza um par de marcadores como unidade de análise ao invés de um marcador somente (Lander and Botstein 1989). O mapeamento por Intervalo Simples (IS) delimita um intervalo entre dois marcadores onde o programa realiza as análises na busca de um QTL. Valores de LOD são gerados dentro do intervalo a cada 1 cM e comparados com o LOD obtido por meio de permutações. No mapeamento por Intervalo Composto (IC), além do intervalo considerado na análise, são considerados também os outros QTLs presentes, ou seja, a correlação não é independente da existência de outros QTLs. Na análise de IC a escolha de cofatores, marcadores supostamente ligados a QTLs a serem incluídos como variáveis independentes no modelo de regressão múltipla, foi feita através de uma regressão *step-wise* (método "*forward*"). O número de cofatores variou para cada característica, de acordo com a recomendação de Zeng (1994) (Zeng 1994) de se testar múltiplos modelos para encontrar aquele que possui o melhor balanço entre os erros tipo I e II. Para todas as características foram feitas análises utilizando 5 e 8 cofatores através do modelo 6 do QTL Cartographer. Com isso, escolheu-se, para cada característica, o modelo cujo valor de cofator proporcionou o maior número de QTLs com maiores significâncias estatísticas (maiores valores de LOD).

4.4. RESULTADOS

4.4.1. Correlação entre as características fenotípicas

A partir dos dados coletados no campo para as características de crescimento e os dados de qualidade da madeira gerados no laboratório, foram estimados os valores

genéticos (VG) de cada uma das características. A análise estatística foi realizada por meio de um modelo misto no qual foram considerados os seguintes fatores: (1) super bloco – efeito aleatório; (2) blocos dentro de superblocos – efeito aleatório; (Brown, Kadel et al. 2001), falhas nas linhas, colunas e diagonais – todas efeitos aleatórios; (4) famílias – efeito fixo; (5) genótipos dentro de famílias – efeito aleatório. Seis dos indivíduos amostrados na genotipagem não foram fenotipados para as características avaliadas neste estudo. Na tabela 4 são apresentadas as estatísticas descritivas para as características fenotípicas avaliadas nos 171 irmãos-completos da família. As estatísticas descritivas contêm os valores máximos e mínimos, média, variância, desvio padrão e coeficiente de variação.

Tabela 4. Estatísticas descritivas das características fenotípicas avaliadas na população de mapeamento de 171 irmãos-completos.

Caraterísticas	Máximo	Mínimo	Média	Variância	Desvio padrão	Coeficiente de variação
DB (Kg/m^3)	469,5	429,0	448,3	70,3	8,4	0,018
Altura (m)	12,2	8,1	10,4	0,6	0,8	0,075
DAP (cm)	12,3	7,2	9,9	1,0	1,0	0,010
RDC (%)	56,6	50,9	53,4	0,7	0,8	0,015
LT (%)	27,2	22,9	25,4	0,5	0,7	0,027
S/G	2,9	2,0	2,4	0,0	0,2	0,065
Pilodyn (mm)	23,8	18,7	21,1	1,0	1,0	0,048

DB - densidade básica da madeira; DAP – diâmetro à altura do peito; RDC – rendimento depurado de celulose; LT – teor de lignina total; S/G – relação siringil/guaiacil; Pilodyn – penetração do Pilodyn.

4.4.2. Mapas genéticos e detecção de QTL (Quantitative Traits Loci)

A detecção de QTLs foi realizada usando um mapa de ligação genética *framework* construído para cada parental com alto suporte de verossimilhança (LOD > 3.0) para ordenamento dos marcadores dentro dos grupos de ligação. O mapa materno (*E. grandis*) teve um total de 825 marcadores (684 DArTs + 141 microssatélites) mapeados, cobrindo 1.993,63 cM do genoma. Já o mapa paterno (*E. urophylla*) constituiu-se de 511 marcadores (410 DArTs + 101 microssatélites), com um tamanho total de recombinação de 1.381,92 cM, ambos os mapas formados por 11 grupos de ligação consistente com o número de cromossomos da espécie e numerados de acordo com o mapa de referência (Brondani, Williams et al. 2006).

Conforme esperado, a densidade de marcadores do mapa construído para análise de QTLs cobriu boa parte do genoma. O fato de haver 314 (62%) marcadores segregando no parental G38 a mais em relação ao U15 é um indicativo de uma maior heterozigosidade no genitor *E. grandis*. Isso por sua vez resultou em maior cobertura genômica no mapa genético de G38 (1.993,63 cM) em relação a U15 (1.381,92cM), devido ao maior número de locos segregantes. Porém, isto aparentemente não teve reflexo no número de QTLs detectados no mapa de cada genitor (18 para G38 e 17 para U15), o que poderia ser explicado pela distribuição uniforme dos marcadores nos mapas individuais.

Nesta população segregante, 18 QTLs foram detectados em *E. grandis*, enquanto que um conjunto de 17 QTLs foi identificado em *E. urophylla* associados às características de crescimento e qualidade da madeira avaliadas via mapeamento por intervalo composto (IC) (Tabela 5). Detectou-se pelo menos um QTL em cada grupo de ligação (GL), com exceção do GL11 no qual não foram identificados QTLs em nenhum dos mapas parentais. Sobre o GL8 foram localizados sete QTLs, um deles para Crescimento em Altura, dois para Densidade Básica, dois locos associados à Lignina Total e outros dois para a relação Siringil/Guaiacil, quatro deles pertencentes ao mapa materno e três ao paterno, sendo, portanto o grupo de ligação com o maior número de QTLs posicionados. Para todas as sete características avaliadas, pelo menos um QTL foi detectado. Especificamente Lignina Total foi a característica para a qual foi detectado o maior número de QTLs, oito em total, cinco de origem materna e três de origem paterna.

Tabela 5. Sumário das informações dos QTLs detectados por Intervalo Composto pela estratégia de pseudo-cruzamento teste. Estão discriminados os genitores, os grupos de ligação e a posição dentro do grupo onde os locos foram detectados, bem como seus valores de LOD e estimativa da porcentagem da variação fenotípica explicada pelos QTLs.

Característica	**Genitor**	**Grupo de ligação**	**Posição (cM)**	**LOD**	**Variação explicada (%)**
CA	G38	1; 2; 6	76; 59; 21	3,4; 3,6; 3,8	6.7; 8.2; 7.3
	U15	7; 8; 10	104; 123; 42	3,7; 3,7; 5,3	6,7; 6,9; 10,4
LT	G38	1; 3; 4; 5; 8	2; 126; 77; 68; 143	3,4; 3,8; 3,4; 6,5; 4	5,5; 8,6; 5,5; 10; 6
	U15	3; 4; 8	39; 51; 189	3,9; 3,9; 3,1	7,7; 7,9; 5,8
S/G	G38	1; 5; 8	1; 31; 15	4,1; 4,2; 3,8	8,1; 8,1; 7,1
	U15	8; 9	69; 61	3,5; 4,5	11,6; 27,8
RDC	G38	4; 5	73; 69	4,7; 6,2	9,3; 22,4
	U15	1; 4; 9	131; 53; 0.2	3.4; 4,2; 3	5,9; 8,5; 6
DB	G38	6; 8; 10	98; 0.01; 26	6,1; 5,1; 5,6	10,9; 10,3; 11,3
	U15	8	73	4,5	9,3
DAP	U15	7; 10	104; 42	3,3; 3,9	6; 7,5
PP	G38	1; 2	105; 133,36	3,4	18,0; 21,22
	U15	1; 2; 10	68; 117; 42	3,2	19,65; 15,44; 18,94

As características analisadas aqui são: Crescimento em Altura (CA); Teor de Lignina Total (LT); relação Siringil/Guaiacil (S/G); Rendimento Depurado de Celulose (RDC); Densidade Básica (DB); Diâmetro à Altura do Peito (DAP) e Penetração do Pilodyn (PP). Os parentais estão representados pelo nome da amostra, sendo G38 é o genitor materno (*E. grandis*) e U15, o genitor paterno (*E. urophylla*).

Foram encontradas correlações fenotípicas significativas ($p < 0,01$) entre Diâmetro à Altura do Peito e outras duas características, uma correlação altamente positiva com o Crescimento em Altura ($r_{xy} = 0,78$) e uma menor com o teor de Lignina Total ($r_{xy} = 0,25$). Por sua vez a Lignina Total também está correlacionada positivamente com Crescimento em Altura ($r_{xy} = 0,20$), porém a correlação foi negativa com a relação Siringil/Guaiacil ($r_{xy} = -0,59$), densidade básica ($r_{xy} = -0,37$) e com o Rendimento Depurado de Celulose ($r_{xy} = -0,85$), sendo esta a maior correlação encontrada. Outra correlação positiva esperada ocorreu entre Siringil/Guaiacil e o Rendimento Depurado de Celulose ($r_{xy} = 0,40$) e também com a Densidade Básica ($r_{xy} = 0,51$). Finalmente e com um alto grau de significância, a capacidade de penetração do Pilodyn foi correlacionada com outras três características de crescimento: correlação positiva com o diâmetro à altura do peito ($r_{xy} = 0,59$) e o Crescimento em Altura ($r_{xy} = 0,32$); e correlação negativa com a densidade básica ($r_{xy} = -0,29$) (Tabela 6).

Tabela 6. Correlação de Pearson entre caraterísticas de crescimento e qualidade da madeira. DAP: diâmetro à altura do peito, CA: Crescimento em Altura, LT: lignina total, S/G: relação Siringil/guaiacil, RDC: rendimento depurado de celulose, DB: densidade básica, PP: penetração de pilodyn.

	DAP	**CA**	**LT**	**S/G**	**RDC**	**DB**	**PP**
DAP	1						
CA	0.7806**	1					
LT	0.2533**	0.2034**	1				
S/G	-0,1157 [a]	-0,0711 [a]	-0.5872**	1			
RDC	-0,1064 [a]	-0,0722 [a]	-0.8482**	0.4012**	1		
DB	-0.2068**	-0,1456 [a]	-0.3757**	0.5152**	0.1697*	1	
PP	0.585**	0.3248**	0.0478	-0,0262	0,0752	-0.2943**	1

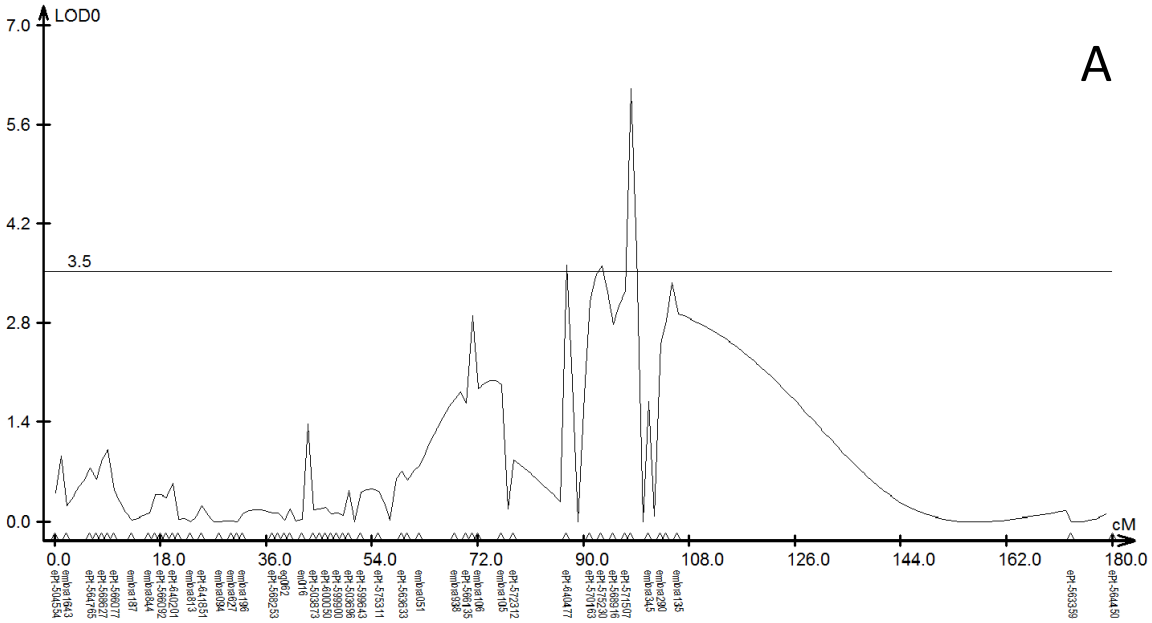
*P<0.05; **P<0.01; [a] não significativo.

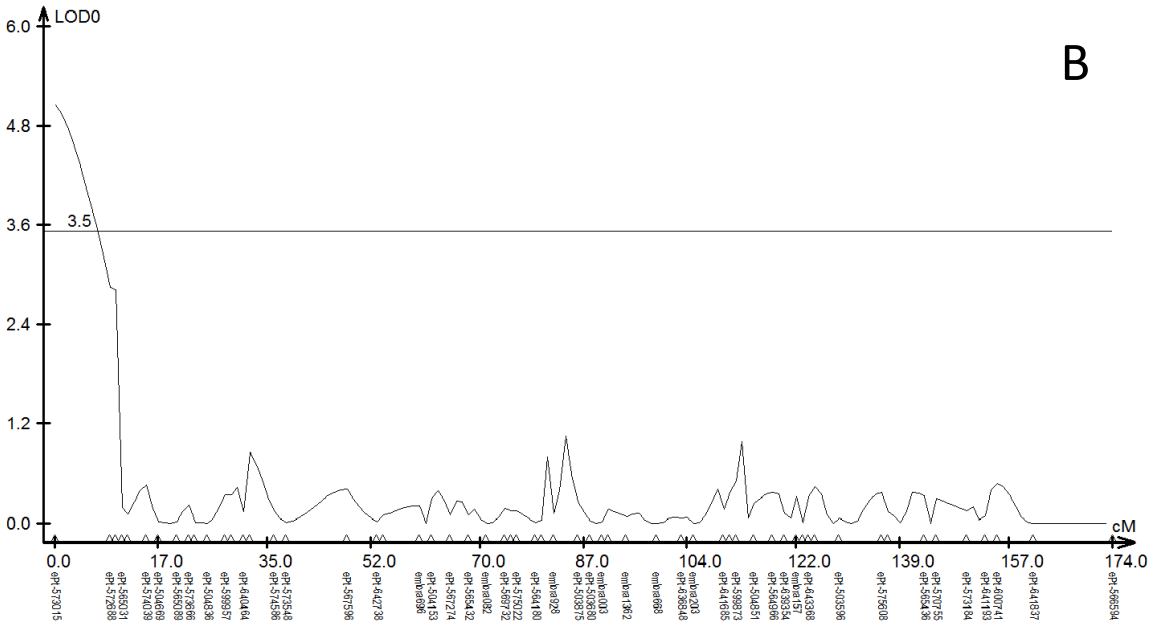
A seguir serão descritos os QTLs identificados pela estratégia de mapeamento por intervalo composto (IC) para cada uma das sete características avaliadas na família de irmãos completos de *E. grandis* x *E. urophylla* analisada neste estudo. Estimativas da proporção de variação fenotípica explicada por cada QTL foram também realizadas com uma análise de intervalo de mapa usando o *software* Cartographer v2.5 (Wang, Basten et al. 2007).

4.4.3. Detecção de QTLs para densidade básica da madeira

A densidade básica é uma medida relacionada com o crescimento secundário da madeira, e também indica a qualidade que esta possui e consequentemente é uma característica de grande interesse econômico para o setor florestal. Quanto maior a densidade menor é a porosidade da madeira e, portanto, maior tende a ser o seu conteúdo de celulose. Porém, a lignina também contribui para a densidade da madeira e, portanto, nem sempre uma madeira mais densa reflete uma matéria-prima de qualidade superior para a indústria de celulose. A média obtida pela população segregante avaliada para densidade básica da madeira foi de 448,3 Kg/m^3, com desvio padrão de 8,4 Kg/m^3.

Quatro QTLs foram detectados como altamente significativos para Densidade Básica da madeira em três grupos de ligação (Figura 10). Três dos QTLs significativamente associados à característica foram identificados no mapa materno (*E. grandis*), nos grupos de ligação GL6, GL8 e GL10. No mapa paterno (*E. urophylla*), um único QTL foi detectado, no GL8. Para o mapa materno determinou-se que os QTLs mapeados explicavam 10,9% (GL6), 10,3% (GL8) e 11,33% (GL10) da variação fenotípica, enquanto que o QTL correspondente ao mapa paterno explicou 9,3% (GL8) da variação em densidade básica da madeira.


A
LOD0
7.0
5.6
4.2
2.8
1.4
0.0
3.5
cM
0.0
18.0
36.0
54.0
72.0
90.0
108.0
126.0
144.0
162.0
180.0


B
LOD0
6.0
4.8
3.6
2.4
1.2
0.0
3.5
cM
0.0
17.0
35.0
52.0
70.0
87.0
104.0
122.0
139.0
157.0
174.0

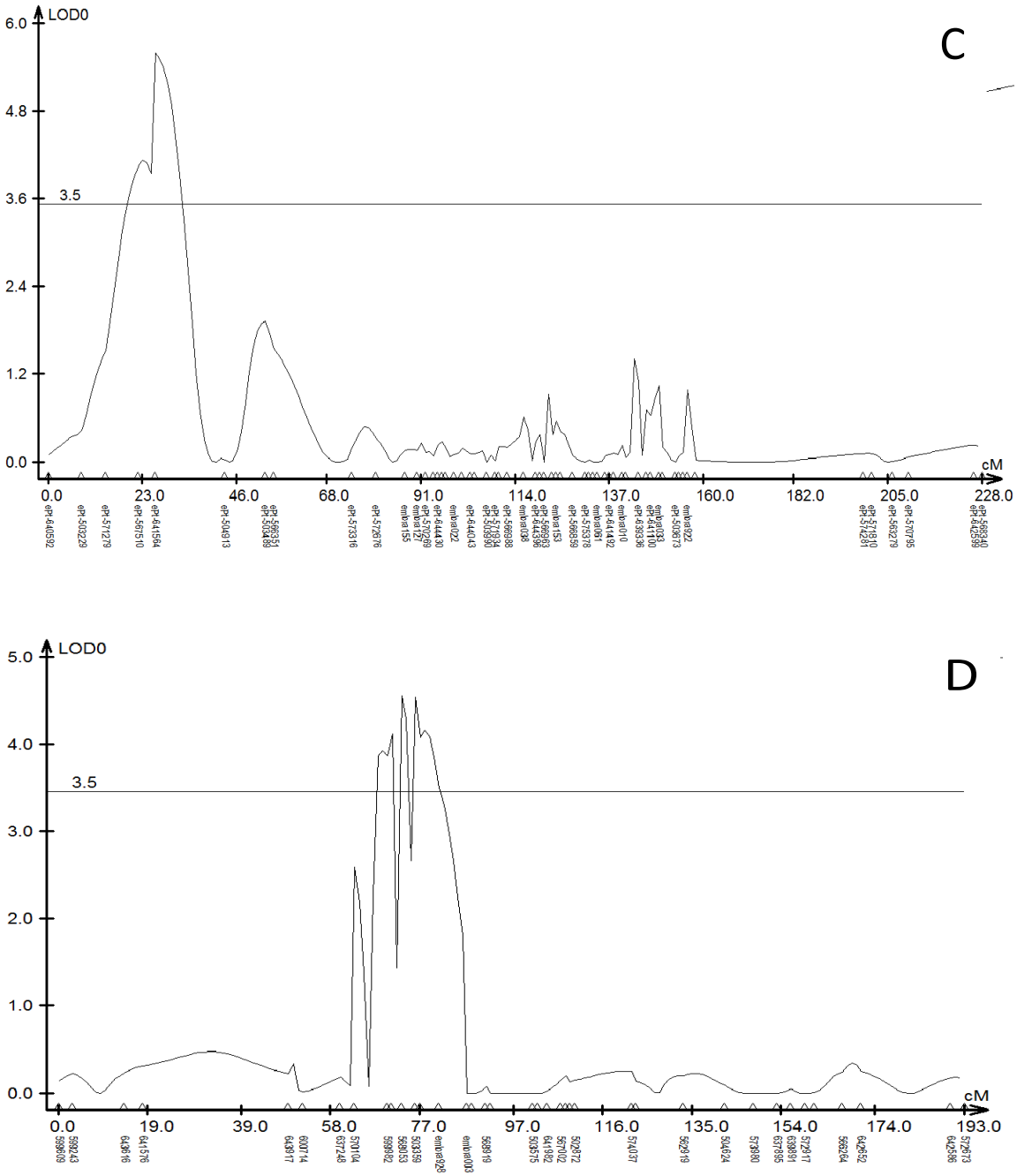
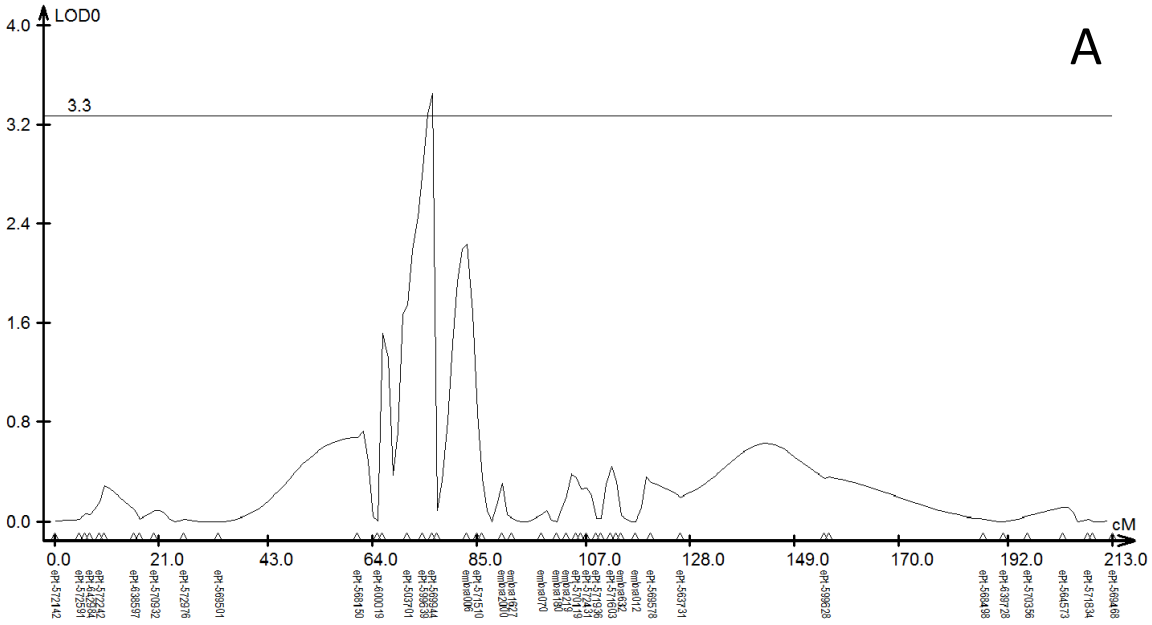


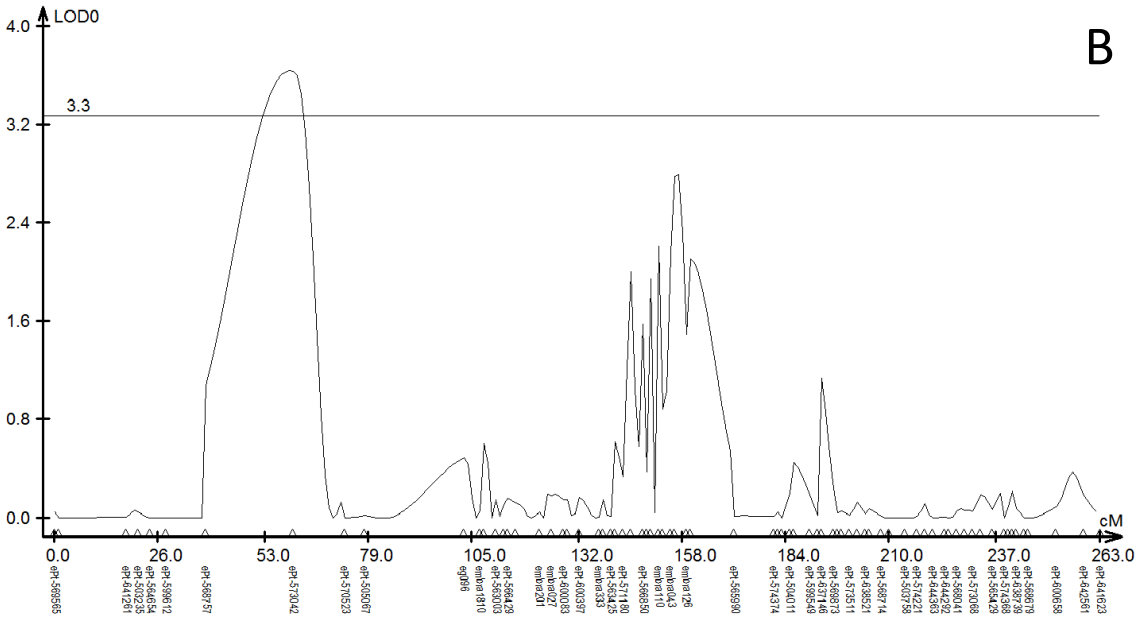
Figura 11. Gráficos gerados pelo programa QTL Cartographer apresentando os QTLs para densidade básica da madeira mapeados por intervalo composto nos grupos de ligação 6 (A), 8 (B) e 10 (C) no genoma do parental materno G38. Em D é mostrado o QTL detectado no grupo de ligação 8 no parental U15. Eixo Y: valor de LOD; Eixo X: intervalo entre os marcadores do grupo de ligação. O valor de LOD definido pelo teste de permutação com 5% de significância foi de 3,5.

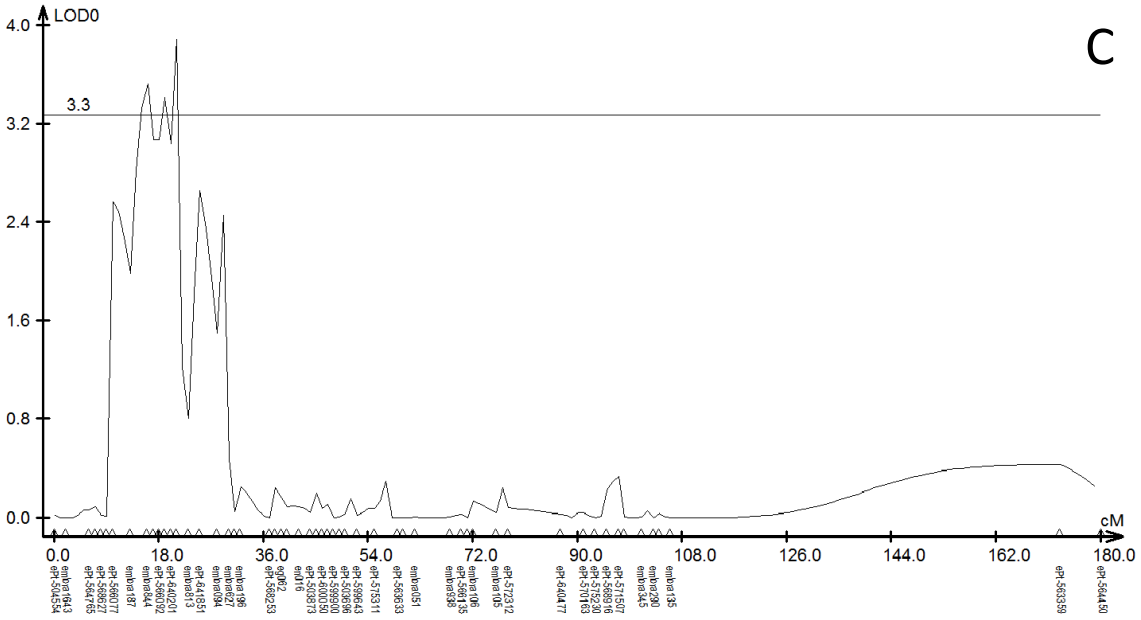
4.4.4. Detecção de QTLs para crescimento em altura

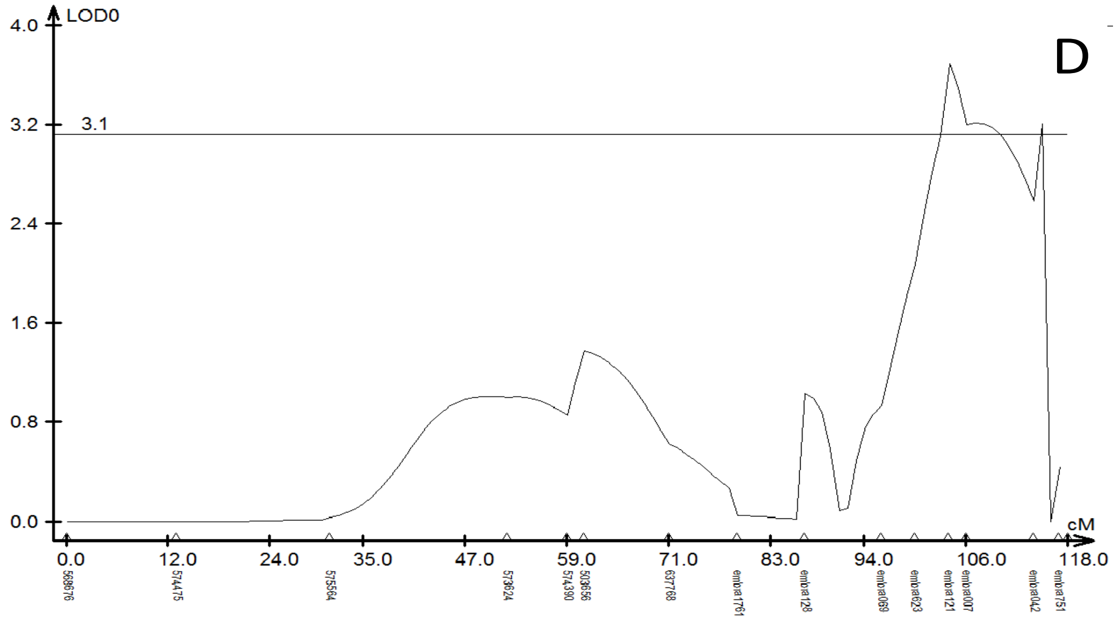
A característica Crescimento em Altura determina em grande parte o crescimento volumétrico final das árvores. Na população segregante do presente trabalho, aos três anos de idade as árvores tiveram altura média de 10,4 m, com desvio padrão de 0,8 m.

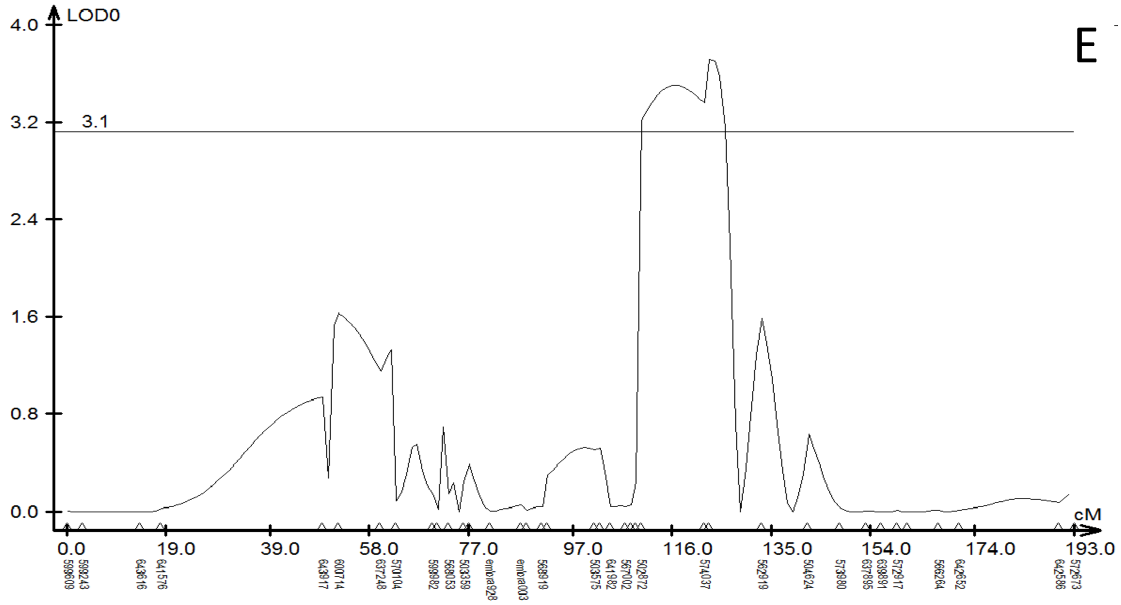
Para esta característica foram detectados seis QTLs através do mapeamento por Intervalo Composto (Figura 11), metade deles no genoma materno e a outra metade no genoma paterno. Os locos identificados em *E. grandis* localizaram-se no GL1, GL2 e GL6, e explicaram 6,7%, 8,2% e 7,3% da variação da Altura, respectivamente. Os três QTLs identificados no genoma de *E. urophylla* foram localizados nos grupos GL7, GL8 e GL10, explicando 6,7%, 6,9% e 10,4% da variação fenotípica desta carasterística, respectivamente.




B
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
3.3
cM
0.0
26.0
53.0
79.0
105.0
132.0
158.0
184.0
210.0
237.0
263.0


C
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
3.3
cM
0.0
18.0
36.0
54.0
72.0
90.0
108.0
126.0
144.0
162.0
180.0


D
LOD0
3.1
cM


E
LOD0
3.1
cM

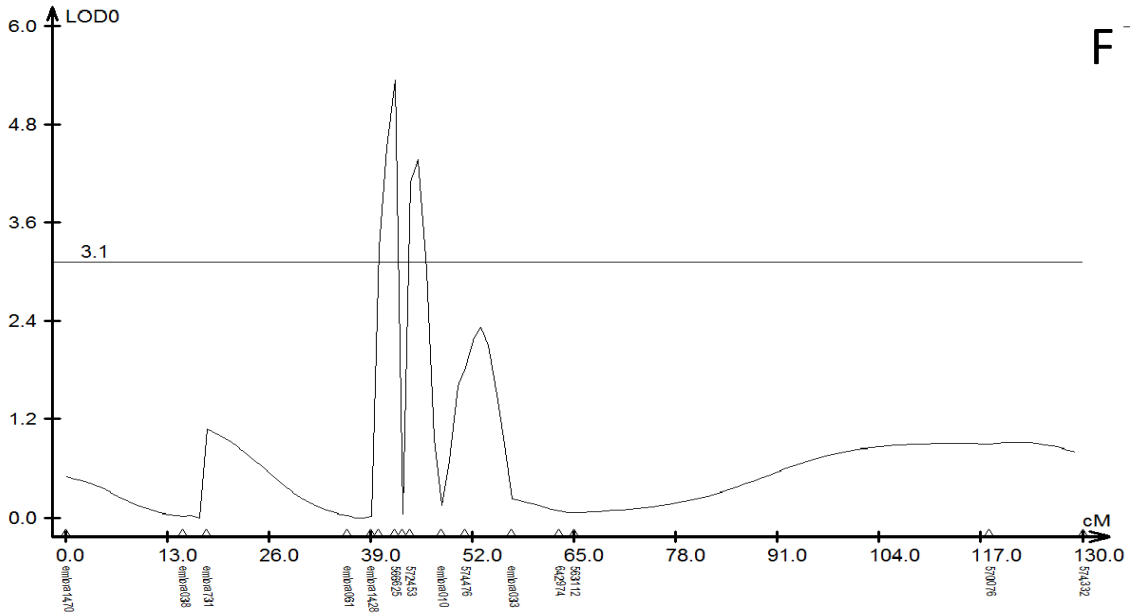


Figura 12. Gráficos gerados pelo programa QTL Cartographer apresentando os três QTLs para altura da árvore mapeados por intervalo composto nos grupos de ligação (GL) 1 (A), 2 (B) e 6 (C) no genoma do parental materno G38. No parental U15 foram detectados outros três QTLs nos GL 7 (D), 8 (E) é 10 (F). Eixo Y: valor de LOD; Eixo X: intervalo entre os marcadores do grupo de ligação. O valor de LOD definido pelo teste de permutação com 5% de significância foi de 3,3 para G38 e 3,1 para U15.

4.4.5. Detecção de QTLs para diâmetro à altura do peito

Esta característica também participa diretamente da determinação do crescimento volumétrico das árvores. Quanto maior o diâmetro das árvores, maior tende a ser o volume final de madeira. Na população segregante do presente trabalho, as árvores tiveram um DAP médio de 9,9 cm, com desvio padrão de 1,0 cm.

Foram detectados dois QTLs (Figura 12), este foram localizados no GL7 e no GL10 apenas no genoma paterno (*E. urophylla*). O QTL identificado no GL7 explicou 6,0% da variação no DAP dentro desta população, enquanto que o loco situado no GL10, explicou 7,5%.

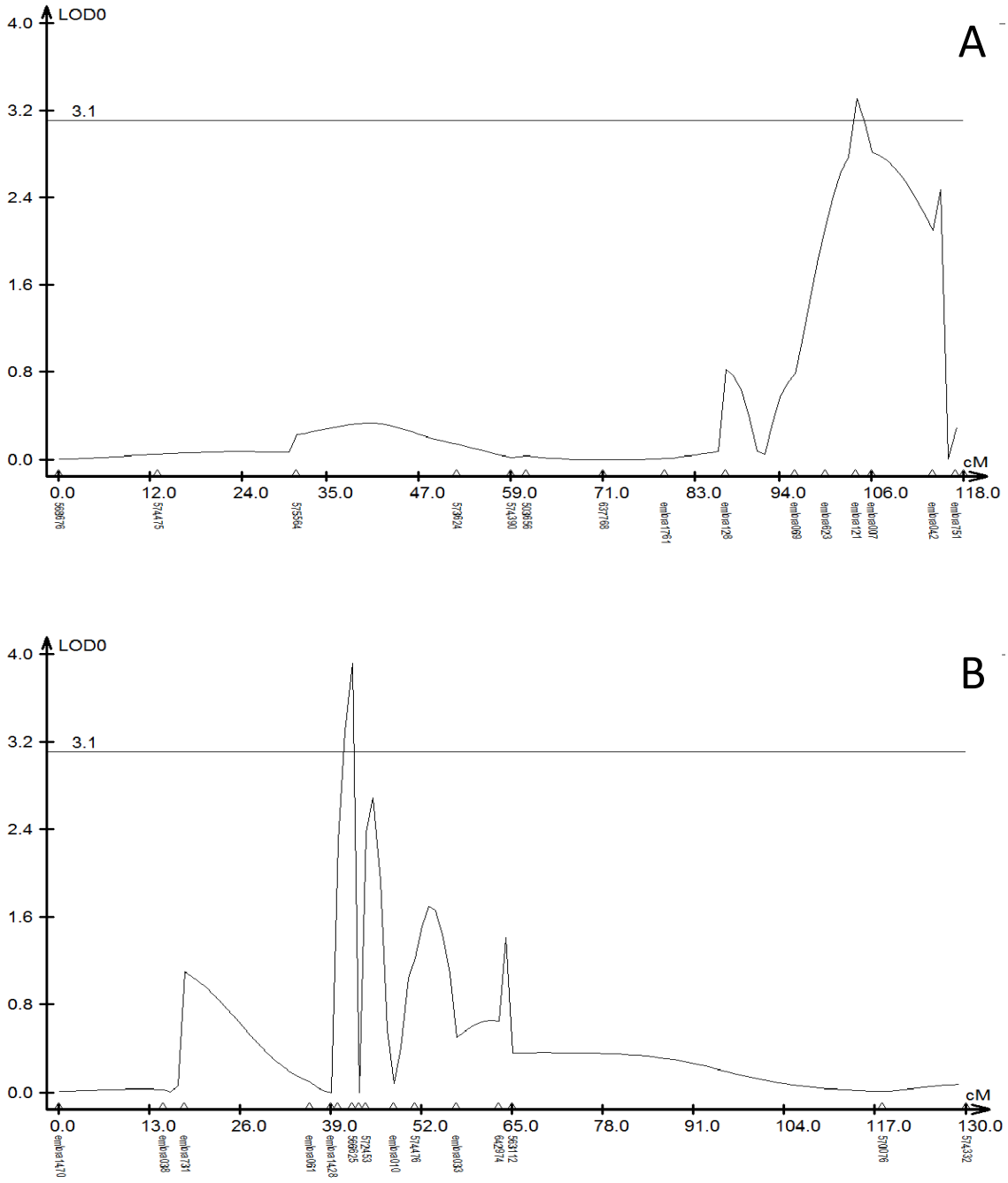
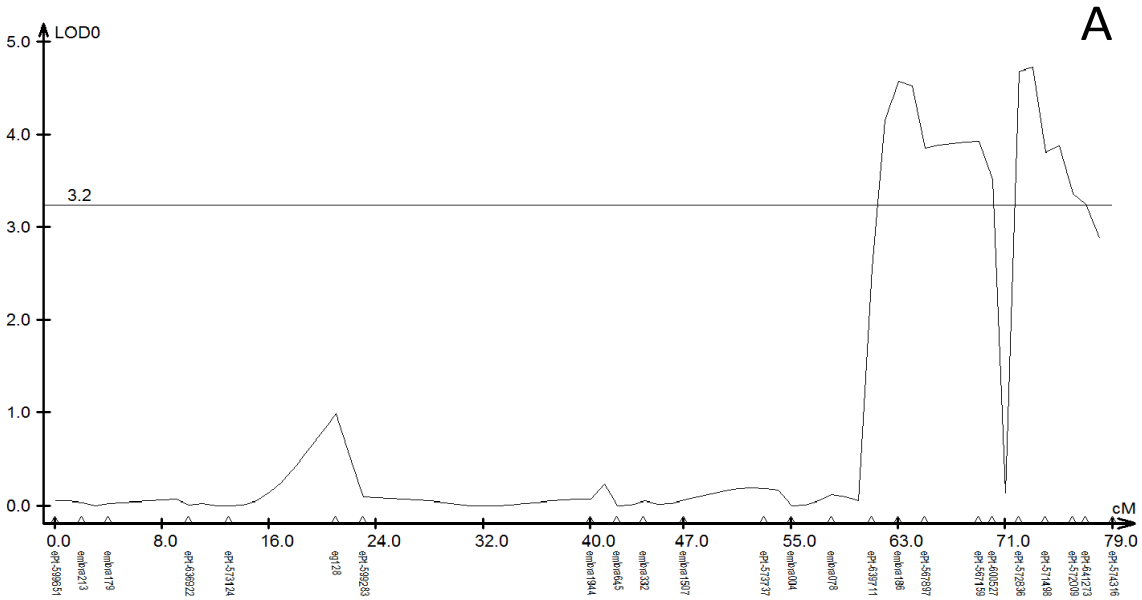


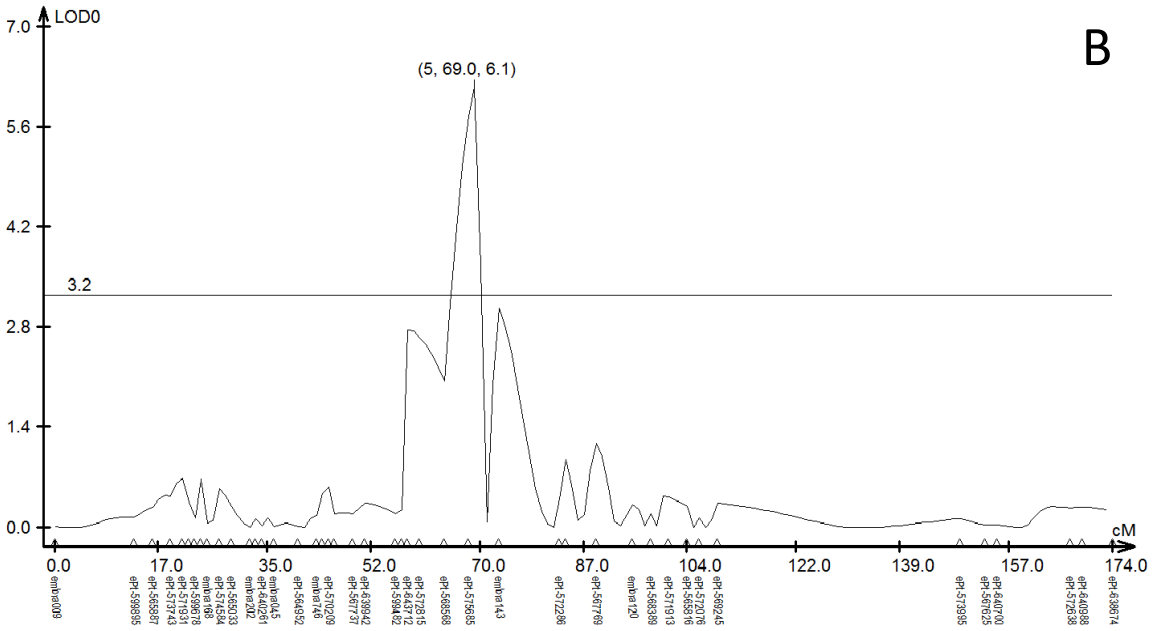
Figura 13. Gráficos gerados pelo programa QTL Cartographer demonstrando os QTLs para diâmetro das árvores à altura do peito (DAP) mapeados por intervalo composto nos grupo de ligação 7 (A) e 10 (B) no genoma do parental U15. Eixo Y: valor de LOD; Eixo X: intervalo entre os marcadores do grupo de ligação. O valor de LOD definido pelo teste de permutação com 5% de significância foi de 3,1.

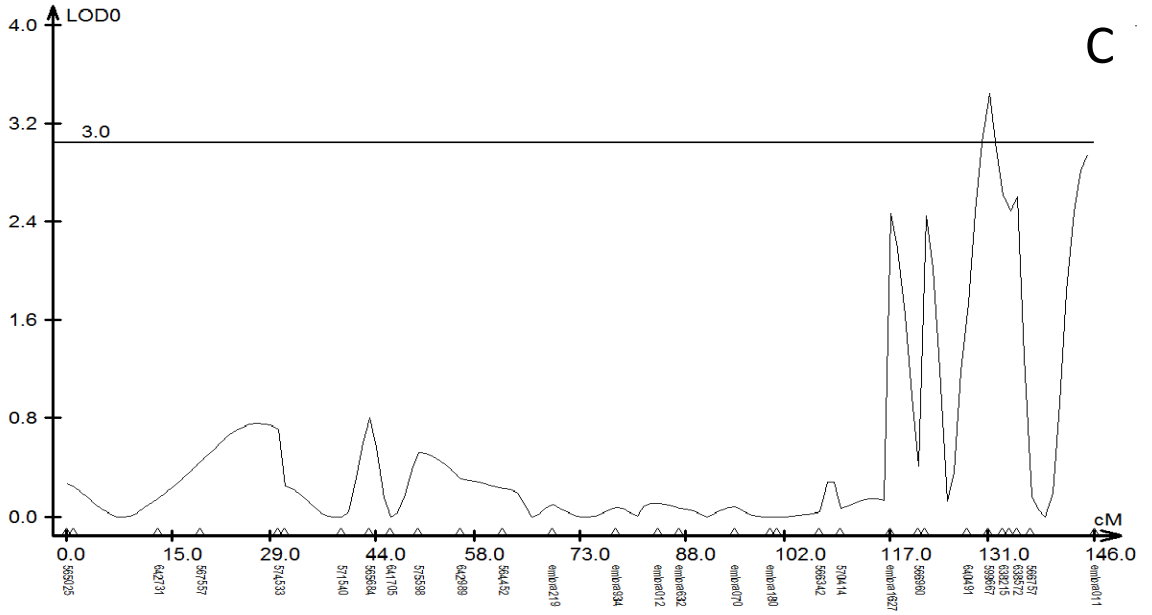
4.4.6. Detecção de QTLs para rendimento depurado de celulose

O rendimento depurado medido percentualmente refere-se à proporção de celulose gerada no final da polpação em relação à quantidade inicial de madeira utilizada no inicio do processo em relação peso/peso. Esse é um dos atributos mais importantes para a indústria de celulose, pois é uma medida que consolida diversos componentes físicos e químicos da madeira que, juntos, impactam o rendimento final de produção de celulose. Na população segregante do presente trabalho, o rendimento da polpa de celulose médio foi de 53,4% com um desvio padrão de 0,8%.

Para esta importante característica foram identificados dois QTLs (Figura 13), localizados nos grupos GL4 e GL5 no do genoma de *E. grandis*. Estes QTLs explicaram respectivamente 9,3% e 22,4% da variação fenotípica da característica dentro da população segregante. No genoma de *E. urophylla* foram três os QTLs localizados nos grupos GL1, GL4 e GL9. O QTL pertencente ao GL1 explicou 5,9% da variação em relação ao rendimento de celulose, enquanto que o QTL mapeado no GL4 explicou 8,5% , e o do GL9, 6,0% da variação de dita característica.




B
LOD0
7.0
5.6
4.2
2.8
1.4
0.0
(5, 69.0, 6.1)
3.2
cM
0.0
17.0
35.0
52.0
70.0
87.0
104.0
122.0
139.0
157.0
174.0


C
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
3.0
cM
0.0
15.0
29.0
44.0
58.0
73.0
88.0
102.0
117.0
131.0
146.0

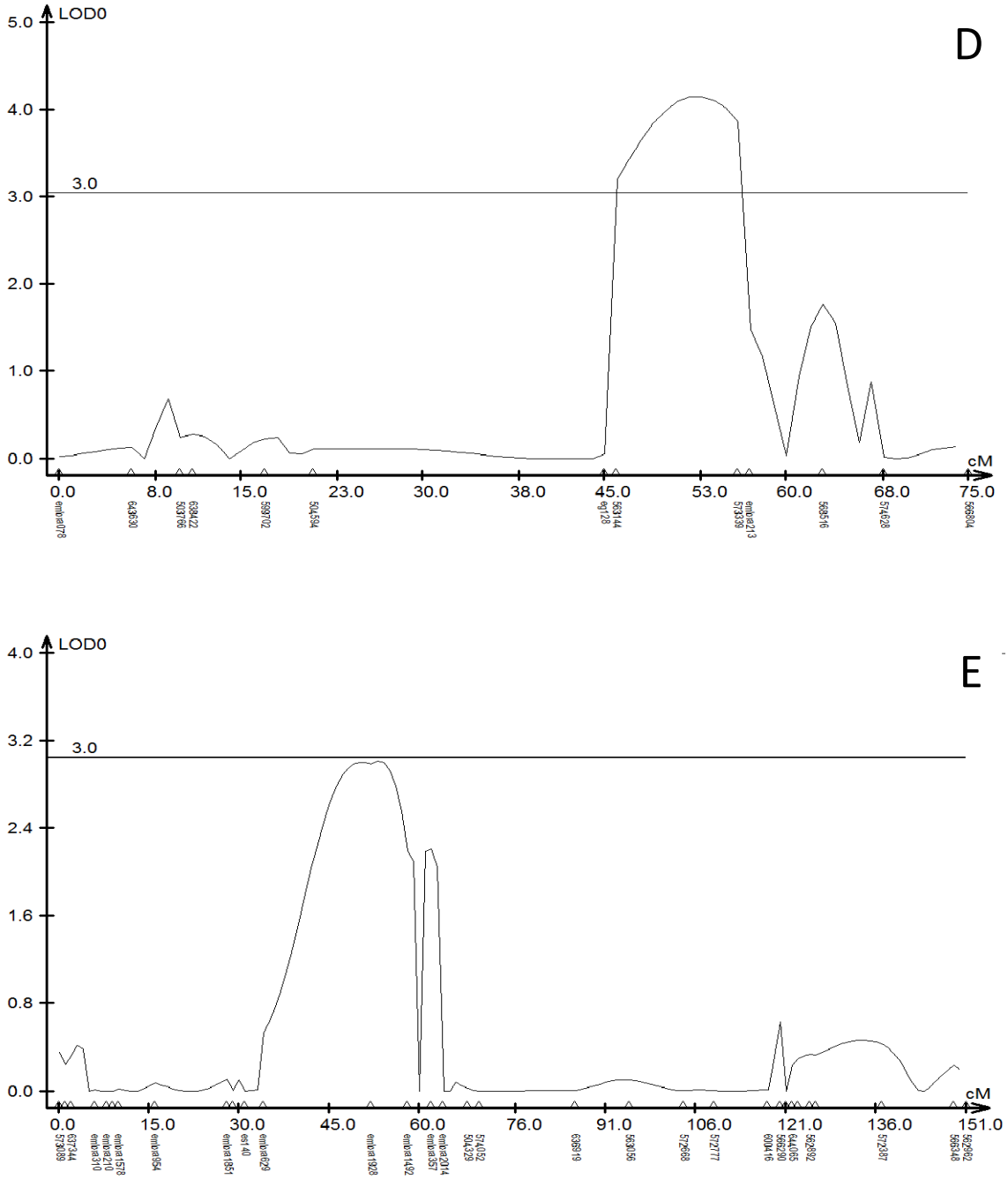
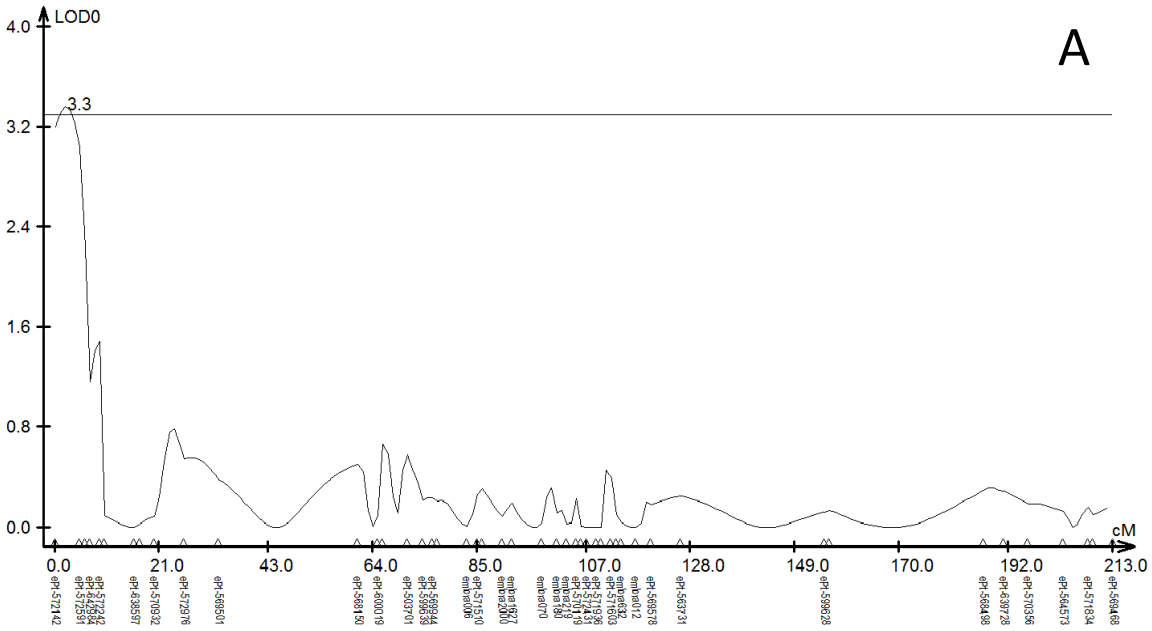


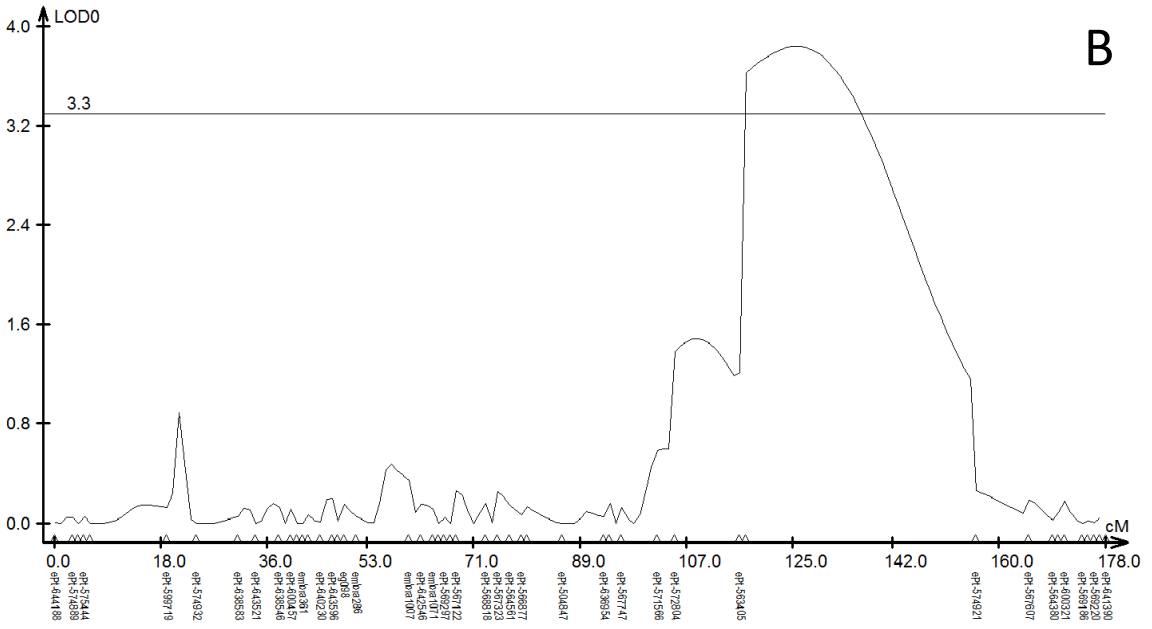
Figura 14. Gráficos gerados pelo programa QTL Cartographer demonstrando os QTLs para rendimento depurado de celulose mapeados por intervalo composto no parental G38 nos grupos de ligação 4 (A) e 5 (B). No genoma do parental U15 foram detectados QTLs nos GL 1 (C), 4 (D) e 9 (E). Eixo Y: valor de LOD; Eixo X: intervalo entre os marcadores do grupo de ligação. O valor de LOD definido pelo teste de permutação com 5% de significância foi de 3,2 para G38 e 3,0 para U15.

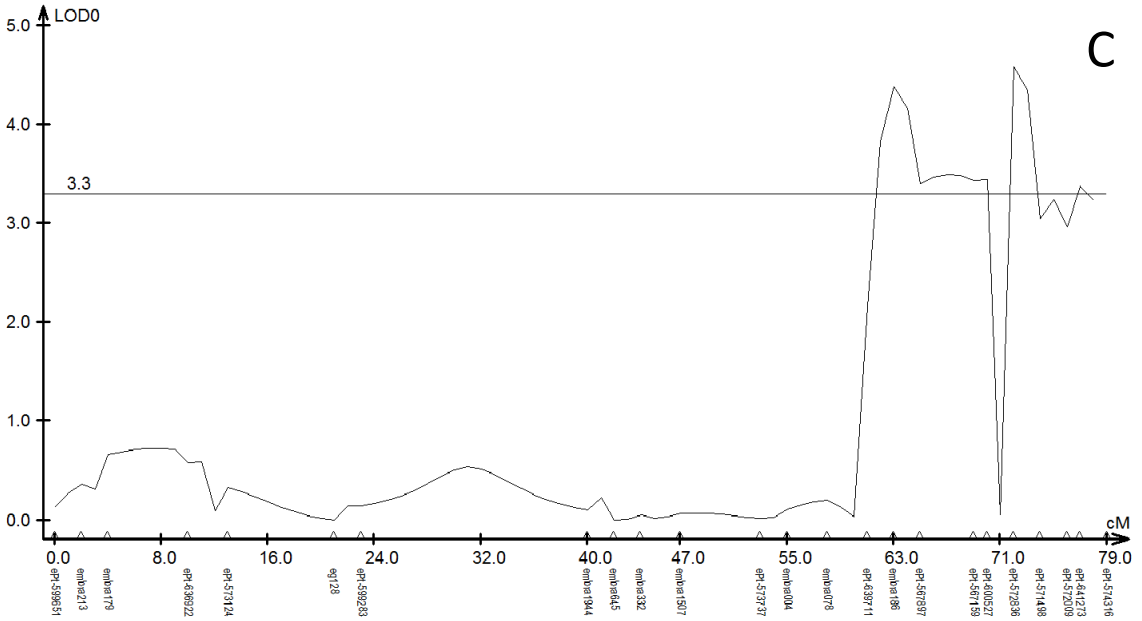
4.4.7. Detecção de QTLs para teor de lignina total da madeira

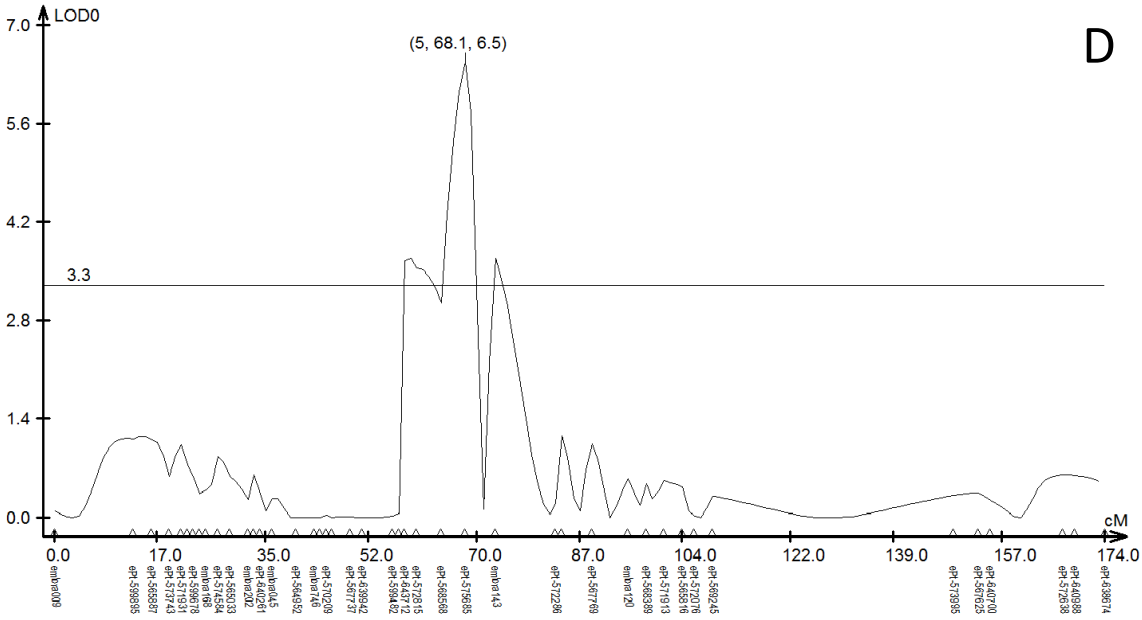
A lignina é, depois da celulose, o constituinte mais abundante da madeira. Para a produção de polpa celulósica, é necessária a desagregação das microfibras de celulose. A lignina pode ser um agente permanente de ligação entre as células e necessita ser eliminada no processo de polpação, exigindo uma grande quantidade de energia e reagentes nocivos ao ambiente para sua solubilização. Além disso, a lignina também dificulta as operações de branqueamento da polpa celulósica. Porém, é importante destacar que lignina extraída durante a polpação é utilizada como energia para realimentar o processo. Tudo isso faz do teor de lignina da madeira um dos atributos mais importantes na definição da qualidade da madeira para a indústria de celulose. Na população segregante do presente trabalho, o teor de lignina total médio foi de 25,4%, com um desvio padrão de 0,7%.

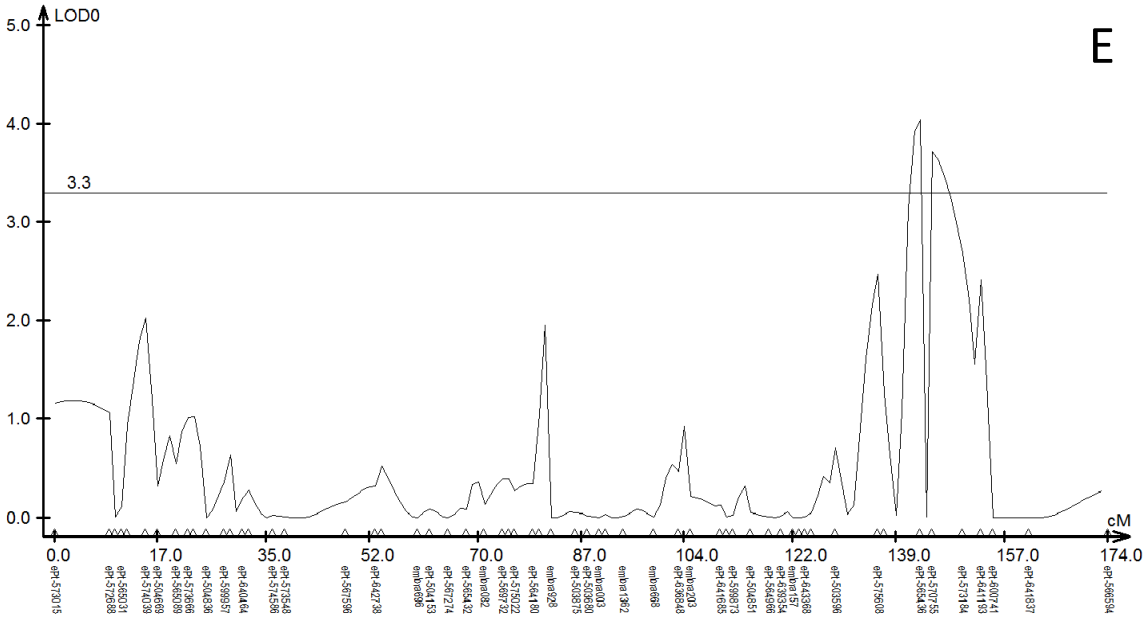
Para esta característica foram detectados oito QTLs (Figura 14), cinco dos quais estavam presentes no mapa do genitor *E. grandis* e três no mapa do genitor *E. urophylla*. Os QTLs maternos foram identificados nos grupos de ligação GL1, GL3, GL4, GL5 e GL8, cada um explicando respectivamente 5,5%, 8,6%, 5,5%, 10,0% e 6,0% da variação de lignina total. Os QTLs do genoma paterno localizaram-se nos grupos de ligação GL3, GL 4 e GL8, cada um deles foi estimado como responsável por 7,7%, 7,9% e 5,8%, da variação fenotípica da característica, respectivamente.

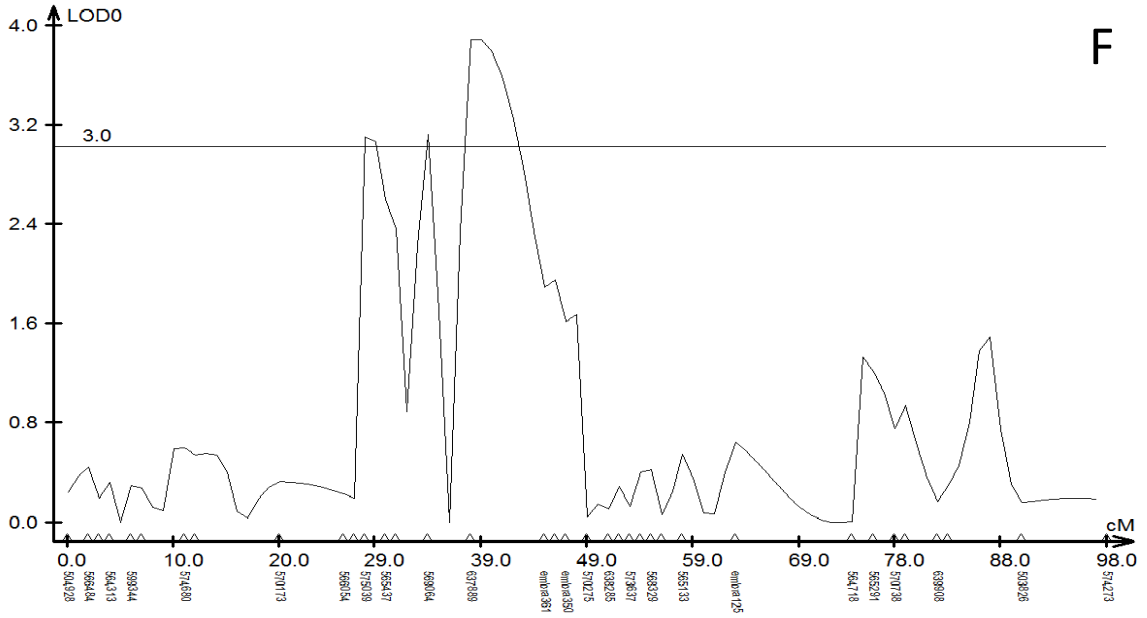

A
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
3.3
cM
0.0
21.0
43.0
64.0
85.0
107.0
128.0
149.0
170.0
192.0
213.0


B
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
3.3
cM
0.0
18.0
36.0
53.0
71.0
89.0
107.0
125.0
142.0
160.0
178.0


C
LOD0
5.0
4.0
3.3
3.0
2.0
1.0
0.0
cM
0.0
8.0
16.0
24.0
32.0
40.0
47.0
55.0
63.0
71.0
79.0
ePt-599651
embra213
embra179
ePt-636922
ePt-573124
ag128
ePt-599283
embra1944
embra945
embra332
embra1507
ePt-573737
embra004
embra078
ePt-639711
embra196
ePt-567697
ePt-600527
ePt-567159
ePt-572836
ePt-571488
ePt-572109
ePt-641273
ePt-574316


D
LOD0
7.0
5.6
4.2
3.3
2.8
1.4
0.0
(5, 68.1, 6.5)
cM
0.0
17.0
35.0
52.0
70.0
87.0
104.0
122.0
139.0
157.0
174.0
embra009
ePt-599935
ePt-568887
ePt-571743
ePt-571911
ePt-598178
embra198
ePt-574584
ePt-565033
ePt-642081
embra202
embra045
ePt-564952
embra746
ePt-570209
ePt-639942
ePt-567737
ePt-572815
ePt-643722
ePt-599442
ePt-568568
ePt-575685
embra143
ePt-572286
ePt-567769
embra120
ePt-568389
ePt-571913
ePt-572076
ePt-568816
ePt-569245
ePt-573995
ePt-567625
ePt-640700
ePt-640938
ePt-572638
ePt-639674


E
LOD0
5.0
4.0
3.0
2.0
1.0
0.0
3.3
cM
0.0
17.0
35.0
52.0
70.0
87.0
104.0
122.0
139.0
157.0
174.0


F
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
3.0
cM
0.0
10.0
20.0
29.0
39.0
49.0
59.0
69.0
78.0
88.0
98.0

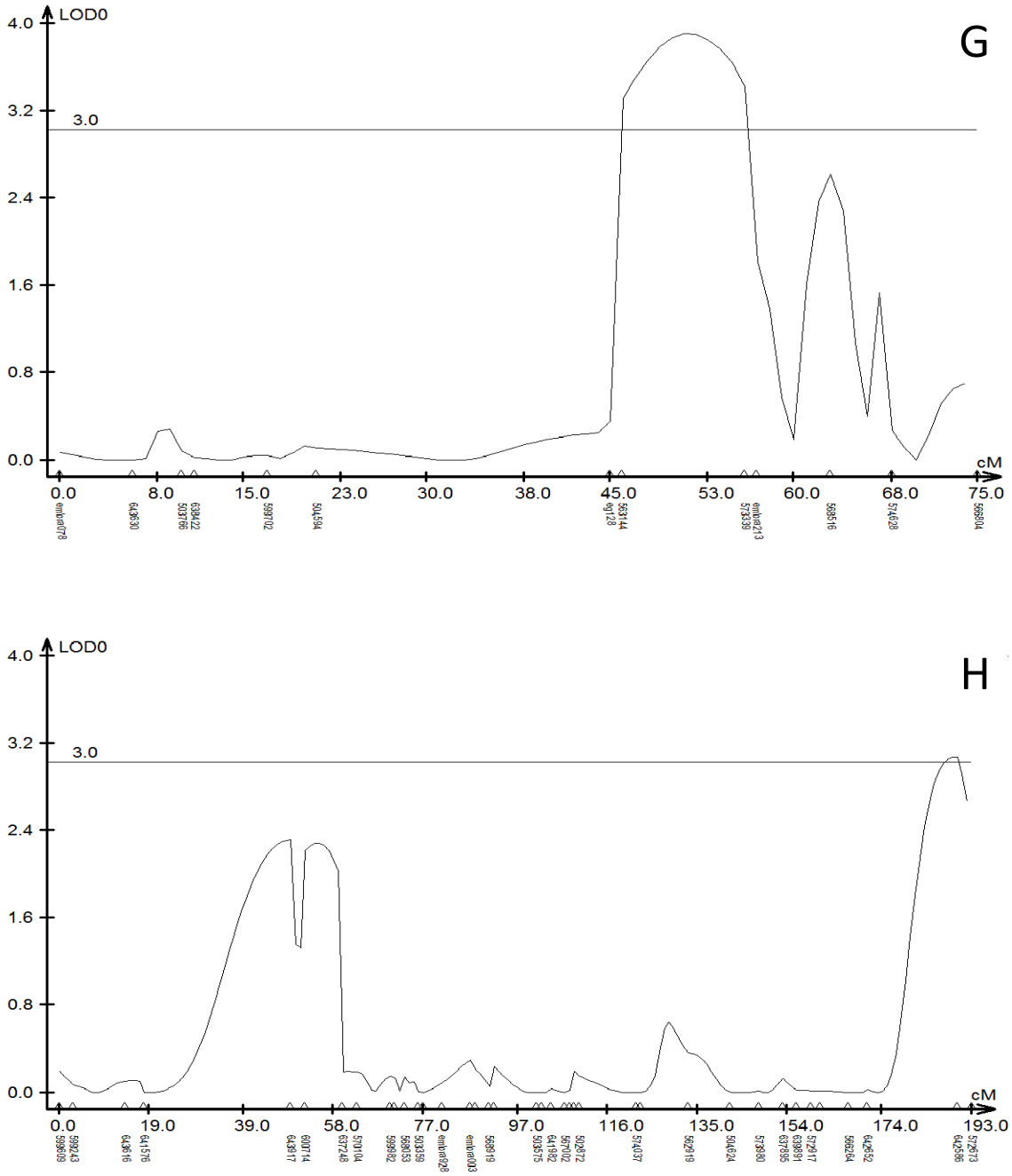
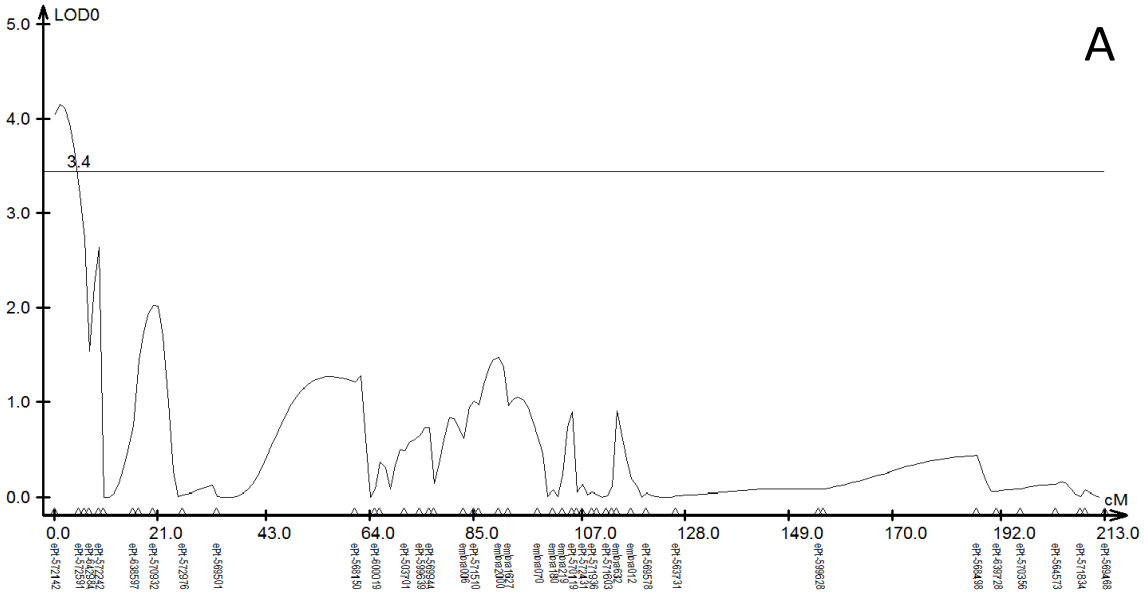


Figura 15. Gráficos gerados pelo programa QTL Cartographer demonstrando os QTLs para teor de lignina total mapeados por intervalo composto no parental G38 nos grupos de ligação 1 (A), 3 (B), 4 (C), 5 (D) e 8 (E). No genoma do parental U15 foram detectados QTLs nos GL 3 (F), 4 (G) e 8 (H). Eixo Y: valor de LOD; Eixo X: intervalo entre os marcadores do grupo de ligação. O valor de LOD definido pelo teste de permutação com 5% de significância foi de 3,3 para G38 e 3,0 para U15.

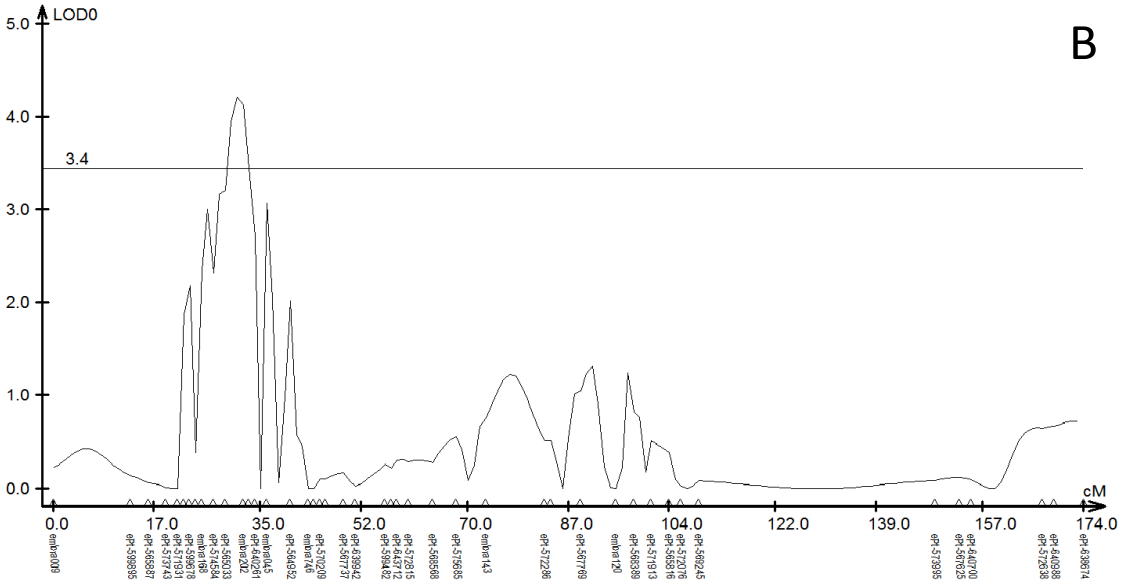
4.4.8. Detecção de QTLs para a relação Siringil/Guaiacil

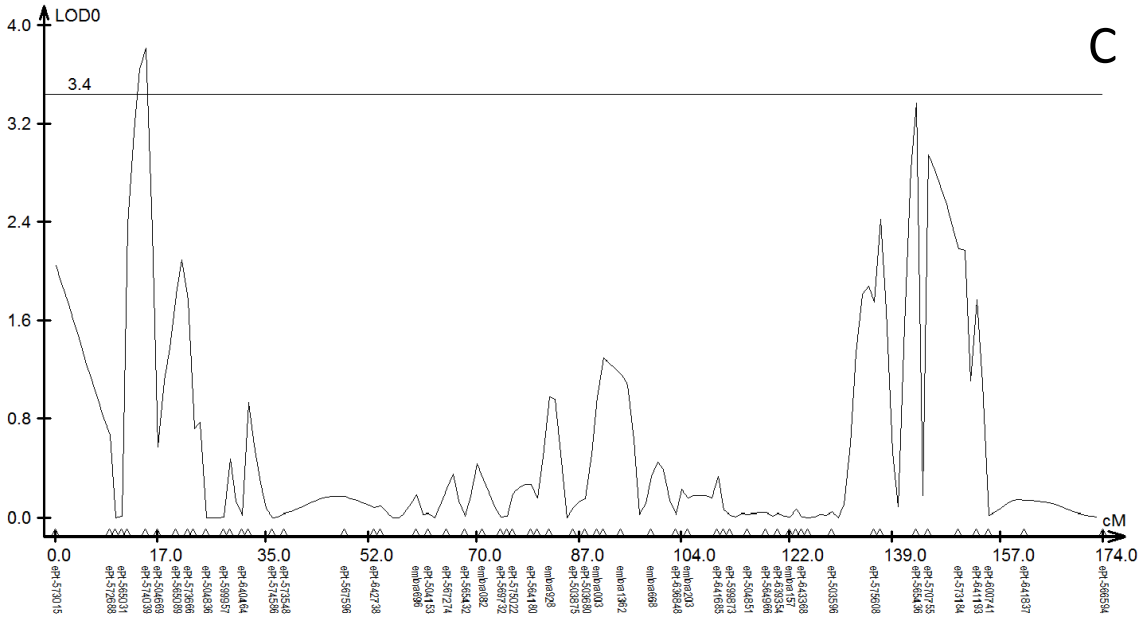
O siringil e guaiacil são dois monolignois, ou seja, alcoóis componentes da lignina. A relação entre esses dois alcoóis é fundamental para determinar a reatividade da lignina. O siringil, por não possuir o carbono C5 disponível para reação, faz com que as estruturas de lignina siringil sejam menos condensadas e, consequentemente, mais favoráveis ao ataque pelo álcali. Assim, quanto maior a relação siringil/guaiacil maior é a reatividade da lignina e, portanto, mais fácil será a sua extração no processo de polpação (Gomide, Colodette et al. 2005). Pela importância da lignina no fornecimento de energia para alimentar o processo de polpação, o grande alvo das pesquisas nas áreas de melhoramento e biologia molecular tem sido o aumento na relação siringil/guaiacil e não a redução do teor de lignina total da madeira. Na população segregante do presente trabalho, a relação siringil/guaiacil média foi de 2,4 com desvio padrão de 0,2.

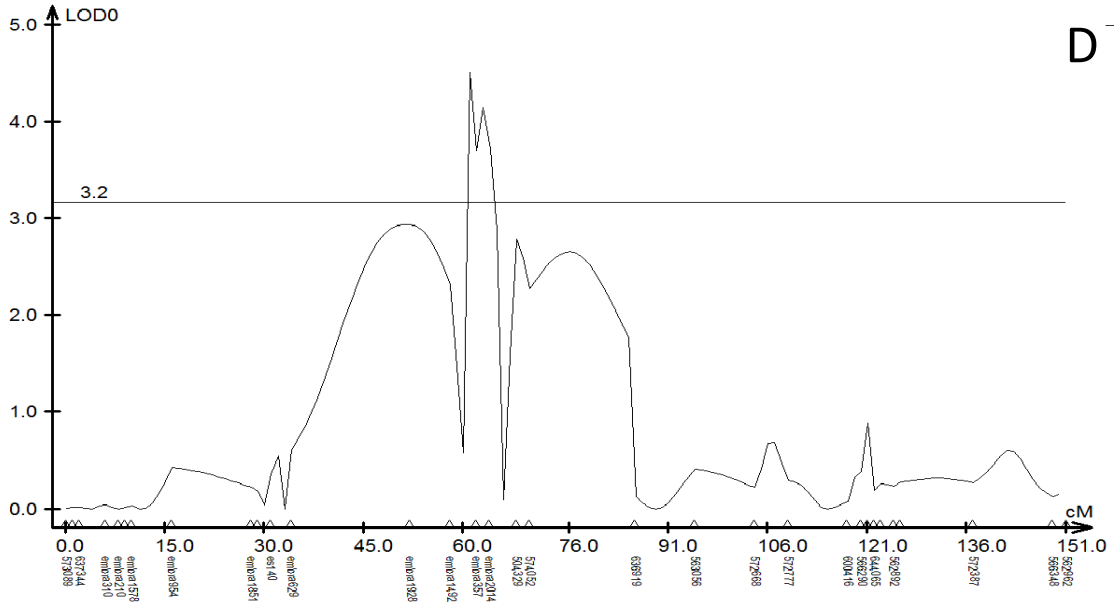
Foram detectados cinco QTLs para a característica relação siringil/guaiacil nesta população interespecífica (Figura 15). A partir do mapa materno foram identificados três QTLs nos grupos de ligação GL1, GL5 e GL8, enquanto os outros dois QTLs eram de origem paterna e estavam posicionados nos grupos GL8 e GL9. Os locos que pertenciam ao mapa *E. grandis* explicaram 8,1%, 8,1% e 7,1% da variação fenotípica desta característica. No mapa de E. *urophylla* determinou-se que os QTL explicavam 11,6% (GL8) e 27,8% (GL9) da variação fenotípica para a relação entre este dois alcoóis.


A
LOD0
5.0
4.0
3.0
2.0
1.0
0.0
3.4
cM
0.0
21.0
43.0
64.0
85.0
107.0
128.0
149.0
170.0
192.0
213.0

]


B
LOD0
5.0
4.0
3.0
2.0
1.0
0.0
3.4
cM
0.0
17.0
35.0
52.0
70.0
87.0
104.0
122.0
139.0
157.0
174.0


C
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
3.4
cM
0.0
17.0
35.0
52.0
70.0
87.0
104.0
122.0
139.0
157.0
174.0


D
LOD0
5.0
4.0
3.0
2.0
1.0
0.0
3.2
cM
0.0
15.0
30.0
45.0
60.0
76.0
91.0
106.0
121.0
136.0
151.0

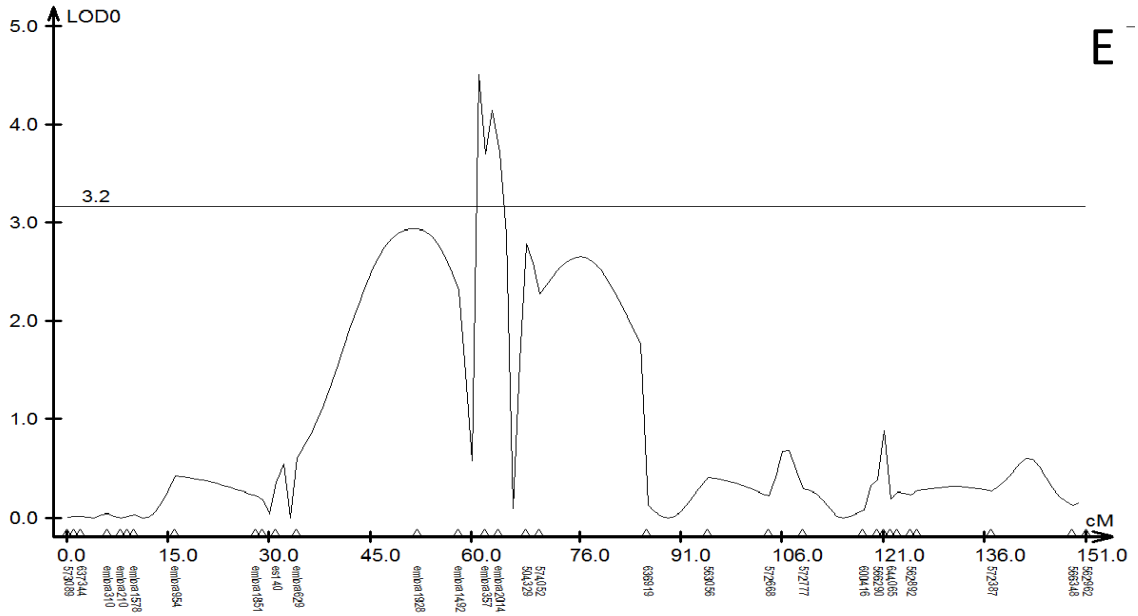


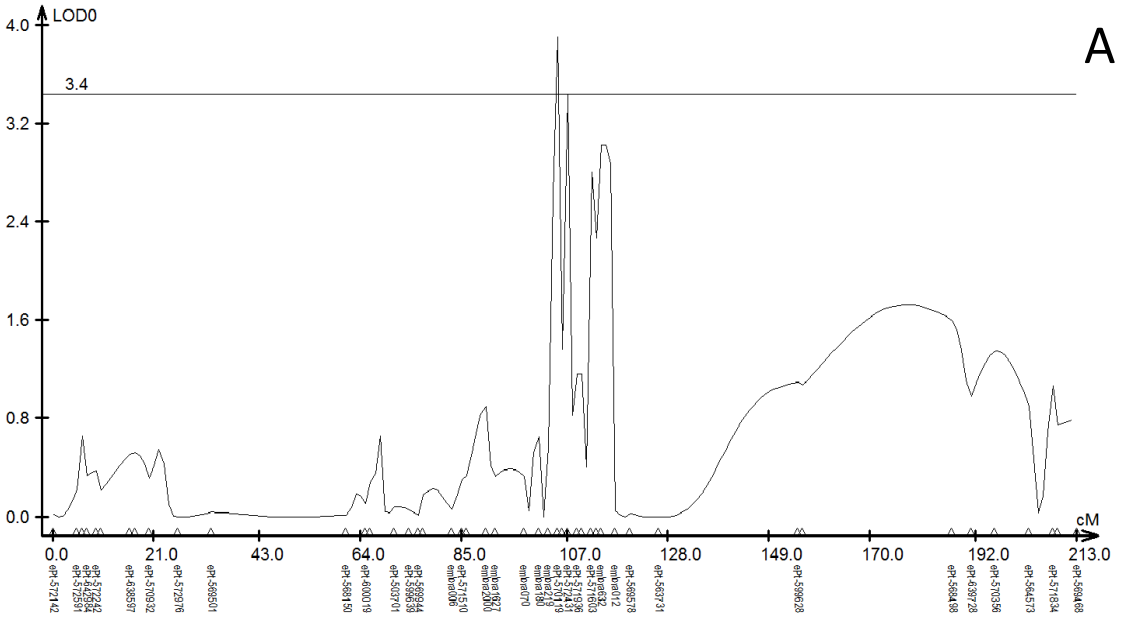
Figura 16. Gráficos gerados pelo programa QTL Cartographer demonstrando os QTLs para relação Siringil/Guaiacil da madeira mapeados por intervalo composto no parental G38 nos grupos de ligação 1 (A), 5 (B) e 8 (C). No genoma do parental U15 foram detectados QTLs nos GL 8 (D), 9 (E). Eixo Y: valor de LOD; Eixo X: intervalo entre os marcadores do grupo de ligação. O valor de LOD definido pelo teste de permutação com 5% de significância foi de 3,4 para G38 e 3,2 para U15.

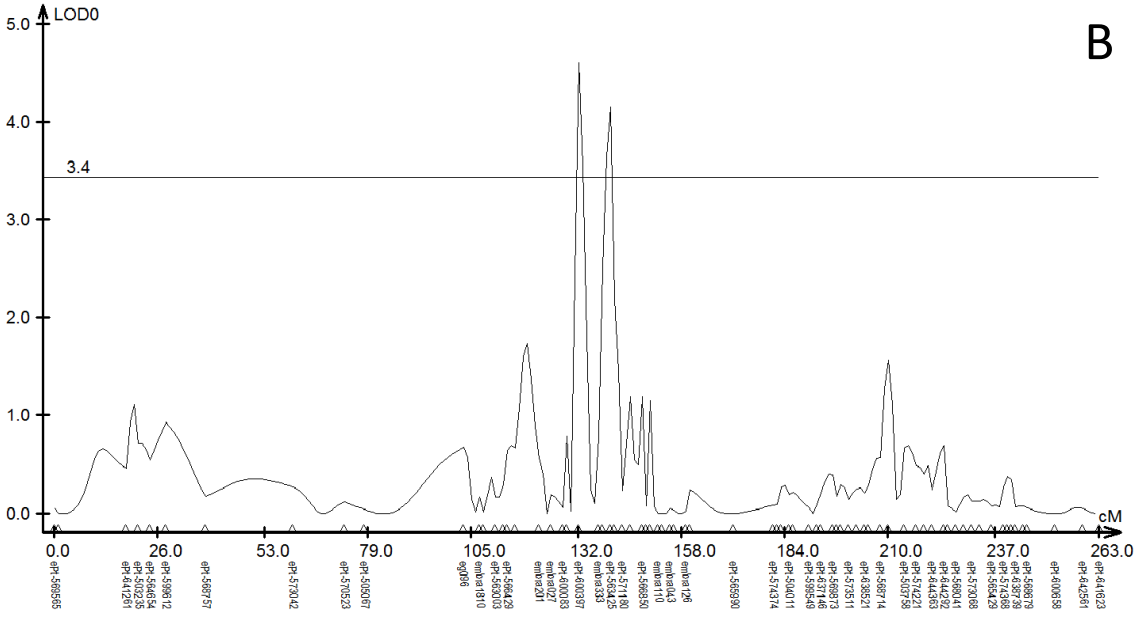
4.4.9. Detecção de QTLs para densidade da madeira medida pela profundidade de penetração do Pilodyn.

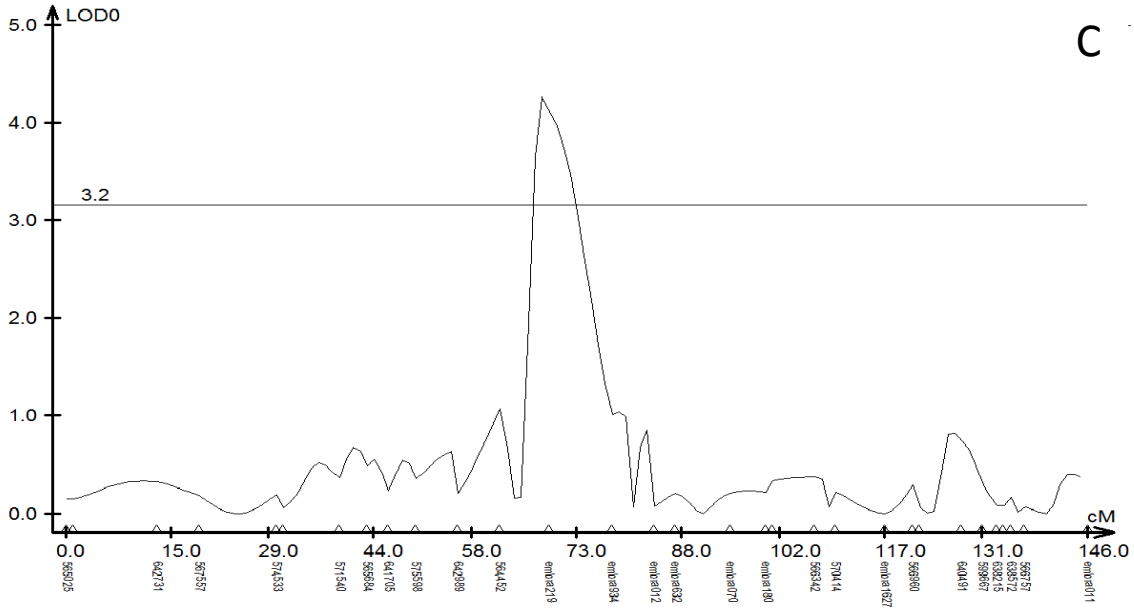
Esta densidade foi medida através da penetração na madeira de um elemento metálico do equipamento Pilodyn. A penetração da haste desse instrumento é utilizada como uma medida indireta da densidade básica da madeira. Quanto maior a penetração dessa haste na madeira, menor é a densidade. Na população segregante do presente trabalho, a profundidade média de penetração do Pilodyn foi de 21,1 mm, com desvio padrão de 1,0 mm.

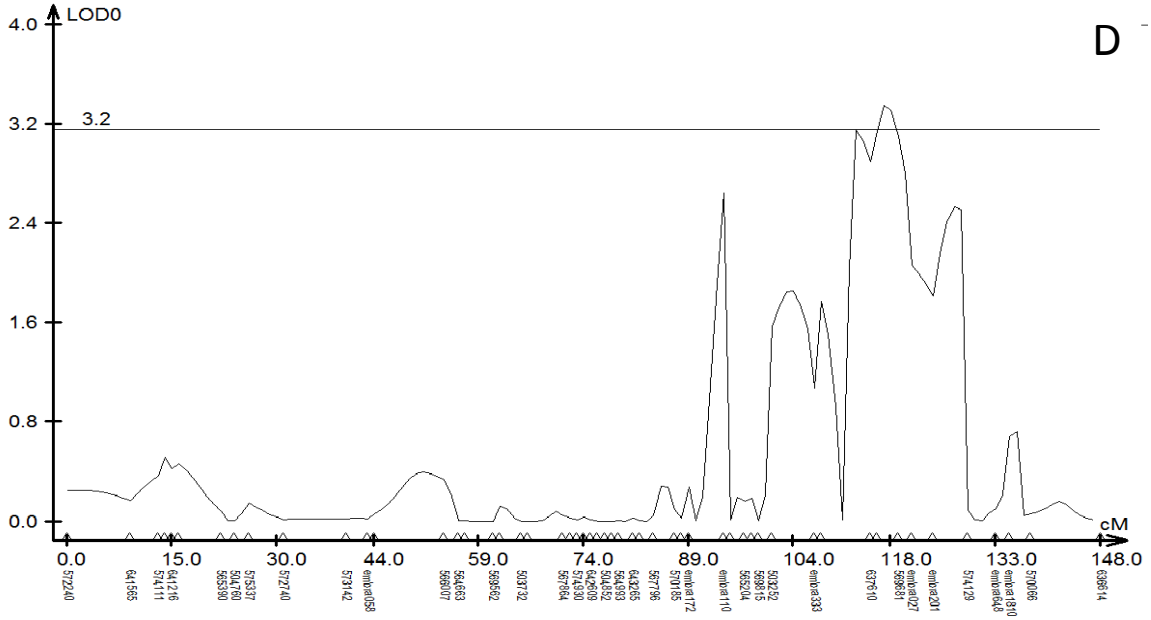
Foram detectados cinco QTLs (Figura 16), dois deles a partir do genoma do genitor *E. grandis* e três a partir do genitor *E. urophylla*. Os locos pertencentes ao mapa

materno foram detectados nos grupos GL1 e GL2, estes dois QTLs explicavam, respectivamente, 18,0% e 21,22% da variação nos valores da característica medida. Aqueles locos identificados no mapa paterno foram mapeados nos grupos GL1, GL2 e GL10, respondendo respectivamente por 19,7%, 15,4% e 18,9% da variação fenotípica em densidade medida pela penetração do Pilodyn.






C
LOD0
5.0
4.0
3.2
3.0
2.0
1.0
0.0
cM
0.0
15.0
29.0
44.0
58.0
73.0
88.0
102.0
117.0
131.0
146.0


D
LOD0
4.0
3.2
2.4
1.6
0.8
0.0
cM
0.0
15.0
30.0
44.0
59.0
74.0
89.0
104.0
118.0
133.0
148.0

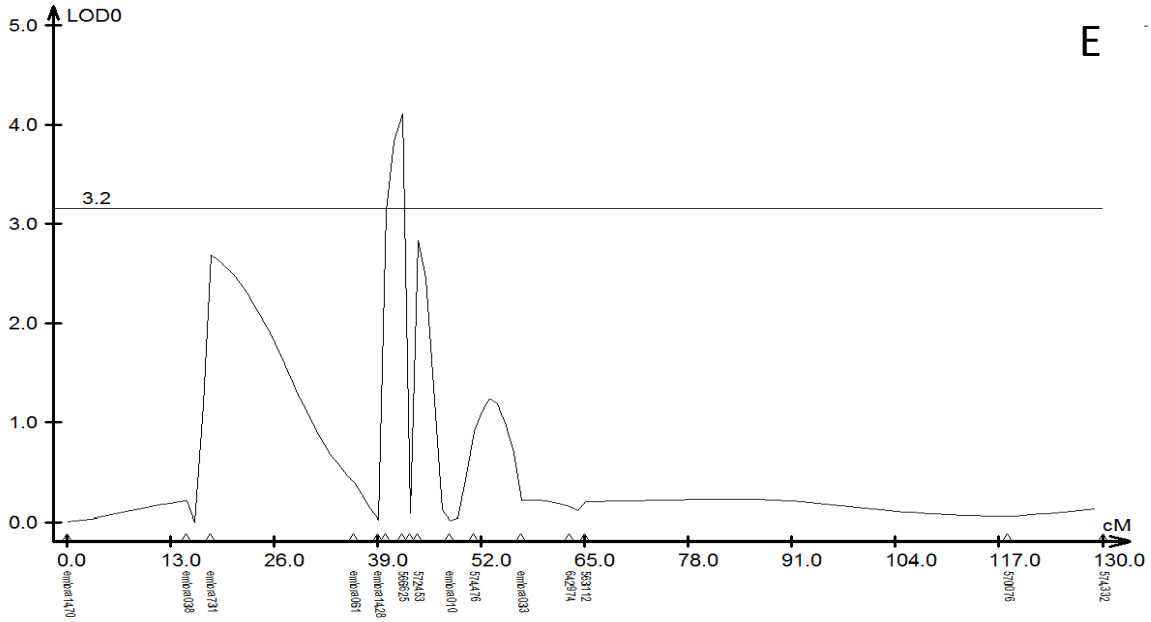


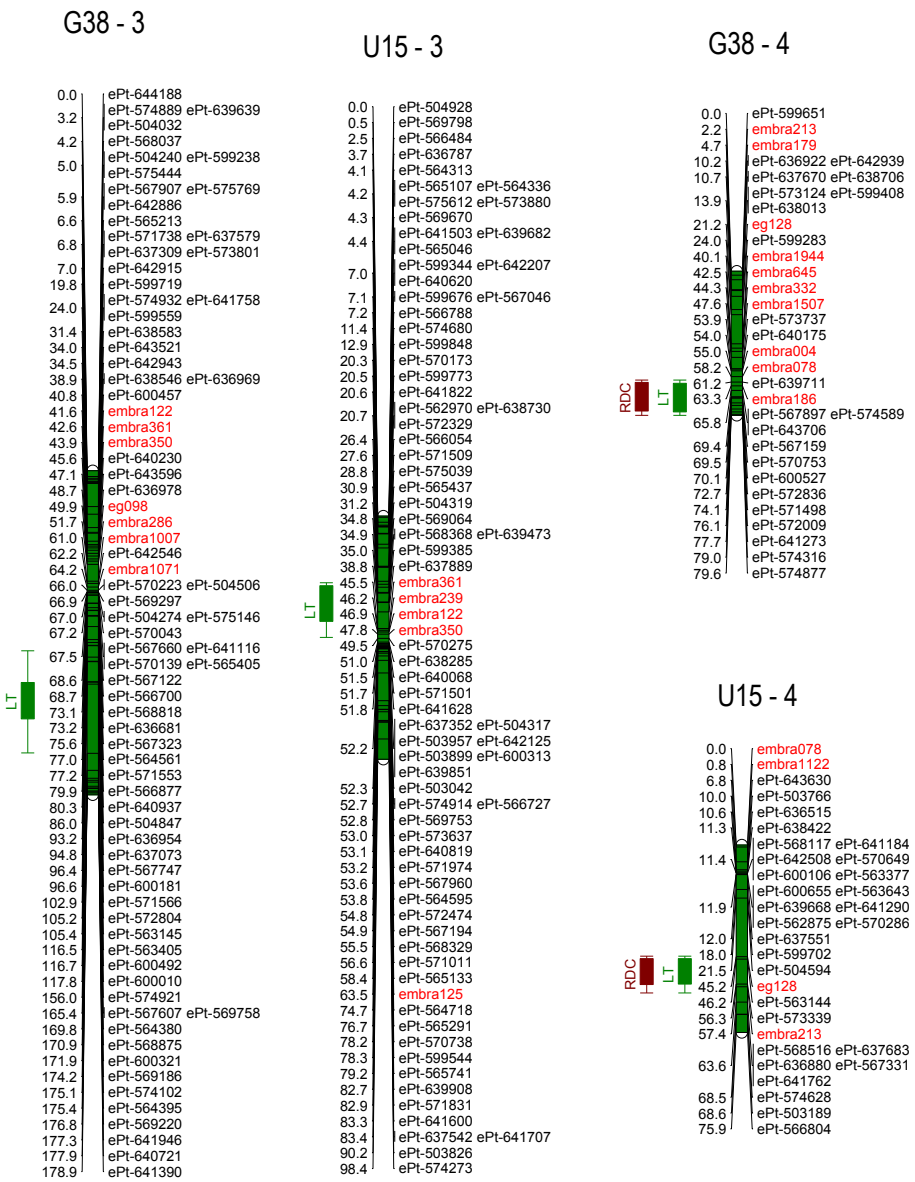
Figura 17. Gráficos gerados pelo programa QTL Cartographer demonstrando os QTLs para penetração do Pilodyn na madeira mapeados por intervalo composto no parental G38 nos grupos de ligação 1 (A) e 2 (B). No genoma do parental U15 foram detectados QTLs nos GL 1 (D), 2 (E) e 10 (E). Eixo Y: valor de LOD; Eixo X: intervalo entre os marcadores do grupo de ligação. O valor de LOD definido pelo teste de permutação com 5% de significância foi de 3,4 para G38 e 3,2 para U15.

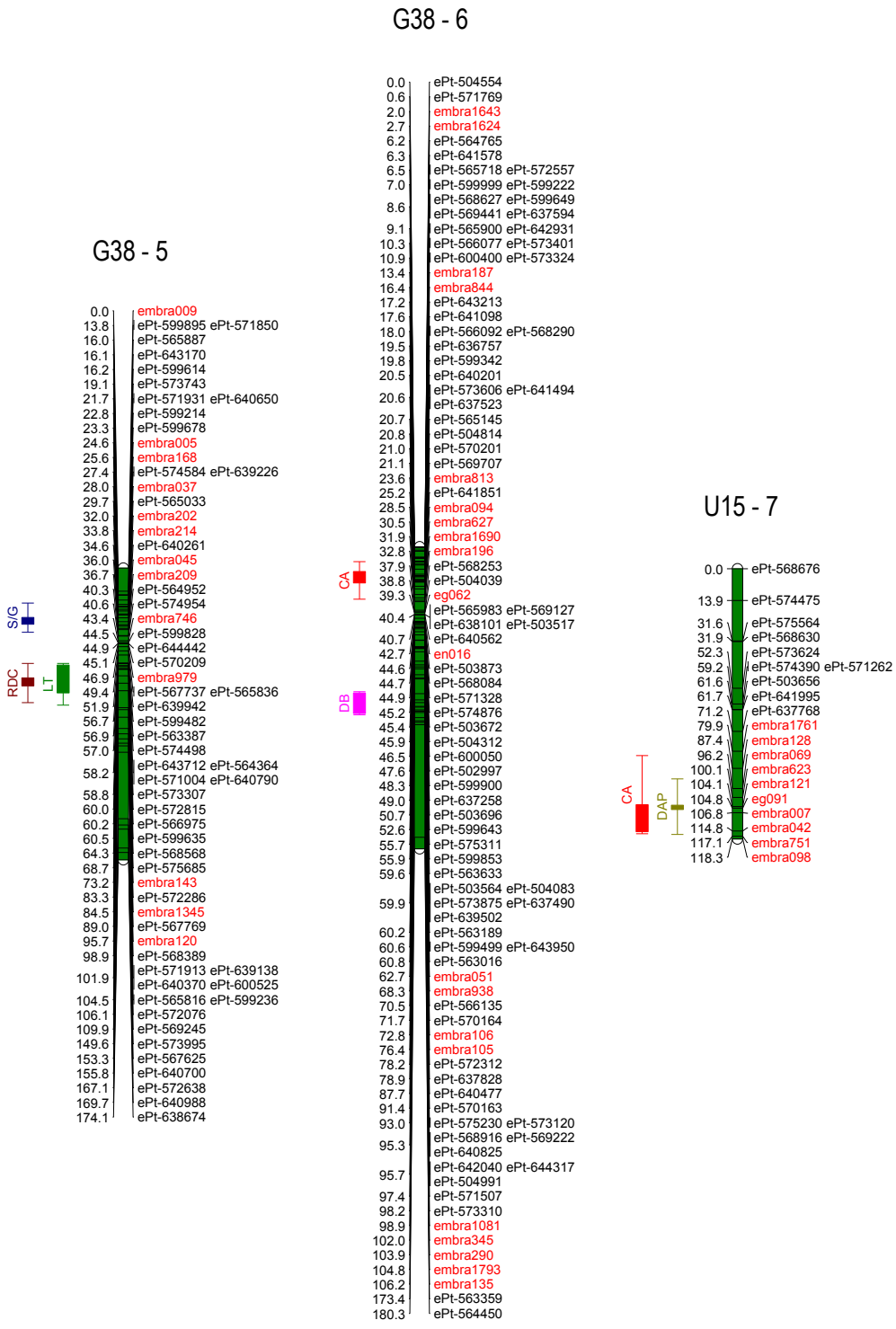
4.4.10. Co-localização dos QTLs para diferentes características

Para todas as sete características fenotípicas analisadas foi detectado pelo menos um QTL pelo método de intervalo composto (IC). Na figura 17 pode ser observada a disposição de todos os QTLs identificados em ambos os mapas parentais, assim como a co-localização, com coberturas diferentes, destes locos nos respectivos grupos de ligação. No mapa de *E. grandis* (G38), três QTLs para teor de Lignina Total foram mapeados nas mesmas posições que um QTL para a relação S/G no GL1 e QTLs para Rendimento Depurado de Celulose (RDC) nos GL4 e GL5. No mesmo mapa parental, outro loco associado à relação S/G foi co-localizado com um QTL relacionado com a Densidade Básica da madeira (GL8). No mapa paterno *E. urophylla* (U15), o QTL para

RDC no GL4 foi co-localizado com um QTL associado ao teor de Lignina Total, coincidindo com o resultado do mapa materno. Mesmo evento se observou entre os QTLs para relação S/G e Densidade Básica da madeira. Outro QTL para S/G co-localizou com QTL para RDC no GL9. Por outro lado, os dois QTLs que influenciam o Diâmetro à Altura do Peito foram co-localizados com dois dos três QTLs associados ao Crescimento em Altura (GL7 e GL10). No GL10, observaram-se, uma tripla sobreposição entre QTLs (Diâmetro à Altura do Peito – Crescimento em Altura – Penetração de Pilodyn).


G38 - 1
U15 - 1
G38 - 2
U15 - 2
S/G
LT
HT
CA
PP
RD
0.0 ePt-572142
5.3 ePt-572591
6.4 ePt-571996
7.4 ePt-642984
9.1 ePt-572242
10.9 ePt-571182
16.1 ePt-638597
17.5 ePt-570448
20.3 ePt-570932
26.1 ePt-572976
33.4 ePt-569501
61.7 ePt-568150
65.8 ePt-600019
66.4 ePt-565892 ePt-640846 ePt-575150
71.0 ePt-503701 ePt-504723 ePt-637966
71.1 ePt-637022 ePt-636685
74.4 ePt-599639
74.9 ePt-571015 ePt-643108
76.5 ePt-569944
77.7 embra147
83.3 embra006
86.0 ePt-571510
86.5 ePt-642723 ePt-570597
90.9 embra2000
92.0 embra1627
98.7 embra070
101.5 embra180
103.4 embra219
105.3 ePt-570119
106.0 embra1639
107.1 ePt-572431
107.9 ePt-642875
110.0 ePt-571936
110.9 embra934
112.6 ePt-571603
113.2 ePt-643939
114.3 embra632
117.5 embra012
120.7 ePt-569578
120.8 ePt-505085
126.4 ePt-563731 ePt-638419 ePt-643325 ePt-644032 ePt-639849 ePt-563014
126.5 ePt-600503
155.7 ePt-599628
156.2 ePt-574568
187.9 ePt-568498
191.2 ePt-639728
191.5 ePt-573462
196.4 ePt-570356
203.0 ePt-564573
208.2 ePt-571834
209.3 ePt-572812
213.2 ePt-569468
0.0 ePt-565025
1.7 ePt-573410
13.1 ePt-642731
19.2 ePt-567557
30.9 ePt-574533
31.2 ePt-573052
39.6 ePt-571540
44.0 ePt-565684
46.9 ePt-641705
50.3 ePt-575598
56.6 ePt-642989
62.1 ePt-564452 ePt-565647 ePt-638607 ePt-563830 ePt-599942 ePt-504436 ePt-600664 ePt-638369 ePt-505124 ePt-570633
70.0 embra219
78.9 embra934
84.0 embra012
87.5 embra632
95.3 embra070
100.6 embra180
101.9 ePt-636576
107.5 ePt-566342 ePt-504519
110.1 ePt-570414
110.5 ePt-566798
110.6 ePt-640361 ePt-641659
118.0 embra1627
121.3 ePt-566960
122.3 ePt-575761
122.5 ePt-567357
122.6 ePt-574232
122.7 ePt-574026
128.5 ePt-640491
131.5 ePt-599667 ePt-638457 ePt-566146 ePt-566200 ePt-638215 ePt-636507 ePt-562913 ePt-640386
133.2 ePt-643505 ePt-644117 ePt-575615 ePt-568160 ePt-565486
134.3 ePt-574816 ePt-639903 ePt-639141
134.4 ePt-638450
134.7 ePt-643522
135.0 ePt-639142
135.5 ePt-638572
137.0 ePt-566757
146.6 embra011
0.0 ePt-569565
1.4 ePt-569253
19.0 ePt-641261
21.2 ePt-503235
24.8 ePt-564654
28.3 ePt-599612
38.2 ePt-568757
60.6 ePt-573042
73.9 ePt-570523
78.8 ePt-505067
103.5 eg096
107.4 embra1810
108.9 embra648
111.4 ePt-563003
112.0 ePt-562922
113.2 ePt-600466
114.6 ePt-566429 ePt-638845 ePt-642027 ePt-565902
116.6 ePt-569525
122.4 embra201
125.9 embra027
128.9 ePt-600083
129.2 ePt-564786
129.8 ePt-565163
132.3 ePt-600397
137.4 embra333
138.5 ePt-599556
140.5 ePt-563425
141.8 ePt-566607
143.5 ePt-571180
145.9 embra072
148.2 ePt-566850
149.7 embra063
150.9 ePt-570024
152.9 embra110
153.4 embra898
155.2 embra043
156.0 embra915
156.8 embra172
159.4 embra126
161.0 embra1332
171.1 ePt-565990 ePt-563419
171.2 ePt-600542
181.8 ePt-574374
182.1 ePt-575243 ePt-639441
183.4 ePt-572763
185.2 ePt-504011
187.0 ePt-575224
190.2 ePt-599549
192.5 ePt-569193
193.1 ePt-637146
193.8 ePt-568083
196.6 ePt-569873
197.0 embra058
197.8 embra068
198.8 ePt-599675
200.4 ePt-573511
202.5 ePt-573171
204.4 ePt-638521
205.7 ePt-570187
208.6 ePt-568714
210.3 ePt-565047
214.3 ePt-503758
217.6 ePt-574221
219.1 ePt-568795
221.7 ePt-644363
222.0 ePt-600734
224.1 ePt-644292
225.6 ePt-503624
227.3 ePt-568041
230.0 ePt-571524
232.0 ePt-573068 ePt-504640
233.4 ePt-639529
236.6 ePt-565429
236.9 ePt-641621
239.1 ePt-574368
239.5 ePt-599975
240.9 ePt-572049
241.3 ePt-570845
242.6 ePt-638739
244.1 ePt-566370
244.8 ePt-574324
245.9 ePt-568679
252.4 ePt-600658
259.4 ePt-642561
263.3 ePt-641623
0.0 ePt-572240
9.8 ePt-641565
13.0 ePt-574111
14.7 ePt-575721
15.4 ePt-641216
16.8 ePt-642222
22.5 ePt-565390
24.7 ePt-504760
26.1 ePt-575337
31.1 ePt-572740
40.8 ePt-573142
43.5 embra058
44.6 embra068
54.9 ePt-566007
56.2 ePt-564663
57.1 ePt-504109
61.6 ePt-569562
61.7 ePt-564805
61.8 ePt-572571
62.3 ePt-643754
65.4 ePt-503732
65.8 ePt-600458
65.9 ePt-563498
66.0 ePt-568161
66.6 ePt-504595
66.9 ePt-599598
71.1 ePt-567864
71.5 ePt-563204
72.9 ePt-637076 ePt-570424
73.5 ePt-574930
73.7 ePt-573117
74.3 ePt-563231
74.6 ePt-569673
74.7 ePt-504602 ePt-641713 ePt-571763 ePt-503249
74.8 ePt-503994
74.9 ePt-503604
75.1 ePt-642609
75.2 ePt-570809
75.5 ePt-570657
76.8 ePt-573566
76.9 ePt-570261
77.0 ePt-504852
77.2 ePt-573877
77.3 ePt-640566
78.7 ePt-574308 ePt-599330 ePt-564993 ePt-639677
79.3 ePt-643189 ePt-599398 ePt-567410 ePt-573651 ePt-644056
81.5 ePt-643265 ePt-599401 ePt-639229 ePt-569480 ePt-642226 ePt-573099
81.6 ePt-567853
82.6 ePt-636833
82.7 ePt-643365
84.9 ePt-567796
87.3 ePt-570185
88.4 embra915
89.8 embra172
94.7 embra110
95.4 ePt-572195 ePt-640323 ePt-644136 ePt-565458 ePt-568856 ePt-639231
95.5 ePt-567973
95.6 ePt-644398
95.8 ePt-573051
97.2 ePt-565204
97.3 ePt-571889
97.5 ePt-563904
98.3 ePt-503579
98.6 ePt-639198
99.1 ePt-569815
101.2 ePt-503252
107.1 embra333
108.2 ePt-503134
115.6 ePt-637610
116.4 ePt-637861
119.1 ePt-569681
121.6 embra027
124.0 embra201
129.4 ePt-574129
133.9 embra648
135.2 embra1810
138.8 ePt-570066
139.0 ePt-570830
148.3 ePt-636614
148.7 ePt-643956


G38 - 3
0.0 ePt-644188
3.2 ePt-574889 ePt-639639
4.2 ePt-504032
5.0 ePt-568037
ePt-504240 ePt-599238
ePt-575444
5.9 ePt-567907 ePt-575769
6.6 ePt-642886
6.8 ePt-565213
ePt-571738 ePt-637579
7.0 ePt-637309 ePt-573801
19.8 ePt-642915
24.0 ePt-599719
ePt-574932 ePt-641758
31.4 ePt-599559
34.0 ePt-638583
34.5 ePt-643521
38.9 ePt-642943
40.8 ePt-638546 ePt-636969
41.6 ePt-600457
42.6 embra122
43.9 embra361
45.6 embra350
47.1 ePt-640230
48.7 ePt-643596
49.9 ePt-636978
51.7 eg098
61.0 embra286
62.2 embra1007
64.2 ePt-642546
66.0 embra1071
66.9 ePt-570223 ePt-504506
67.0 ePt-569297
67.2 ePt-504274 ePt-575146
67.5 ePt-570043
ePt-567660 ePt-641116
68.6 ePt-570139 ePt-565405
68.7 ePt-567122
73.1 ePt-566700
73.2 ePt-568818
75.6 ePt-636681
77.0 ePt-567323
77.2 ePt-564561
79.9 ePt-571553
80.3 ePt-566877
86.0 ePt-640937
93.2 ePt-504847
94.8 ePt-636954
96.4 ePt-637073
96.6 ePt-567747
102.9 ePt-600181
105.2 ePt-571566
105.4 ePt-572804
116.5 ePt-563145
116.7 ePt-563405
117.8 ePt-600492
156.0 ePt-600010
165.4 ePt-574921
169.8 ePt-567607 ePt-569758
170.9 ePt-564380
171.9 ePt-568875
174.2 ePt-600321
175.1 ePt-569186
175.4 ePt-574102
176.8 ePt-564395
177.3 ePt-569220
177.9 ePt-641946
178.9 ePt-640721
ePt-641390
LT
U15 - 3
0.0 ePt-504928
0.5 ePt-569798
2.5 ePt-566484
3.7 ePt-636787
4.1 ePt-564313
4.2 ePt-565107 ePt-564336
4.3 ePt-575612 ePt-573880
4.4 ePt-569670
ePt-641503 ePt-639682
7.0 ePt-565046
ePt-599344 ePt-642207
7.1 ePt-640620
7.2 ePt-599676 ePt-567046
11.4 ePt-566788
12.9 ePt-574680
20.3 ePt-599848
20.5 ePt-570173
20.6 ePt-599773
20.7 ePt-641822
ePt-562970 ePt-638730
26.4 ePt-572329
27.6 ePt-566054
28.8 ePt-571509
30.9 ePt-575039
31.2 ePt-565437
34.8 ePt-504319
34.9 ePt-569064
35.0 ePt-568368 ePt-639473
38.8 ePt-599385
45.5 ePt-637889
46.2 embra361
46.9 embra239
47.8 embra122
49.5 embra350
51.0 ePt-570275
51.5 ePt-638285
51.7 ePt-640068
51.8 ePt-571501
ePt-641628
ePt-637352 ePt-504317
52.2 ePt-503957 ePt-642125
ePt-503899 ePt-600313
ePt-639851
52.3 ePt-503042
52.7 ePt-574914 ePt-566727
52.8 ePt-569753
53.0 ePt-573637
53.1 ePt-640819
53.2 ePt-571974
53.6 ePt-567960
53.8 ePt-564595
54.8 ePt-572474
54.9 ePt-567194
55.5 ePt-568329
56.6 ePt-571011
58.4 ePt-565133
63.5 embra125
74.7 ePt-564718
76.7 ePt-565291
78.2 ePt-570738
78.3 ePt-599544
79.2 ePt-565741
82.7 ePt-639908
82.9 ePt-571831
83.3 ePt-641600
83.4 ePt-637542 ePt-641707
90.2 ePt-503826
98.4 ePt-574273
LT
G38 - 4
0.0 ePt-599651
2.2 embra213
4.7 embra179
10.2 ePt-636922 ePt-642939
10.7 ePt-637670 ePt-638706
13.9 ePt-573124 ePt-599408
ePt-638013
21.2 eg128
24.0 ePt-599283
40.1 embra1944
42.5 embra645
44.3 embra332
47.6 embra1507
53.9 ePt-573737
54.0 ePt-640175
55.0 embra004
58.2 embra078
61.2 ePt-639711
63.3 embra186
65.8 ePt-567897 ePt-574589
ePt-643706
69.4 ePt-567159
69.5 ePt-570753
70.1 ePt-600527
72.7 ePt-572836
74.1 ePt-571498
76.1 ePt-572009
77.7 ePt-641273
79.0 ePt-574316
79.6 ePt-574877
RDC
LT
U15 - 4
0.0 embra078
0.8 embra1122
6.8 ePt-643630
10.0 ePt-503766
10.6 ePt-636515
11.3 ePt-638422
ePt-568117 ePt-641184
11.4 ePt-642508 ePt-570649
ePt-600106 ePt-563377
ePt-600655 ePt-563643
11.9 ePt-639668 ePt-641290
ePt-562875 ePt-570286
12.0 ePt-637551
18.0 ePt-599702
21.5 ePt-504594
45.2 eg128
46.2 ePt-563144
56.3 ePt-573339
57.4 embra213
ePt-568516 ePt-637683
63.6 ePt-636880 ePt-567331
ePt-641762
68.5 ePt-574628
68.6 ePt-503189
75.9 ePt-566804
RDC
LT


G38 - 6
0.0 ePt-504554
0.6 ePt-571769
2.0 embra1643
2.7 embra1624
6.2 ePt-564765
6.3 ePt-641578
6.5 ePt-565718 ePt-572557
7.0 ePt-599999 ePt-599222
8.6 ePt-568627 ePt-599649
ePt-569441 ePt-637594
9.1 ePt-565900 ePt-642931
10.3 ePt-566077 ePt-573401
10.9 ePt-600400 ePt-573324
13.4 embra187
16.4 embra844
17.2 ePt-643213
17.6 ePt-641098
18.0 ePt-566092 ePt-568290
19.5 ePt-636757
19.8 ePt-599342
20.5 ePt-640201
20.6 ePt-573606 ePt-641494
ePt-637523
20.7 ePt-565145
20.8 ePt-504814
21.0 ePt-570201
21.1 ePt-569707
23.6 embra813
25.2 ePt-641851
28.5 embra094
30.5 embra627
31.9 embra1690
32.8 embra196
37.9 ePt-568253
38.8 ePt-504039
39.3 eg062
40.4 ePt-565983 ePt-569127
ePt-638101 ePt-503517
40.7 ePt-640562
42.7 en016
44.6 ePt-503873
44.7 ePt-568084
44.9 ePt-571328
45.2 ePt-574876
45.4 ePt-503672
45.9 ePt-504312
46.5 ePt-600050
47.6 ePt-502997
48.3 ePt-599900
49.0 ePt-637258
50.7 ePt-503696
52.6 ePt-599643
55.7 ePt-575311
55.9 ePt-599853
59.6 ePt-563633
ePt-503564 ePt-504083
59.9 ePt-573875 ePt-637490
ePt-639502
60.2 ePt-563189
60.6 ePt-599499 ePt-643950
60.8 ePt-563016
62.7 embra051
68.3 embra938
70.5 ePt-566135
71.7 ePt-570164
72.8 embra106
76.4 embra105
78.2 ePt-572312
78.9 ePt-637828
87.7 ePt-640477
91.4 ePt-570163
93.0 ePt-575230 ePt-573120
95.3 ePt-568916 ePt-569222
ePt-640825
95.7 ePt-642040 ePt-644317
ePt-504991
97.4 ePt-571507
98.2 ePt-573310
98.9 embra1081
102.0 embra345
103.9 embra290
104.8 embra1793
106.2 embra135
173.4 ePt-563359
180.3 ePt-564450
CA
DB
G38 - 5
0.0 embra009
13.8 ePt-599895 ePt-571850
16.0 ePt-565887
16.1 ePt-643170
16.2 ePt-599614
19.1 ePt-573743
21.7 ePt-571931 ePt-640650
22.8 ePt-599214
23.3 ePt-599678
24.6 embra005
25.6 embra168
27.4 ePt-574584 ePt-639226
28.0 embra037
29.7 ePt-565033
32.0 embra202
33.8 embra214
34.6 ePt-640261
36.0 embra045
36.7 embra209
40.3 ePt-564952
40.6 ePt-574954
43.4 embra746
44.5 ePt-599828
44.9 ePt-644442
45.1 ePt-570209
46.9 embra979
49.4 ePt-567737 ePt-565836
51.9 ePt-639942
56.7 ePt-599482
56.9 ePt-563387
57.0 ePt-574498
58.2 ePt-643712 ePt-564364
ePt-571004 ePt-640790
58.8 ePt-573307
60.0 ePt-572815
60.2 ePt-566975
60.5 ePt-599635
64.3 ePt-568568
68.7 ePt-575685
73.2 embra143
83.3 ePt-572286
84.5 embra1345
89.0 ePt-567769
95.7 embra120
98.9 ePt-568389
101.9 ePt-571913 ePt-639138
ePt-640370 ePt-600525
104.5 ePt-565816 ePt-599236
106.1 ePt-572076
109.9 ePt-569245
149.6 ePt-573995
153.3 ePt-567625
155.8 ePt-640700
167.1 ePt-572638
169.7 ePt-640988
174.1 ePt-638674
S/G
RDC
LT
U15 - 7
0.0 ePt-568676
13.9 ePt-574475
31.6 ePt-575564
31.9 ePt-568630
52.3 ePt-573624
59.2 ePt-574390 ePt-571262
61.6 ePt-503656
61.7 ePt-641995
71.2 ePt-637768
79.9 embra1761
87.4 embra128
96.2 embra069
100.1 embra623
104.1 embra121
104.8 eg091
106.8 embra007
114.8 embra042
117.1 embra751
118.3 embra098
CA
DAP

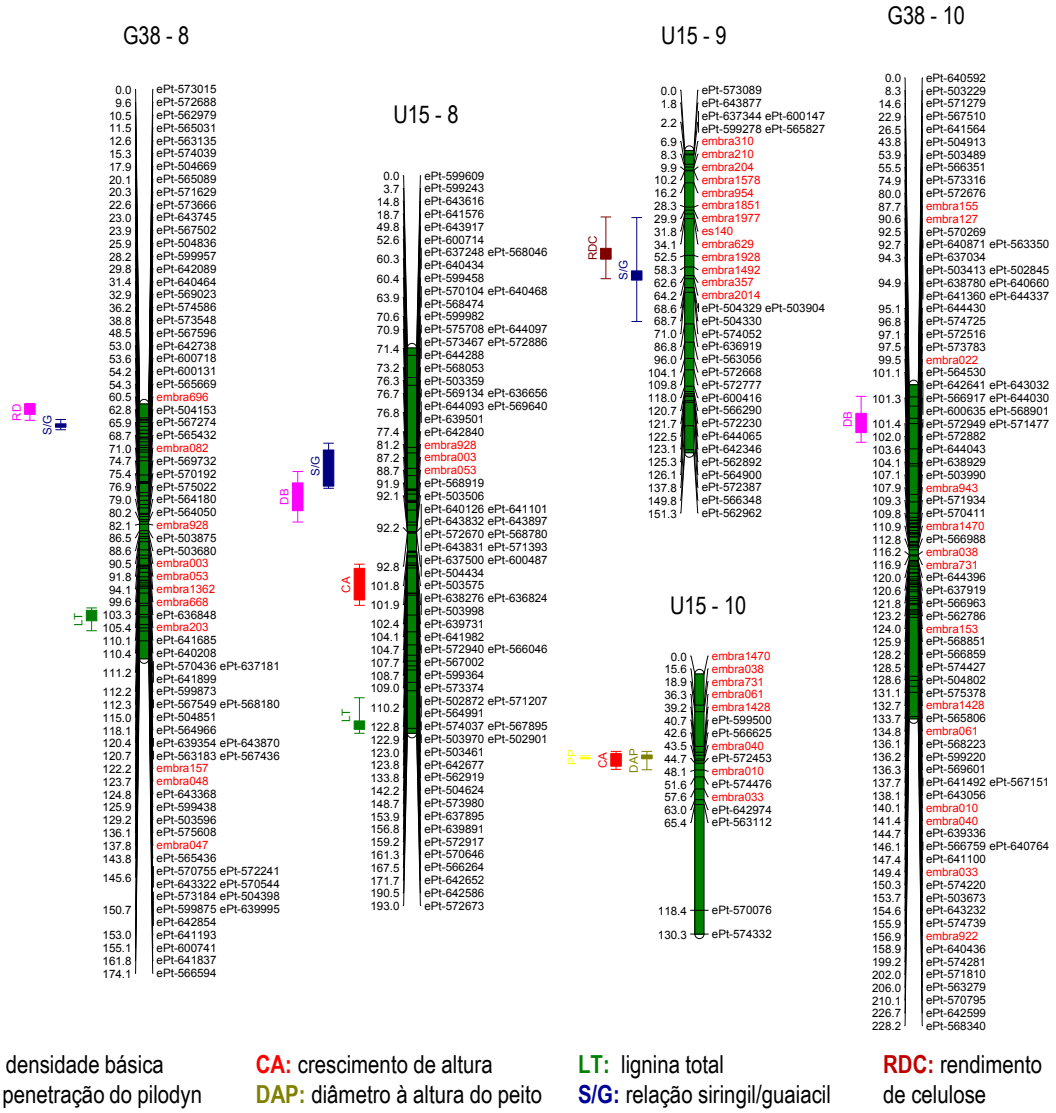


Figura 18. Localização dos QTLs para todas as características estudadas nos diferentes grupos de ligação (GL) de ambos os parentais (*E. grandis* e *E. urophylla*). As características analisadas foram: Densidade Básica da madeira (DB); Penetração do Pilodyn (PP); Crescimento em altura (CA); Diâmetro à Altura do Peito (DAP); Lignina Total (LT); relação Siringil/Guaiacil (S/G) e Rendimento Depurado de Celulose (RDC). Apenas os GL onde foram mapeados QTLs são apresentados. As barras representam as regiões onde o valor de LOD ultrapassou o limiar definido pelo teste de permutação com um nível de significância de 5%. Os marcadores DArTs estão representados pela cor preta e os marcadores microssatélites estão designados com a cor vermelha.

4.4.11. Conteúdo de genes nos intervalos de QTLs

As sequências de marcadores DArT e microssatélites flanqueando os intervalos nos quais os QTLs foram mapeados, foram usadas para extrair os genes anotados na atual versão do genoma referência de *Eucalyptus* (http://www.phytozome.net/). No total 8.036 modelos gênicos anotados foram encontrados dentro dos amplos intervalos genômicos compreendidos por todos os 18 QTLs de origem materna (*E. grandis*), com uma média de 446,4 genes por QTL. Para os 17 QTLs identificados no mapa paterno (*E. urophylla*), 6.678 genes foram encontrados, com uma média de 692,8 genes por QTL (Tabela 7). Na maioria dos intervalos genômicos analisados, os QTLs de origem paterna (U15) co-localizaram com um número maior de modelos gênicos anotados do que aqueles de origem materna (G38). Um total de 3.294 genes previamente identificados no genoma de referência encontraram-se co-localizados com o grupo de três QTLs maternos identificados para Densidade Básica. Este é o maior número de genes localizados na mesma região genômica entre todos os QTLs descobertos nesta análise. Por outro lado, nos dois QTLs de origem paterna para a característica Diâmetro à Altura do Peito, foram identificados apenas 228 genes no genoma referência de *Eucalyptus.* A característica com maior número de QTLs identificados neste estudo foi Lignina Total. Os cinco QTLs de origem materna e os três QTLs provenientes do genitor paterno encontraram-se co-localizados com uma quantidade similar de genes (Tabela 7).

Tabela 7. Número de QTLs detectados para cada característica e genitor e o número de modelos gênicos preditos observados nos intervalos genômicos correspondentes na versão atual do genoma de referência de *E. grandis*.

Característica	Parental	# de QTLs	# total de genes
CA	G38	3	914
	U15	3	1.599
LT	G38	5	1.698
	U15	3	1.749
S/G	G38	3	719
	U15	2	499
RDC	G38	2	790
	U15	3	496
DB	G38	3	3.294
	U15	1	505
PP	G38	2	890
	U15	3	1.602
DAP*	U15	2	228

*Genitores: G38, materno (E. grandis) e U15, paterno (E. urophylla). Número (#) de genes dentro do mesmo intervalo genômico e anotados previamente no genoma de Eucalyptus. Características avaliadas: Crescimento da Altura (CA); Lignina Total (LT); relação Siringil/Guaiacil (S/G); Rendimento Depurado de Celulose (RDC); Densidade Básica (DB); Penetração do Pylodin (PP) e Diâmetro à Altura do Peito (DAP). * Não foram identificados QTLs para DAP no genoma materno (G38 – E. grandis).*

A seguir, a Figura 18 representa o número de genes observados no pseudo-cromossomo 3 do genoma de referência contidos no intervalo de um dos cinco QTLs para Lignina Total em *E. grandis*. Na distância de 11 Mpb coberta pelo intervalo de confiança do QTL detectado envolvendo 9 marcadores DArT e 4 microssatélites, encontram-se distribuídos 895 modelos gênicos preditos. Uma vez reduzido esse espaço genômico a 100 kpb, o número de genes dessa região diminuiu para nove.

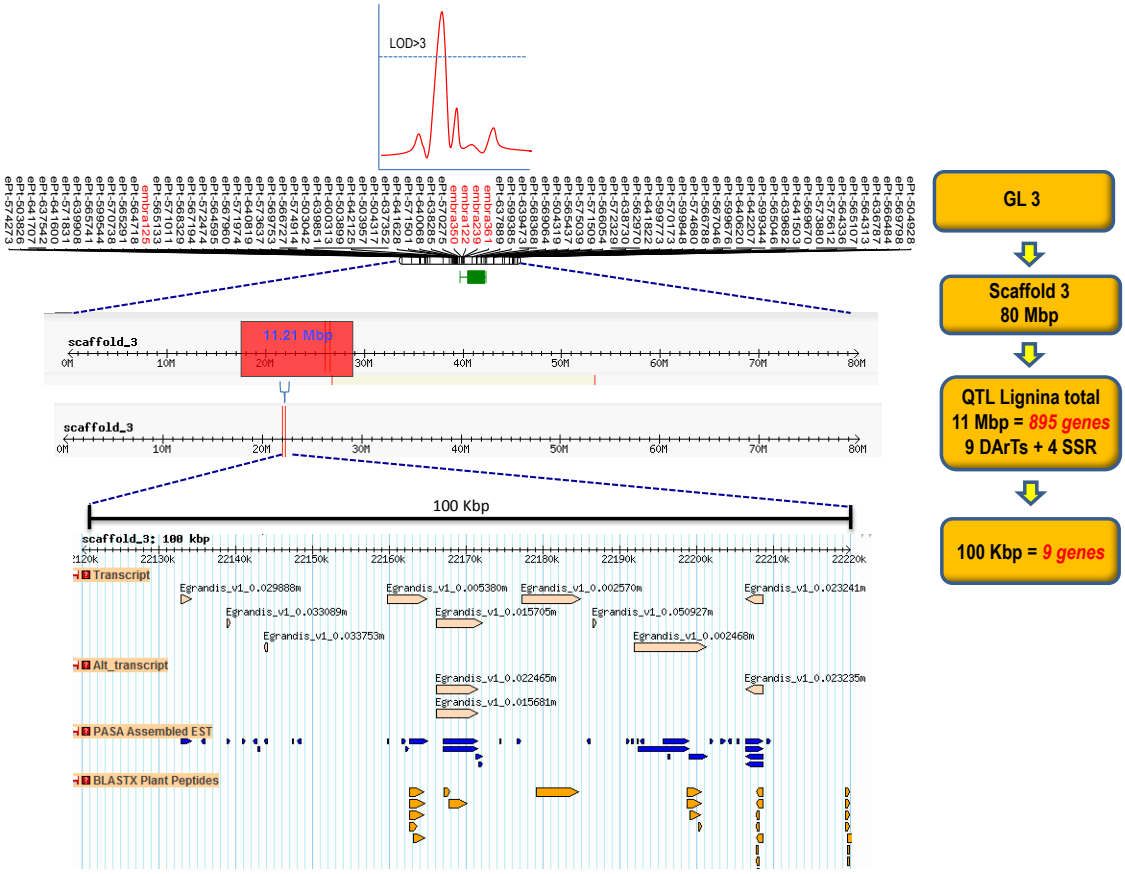


Figura 19. Representação do número de modelos gênicos preditos no genoma referência de *Eucalyptus* detectados na mesma região do QTL de maior efeito para teor de Lignina Total mapeado no grupo de ligação 3.

4.5. DISCUSSÃO

4.5.1. Análise comparativa com QTLs detectados em outros estudos em *Eucalyptus*

Diversos estudos de mapeamento comparativo demonstraram que os genomas das espécies de *Eucalyptus* comercialmente utilizadas, pertencentes ao subgênero *Symphyomyrtus* são sintênicos e colineares (Marques, Brondani et al. 2002; Brondani, Williams et al. 2006; Freeman, Potts et al. 2006; Hudson, Kullan et al. 2011; Kullan, van Dyk et al. 2012). Isto sugere que o posicionamento de QTLs, uma vez detectados, podem ser comparados entre espécies do gênero, embora não necessariamente sua magnitude e direção do efeito na característica alvo. Para realizar um mapeamento

comparativo de QTLs entre espécies diferentes, é necessário utilizar marcadores comuns entre mapas de ligação diferentes (Thumma, Baltunis et al. 2010). Porém, até hoje poucas são as análises de QTLs realizadas com a utilização de marcadores DArT no gênero *Eucalyptus* (Kullan, van Dyk et al. 2012; Freeman, Potts et al. 2013), e mesmo com microssatélites, em geral apenas algumas dezenas de marcadores comuns têm sido utilizadas. Isto evidentemente dificulta esta avaliação comparativa, permitindo apenas fazer comparações tentativas em nível de grupo de ligação, e apenas ocasionalmente de forma mais precisa em nível de mesmos intervalos genômicos ou flanqueados pelos mesmos marcadores. As análises comparativas abaixo descritas são, portanto, em sua maioria, realizadas apenas em nível de cromossomo e devem ser consideradas apenas sugestivas de homologia.

Um total de 35 QTLs foi encontrado para as sete características de crescimento e qualidade da madeira avaliadas (Tabela 5). Para todas as características, mais de um QTL foi detectado, cada um explicando uma proporção relativamente pequena, e provavelmente superestimada (veja a seguir), da variação. Este resultado corrobora o controle poligênico destas características conforme destacado em outro estudo recentes (Gion, Carouche et al. 2011). Uma arquitetura poligênica evidentemente também dificulta a comparação de QTLs entre diferentes estudos, tendo em vista que os efeitos são pequenos e distribuidos por todo o genoma.

A localização dos QTLs descritos para Densidade Básica da madeira coincide parcialmente com QTLs mapeados em estudos anteriores em *Eucalyptus.* Alguns QTLs têm demonstrado coincidência com análises que envolveram espécies como *E. globulus* (Freeman, Whittock et al. 2009), *E. nitens* (Thumma, Baltunis et al. 2010), *E. grandis* e *E. urophylla* (Grattapaglia, Bertolucci et al. 1996; Verhaegen, Plomion et al. 1997; Rocha, Barros et al. 2007; Gion, Carouche et al. 2011; Kullan, van Dyk et al. 2012). Nosso estudo revelou também um total de 5 QTLs para a penetração do pilodyn, os quais foram sintênicos com outros trabalhos em populações de *E. grandis* e *E. urophylla* (Verhaegen, Plomion et al. 1997; Gion, Carouche et al. 2011; Kullan, van Dyk et al. 2012).

Em ambos os mapas foram identificados QTLs associados ao Rendimento Depurado de Celulose no GL4. Este resultado condiz com o posicionamento, em nível cromossômico, de um QTL que foi detectado para o mesmo caráter em *E. globulus* (Moran, Thamarus et al. 2002; Thamarus, Groom et al. 2004) e *E. nitens* (Thumma, Baltunis et al. 2010). Entretanto, esta coincidência não foi observada para os demais QTLs detectados para esta característica nos outros grupos de ligação (GL1, GL5 e GL9). Também no GL4 de ambos os parentais, foram localizados QTLs responsáveis pela variação no teor de Lignina Total, consistente com o fato destas duas características estarem associadas, embora de forma inversa. QTLs associados ao teor de lignina foram também descritos anteriormente no GL4 em *E. nitens* (Thumma, Baltunis et al. 2010). Curiosamente, em *E. grandis* nosso estudo identificou o mesmo número de QTLs para a relação S/G que o trabalho publicado recentemente para o mesmo tipo de cruzamento interespecífico (Gion, Carouche et al. 2011), um deles mapeado também no GL1, enquanto os outros dois em cromossomos distintos.

Geralmente, as características de crescimento possuem uma herdabilidade menor do que as características de qualidade da madeira (Raymond 2002; Hamilton and Potts 2008), consequência, entende-se, do maior número de genes envolvidos no controle e do maior efeito ambiental sobre estas características (Raymond and Apiolaza 2004; Freeman, Potts et al. 2011). Para as características de crescimento avaliadas foram descritos no total 8 QTLs, três de origem materna (*E. grandis*) e outros cinco de origem paterna (*E. urophylla*). Especificamente para Diâmetro à Altura do Peito, foi localizado um par de QTLs unicamente a partir do parental *E. urophylla*. Um QTL foi posicionado no GL7,assim como no estudo realizado no primeiro mapeamento de QTLs para o gênero *Eucalyptus*, embora aquele QTL tenha sido identificado em *E. grandis* (Grattapaglia, Bertolucci et al. 1996). O segundo QTL foi mapeado no GL10, este coincide na sua localização cromossômica com outra análise feita também em *E. urophylla* (Kullan, van Dyk et al. 2012) e em *E. globulus* (Bundock, Potts et al. 2008; Freeman, Whittock et al. 2009). Não foram observadas coincidências de QTLs com outros estudos (Verhaegen, Plomion et al. 1997; Kirst, Myburg et al. 2004; Gion, Carouche et al. 2011). Por outro lado, dois dos três QTLs para o Crescimento em Altura foram mapeados nos GL1 e GL2 em *E. grandis*, e estes foram sintênicos aos obtidos

para a mesma espécie (Gion, Carouche et al. 2011). O posicionamento no GL10 do QTL identificado a partir de *E. urophylla* e que controla a mesma característica foi coincidente na sua localização cromossômica com outra análise realizada em *E. globulus* (Bundock, Potts et al. 2008). O fato dos QTLs para este segundo caráter de crescimento terem sido encontrados em cinco dos 11 grupos de ligação (GL2, GL6, GL7, GL8 e GL10) (Tabela 5, Figura 17), sugere que provavelmente várias regiões genômicas estejam envolvidas no controle do crescimento, consistente com resultados recentes de experimentos com maior poder de detecção em uma abordagem de seleção genômica (Resende, Resende et al. 2012).

A análise tentativa de mapeamento comparativo, em grande parte em nível de grupo de ligação, indica que mais de 45% dos QTLs detectados nesta análise foram também detectados de forma independente em outros experimentos de mapeamento de QTLs envolvendo famílias não relacionadas das mesmas espécies ou de espécies diferentes. Estes resultados sugerem que algum nível de coincidência de QTLs pode ser esperado embora testes estatísticos sejam necessários para avaliar esta hipótese contra a alternativa de uma coincidência devida ao simples acaso. Esta análise se torna, entretanto, impossível de ser realizada tendo em vista a baixa precisão do posicionamento dos QTLs pela falta de marcadores comuns aos vários experimentos de mapeamento.

Talvez a melhor análise comparativa de QTL detectados entre diferentes populações de uma espécie florestal foi recentemente realizada envolvendo duas populações geneticamente não relacionadas de *Eucalyptus* com 738 e 920 indivíduos, respectivamente, no âmbito de um experimento de seleção genômica (SG) (Resende, Resende et al. 2012). Essas populações, pertencentes a duas empresas distintas e compostas por indivíduos pertencentes a algumas dezenas de famílias, foram avaliadas para quatro características fenotípicas, crescimento em altura, diâmetro, densidade básica e rendimento em celulose e genotipadas com a mesma plataforma DArT utilizada neste estudo, para cerca de 2500 marcadores, sendo cerca de 1600 em comum entre as duas populações. Os resultados demonstraram uma coincidência altamente significativa no que diz respeito ao posicionamento genômico das centenas de marcadores associados a QTLs, todos de muito pequeno efeito. Cerca de 50% dos

QTLs apresentaram-se conservados em sua localização delimitada em bins genômicos de 500 kpb. Entretanto, a magnitude e direção dos efeitos (positivo ou negativo no caráter) dos QTLs variaram amplamente entre as duas populações, fazendo com que modelos de SG tivessem baixa acurácia preditiva entre as populações. Este resultado foi consistente com as expectativas teóricas de existência de variabilidade na fase de ligação entre alelos aos marcadores e alelos aos QTLs, ou seja, na estrutura do desequilíbrio de ligação. Além disso, as duas populações foram avaliadas em ambientes distintos, o que introduziu um efeito confundido da interação QTL com ambiente, contribuindo para as diferenças na magnitude e direção dos efeitos. A conclusão daquela análise indica que embora boa parte dos locos que controlam características complexas possa ser conservada entre populações quanto ao posicionamento genômico, os alelos e seus efeitos dificilmente o serão em função da heterogeneidade genética da espécie, impossibilitando a transferência de informação entre populações para fins práticos de seleção assistida (Resende, Resende et al. 2012). A inconsistência de QTLs entre diferentes estudos evidentemente dificulta a utilização da informação de QTLs para seleção em programas de melhoramento. Além disso, os dados mostram ainda que algumas dezenas ou centenas de locos estão envolvidos no controle da variação de características complexas como crescimento, de forma que a abordagem de localizar e tentar introgredir alelos favoráveis a múltiplos QTLs tem aplicabilidade prática muito limitada para não dizer nula (Bernardo 2008; Grattapaglia, Plomion et al. 2009).

4.5.2. Estimativas das proporções da variação explicada pelos QTLs

Da mesma forma que em todos os estudos de mapeamento anteriormente publicados para *Eucalyptus*, neste estudo, também, estimativas das proporções da variação explicada por todos os QTLs fornecidas pelo QTL Cartographer foram elevadas, variando de 5,5% para LT até incríveis 27,8% para S/G (Tabela 5). O QTL de maior efeito mapeado no parental *E. urophylla* controlaria assim quase 28% da variação fenotípica na relação Siringil/Guaiacil. Um loco com esta característica seria potencialmente muito interessante, uma vez que um aumento na relação S/G facilita a

extração da lignina no processo de polpação. Da mesma forma, um dos QTLs detectados a partir do parental *E. grandis* para Rendimento Depurado de Celulose explicaria 22,5% da variação fenotípica. Uma baixa concentração de lignina total e uma elevada relação Siringil/Guaiacil estão ligadas a uma maior eficiência na extração química de polpa de celulose, portanto são propriedades interessantes para seleção. QTLs de grande efeito também seriam aqueles detectados para a penetração do Pilodyn explicando entre 18,0% e 21,2% no genitor materno e 15,4% e 19,6% no genitor paterno (Tabela 5). Em *E. urophylla*, foi detectado um QTL de grande efeito responsável por estimados 25,4% da variação na capacidade de penetração do Pilodyn (Gion, Carouche et al. 2011). QTLs para Densidade Básica explicaram individualmente até 11,3% da variação fenotípica em *E. urophylla* (Grattapaglia, Bertolucci et al. 1996; Kullan, van Dyk et al. 2012) e até mesmo 29,4% em uma população hibrida *E. grandis* x *E. urophylla* (Rocha, Barros et al. 2007). Já para as características de crescimento, os 8 QTLs detectados explicariam individualmente proporções menores da variação, entre 6% e pouco mais de 10%, ou seja, cerca da metade do que QTLs para características de qualidade da madeira. Estes resultados coincidem com outros estudos em *Eucalyptus*, nos quais QTLs explicaram <10% da variação (Grattapaglia, Bertolucci et al. 1996; Freeman, Whittock et al. 2009; Thumma, Baltunis et al. 2010; Gion, Carouche et al. 2011; Kullan, van Dyk et al. 2012).

Infelizmente, entretanto, todos os estudos de mapeamento de QTLs em *Eucalyptus* publicados até o momento, inclusive este do presente estudo, sofrem do bem conhecido "efeito Beavis", em seguimento ao famoso experimento de simulação publicado por William Beavis para avaliar a eficiência do mapeamento de intervalo para detectar e estimar o efeito de poligenes (Beavis 1998). Aquele estudo demonstrou que quando uma progênie segregante de apenas 100 indivíduos é utilizada, o poder de detecção de QTLs de menor efeito é de apenas 3%. Além disso, a magnitude do efeito dos poucos QTLs que ultrapassam com sucesso o limite de detecção é superestimado em cinco ou até dez vezes, a depender da herdabilidade do caráter. Além disso, o efeito Beavis também causa um viés importante ao subestimar o número total de QTLs controlando a característica, justamente pelo fato do experimento não conseguir detectar QTLs de efeitos médios e pequenos.

O estudo de Beavis, baseado em simulações, foi mais tarde validado com um desenvolvimento estatístico teórico mostrando que este efeito decorre fundamentalmente do fato de que os QTLs detectados são amostrados de uma distribuição truncada ao se aplicar um limite crítico para a declaração do QTL (Xu 2003). Em seguida, um amplo estudo experimental em milho mapeando QTLs com 990 linhagens F_5 em 19 ambientes demonstrou que os 18 QTLs detectados para produtividade explicaram somente 42% da variância genética e mesmo estes estavam superestimados. Ao realizar validação cruzada com subconjuntos de indivíduos, somente 30% da variação pôde ser explicada. O problema da superestimativa dos efeitos se tornou exacerbado ao reduzir o tamanho da população de mapeamento para menos de 200 indivíduos. Estes resultados sugerem que o mapeamento de QTLs para características complexas (ex. produtividade, crescimento volumétrico), características estas possivelmente controladas por muitos genes de pequeno efeito (modelo infinitesimal), dificilmente, ou provavelmente nunca produzem resultados confiáveis do ponto de vista da variância fenotípica ou genética explicada, ao utilizar progênies de tamanho limitado de apenas poucas centenas de indivíduos. Os resultados apresentados na Tabela 5 deste estudo devem, portanto, ser considerados sob esta ótica. Métodos que buscam corrigir retrospectivamente essas estimativas para valores mais próximos da realidade tem sido propostos (Sun, Dimitromanolakis et al. 2011) e poderão ser utilizados nos nossos dados futuramente.

4.5.3. Co-localização entre QTLs mapeados para diferentes características

Vários dos caracteres estudados apresentaram correlações fenotípicas significativas (Tabela 6) e vários QTLs para diferentes características foram co-localizados no mesmo intervalo em ambos os mapas parentais. Dos 35 QTLs, 19 apresentaram algum tipo de sobreposição, em geral entre dois QTLs ou mesmo três em alguns casos (Figura 17). As três propriedades químicas da madeira avaliadas (LT, S/G e RDC) tiveram a maior quantidade de QTLs co-localizados, o que é esperado, tendo em vista que o teor de lignina e a relação dos tipos de lignina S/G são naturalmente correlacionados e ambos impactam diretamente o rendimento em

celulose. As correlações fenotípicas entre estas três características foram elevadas, entre LT e RDC (r_{xy} = -0,85; P<0,01), entre LT e S/G (r_{xy} = -0,58; P<0,01) e entre S/G e RDC (r_{xy} = 0,40; P<0,01), consistente com as estimativas disponíveis para *E. globulus* (Stackpole, Vaillancourt et al. 2011). Kullan *et al*. (Kullan, van Dyk et al. 2012) reportaram a co-localização entre QTLs para características de crescimento e qualidade da madeira, fato este que não foi observado neste estudo, com exceção da co-localização no GL 10 para densidade, crescimento em altura e diâmetro. A correlação fenotípica entre estas três características foi de baixa a moderada, mas como era de se esperar, a correlação entre Crescimento em Altura e Diâmetro à Altura do Peito foi positiva e elevada (r_{xy} = 0,78).

A co-localização de QTLs para características correlacionadas fornece evidência experimental adicional da validade dos QTLs detectados do ponto de vista de posicionamento no genoma. Esta observação pode ser devida a QTLs pleiotrópicos, ou seja, um mesmo QTL controlando diferentes características, ou à ligação física de diferentes genes que controlam as diferentes características e a resolução do mapeamento não permite separar os efeitos individuais. A hipótese mais parsimoniosa neste caso, entretanto, parece ser a de QTLs pleiotrópicos, considerando a detecção de várias co-localizações para as mesmas duas características e as elevadas correlações fenotípicas. Esta mesma conclusão foi também proposta no estudo semelhante ao nosso, recentemente publicado para uma população de mapeamento de E. *grandis x E. urophylla* (Gion, Carouche et al. 2011).

4.5.4. De QTLs para genes e seu uso no melhoramento

A abordagem de mapeamento e co-localização entre QTLs e genes expressos tem sido utilizada historicamente em estudos de espécies florestais como uma maneira de sugerir possíveis candidatos ou por vezes de tentar validar genes candidatos de forma indireta. Isto foi publicado para *Pinus taeda* (Brown, Bassoni et al. 2003), *Pinus pinaster* (Pot, Rodrigues et al. 2006) e *Picea glauca* (Pelgas, Bousquet et al. 2011) por exemplo, casos nos quais não existem genomas de referência e provavelmente continuarão sem contar com este recurso por algum tempo, apesar de esforços recentes terem sido

iniciados neste sentido (Neale and Kremer 2011). Em *Eucalyptus*, os primeiros relatos de co-localização tentativa de genes e QTLs foram em *E. globulus* com base em sondas de RFLP derivadas de cDNA para genes que codificam para enzimas da via de lignificação (Thamarus, Groom et al. 2004). Esta mesma abordagem foi utilizada para QTLs para ângulo da microfibra em *E. nitens* (Thumma, Baltunis et al. 2010). Recentemente esta abordagem foi novamente utilizada para sugerir conexão entre QTLs e o gene *ccr* que codifica para uma importante enzima da via de lignificação e tem sido proposto como um forte gene candidato para o controle do teor de lignina (Gion, Carouche et al. 2011), apesar das evidências serem somente indiretas.

Com a disponibilização da sequência de referência do genoma de *E. grandis*, a análise de QTL ganhou mais relevância ainda e tem sido proposta como uma abordagem chave para a identificação posicional de genes candidatos responsáveis pela variação nas características quantitativas de qualidade da madeira (Freeman, Potts et al. 2011). Partindo da premissa de que esta co-localização não é um evento aleatório, a co-localização entre QTLs e genes, poderia apontar para diversos genes, com funções conhecidas ou desconhecidas (Gion, Carouche et al. 2011). Os estudos de QTLs, entretanto, pelas limitações inerentes aos experimentos, tipicamente têm resultado em QTLs que cobrem amplas regiões genômicas envolvendo vários centiMorgans. Considerando que, em média, 1 cM corresponde entre 350 e 550 kpb, cada QTL provavelmente compreenderá algumas centenas de genes ou elementos reguladores *cis*. Por isso, diferentemente do que às vezes é proposto, o mapeamento de QTL, embora uma abordagem útil e não enviesada de conectar fenótipo ao genoma, representa apenas um passo inicial, de muito baixa resolução, rumo à identificação do(s) gene(s) ou sequências genômicas relevantes.

A avaliação realizada neste estudo demonstrou exatamente isso. O experimento de mapeamento de QTL realizado com uma progênie de 171 indivíduos representa um tamanho bastante comum em estudos deste tipo. A densidade de marcadores utilizada, entretanto, foi consideravelmente maior do que a que normalmente se usa. Além disso, o genoma de *Eucalyptus* além de ter um tamanho de moderado a pequeno para plantas em geral, conta com uma sequência de referência de altíssima qualidade, gerada por sequenciamento Sanger. Mesmo assim, ao proceder com esta abordagem,

alardeada como sendo de grande utilidade por alguns, centenas de modelos gênicos são observados, vários dos quais poderiam ser provisoriamente sugeridos como responsáveis pelo QTL. A validação de alguma proposta de gene candidato dependeria de experimentos adicionais complexos de validação. Isto, entretanto, seria um grande desafio em espécies florestais geneticamente heterogêneas, para as quais não existem populações de linhagens puras recombinantes ou linhagens quase isogênicas e os fenótipos são quantitativos. Além disso, mesmo se um ou mais genes forem identificados com sucesso, estes certamente explicarão apenas uma pequena proporção da variação fenotípica total na característica o que, do ponto de vista prático do melhoramento, será pouco útil.

4.6. CONCLUSÃO

Neste estudo foram identificados vários QTLs que controlam proporções aparentemente elevadas da variação fenotípica para um conjunto de características economicamente importantes em *Eucalyptus*. Estas estimativas são, entretanto, superestimadas pelas limitações inerentes ao experimento. Uma análise comparativa entre o nosso estudo e outros estudos de mapeamento de QTLs em espécies de *Eucalyptus* permitiu sugerir a sintenia de parte dos QTLs detectados, embora apenas em nível de cromossomo, sem resolução suficiente para efetivamente declarar homologias robustas. Uma avaliação da co-localização de QTLs com genes anotados no genoma de *Eucalyptus* demonstrou que centenas de genes poderiam ser tentativamente sugeridos como candidatos ao controle da variação nas características alvo do estudo, cuja validação experimental envolveria um esforço enorme e muito provavelmente inútil do ponto de vista prático, pois estes genes explicariam muito pouco da variação genética total.

Apesar do grande volume de informação de mapeamento de QTLs em *Eucalyptus*, não existe demonstração da utilização efetiva de QTLs no melhoramento operacional de espécies do gênero e em espécies florestais em geral. O mesmo vale para resultados de experimentos de genética de associação que, da mesma forma, resultam em algumas poucas associações, cujos efeitos, também superestimados (Sun,

Dimitromanolakis et al. 2011), explicam uma porção muito pequena da variação genética (Thumma, Nolan et al. 2005; Thumma, Matheson et al. 2009; Wegrzyn, Eckert et al. 2010). Várias são as razões para isso, já amplamente discutidas para o caso de espécies florestais (Grattapaglia, Plomion et al. 2009; Grattapaglia and Resende 2011),

É tácito entre melhoristas que características complexas são produto da interação coletiva de múltiplos genes dinâmicos ao longo do tempo e no espaço, e cujos efeitos são modificados por influências ambientais, principalmente em espécies perenes. Este ponto de vista tem sido cada vez mais corroborado por evidências experimentais nos mais diversos organismos. Em vista das dificuldades mencionadas e a falta de evidências ao longo dos últimos 25 anos que justifiquem o grande esforço que envolve a detecção de QTLs, é necessario capitalizar a utilização de outros métodos que possam correlacionar genótipos e fenótipos de uma maneria mais global, simples e rápida. Neste sentido, a seleção genômica ampla (Genome Wide Selection - GWS) (Meuwissen, Hayes et al. 2001) tem surgido como uma alternativa prática para capturar, simultaneamente, a maior parte da variação para múltiplos caracteres quantitativos. Esta abordagem preditiva dispensa a necessidade de mapear e localizar QTLs ou genes, mas foca exclusivamente nos aspectos de eficiência operacional e ganho genético. Recentemente tem sido proposta para espécies florestais (Grattapaglia, Plomion et al. 2009; Grattapaglia and Resende 2011; Iwata, Hayashi et al. 2011; Denis and Bouvet 2012) e seu potencial demonstrado em experimentos em *Eucalyptus* (Resende, Resende et al. 2012) e *Pinus* (Resende, Munoz et al. 2012). Estes resultados indicam que a seleção genômica ampla deverá ser tema de intensa pesquisa com grande potencial de aplicação nos próximos anos.

5. BIBLIOGRAFIA

ABRAF. (2012). "Associação Brasileira de Produtores de Florestas Plantadas." from www.abraflor.org.br.

Akbari, M., P. Wenzl, et al. (2006). "Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome." TAG Theoretical and Applied Genetics **113**: 1409 - 1420.

Alves-Freitas, D. M. T., A. Kilian, et al. (2010). Towards a high-density DArT (Diversity Arrays Technology) microarray for high-throughput genotyping of Pinus taeda and closely related species. Resumos do 56º Congresso Brasileiro de Genética, Santos.

Alves, A., C. Rosado, et al. (2011). "Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of *Eucalyptus* to *Puccinia psidii* rust infection." Euphytica: 1-12.

Bailey, K., V. Cevik, et al. (2011). "Molecular Cloning of ATR5(Emoy2) from Hyaloperonospora arabidopsidis, an Avirulence Determinant That Triggers RPP5-Mediated Defense in Arabidopsis." Molecular Plant-Microbe Interactions **24**(7): 827-838.

Bariana, H., U. Bansal, et al. (2010). "Molecular mapping of adult plant stripe rust resistance in wheat and identification of pyramided QTL genotypes." Euphytica **176**(2): 251-260.

Bartos, J., S. R. Sandve, et al. (2011). "Genetic mapping of DArT markers in the *Festuca-Lolium* complex and their use in freezing tolerance association analysis." Theoretical and Applied Genetics **122**(6): 1133-1147.

Beavis, W. D. (1998). QTL analyses: power, precision, and accuracy. Molecular dissection of complex traits. A. H. Patterson. Boca Raton, Florida, CRC Publishing**:** 145-162.

Bedo, J., P. Wenzl, et al. (2008). "Precision-mapping and statistical validation of quantitative trait loci by machine learning." BMC Genetics **9**(1): 35.

Belaj, A., M. d. C. Dominguez-Garcia, et al. (2012). "Developing a core collection of olive (Olea europaea L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits." Tree Genetics & Genomes **8**(2): 365-378.

Bernardo, R. (2008). "Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years." Crop Sci. **48**(5): 1649-1664.

Bison, O., M. A. P. Ramalho, et al. (2009). "Dialelo parcial entre clones de Eucalyptus camaldulensis e clones de *E. urophylla*, *E. grandis* e *E. saligna*." Revista Árvore **33**: 395-402.

Bolibok-Bragoszewska, H., K. Heller-Uszynska, et al. (2009). "DArT markers for the rye genome - genetic diversity and mapping." BMC Genomics **10**(1): 578.

Bradshaw, H. D., Jr. and R. F. Stettler (1995). "Molecular genetics of growth and development in populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree." Genetics **139**(2): 963-73.

Brondani, R. P., C. Brondani, et al. (2002). "Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers." Molecular Genetics and Genomics **267**(3): 338-347.

Brondani, R. P., E. R. Williams, et al. (2006). "A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus." BMC Plant Biol **6**: 20.

Brondani, R. P. V., C. Brondani, et al. (1998). "Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*." TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **97**(5/6): 816-827.

Brondani, R. P. V. and D. Grattapaglia (2001). "Cost-effetive method to synthetise a fluorescent internal DNA standard for automated fragment sizing. ." Biotechniques **31**: 793-795.

Brooker, M. I. H. (2000). "A new classification of genus Eucalyptus L´Her. (Myrtaceae)." Australian Systematic Botany **13**: 79-148.

Brown, G. R., D. L. Bassoni, et al. (2003). "Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL Verification and candidate gene mapping." Genetics **164**: 1537-46.

Brown, G. R., E. E. Kadel, 3rd, et al. (2001). "Anchored reference loci in loblolly pine (Pinus taeda L.) for integrating pine genomics." Genetics **159**(2): 799-809.

Brown, R. G., G. P. Gill, et al. (2004). "Nucleotide diversity and linkage disequilibrium in loblolly pine." Proc Natl Acad Sci U S A **101**(42): 15255-15260.

Bundock, P. C., M. Hayden, et al. (2000). "Linkage maps of Eucalyptus globulus using RAPD and microsatellite markers." Silvae Genetica **49**(4-5): 223-232.

Bundock, P. C., B. M. Potts, et al. (2008). "Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*." Tree Genetics & Genomes **4**: 85-95.

Burgess, I. P. and J. C. Bell (1983). "Comparative morphology and allozyme frequencies of *Eucalyptus grandis* Hill ex Maiden and *Eucalyptus saligna* SM." Aust. Forest Res. **13**: p.133-149.

Butcher, P., M. McDonald, et al. (2009). "Congruence between environmental parameters, morphology and genetic structure in Australia's most widely distributed eucalypt, *Eucalyptus camaldulensis*." Tree Genetics & Genomes **5**(1): 189-210.

Butcher, P. A., A. Otero, et al. (2002). "Nuclear RFLP variation in Eucalyptus camaldulensis Dehnh. from northern Australia." Heredity **88**(5): 402-12.

Byrne, M. and B. Hines (2004). "Phylogeographical analysis of cpDNA variation in *Eucalyptus loxophleba* (Myrtaceae)." Aust. J. Bot. **52**: p.459-470.

Byrne, M., M. I. Marquezgarcia, et al. (1996). "Conservation and Genetic Diversity of Microsatellite loci in the Genus Eucalyptus." Australian Journal of Botany **44**: 331-341.

Byrne, M., J. C. Murrell, et al. (1995). "An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers." Theoretical and Applied Genetics **91**: 869 - 875.

Byrne, M., J. C. Murrell, et al. (1997). "Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in Eucalyptus nitens." Theoretical and Applied Genetics **94**(5): 674-681.

Byrne, M., J. C. Murrell, et al. (1997). "Mapping of quantitative trait loci influencing frost tolerance in Eucalyptus nitens." Theoretical and Applied Genetics **95**(5-6): 975-979.

Cao, J., K. Schneeberger, et al. (2011). "Whole-genome sequencing of multiple *Arabidopsis thaliana* populations." Nature Genetics **43**: 956-963.

Carlson, J. E., L. K. Tulsieram, et al. (1991). "Segregation of random amplified DNA markers in F1 progeny of conifers." Theoretical and Applied Genetics **83**: 194-200.

Carneiro, M. S. and M. L. C. Vieira (2002). "Mapas genéticos em plantas." Bragantia **61**: 89-100.

Cheema, J. and J. Dicks (2009). "Computational approaches and software tools for genetic linkage map estimation in plants." Briefings in Bioinformatics **10**(6): 595-608.

Churchill, G. A. and R. W. Doerge (1994). "Empirical threshold values for quantitative trait mapping." Genetics **138**(3): 963-71.

Coelho, A. S. G. (2000). CONSIDERAÇÕES GERAIS SOBRE A ANÁLISE DE QTL's. **Análise de QTL no Melhoramento de Plantas**. Goiânia, Brasil, Pinheiro, J.B.;Carneiro, I.F.**:** 4.

Coelho, A. S. G. and H. D. Silva (2002). Construção de Mapas Genéticos e Mapeamento de QTL's. Piracicaba, ESALQ, apostila**:** p.77.

Coelho, A. S. G. and H. D. Silva (2005). Métodos biométricos aplicados a análise de QTL's. 11 Simpósio de Estatística Aplicada à Experimentação Agronômica & 50 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometría., Paraná: Londrina.

Collard, B., M. Jahufer, et al. (2005). "An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts." Euphytica **142**(1): 169-196.

Condit, R. and S. P. Hubbell (1991). "Abundance and DNA sequence of two-base repeat regions in tropical tree genomes." Genome **34**(1): 66-71.

Creighton, H. B. and B. McClintock (1931). "A correlation of cytological and genetical crossing-over in *Zea mays*." Electronic Scholarly Publishing. Botany Department, Cornell University, Ithaca, New York. **17**: 492-497.

Crossa, J., J. Burgueno, et al. (2007). "Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure." Genetics **177**(3): 1889-913.

Crossa, J., G. d. l. Campos, et al. (2010). "Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers." Genetics **186**(2): 713-724.

Denis, M. and J.-M. Bouvet (2012). "Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding." Tree Genetics & Genomes: 1-15.

Devey, M. E., K. D. Jermstad, et al. (1991). "Inheritance of RFLP loci in a loblolly pine three-generation pedigree." TAG Theoretical and Applied Genetics **83**(2): 238-242.

Doerge, R. W. (2002). "Mapping and analysis of quantitative trait loci in experimental populations." Nat Rev Genet **3**(1): 43-52.

Drost, D. R., E. Novaes, et al. (2009). "A microarray-based genotyping and genetic mapping approach for highly heterozygous outcrossing species enables localization of a large fraction of the unassembled *Populus trichocarpa* genome sequence." The Plant Journal **58**(6): 1054-1067.

Drummond, A. J., B. Ashton, et al. (2011). Geneious v5.4 http://www.geneious.com/.

Eldridge, K., J. Davidson, et al. (1994). Eucalypt Domestication and Breeding, Oxford University Press, USA.

Elliott, C. P. and M. Byrne (2004). "Phylogenetics and the conservation of rare taxa in the *Eucalyptus angustissima* complex in Western Australia." Conservation Genetics **5**(1): 39-47.

Elshire, R. J., J. C. Glaubitz, et al. (2011). "A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species." Plos One **6**(5): e19379.

Faria, D., E. Mamani, et al. (2011). "Genotyping systems for Eucalyptus based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests." Tree Genetics & Genomes **7**(1): 63-77.

Faria, D. A., E. M. Mamani, et al. (2010). "A selected set of EST-derived microsatellites, polymorphic and transferable across 6 species of eucalyptus." J Hered **101**(4): 512-20.

Ferreira, G. W., J. V. Gonzaga, et al. (1997). "Qualidade da Celulose Kraft-Antraquinona de *Eucalyptus dunni* Plantado em Cinco Espaçamentos em Relação ao *Eucalyptus grandis* e *Eucalyptus saligna*." Ciência Florestal, Santa Maria: v.7, n.1, p.41-63.

Ferreira, M. and P. E. T. Santos (1997). Melhoramento Genético Florestal do *Eucalyptus* no Brasil - Breve Histórico e Perspectivas. INTERNATIONAL IUFRO CONFERENCE ON EUCALYPTUS GENETICS AND SILVICULTURE Salvador, Brazil.

Ferreira, M. E. and D. Grattapaglia (1998). Introdução ao uso de marcadores moleuculares em análise genética. Brasília, Embrapa.

Firmino-Winckler, D. C., C. F. Wilcken, et al. (2009). "Biologia do psiládeo-de-concha *Glycaspis brimblecombei Moore* (Hemiptera, Psyllidae) em *Eucalyptus spp*." Revista Brasileira de Entomologia **53**: 144-146.

Fisher, R. A. (1918). "The correlation between relatives on the supposition of Mendelian inheritance." **Trans. Roy. Soc. Edinb.**(**52**): p.399-433.

Freeman, J., B. Potts, et al. (2011). "QTL analysis for growth and wood properties across multiple pedigrees and sites in *Eucalyptus globulus*." BMC Proceedings **5**(Suppl 7): O8.

Freeman, J. S., J. M. O'Reilly-Wapstra, et al. (2008). "Quantitative trait loci for key defensive compounds affecting herbivory of eucalypts in Australia." New Phytologist **178**(4): 846-851.

Freeman, J. S., B. M. Potts, et al. (2013). "Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in Eucalyptus globulus." New Phytol.

Freeman, J. S., B. M. Potts, et al. (2006). "Parental and consensus linkage maps of *Eucalyptus globulus* using AFLP and microsatellite markers." Silvae Genetica **55**: 202 - 217.

Freeman, J. S., S. P. Whittock, et al. (2009). "QTL influencing growth and wood properties in *Eucalyptus globulus*." Tree Genetics & Genomes **5**: 713 - 722.

Gaiotto, F. A., M. Bramucci, et al. (1997). "Estimation of outcrossing rate in a breeding population of Eucalyptus urophylla with dominant RAPD and AFLP markers." TAG Theoretical and Applied Genetics **95**(5): 842-849.

Ganal, M. W., T. Altmann, et al. (2009). "SNP identification in crop plants." Current Opinion in Plant Biology **12**(2): 211-217.

Gardner, E. J. and D. P. Snustad (1986). **Genética**. Rio de Janeiro, Guanabara Koogan, cap.9, p.497.

Gindl, W., H. S. Gupta, et al. (2004). "Mechanical properties of spruce wood cell walls by nanoindentation." Applied Physics A: Materials Science & Processing **79**(8): 2069-2073.

Gindl, W. and A. Teischinger (2002). "Axial compression strength of Norway spruce related to structural variability and lignin content." Composites Part A: Applied Science and Manufacturing **33**(12): 1623-1628.

Gion, J. M., A. Carouche, et al. (2011). "Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: Eucalyptus." Bmc Genomics **12**(1): 301.

Gion, J. M., P. Rech, et al. (2000). "Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes." Molecular Breeding **6**: 441-449.

Golle, D. P., L. R. S. Reiniger, et al. (2009). "Melhoramento florestal: ênfase na aplicação da biotecnologia." Ciência Rural **39**(5): 1607-1614.

Gomide, J. L., J. L. Colodette, et al. (2005). "Caracterização tecnológica, para produção de celulose, da nova geração de clones de *Eucalyptus* do Brasil." Revista Árvores **29**: 129-137.

Gonçalves, F. M. A., G. D. Rezende, et al. (2001). "Progresso Genético por Meio da Seleção de Clones de Eucalipto em Plantios Comerciais." Revista Árvore **v.25, n.3**: p.289-293.

Gonzalez-Martinez, S. C., E. Ersoz, et al. (2006). "DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in Pinus taeda L." Genetics **172**(3): 1915-26.

Grattapaglia, D. (2004). "Integrating genomics into *Eucalyptus* breeding." Genetics and Molecular Research **3**(3): 369-379.

Grattapaglia, D., F. L. Bertolucci, et al. (1996). "Genetic mapping of quantitative trait loci controlling growth and wood quality traits in Eucalyptus grandis using a maternal half-sib family and RAPD markers." Genetics **144**(3): 1205-14.

Grattapaglia, D., F. L. Bertolucci, et al. (1995). "Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers." Theoretical and Applied Genetics **90**: 933-947.

Grattapaglia, D. and H. D. Bradshaw (1994). "Nuclear DNA content of commercially important *Eucalyptus* species and hybrids." Canadian Journal of Forest Research **24**(5): 1074-1078.

Grattapaglia, D. and M. Kirst (2008). "Eucalyptus applied genomics: from gene sequences to breeding tools." New Phytol **179**(4): 911-29.

Grattapaglia, D., C. Plomion, et al. (2009). "Genomics of growth traits in forest trees." Current Opinion in Plant Biology **12**(2): 148-156.

Grattapaglia, D. and M. D. V. Resende (2011). "Genomic selection in forest tree breeding." Tree Genetics & Genomes **7**(2): 241-255.

Grattapaglia, D., V. J. Ribeiro, et al. (2004). "Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for Eucalyptus." Theor Appl Genet **109**(1): 192-9.

Grattapaglia, D., C. P. Sansaloni, et al. (2010). Genomic Selection In Eucalyptus: Marker Assisted Selection Coming To Reality In Forest Trees. Plant and Animal Genome XVIII Conference. San Diego**:** Abstract W 237.

Grattapaglia, D. and R. Sederoff (1994). "Genetic-Linkage Maps of Eucalyptus-Grandis and Eucalyptus-Urophylla Using a Pseudo-Testcross - Mapping Strategy and Rapd Markers." Genetics **137**(4): 1121-1137.

Grattapaglia, D., O. B. Silva-Junior, et al. (2011). "High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: assay success, polymorphism and transferability across species." BMC plant biology **11**(1): 65.

Grattapaglia, D., R. E. Vaillancourt, et al. (2012). "Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus." Tree Genetics & Genomes **8**(3): 463-508.

Grattapaglia, D., P. Wilcox, et al. (1991). "A RAPD map of loblolly pine in 60 days." Third International Congress of the International Society for Plant Molecular Biology: abstract 2224.

Griffin, A. R., I. P. Burgess, et al. (1988). "Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit. - a review." Australian Journal of Botany **36**: 41-66.

Groover, A., M. Devey, et al. (1994). "Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine." Genetics **138**(4): 1293-300.

Hackett, C. and L. Broadfoot (2003). "Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps." Heredity **90**(1): 33-38.

Haley, C. S. and S. A. Knott (1992). "A simple regression method for mapping quantitative trait loci in line crosses using flanking markers." Heredity **69**(4): 315-324.

Hamilton, M. G. and B. M. Potts (2008). "Review of Eucalyptus nitens genetic parameters." New Zealand Journal of Forestry Science **38**(1): 102–119.

He, X. and Å. Bjørnstad (2012). "Diversity of North European oat analyzed by SSR, AFLP and DArT markers." TAG Theoretical and Applied Genetics: 1-14.

Henery, M. L., G. F. Moran, et al. (2007). "Identification of quantitative trait loci influencing foliar concentrations of terpenes and formylated phloroglucinol compounds in Eucalyptus nitens." New Phytologist **176**(1): 82-95.

Hippolyte, I., F. Bakry, et al. (2010). "A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas." BMC Plant Biology **10**: 65.

Hori, R., H. Suzuki, et al. (2003). "Variation of microfibril angles and chemical composition: Implication for functional properties." Journal of Materials Science Letters **22**(13): 963-966.

Howard, E. L., S. P. Whittock, et al. (2011). "High-throughput genotyping of hop (*Humulus lupulus* L.) utilising diversity arrays technology (DArT)." Theoretical and Applied Genetics **122**(7): 1265-1280.

Hudson, C., A. Kullan, et al. (2011). "High synteny and colinearity among Eucalyptus genomes revealed by high-density comparative genetic mapping." Tree Genetics & Genomes: 1-14.

Hudson, C. J., J. S. Freeman, et al. (2012). "A reference linkage map for eucalyptus." BMC Genomics **13**(1): 240.

Huynh, B.-L., H. Wallwork, et al. (2008). "Quantitative trait loci for grain fructan concentration in wheat (Triticum aestivum L.)." TAG Theoretical and Applied Genetics **117**(5): 701-709.

Iwata, H., T. Hayashi, et al. (2011). "Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*." Tree Genetics & Genomes **7**(4): 747-758-758.

Jaccoud, D., K. Peng, et al. (2001). "Diversity arrays: a solid state technology for sequence information independent genotyping." Nucleic Acids Res **29**(4): E25.

Jackson, Steane, et al. (1999). "Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae)." Molecular Ecology **8**(5): 739-751.

James, K. E., H. Schneider, et al. (2008). "Diversity Arrays Technology (DArT) for Pan-Genomic Evolutionary Studies of Non-Model Organisms." PLoS One **3**(2): e1682.

Jansen, R. C. (1993). "Interval Mapping of Multiple Quantitative Trait Loci." Genetics **135**(1): 205-211.

Jones, M., M. Shepherd, et al. (2008). "Pollen flow in *Eucalyptus grandis* determined by paternity analysis using microsatellite markers." Tree Genetics & Genomes **4**(1): 37-47.

Jones, N., H. Ougham, et al. (2009). "Markers and mapping revisited: finding your gene." New Phytologist **183**(4): 935-966.

Junghans, D. T., A. C. Alfenas, et al. (2003). "Resistance to rust (Puccinia psidii Winter) in eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers." Theor Appl Genet **108**(1): 175-80.

Keats, B. J. B., S. L. Sherman, et al. (1991). "Guidelines for human linkage maps An International System for Human Linkage Maps (ISLM, 1990)." Annals of Human Genetics **55**(1): 1-6.

Keil, M. and A. R. Griffin (1994). "Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*." TAG Theoretical and Applied Genetics **89**(4): 442-450.

Kilian, A. (2009). "Diversity Arrays Technology Pty Ltd (DArT). Applying the open source philosophy in agriculture. In: Van Overwalle G, editor. Gene patents and collaborative licensing models: patent pools, clearinghouses, open source models and liability regimes." Cambridge University Press: 204 - 213.

Kirst, M., C. M. Cordeiro, et al. (2005). "Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations." J Hered **96**(2): 161-6.

Kirst, M., A. A. Myburg, et al. (2004). "Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus." Plant Physiol **135**(4): 2368-78.

Külheim, C., S. H. Yeoh, et al. (2009). "Comparative SNP diversity among four Eucalyptus species for genes from secondary metabolite biosynthetic pathways." BMC genomics **10**(452).

Kullan, A. R., M. van Dyk, et al. (2012). "Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*." BMC Genetics **13**(1): 60.

Kullan, A. R. K., M. M. van Dyk, et al. (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 x E. urophylla." Tree Genetics & Genomes **8**(1): 163-175.

Lander, E. S. and D. Botstein (1989). "Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps." Genetics **121**(1): 185-199.

Li, H. and R. Durbin (2010). "Fast and accurate long-read alignment with Burrows-Wheeler transform." Bioinformatics **26**(5): 589-595.

Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, W. and A. Godzik (2006). "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences." Bioinformatics **22**(13): 1658-1659.

Lillemo, M., B. Asalf, et al. (2008). "The adult plant rust resistance loci *Lr34/Yr18* and *Lr46/Yr29* are important determinants of partial resistance to powdery mildew in bread wheat line Saar." TAG Theoretical and Applied Genetics **116**(8): 1155-1166.

Lima, B., O. Silva-Junior, et al. (2011). "Assessment of SNPs for linkage mapping in *Eucalyptus*: construction of a consensus SNP/microsatellite map from two unrelated pedigrees." BMC Proceedings **5**(Suppl 7): P31.

Mace, E. S., J. F. Rami, et al. (2009). "A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers." Bmc Plant Biology **9**: -.

Mackay, T. F. C. (2001). "THE GENETIC ARCHITECTURE OF QUANTITATIVE TRAITS." Annual Review of Genetics **35**(1): 303-339.

Mamani, E. M. C., N. W. Bueno, et al. (2010). "Positioning of the major locus for *Puccinia psidii* rust resistance (*Ppr1*) on the *Eucalyptus* reference map and its validation across unrelated pedigrees." Tree Genetics & Genomes **6**(6): 953-962.

Mantovani, P., M. Maccaferri, et al. (2008). "An integrated DArT-SSR linkage map of durum wheat." Molecular Breeding **22**(4): 629-648.

Marcucci Poltri, S. N., N. Zelener, et al. (2003). "Selection of a seed orchard of *Eucalyptus dunnii* based on genetic diversity criteria calculated using molecular markers." Tree Physiol **23**(9): 625-32.

Marques, C. M., J. A. Araújo, et al. (1998). "AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*." Theoretical and Applied Genetics **96**: 727-737.

Marques, C. M., V. J. Carocha, et al. (2005). "Verification of QTL linked markers for propagation traits in *Eucalyptus*." Tree Genetics & Genomes **1**(3): 103-108.

Marques, C. M., J. Vasques-Kool, et al. (1999). "Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*." Theoretical and Applied Genetics **99**: 936-946.

Marques, M., V. Brondani, et al. (2002). "Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species." Theor Appl Genet **105**(2-3): 474-478.

Matheson, A. C. (1990). Breeding strategies for MTPs. Tree Improvement of Multipurpose Species, Multipurpose Tree Species Network. **2:** 67-99.

McCartney, C., R. Stonehouse, et al. (2011). "Mapping of the oat crown rust resistance gene Pc91." TAG Theoretical and Applied Genetics **122**(2): 317-325.

McKinnon, G. E., R. E. Vaillancourt, et al. (2008). "An AFLP marker approach to lower-level systematics in *Eucalyptus* (Myrtaceae)." American Journal of Botany **95**(3): 368-380.

Meuwissen, T. H., B. J. Hayes, et al. (2001). "Prediction of total genetic value using genome-wide dense marker maps." Genetics **157**(4): 1819-29.

Milczarski, P., H. Bolibok-Bragoszewska, et al. (2011). "A high density consensus map of rye (*Secale cereale* l.) based on DArT Markers." PLoS One **6**(12): e28495.

Missiaggia, A. A. (2005). Mapeamento genético de QTL para qualidade da madeira e florecimento precoce e estudos de expressão gênica alelo específica em *Eucalyptus* spp. Piracicaba, Universidade de São Paulo. **Doutorado:** 236.

Mohan, M., S. Nair, et al. (1997). "Genome mapping, molecular markers and marker-assisted selection in crop plants." Molecular Breeding **3**(2): 87-103.

Mora, A. L. and C. H. Garcia (2000). "A cultura do eucalipto no Brasil." São Paulo - SBS: 1.

Moran, G. F. and J. C. Bell (1983). Eucalyptus. Isozymes in plant genetics and breeding. S. D. Tanksley and T. J. Orton. Amsterdam, Elsevier**:** 423-441.

Moran, G. F., K. A. Thamarus, et al. (2002). "Genomics of *Eucalyptus* wood traits." Annals of Forest Science **59**(5-6): 645-650.

Morgante, M., E. De Paoli, et al. (2007). "Transposable elements and the plant pan-genomes." Curr Opin Plant Biol **10**(2): 149-55.

Myburg, A. A., A. R. Griffin, et al. (2003). "Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach." Theor Appl Genet **107**(6): 1028-42.

Myburg, A. A., C. Vogl, et al. (2004). "Genetics of Postzygotic Isolation in *Eucalyptus*: Whole-Genome Analysis of Barriers to Introgression in a Wide Interspecific Cross of *Eucalyptus grandis* and *E. globulus*." Genetics **166**(3): 1405-1418.

Neale, D. B. and A. Kremer (2011). "Forest tree genomics: growing resources and applications." Nature Reviews Genetics **12**(2): 111-122.

Nesbitt, K. A., B.M. Potts, et al. (1995). "Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae)." Heredity **74**: 628-637.

Neves, L., E. M. C. Mamani, et al. (2011). "A high-density transcript linkage map with 1,845 expressed genes positioned by microarray-based Single Feature Polymorphisms (SFP) in *Eucalyptus*." BMC Genomics **12**(1): 189.

Novaes, E., D. R. Drost, et al. (2008). "High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome." BMC Genomics **9**: 312.

Oda, S., A. L. M. Menck, et al. (1989). "Problemas no melhoramento clássico de eucalipto em função da alta intensidade de seleção." **IPEF**, Piracicaba: v.41/42, p. 8-17.

Oliver, R., E. Jellen, et al. (2011). "New Diversity Arrays Technology (DArT) markers for tetraploid oat (Avena magna Murphy et Terrell) provide the first complete oat linkage map and markers linked to domestication genes from hexaploid *A. sativa* L." TAG Theoretical and Applied Genetics **123**(7): 1159-1171.

Paterson, A. H. (1996). **Genome Mapping in Plants**. San Diego, California, R.G. Landes Company; Austin - Texas: Academic Press, p.330.

Paula, R. C. d., I. E. Pires, et al. (2002). "Predição de ganhos genéticos em melhoramento florestal." Pesquisa Agropecuária Brasileira **37**(2): 159-165.

Paux, E., P. Sourdille, et al. (2008). "A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B." Science **322**(5898): 101-104.

Pelgas, B., J. Bousquet, et al. (2011). "QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments." BMC Genomics **12**(1): 145.

Pereira, M. G. and T. N. S. Pereira (2006). Marcadores Moleculares no Pré-Melhoramento de plantas. **Marcadores Moleculares**. A. Borém and E. T. Caixeta. Viçosa, MG**:** 85-106.

Petroli, C. D., C. P. Sansaloni, et al. (2012). "Genomic Characterization of DArT Markers Based on High-Density Linkage Analysis and Physical Mapping to the Eucalyptus Genome." Plos One **7**(9).

Pevzner, P. A., H. Tang, et al. (2001). "An Eulerian path approach to DNA fragment assembly." Proceedings of the National Academy of Sciences **98**(17): 9748-9753.

Poke, F. S., R. E. Vaillancourt, et al. (2005). "Genomic research in *Eucalyptus*." Genetica **125**(1): 79-101.

Poland, J. A., P. J. Brown, et al. (2012). "Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach." PLoS One **7**(2): e32253.

Pot, D., J.-C. Rodrigues, et al. (2006). "QTLs and candidate genes for wood properties in maritime pine (Pinus pinaster Ait.)." Tree Genetics & Genomes **2**(1): 10-24.

Potts, B. M. (2004). Genetic improvement of eucalypts. Encyclopedia of forest sciences. Oxford UK, Elsevier**:** 1480-1490.

Pozniak, C., R. Knox, et al. (2007). "Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat." TAG Theoretical and Applied Genetics **114**(3): 525-537.

Price, A. H. (2006). "Believe it or not, QTLs are accurate!" Trends Plant Sci. **11**: 213-216.

Rafalski, A. and M. Morgante (2004). "Corn and humans: recombination and linkage disequilibrium in two genomes of similar size." Trends in Genetics **20**(2): 103-111.

Raymond, C. A. (2002). "Genetics of *Eucalyptus* wood properties." Annals of Forest Science **59**(5–6): 525–531.

Raymond, C. A. and L. A. Apiolaza (2004). "Incorporating wood quality and deployment traits in *Eucalyptus globulus* and *Eucalyptus nitens*." In Plantation forest biotechnology for the 21st Century. Edited by Walter C, Carson M. Rotorua, New Zealand: Forest Research New Zealand.: 87-99.

Reddy, U. K., J.-k. Rong, et al. (2011). "Use of diversity arrays technology markers for integration into a cotton reference map and anchoring to a recombinant inbred line map." Genome **54**(5): 349-359.

Remington, D. L., R. W. Whetten, et al. (1999). "Construction of an AFLP genetic map with nearly complete genome coverage in Pinus taeda." Theor Appl Genet **98**(8): 1279-92.

Resende, M. D. V. (2001). Melhoramento de espécies perenes. Recursos genéticos e melhoramento de plantas. L. L. NASS. Rondonópolis, MT, Fundação MT**:** 357-421.

Resende, M. D. V., M. F. R. Resende, et al. (2012). "Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees." New Phytologist **194**(1): 116-128.

Resende, M. F. R., P. Munoz, et al. (2012). "Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments." New Phytologist **193**(3): 617-624.

Rocha, R. B., J. I. M. Abad, et al. (2002). "Fingerprint and genetic diversity analysis of *Eucalyptus* ssp. genotypes using RAPD and SSR markers." Scientia Florestais(62): p.24-31.

Rocha, R. B., E. G. Barros, et al. (2007). "Mapping of QTLs related with wood quality and developmental characteristics in hybrids (Eucalyptus grandis x Eucalyptus urophylla)." Revista Árvore **31**: 13-24.

Rodríguez-Suárez, C., M. Giménez, et al. (2012). "Development of wild barley (Hordeum chilense) - derived DArT markers and their use into genetic and physical mapping." TAG Theoretical and Applied Genetics **124**(4): 713-722.

Sadeque, A. and M. A. Turner (2010). "QTL Analysis of Plant Height in Hexaploid Wheat Doubled Haploid Population." Thai Journal of Agricultural Science **43**(2): 91-96.

Sale, M. M., B. M. Potts, et al. (1993). "Relationships within *Eucalyptus* using chloroplast DNA. ." Aust. Syst. Bot. **6**: p.127-138.

Sansaloni, C., C. Petroli, et al. (2011). "Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*." BMC Proceedings C7 - P54 **5**(7): 1-2.

Sansaloni, C. P., C. D. Petroli, et al. (2010). "A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*." Plant Methods **6**: 16.

Sax, K. (1923). "THE ASSOCIATION OF SIZE DIFFERENCES WITH SEED-COAT PATTERN AND PIGMENTATION IN PHASEOLUS VULGARIS." Genetics **8**(6): 552-560.

Schouten, H., W. E. Weg, et al. (2012). "Diversity arrays technology (DArT) markers in apple for genetic linkage maps." Molecular Breeding **29**(3): 645-660.

Shepherd, M., J. X. Chaparro, et al. (1999). "Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid." TAG Theoretical and Applied Genetics **99**(7 - 8): 1207-1215.

Stackpole, D. J., R. E. Vaillancourt, et al. (2011). "Genetic Variation in the Chemical Components of Eucalyptus globulus Wood." G3: Genes, Genomes, Genetics **1**(2): 151-159.

Staub, J. E., F. C. Serquen, et al. (1996). "Genetic markers, map construction, and their application in plant breeding." Hort. Sci.(31): p.729-740.

Steane, D., N. Conod, et al. (2006). "A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits." Tree Genetics & Genomes **2**: 30 - 38.

Steane, D. A., D. Nicolle, et al. (2011). "Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping." Mol Phylogenet Evol **59**(1): 206-224.

Steane, D. A., D. Nicolle, et al. (2002). "Higher-level relationships among the eucalypts are resolved by ITS-sequence data." Australian Systematic Botany **15**: 49 - 62.

Steane, D. A., A. K. West, et al. (1992). "Restriction fragment length polymorphisms in chloroplast DNA from six species of *Eucalyptus*." Aust. J. Bot. **39**: p.399-414.

Strauss, S. H., R. Lande, et al. (1992). "Limitations of Molecular-Marker-Aided Selection in Forest Tree Breeding." Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere **22**(7): 1050-1061.

Sugimoto, K., Y. Takeuchi, et al. (2010). "Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice." Proceedings of the National Academy of Sciences of the United States of America **107**(13): 5792-5797.

Sun, L., A. Dimitromanolakis, et al. (2011). "BR-squared: a practical solution to the winner's curse in genome-wide scans." Human Genetics **129**(5): 545-552.

Supriya, A., S. Senthilvel, et al. (2011). "Development of a molecular linkage map of pearl millet integrating DArT and SSR markers." Theoretical and Applied Genetics **123**(2): 239-250.

Tanksley, S. D. (1993). "Mapping polygenes." Annu Rev Genet **27**: 205-33.

Tanksley, S. D., N. D. Young, et al. (1989). "RFLP Mapping in Plant Breeding: New Tools for an Old Science." Nat Biotech **7**(3): 257-264.

Thamarus, K., K. Groom, et al. (2004). "Identification of quantitative trait loci for wood and fibre properties in two full-sib properties of Eucalyptus globulus." Theor Appl Genet **109**(4): 856-864.

Thamarus, K., K. Groom, et al. (2002). "A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits." Theor Appl Genet **104**: 379-387.

Thoday, J. M. (1961). "Location of Polygenes." Nature **191**(4786): 368-370.

Thudi, M., A. Bohra, et al. (2011). "Novel SSR Markers from BAC-End Sequences, DArT Arrays and a Comprehensive Genetic Map with 1,291 Marker Loci for Chickpea (*Cicer arietinum* L.)." PLoS One **6**(11): e27275.

Thumma, B., B. Baltunis, et al. (2010). "Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families." Tree Genetics & Genomes **6**(6): 877-889.

Thumma, B. R., B. A. Matheson, et al. (2009). "Identification of a cis-acting regulatory polymorphism in a eucalypt COBRA-like gene affecting cellulose content." Genetics **183**(3): 1153-1164.

Thumma, B. R., M. R. Nolan, et al. (2005). "Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in Eucalyptus spp." Genetics **171**(3): 1257-1265.

Tinker, N. A., A. Kilian, et al. (2009). "New DArT markers for oat provide enhanced map coverage and global germplasm characterization." BMC Genomics **10**: 39.

Tzfira, T., A. Zuker, et al. (1998). "Forest-tree biotechnology: genetic transformation and its application to future forests." Trends in biotechnology **16**(10): 439-446.

Van Ooijen, J. W. and R. E. Voorrips (2001). JoinMap 3.0, Software for the calculation of genetic linkage maps. P. R. International. Wageningen, The Netherlands.

Van Schalkwyk, A., P. Wenzl, et al. (2012). "Bin mapping of tomato diversity array (DArT) markers to genomic regions of *Solanum lycopersicum* × *Solanum pennellii* introgression lines." TAG Theoretical and Applied Genetics **124**(5): 947-956.

Vencovsky, R. and M. A. P. Ramalho (2000). Contribuição do Melhoramento Genético de Plantas no Brasil. **Agricultura Brasileira e Pesquisa Agropecuária**. E. Paterniani. Brasília**:** EMBRAPA Comunicação para Transferência de Tecnologia, p.57-89.

Verhaegen, D. and C. Plomion (1996). "Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers." Genome **39**: 1051-1061.

Verhaegen, D., C. Plomion, et al. (1997). "Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers, I. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages." Theoretical and Applied Genetics **95**: 597-608.

Vision, T. J., D. G. Brown, et al. (2000). "Selective Mapping: A strategy for optimizing the construction of high-density linkage maps." Genetics **155**(1): 407-420.

Vos, P., R. Hogers, et al. (1995). "AFLP: a new technique for DNA fingerprinting." Nucleic Acids Research **23**(21): 4407-4414.

Wang, S., C. J. Basten, et al. (2007). Windows QTL Cartographer 2.5. Raleigh - NC, North Carolina State University.

Wegrzyn, J. L., A. J. Eckert, et al. (2010). "Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem." New Phytologist **188**(2): 515-532.

Wenzl, P., J. Carling, et al. (2004). "Diversity Arrays Technology (DArT) for whole-genome profiling of barley." Proc Natl Acad Sci U S A **101**(26): 9915-20.

Wenzl, P., H. Li, et al. (2006). "A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits." BMC Genomics **7**: 206.

White, J., J. R. Law, et al. (2008). "The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome." Theoretical and Applied Genetics **116**(3): 439-453.

Whittock, S., Steane, DA, Vaillancourt, RE, Potts, BM (2003). "Is *Eucalyptus* subgenus *Minutifructa* over-ranked?" Transactions of the Royal Society of South Australia **127**: 27-32.

Williams, J. G. K., A. R. Kubelik, et al. (1990). "DNA polymorphisms amplified by arbitrary primers are useful as genetic markers." Nucleic Acids Research **18**(22): 6531-6535.

Wittenberg, A., T. Lee, et al. (2005). "Validation of the high-throughput marker technology DArT using the model plant Arabidopsis thaliana." Molecular Genetics and Genomics **274**: 30 - 39.

Xia, L., K. Peng, et al. (2005). "DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives." TAG Theoretical and Applied Genetics **110**: 1092 - 1098.

Xu, S. Z. (2003). "Theoretical basis of the Beavis effect." Genetics **165**(4): 2259-2268.

Yang, S. Y., R. K. Saxena, et al. (2011). "The first genetic map of pigeon pea based on diversity arrays technology (DArT) markers." J Genet **90**(1): 103-9.

Zeng, Z. B. (1993). "Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci." Proc. Natl. Acad. Sci. USA. **90**(23): 10972-6.

Zeng, Z. B. (1994). "Precision mapping of quantitative trait loci." Genetics **136**(4): 1457-68.

Zhang, C., C. Dong, et al. (2011). "Inheritance and QTL analysis of dough rheological parameters in wheat." Frontiers of Agriculture in China **5**(1): 15-21.

Zhang, L., D. Liu, et al. (2011). "Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers." BMC Genetics **12**(1): 42.

OPEN ACCESS Freely available online

PLOS ONE

Genomic Characterization of DArT Markers Based on High-Density Linkage Analysis and Physical Mapping to the *Eucalyptus* Genome

César D. Petroli[1,2], Carolina P. Sansaloni[1,2], Jason Carling[3], Dorothy A. Steane[4,5], René E. Vaillancourt[4], Alexander A. Myburg[6], Orzenil Bonfim da Silva Jr.[1], Georgios Joannis Pappas Jr.[1], Andrzej Kilian[3], Dario Grattapaglia[1,2,7]*

1 Plant Genetics Laboratory, EMBRAPA Genetic Resources and Biotechnology, Brasilia, Brazil, **2** Department of Cell Biology, Universidade de Brasilia, Brasília – DF, Brazil, **3** Diversity Arrays Technology Pty Ltd., Yarralumla, Australia, **4** School of Plant Science and CRC for Forestry, University of Tasmania, Hobart, Tasmania, Australia, **5** Faculty of Science, Health, Education and Engineering - ML12, University of the Sunshine Coast, Maroochydore DC, Queensland, Australia, **6** Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa, **7** Genomic Sciences Program - Universidade Católica de Brasília, Brasília – DF, Brazil

Abstract

Diversity Arrays Technology (DArT) provides a robust, high throughput, cost-effective method to query thousands of sequence polymorphisms in a single assay. Despite the extensive use of this genotyping platform for numerous plant species, little is known regarding the sequence attributes and genome-wide distribution of DArT markers. We investigated the genomic properties of the 7,680 DArT marker probes of a *Eucalyptus* array, by sequencing them, constructing a high density linkage map and carrying out detailed physical mapping analyses to the *Eucalyptus grandis* reference genome. A consensus linkage map with 2,274 DArT markers anchored to 210 microsatellites and a framework map, with improved support for ordering, displayed extensive collinearity with the genome sequence. Only 1.4 Mbp of the 75 Mbp of still unplaced scaffold sequence was captured by 45 linkage mapped but physically unaligned markers to the 11 main *Eucalyptus* pseudochromosomes, providing compelling evidence for the quality and completeness of the current *Eucalyptus* genome assembly. A highly significant correspondence was found between the locations of DArT markers and predicted gene models, while most of the 89 DArT probes unaligned to the genome correspond to sequences likely absent in *E. grandis*, consistent with the pan-genomic feature of this multi-*Eucalyptus* species DArT array. These comprehensive linkage-to-physical mapping analyses provide novel data regarding the genomic attributes of DArT markers in plant genomes in general and for *Eucalyptus* in particular. DArT markers preferentially target the gene space and display a largely homogeneous distribution across the genome, thereby providing superb coverage for mapping and genome-wide applications in breeding and diversity studies. Data reported on these ubiquitous properties of DArT markers will be particularly valuable to researchers working on less-studied crop species who already count on DArT genotyping arrays but for which no reference genome is yet available to allow such detailed characterization.

Citation: Petroli CD, Sansaloni CP, Carling J, Steane DA, Vaillancourt RE, et al. (2012) Genomic Characterization of DArT Markers Based on High-Density Linkage Analysis and Physical Mapping to the *Eucalyptus* Genome. PLoS ONE 7(9): e44684. doi:10.1371/journal.pone.0044684

Editor: Tongming Yin, Nanjing Forestry University, China

Received June 1, 2012; **Accepted** August 6, 2012; **Published** September 11, 2012

Funding: The project was supported by the Brazilian Ministry of Science and Technology (CNPq Project Grants 559245/2008-4 and 560831/2008-0) and PRONEX FAP-DF Project Grant "NEXTREE" 193.000.570/2009 to DG. This study was carried out as part of Cesar Petroli's doctoral thesis at the University of Brasilia. CPS and CDP had doctoral fellowships from CAPES (Brazilian Ministry of Education) and DG a productivity research fellowship from CNPq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors acknowledge the help and technical support from many employees of Diversity Arrays Technology Pty. Ltd. Two of the authors on this manuscript, namely Andrzej Kilian and Jason Carling, are employees of Diversity Arrays Technology Pty Ltd, which offers genome profiling service using the microarray whose genomic characterization is described in this report. This fact, however, has not interfered whatsoever with the full, objective, transparent and unbiased presentation of the research results described in the manuscript nor alters the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: dario.grattapaglia@embrapa.br

Introduction

DNA marker technologies for high throughput genome-wide genotyping at affordable costs have become indispensable in the plant geneticist's toolbox. A large array of methods to detect DNA sequence polymorphisms among individual plants have been developed and used widely in the last twenty five years. Although DNA based hybridization inaugurated this journey with RFLP markers [1], PCR-based methods [2,3] were responsible for removing the barrier to entry in plant genomic analysis for a large number of species, including orphan crops and many forest trees. Most PCR-based molecular marker methods, however, are low throughput and mobility-based, and therefore too time consuming and costly for applications that require genotyping thousands of samples for thousands of markers within modest budgets. Although large SNP arrays have been developed for an increasing number of plant species [4], they still remain largely limited to the major crops and their costs per sample are unaffordable for most plant breeding and germplasm conservation programs.

Diversity Arrays Technology (DArT) was described over a decade ago [5] and has experienced increasing interest in recent

years as a robust, high throughput, cost-effective genome-wide method to assay thousands of presence/absence polymorphisms in a single assay. Although proprietary, this technique is licensed freely under an open-source model [6], a condition that has stimulated the development of genotyping arrays for more than 60 organisms including many less privileged crops [7,8,9,10,11,12,13,14,15,16,17,18,19,20,21]. DArT involves the isolation and cloning of a random set of DNA fragments from a complexity-reduced DNA sample assembled by pooling several germplasm accessions so that a representative collection of variable genomic sequences of one or more target species is captured. Several thousand of these DNA clones are arrayed on a glass slide and interrogated with a similarly complexity-reduced, PCR-amplified genomic sample. Being a DNA-DNA hybridization-based method using relatively long probes (~300–500 bp), DArT provides high and consistent signal to noise ratio even across related taxa [22].

In spite of the extensive use of this genotyping platform for many plant species, very little is known regarding the genomic attributes of the DArT array probes that generate the several thousand markers genotyped. With the exception of a study in oats [23], and recent small scale surveys of a few hundred DArT probe sequences in tomato [16] and apple [24], to the best of our knowledge complete DArT arrays have not yet been examined at the sequence level for redundancy, genome coverage and gene content. Additionally, no information is available about the distribution of DArT markers across a genome, mainly because no reference assembly has yet been available for most species where this technology has been used.

A high density DArT genotyping microarray with 7,680 selected probes from a wide representation of 64 *Eucalyptus* species was recently developed [25]. The genus *Eucalyptus* includes over 700 species some of which are the most widely planted hardwood trees worldwide [26]. A particularly outstanding feature of this hybridization-based genotyping tool has been its genus-wide transferability across species, an attribute hardly offered by microsatellites or SNPs [27]. DArT has provided a standardized high-throughput genotyping platform, whereby thousands of markers can be readily assayed in parallel for thousands of samples across *Eucalyptus* species. This DArT array has demonstrated excellent performance for complex phylogenetic and diversity analyses [22], genomic selection [28] and linkage mapping [25,29,30]. A detailed understanding regarding the sequence content and genome-wide distribution of the DNA probes that compose this DArT array should greatly expand its value for comparative QTL mapping studies, to navigate from linkage maps to the reference sequence in positional cloning projects and to extract additional genomic information from uniquely informative markers identified in phylogenetic, population genetics and Genomic Selection studies.

Genetic linkage maps have been pivotal tools for examining the inheritance of qualitative and quantitative traits, for comparative mapping, whole genome assembly and for molecular breeding applications, including germplasm analyses, marker-assisted selection and map-based cloning [31]. Linkage maps for species of *Eucalyptus* have been reported for several pedigrees, both intra- and inter-specific, using different molecular marker technologies [32,33]. Extensive linkage mapping data of anonymous markers has been accumulated with dominant RAPD and AFLP technologies [34,35,36,37,38,39], while RFLPs [40,41] and a recent Single Feature Polymorphisms (SFP) genotyping array [42] have allowed positioning hundreds of genes on existing maps. In spite of all these advances, these marker technologies have not provided a widely applicable tool that can be used to link genotypes to phenotypes in a broader and more sustainable way that includes comparative mapping, gene discovery and genome assisted breeding. Recently, DArT markers have provided the coverage and high-density mapping required to move in that direction [29,30], although they are still lacking a deeper characterization of their genomic content.

In this study we investigated the genomic properties of the 7,680 DArT marker probes that populate the *Eucalyptus* array by sequencing them, constructing a high density linkage map and carrying out detailed physical mapping analyses using the recently released *Eucalyptus grandis* reference genome sequence (www.phytozome.net). We were specifically interested in: (1) verifying DArT marker performance for linkage mapping, i.e. level of polymorphism, locus ordering and genome coverage; (2) characterizing the sequence composition of the DArT array probes regarding sequence redundancy and gene content; (3) assessing the physical distribution of the DArT marker probes in terms of overall genome coverage and distance from predicted gene models; (4) aligning the linkage map to the corresponding pseudochromosome scaffolds to assess the consistency of physical-*versus* recombination-based locus ordering; and (5) providing pseudochromosome level and genome wide estimates of the relationship between physical and recombination distances.

Materials and Methods

Plant Material

A mapping population of 177 F_1 individuals was derived from an inter-specific cross between two highly heterozygous elite trees, *E. grandis* (clone G38) and *E. urophylla* (clone U15). Both species are widely planted in the tropics and belong to the same subgenus, *Symphyomyrtus*. This mapping pedigree, named GxU-IP was selected as a reference pedigree for mapping purposes in the Genolyptus project [43], immortalized by mini-cutting propagation and planted in a replicated trial in five locations in July 2003 in randomized blocks with single tree plot with five replicates per location. Genomic DNA was extracted from both parents and all F_1 individuals using 150 mg of leaf tissue stored at –20°C as described previously [34]; the resulting DNA samples were of consistent quality and suitable for DArT and microsatellite genotyping.

Microsatellite Genotyping

Screening of 300 EMBRA microsatellite markers [44,45,46] for polymorphism between the two parents with the additional analysis of six F1 progeny individuals to verify segregation, resulted in the selection of 222 informative microsatellites**.** Microsatellite genotyping was carried out in multiplexed systems with multi-fluorescence detection in an ABI 3100XL as described earlier [45,47].

DArT Genotyping

A detailed account of the methods used to prepare the high density *Eucalyptus* DArT array was reported earlier [25]. Briefly, 18 reduced representation *Pst*I/*Taq*I genomic libraries involving a total of 64 different *Eucalyptus* species were built and 23,808 DNA probes were screened in a panel of 96 individuals. A set of 7,680 probes that revealed robust polymorphisms was selected and used to construct the operational DArT genotyping array. This procedure optimized (1) sampling of a large collection of sequence variants to increase recovery of polymorphic clones; and (2) inter-specific transferability of the scored markers. Genomic representations of the two parents and 177 F_1 individuals of the mapping population were generated with the same complexity reduction

method used to prepare the library to generate 'targets' for hybridizing to the arrays. After hybridization, microarray slides were washed and scanned using a TECAN LS300 confocal laser microarray scanner at a resolution of 20 μm per pixel with sequential acquisition of 3 images for each microarray slide. The signal from the FAM-labeled vector polylinker provided a reference value for quantity of amplified DNA fragment present in each 'spot' of the microarray. The resulting images were analyzed using *DArTSoft* version 7.44, a program created by *Diversity Arrays Technology Pty. Ltd.* for microarray image data extraction, polymorphism detection, and marker scoring. A relative hybridization intensity value was then calculated for all accepted spots as log [Cy-3 signal/FAM signal] for the targets labelled with Cy-3, and log [Cy-5 signal/FAM signal] for targets labelled with Cy-5. *DArTSoft* then compared the relative intensity values obtained for each clone across all slides/targets to detect the presence of clusters of higher and lower values corresponding to marker scores of '1' and '0' respectively. Targets with relative intensity values that could not be assigned to either of the clusters were recorded as missing data. Standard methods of marker discovery were deployed using a combination of parameters automatically extracted from the array data using DArTsoft. The following parameters were used: (1) reproducibility ≥95% as measured by the concordance of the genotype call between technical replicates (replicated targets processed for a minimum of 30% of the DNA samples genotyped); (2) marker quality Q ≥65, which measures between-cluster variance as a percentage of total variance in fluorescent signal distribution among tested samples; and (3) marker call rate ≥75% (percentage of targets able to be scored as '0' or '1').

Genetic Map Construction

A single integrated genetic linkage map was constructed using both the co-dominant microsatellite data and the dominant DArT marker data using JoinMap v3.0 [48]. Microsatellite markers segregated either from each single parent in a 1:1 ratio, from both parents in a 1:2:1 ratio following a phase-unknown F2 configuration with both parents equally heterozygous for the same genotype, or in a fully informative 1:1:1:1 ratio with three or four different alleles segregating from the two parents. Dominant DArT markers, on the other hand, segregated either in a 1:1 pseudo-testcross configuration from each single parent or in a 3:1 ratio when both parents where heterozygous. For both the microsatellite and DArT data, markers that showed ≥75% call rate and fitted one of the expected segregation ratios at $\alpha \geq 0.01$ were used for linkage analysis. The grouping and ordering of the markers were established initially by applying the maximum likelihood algorithm of JoinMap with population type CP; grouping at LOD>15; recombination fraction ≤0.4; ripple value = 1; jump in goodness-of-fit threshold (the normalized difference in goodness-of-fit chi-square before and after adding a locus) equal to 5 under a Kosambi mapping function. Marker ordering with JoinMap was carried out by simulated annealing, excluding markers that contributed to unstable marker orders in the first two ordering rounds to yield a higher likelihood support framework map. Additional segregating markers were then fitted to the linkage maps at lower stringency by the third and final round of JoinMap to provide map position for a larger number of segregating DArT markers.

Comparative Analysis between the Linkage Map and the Assembled Genome Sequence

A genome-wide assessment of the consistency of the marker order estimated in the linkage map derived from this particular pedigree with the physical position of the markers in the currently assembled genome sequence was carried out by aligning the higher confidence framework linkage map to the 11 main scaffolds of the current assembly of the *E. grandis* genome sequence (version 1.0 available in Phytozome 6.0) produced for the one-generation selfed tree 'BRASUZ1' (Brazil Suzano S_1). This alignment was also used to provide pseudochromosome-specific and genome-wide estimates of the correspondence between physical distance and recombination fraction in the *Eucalyptus* genome, as well as an estimate of the effective genome coverage provided by the framework map.

Genomic Characterization of DArT Marker Probes

E. coli clones containing the 7,680 *Eucalyptus* DArT probes [25] were re-arrayed in twenty 384-deep-well-plates and submitted for bi-directional Sanger sequencing to the genomics facility of Purdue University (www.genomics.purdue.edu). Following quality trimming and clipping of vector regions and *Pst*I sites, sequences obtained were deposited in GenBank (accession numbers HR865291-HR872186). Redundancy of DArT probes at the sequence level was investigated using Geneious Pro 5.1.7 [49] using a minimum sequence overlap of 50 bp for a sequence to be assembled into a contig and an overlap identity of 98% (the "overlap identity" is the minimum percentage of bases that must be identical in the region of overlap in order for a sequence to be assembled). The numbers of unique and redundant DArT probes were then assessed by applying four different sets of sequence assembly parameters, from a most stringent assembly (A1) to the most liberal one (A4). These parameters were: (a) word length, i.e. the minimum number of consecutive bases that must match perfectly in order to find a match between two sequences; (b) maximum number or single base mismatches allowed per reads as a percentage of the size of the overlap between two reads; (c) maximum number of base ambiguities allowed in word matches; (d) maximum number of gaps that may be inserted into each read as a percentage of the size of the overlap between two reads; (e) maximum size of each gap that may be inserted into reads.

After preprocessing to remove contaminants, sequencing artifacts and low quality sequences, all DArT probes for which sequences were obtained were mapped to the assembled *Eucalyptus grandis* reference genome (version 1.0 available in Phytozome 6.0). Mapping was carried out using the BWA-SW (version 0.5.8) component from the Burrows-Wheeler Alignment tool [50] to produce a BAM [51] file. As the BWA tool can detect chimerical reads reporting two or more hits, parameters were set up such that non-optimal mapping was avoided. The threshold for a probe sequence hit to be retained was set to a fixed value (T = 70), corresponding to twice the median value of the numeric distribution obtained from the formula $5.5 \times \log(L)$. This formula was applied to each one of the DArT probe sequences with length L, accounting for the fact that this formula is used by BWA as a coefficient for threshold adjustments. As a consequence, BWA did not search for suboptimal hits with a score lower than the alignment score minus T. The BWA options for the alignment score calculation including the score of a match (a), mismatch penalty (b), gap open penalty (q), and gap extension penalty (r) were left at their default settings (a = 1; b = 3; q = 5; r = 2). As an additional evaluation of the quality of the mapping procedure, sequence alignment information was extracted from the BAM file using an in-house Perl script designed to report all queried sequence hits and sub-optimal alignment scores. We considered up to two hits to be a "successful mapping" and used the sub-optimal scores for each one of the hits to classify the "mapping reliability" and "expected mapping error rate" of the procedure. Finally, to

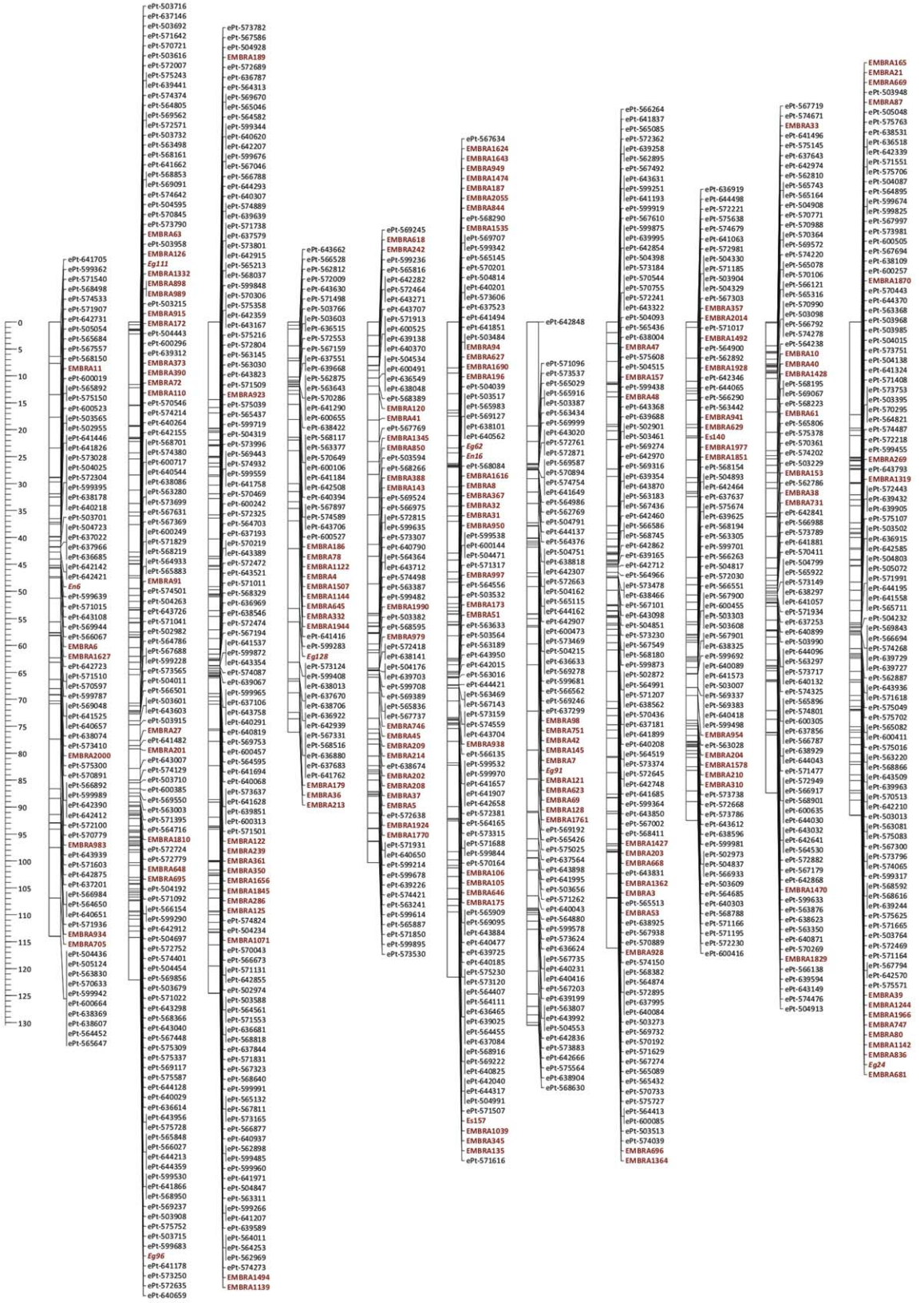

LG 1
LG 2
LG 3
LG 4
LG 5
LG 6
LG 7
LG 8
LG 9
LG 10
LG 11

Figure 1. Framework DArT/microsatellite linkage map for *Eucalyptus*. The map includes 1,029 markers positioned with high confidence for locus order, involving 861 DArT (in black) and 168 microsatellites (in red) with a centiMorgan scale on the left.
doi:10.1371/journal.pone.0044684.g001

inspect the genomic features of the DArT markers relative to predicted gene models in version 1.0 of the *Eucalyptus grandis* genome, the 11 scaffolds were partitioned into 5 Mbp bins. A Spearman rank correlation between the number of DArT markers and number of gene models annotated in each bin was estimated for each pseudochromosome scaffold. Additionally the physical distance in base pairs from each sequenced DArT probe to the closest gene model was estimated to provide a genome-wide picture of the gene-space coverage of the *Eucalyptus* genome provided by the DArT array.

Results

DArT Marker Genotyping

The distribution of markers across the different levels of reproducibility was skewed towards the highest quality classes. For example, out of the 3,933 markers that had reproducibility ≥95%, 70% of them had reproducibility equal to or greater than 99%. For the 4,884 markers that passed the threshold quality score, 61% had Q ≥70, while out of the 5,415 markers with a call rate ≥75%, 36% had a call rate ≥90% (Figure S1). While reproducibility and Q score are measures that directly appraise the quality of the genotyping, the call rate essentially reflects the percent missing data tolerated. A relatively less stringent marker call rate of ≥75% was adopted to maximize the number of markers positioned on the linkage map, since such a threshold would still yield good quality marker data for ~128 informative recombinant gametes that allow satisfactory marker linkage and ordering analyses during map construction. The 7,680 probe *Eucalyptus* microarray yielded a total of 3,191 markers that simultaneously passed all marker quality and call rate filtering parameters (Figure S1).

Out of 3,191 markers tested for Mendelian behavior only 215 did not fit either a 1:1 or a 3:1 ratio and were excluded from further analyses. Out of the 2,976 DArT markers that showed Mendelian behavior and were used in the linkage analysis, 1,777 segregated in a 1:1 pseudo-testcross configuration and 1,199 in a 3:1 fashion, i.e. were heterozygous loci segregating simultaneously from the two parents. Regarding the microsatellite dataset, 166 loci were fully informative with three or four alleles segregating in four distinct genotypic classes providing valuable anchor loci for the construction of an integrated linkage map. Forty-two microsatellites segregated from one of the parents only, 25 from *E. grandis* and 17 from *E. urophylla*, while 14 segregated in a 1:2:1 F2 phase-unknown configuration.

Linkage Mapping

A dataset with 3,198 markers (2,976 DArT and 222 microsatellites) was subjected to a mapping analysis. Grouping analysis at LOD>15.0 resulted in 2,980 markers grouped in 11 *bona fide* groups (numbered as established earlier [44]) and assessed by the presence of anchoring microsatellite markers. The linkage map built with higher likelihood support for marker order following the second round of JoinMap is presented as a "Framework map" (Figure 1). The linkage map obtained after the third round of ordering is hereafter called the "Full map" and is presented as a way to provide a preliminary position from which all the informative markers fed into the subsequent genomic characterization analysis (Table 1 and Figure S2). The Framework map, built following the second round of JoinMap, resulted in 1,029 markers positioned with higher confidence, 861 DArT and 168 microsatellites. When a more liberal marker ordering was allowed, a total of 2,484 markers were mapped (2,274 DArT markers and 210 microsatellites). The remaining 496 markers could not be mapped, even using a relaxed stringency, possibly as a result of redundancy of DArT markers at the sequence level (see below) or due to very close linkage so that not enough recombinants could be sampled to resolve relative ordering along the map.

A much larger proportion of segregating microsatellites (80%) could be fitted in the Framework map than DArT markers (45%), most likely due to the higher information content of the fully segregating multiallelic microsatellites that provide higher power to categorize recombinant *versus* parental haplotypes and thus determine order. However, the final size of the Framework map was only 9.5% smaller than the Full map (1,176.7 versus 1,303.9 cM) (Table 1 and Figure S2). Marker orders of the Framework map and the Full map were generally consistent, although some inverted sets of markers were observed, mainly on linkage groups 1, 2, 7 and 9. Furthermore, although the expectation was that all framework markers would be contained in the Full map this was not always the case. The Framework map contained 99 markers that were excluded when a relaxed ordering threshold was allowed upon Full map construction. They were concentrated on linkage groups 2 (42 markers), 3 (26 markers), 1 (19 markers) and 11 (8 markers) (Figure S2).

The Full map of DArT markers contained on average 226 markers per linkage group positioned at a sub-centiMorgan average inter-marker distance of 0.5 while the Framework map had on average 93.5 markers per linkage group and yielded a 1.1 cM average inter-marker distance (Table 1). The distribution of map distances between consecutive markers in the Framework map was significantly different from the one in the Full map ($p = 0.021$ in a non-parametric Komolgorov-Smirnov test) (Figure S3). This result demonstrates that (i) a Framework map spreads out well-supported markers to provide robust locus ordering and (ii) reduces the proportion of short inter-marker distances (<1 cM) relative to a Full map (87% in the Full map and 65% in the Framework map).

A tally of the origin of the 861 DArT markers mapped on the Framework map showed that 197 (23%) markers segregated 1:1 from *E. urophylla*, 298 (35%) from *E. grandis*, while 366 (42%) were heterozygous in both parents segregating 3:1. Very similar proportions were observed when all 2,274 DArT markers were examined. These results suggest a higher sequence heterozygosity in the *E. grandis* parent than in the *E. urophylla* parent. Out of the 7,680 marker probes in the array, 2,274 (i.e. approximately 30%), were ultimately mapped in this single segregating family. However, if the 3,191 DArT markers that passed the genotyping quality filters for this experiment were considered, 71% of the markers could be mapped. Although the proportion of markers that can be mapped depends largely on the sequence heterozygosity of the parents and their genetic divergence, this result corroborates the outstanding performance of the DArT array for linkage mapping purposes in *Eucalyptus*.

Recombination and Physical Distances in the *Eucalyptus* Genome

The alignment of the Framework linkage map to the *Eucalyptus* genome sequence indicates that the relative order of linkage mapped markers by and large agrees with their relative physical

Table 1. Mapping statistics of the DArT/microsatellite consensus maps of *Eucalyptus grandis x E. urophylla.*

Linkage Group/ Pseudochromosome	1	2	3	4	5	6	7	8	9	10	11	Total	Mean	St.dev.
Full map[a]														
Total # markers	207	244	270	106	189	271	224	275	220	263	215	2,484	225.8	49.34
# DArT markers	191	219	256	93	166	231	210	262	204	246	196	2,274	206.7	47.68
# Microsatellites	16	25	14	13	23	40	14	13	16	17	19	210	19.1	7.98
Total size (cM)	167.8	129	102.4	86.6	130.7	116.5	117.3	118.6	118.3	117.4	99.5	1,303.9	118.5	20.8
Average inter-marker distance	0.8	0.5	0.4	0.8	0.7	0.4	0.5	0.4	0.5	0.4	0.5	0.5	–	–
Framework Map[b]														
Total # markers	80	131	128	57	74	104	75	107	78	92	103	1,029	93.5	23.4
# DArT markers	72	112	115	44	54	72	64	95	64	82	87	864	78.3	22.6
# Microsatellites	8	19	13	13	20	32	11	12	14	10	16	168	15.3	6.6
Total size in cM	114.0	122.1	124.6	76.1	100.3	121.6	130.5	116.2	92.5	87.4	91.5	1,176.8	107	18.1
Average inter-marker distance	1.4	0.9	1.0	1.3	1.4	1.2	1.7	1.1	1.2	1.0	0.9	1.1	-	0.3
Framework to genome map[c]														
Total # framework markers	62	98	102	52	68	88	68	94	66	82	89	869	79	16.5
Physical dist. covered (Mbp)	40.7	63.8	79.7	41.1	73.8	50.3	51.9	68.4	38.4	38.6	40.8	587.5	–	–
Ratio kbp/centiMorgan	357.3	522.1	639.2	539.9	736.1	413.7	548.4	588.9	415.1	441.4	445.4	–	513.4	112.7

[a]Full map: all markers mapped at relaxed support for order.
[b]Framework map: markers ordered with higher statistical support.
[c]Framework to genome map: framework markers were positioned onto the assembled *Eucalyptus grandis* genome sequence to provide a correspondence between physical distance and recombination fraction for each pseudochromosome and at the genome-wide level.
doi:10.1371/journal.pone.0044684.t001

positions (Figure 2). Typically only a few sparse markers or small blocks of markers (e.g. LG1 and LG4) show a locally inconsistent order with the one estimated in the genome sequence. Upon further inspection of the segregation data of the few scattered markers showing discrepancy between their physical- and recombination-based positions, several of them were borderline in terms of marker quality and call rate parameters, which could possibly explain the observed inconsistencies. Out of the 1,029 framework-mapped markers, 869 could be positioned on the genome sequence while the remaining 160 either had no sequence available for mapping or mapped to the 4,941 smaller additional unanchored scaffolds of the current *Eucalyptus* genome assembly. The 869 linkage- and physically-mapped markers covered a total of 587.5 Mbp of sequence (Table 1) thereby providing 97% coverage of the 605.8 Mbp currently assembled in the 11 main scaffolds of the *Eucalyptus* genome. Pseudochromosome-specific estimates of the relationship between physical distance in kbp and recombination fraction in cM varied between 357.3 for pseudochromosome 1 and 736.1 for pseudochromosome 5, with a genome-wide average of 513.4 kbp/cM (Table 1). When the full linkage map was aligned to the genome sequence (data not shown), out of the 2,274 genetically mapped segregating DArT markers, 1,986 aligned to the 11 pseudochromosomes, while 45 markers mapped to 31 unanchored scaffolds and 243 DArT markers had no sequence available or did not map to the current assembly. Based on the linkage-mapped DArT markers, the 31 unanchored scaffolds, adding up 1.4 Mbp of sequence, could be assigned to the 11 main pseudochromosomes (Table S1).

DArT probe sequence redundancy analysis. Sequences were obtained for 6,918 of the 7,680 DArT probes (90%), with average size of 534 bp. Under the most stringent assembly parameters (see Material & Methods), out of the 6,918 sequences, 3,709 fell into multi-sequence clusters with two or more sequences per cluster. These were merged into 1,374 unique clusters of non-redundant sequences, while 3,209 sequences were unique, unmatched singletons. In total, the 6,918 probes for which sequences were obtained represented effectively 4,583 unique loci, i.e. a low bound estimate of the rate of redundancy of 33.75%. Under more liberal assembly parameters, the total number of unique loci was reduced to a total of 3,864, providing a high-end estimate of the rate of sequence redundancy at 44.14% (Table 2). If an equivalent rate of redundancy is assumed for the 762 DArT probes for which no sequences could be obtained, the 7,680 probes in the DArT array effectively sample between 4,289 and 5,087 unique loci in the *Eucalyptus* genome. All 6,918 sequences were submitted to GenBank and 6,896 were eventually accepted and deposited (22 were trimmed to sizes smaller than acceptable by the NCBI), receiving accession numbers HR865291-HR872186, with clone identifiers corresponding to the DArT marker naming convention used in this report. Searched against the complete NCBI EST database, 3,703 (53.6%) returned with positive BLASTn hits (Table S2).

DArT Marker Probe Alignment to the *Eucalyptus* Genome

Out of the 6,896 DArT probes for which quality sequences were obtained, 6,631 (96%) could be successfully aligned to the assembly of the *Eucalyptus grandis* genome sequence (version 1.0 in Phytozome 6.0) while 265 DArT probes could not be mapped using high stringency parameters. Of the mapped probes, 6,390 of them aligned to the 11 main pseudochromosomes and 241 to the additional 4,941 small unanchored scaffolds. When these mapping results were used to assess the quality of the sequence alignment parameters adopted (see Material and Methods) a mapping error rate of 0.002 was estimated by observing 12 unsuccessfully mapped sequences in 6,631, i.e. a reliability $\geq$99.8%. Interestingly, this number matches precisely the average scoring reproducibility estimated by DArTsoft following the standard

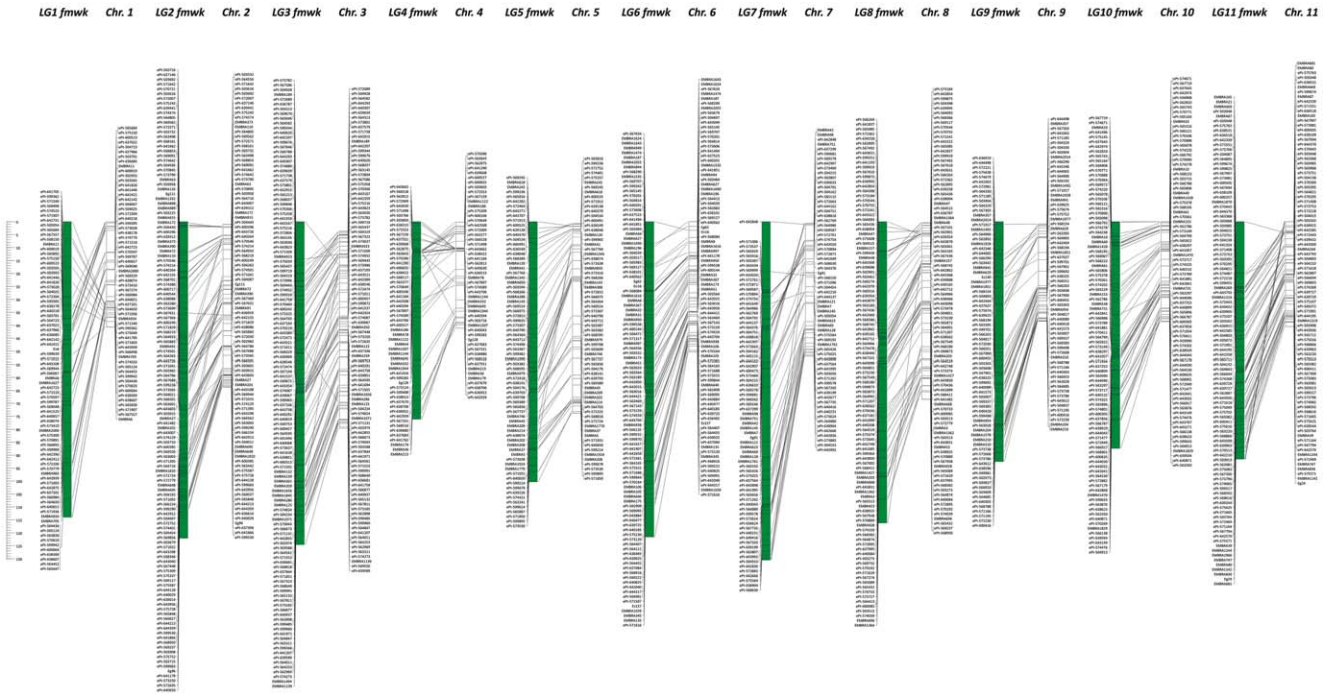


Figure 2. Alignment of the Framework map to the *Eucalyptus grandis* reference genome. Correspondence of the DArT and microsatellite marker positions on the Framework linkage map (green bars) with their location on the 11 *Eucalyptus grandis* pseudochromosome scaffolds (white bars). The scale on the left corresponds simultaneously to centiMorgan distances for the linkage map and to Mbp of sequence for the pseudochromosome scaffolds.
doi:10.1371/journal.pone.0044684.g002

marker selection thresholds used. When the threshold for a probe sequence hit to be retained was set down to the BWA default level (T = 37), 166 of the 265 unmapped probes could be additionally aligned to the 11 main pseudochromosomes and 91 of these 166 probes were also linkage-mapped onto the Full map. Additionally, out of the 89 probes that remained physically unmapped to the *E. grandis* genome assembly, 36 were successfully linkage-mapped as well.

A further examination of the alignment of the 6,631 DArT probes to the *Eucalyptus* genome assembly, including all 4,952 scaffolds, was carried out. A total of 4,189 probes were confidently aligned to a single and unique position in the genome as their mapping produced a single hit with no subalignment score. For 2,252 of the 2,442 remaining probes, a second subalignment was retained which overlapped the same locus as the one of the first best alignment. Therefore in total 6,441 DArT probes out of the 6,631 evaluated (97.1%) were considered to be aligned to a single locus in the genome. For the 190 probes in which a second hit was reported by BWA, we carried out an analysis using chimeric tools detection provided by the CD-HIT [52] and EULER DNA [53] fragment assembly softwares with default parameters, resulting in no detectable chimeric reads. For 135 probes, the retained subalignment was located in a different position on the same scaffold, and in 95 of these cases the distance between alignments was smaller than 1 kb, suggesting a contiguous tandem duplication. For 40 DArT probes the distances between the first and second hit alignments were larger than 1 kb, with 13 of them larger than 10 kb. Finally, only 55 probes were aligned to positions on different pseudochromosomes. In 37 of these cases both loci were found in the 11 main pseudochromosomes and in 18 cases one of the loci was found in an unanchored scaffold. The frequency of multilocus DArT probes in different chromosomes observed in *Eucalyptus* (55 in 6,631, i.e. 0.83%) is consistent with the 1.4% frequency observed in a linkage mapping study in barley [54].

DArT Marker Coverage of the *Eucalyptus* Gene-space

To characterize the *Eucalyptus* gene-space covered by the DArT array, we considered only the probes that were aligned to the 11 annotated pseudochromosomes. These totaled 6,390 probes which aligned to a total of 6,571 positions, given that 181 probes also aligned to a second position according to the BWA threshold adopted. The distribution of both the 6,571 DArT probes positions and the 1,986 genetically mapped DArT markers, were plotted together with the distribution of the 41,204 predicted gene models in the *Eucalyptus* genome (version 1.0) partitioned into 122 bins of 5 Mbp each, which on average correspond to ~10 cM map distance bins, assuming a ~1200 cM total recombination distance (Figure 3). The histogram indicates that the DArT microarray provides a homogeneous genome-wide coverage of markers and suggests a monotonic relationship between the number of gene models and the number of DArT markers. In fact a relatively strong and highly significant Spearman rank correlation was found between the number of predicted gene models and the total number of DArT markers found in a genome bin ($\rho = 0.682$; p = 3.79e-18), and likewise with the number of mapped DArT markers ($\rho = 0.467$; p = 5.19e-8) (Figure 4). These results show that the DArT array tends to provide segregating markers in essentially all 5 Mbp genomic bins with the number of DArT markers scaling with the number of genes in the bin. On average each bin contains 16±9.0 genetically mapped markers, 53±21.6 DArT marker probes and 334±100.9 predicted gene models. Only four bins had fewer than 20 DArT marker probes mapping to them and only 11 out of the 122 bins had fewer than 5

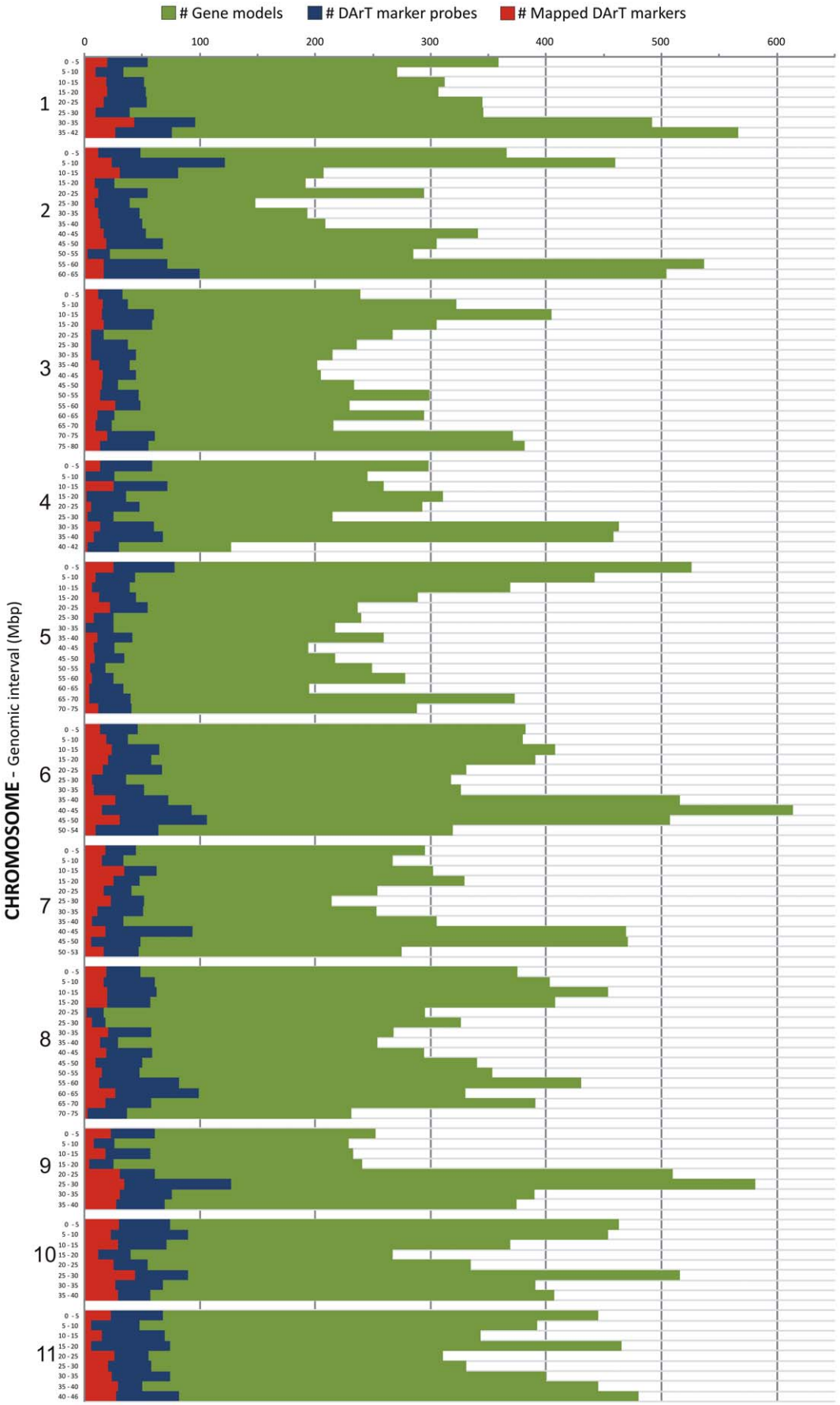

Gene models
DArT marker probes
Mapped DArT markers
0
100
200
300
400
500
600
CHROMOSOME - Genomic interval (Mbp)
1
2
3
4
5
6
7
8
9
10
11

Figure 3. Genome-wide correspondence of DArT markers and predicted gene models in the *Eucalyptus grandis* genome. The 11 pseudochromosomes of the *Eucalyptus grandis* genome (Version 1.0 in Phytozome 6.0), were partitioned into 122 bins of 5 Mbp. For each bin the numbers of DArT marker probe positions (blue bars), the number of genetically mapped DArT markers (red bars) and the number of predicted gene models (green bars) were plotted.
doi:10.1371/journal.pone.0044684.g003

genetically mapped markers. In addition, almost 70% of the DArT marker sequences were mapped at zero bp from the closest predicted gene model and less than 10% were located further than 10 kbp from predicted gene models (Figure 5).

Discussion

This study provides unprecedented data regarding the detailed genomic attributes of DArT markers in a plant genome. Following the development of a high performance *Eucalyptus* DArT array [25] we have now examined the genomic properties of the DArT marker probes that populate this array by sequencing them, constructing a high density linkage map and carrying out physical comparisons to the recently released *Eucalyptus grandis* BRASUZ1 annotated genome sequence. We have shown that DArT marker probes preferentially target the gene space and display a uniform distribution across the genome, providing excellent coverage for genome-wide applications in breeding and diversity studies. Such ubiquitous DArT marker properties had not been described

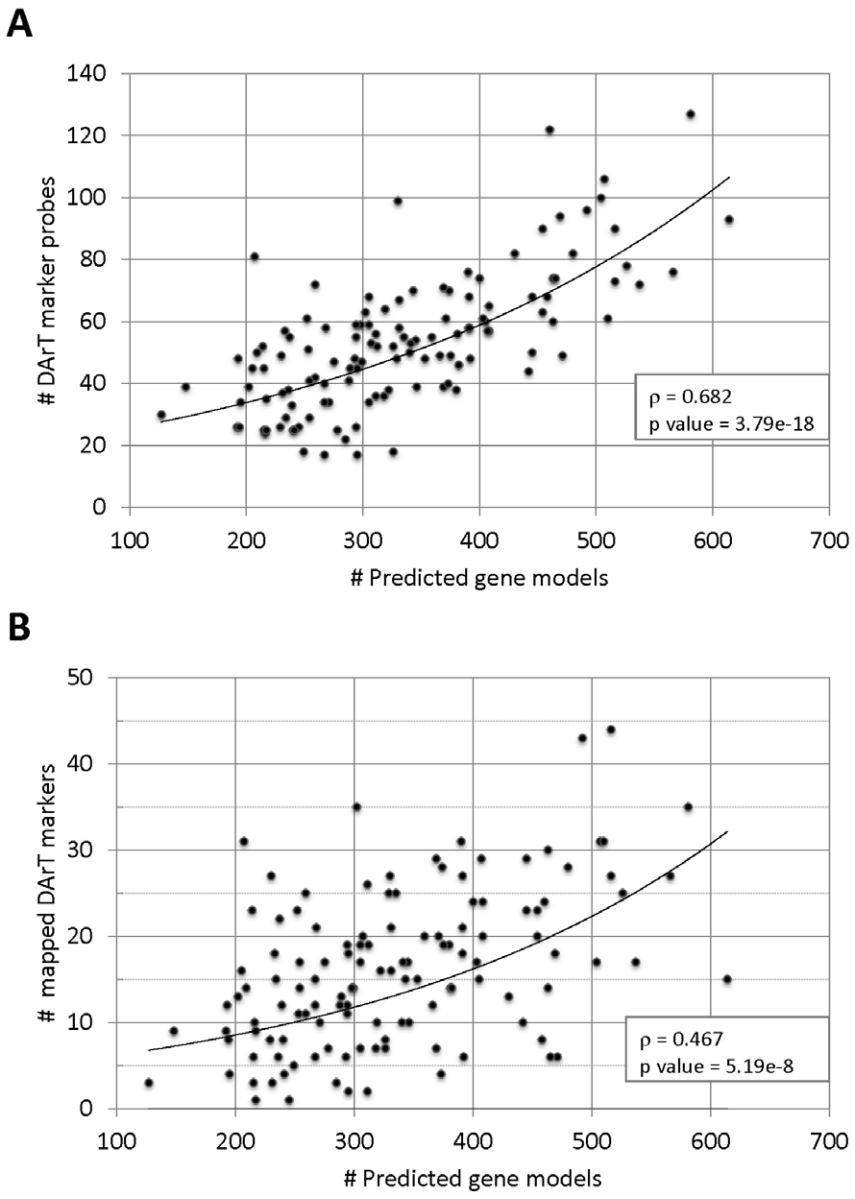


Figure 4. Correlations between DArT markers probes, mapped DArT markers and gene models. Spearman Rank correlations were estimated between: (A) the number of DArT marker probes and the number of gene models; and (B) the number of mapped DArT markers and the number of gene models, for every 5 Mbp genome bin.
doi:10.1371/journal.pone.0044684.g004

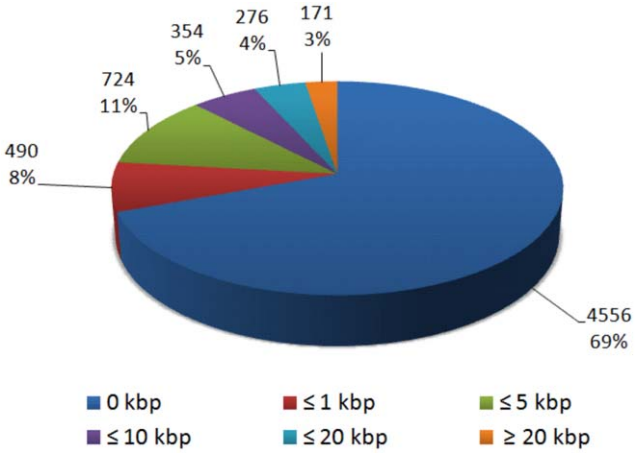


Figure 5. Distribution of the physical distance between DArT markers and gene models in the *Eucalyptus* genome. Distribution of the proportions of the 6,571 DArT marker probe positions according to distance classes in kbp from the closest predicted gene model in the *Eucalyptus grandis* genome (version 1.0).
doi:10.1371/journal.pone.0044684.g005

previously in spite of several DArT marker-based applied studies published to date for a large number of plant species.

DArT Marker Genotyping Efficiency for Genetic Analysis in *Eucalyptus*

A set of relatively strict filtering parameters was applied to the signal intensities obtained from the 7,680 probes in the *Eucalyptus* microarray. A total of 3,191 markers (41.5%) passed all marker quality and call rate thresholds (Figure S1), a proportion consistent with the original estimates reported during array validation [25] and recent mapping studies in similar interspecific pedigrees [29,30]. Besides marker quality filtering, a strict screening for adherence to Mendelian expectations was applied. Segregation distortion has been reported in some previous *Eucalyptus* mapping studies although, most of the time, at a rate no different from the one expected by chance alone [34,44]. This became a topic of interest as a way to assess heterospecific interactions in the F1 hybrid affecting introgression rates between distant species [55]. In this particular pedigree, however, just as in the first linkage map study in *Eucalyptus* [34], no segregation distortion would be expected in principle, since marker segregation was observed from each pure species parent and not in the gametes derived from a F1 hybrid where distortion could be expected. Accordingly, only 215 markers (6.7%) were excluded due to departures from expected segregation ratios, a proportion close to the 5% expected by chance alone. Besides sampling error, these distorted markers could include cases of duplicated loci such as the 55 DArT probes shown to confidently align to positions on different pseudochromosomes, and the 13 probes aligning at distances greater than 10 kbp, which in both cases would give rise to mixed hybridization signals and distorted segregation ratios. The 97% genome-wide physical coverage provided by the linkage map built in this study indicates that keeping distorted markers in the linkage analysis would not improve coverage but, rather, could complicate marker ordering if distortions derived from excessive missing data or mistyping were included [56].

Out of the 2,976 DArT markers included in the linkage analysis 2,274 were eventually mapped and ordered in the full consensus map and 1,029 in the framework version (Table 1). The proportion of DArT markers mapped is well within the predicted number of useful markers for mapping, between 1,818 and 2,553, as originally reported when the *Eucalyptus* DArT array was developed [25]. This number is also comparable to the 2,229 DArT markers mapped in the consensus map of two related backcross families of *E. grandis x E. urophylla* [30] and to the 1,845 EST-based Single Feature Polymorphism markers map reported earlier for the same pedigree used in this study [42]. The proportion of informative markers in such interspecific mapping pedigrees has been considerably higher than the number observed in intraspecific pedigrees. Hudson et al. [29] could only map 1,060 DArT markers in an outcrossed F2 family derived from two inter-provenance F1 individuals and only 569 in an inter-provenance cross of *E. globulus*. Genetic divergence between species and corresponding levels of differential sequence heterozygosity at the DArT loci determine in large part the proportion of informative markers ultimately captured. The DArT genotyping platform has provided an order of magnitude larger number of markers for mapping in *Eucalyptus* than previous technologies such as RAPD, AFLP and microsatellites [32]. The undomesticated nature of *Eucalyptus* resulted in a larger number of markers mapped than most DArT-based linkage maps built with extensively optimized DArT arrays such as those for wheat, oats [23], sorghum [57], and barley [54].

Recently, between 3,100 and 3,500 high quality polymorphic DArT markers were scored in breeding populations composed of several hundred individuals of *E. grandis* and *E. urophylla* in the context of Genomic Selection experiments [28]. As expected, when compared to an average of 2,200 to 2,300 markers captured and mapped in biparental pedigrees, the DArT array provides between 40 and 50% more markers at the population level. If similar proportions are kept, one can anticipate that in *E. globulus* breeding populations, the DArT array will provide between 750 and 1,500 informative markers depending on the general variability of the population and provenance composition.

Probe Redundancy is a useful Property of the DArT Array

Reported estimates of DArT marker redundancy obtained by comparing the segregation pattern in mapping population or estimating Hamming distances between markers have varied from 38% in barley [54], to 43% in *Arabidopsis* [58]. After sequencing all DArT probes on the array, a redundancy of between 33.75 and 44.14% was estimated (Table 2). This is consistent with the 46% redundancy rate reported for several thousand sequenced DArT probes from an oat genotyping array [23]. Under the same genome complexity reduction protocol, redundancy will vary with the particular genome structure of the target species, the diversity of samples used to build genomic representations and, largely, with the probe screening criteria and the final number of selected probes. As more probes are surveyed, a higher redundancy will result. Redundancy from sequencing only a few hundred DArT probes, previously selected for polymorphism, was estimated at 15% in an apple array, although a potential redundancy of 50% was acknowledged had all clones on the array been sequenced [24]. Considering that a fairly uniform physical and mapping distribution of the DArT probes was achieved across the *Eucalyptus* genome (Figures 2 and 3), a certain level of probe redundancy in the DArT array is actually a desirable feature. *Eucalyptus* DArT probes vary in size (534±215 bp) and, although sharing portions of DNA sequence, will have variable abilities to detect sequence polymorphism across individuals and populations, thereby providing improved power and flexibility for genome-wide genotyping.

An interesting aspect of the DArT probe sequence redundancy emerged when comparing the alignment of DArT probes to the reference genome. Between 3,864 and 4,583 unique sequences

Table 2. Results of redundancy analysis of the 6,918 DArT marker probe sequences under four different sets of assembly parameters from the most stringent (A1) to the most relaxed (A4) (see Material and Methods for details).

Parameter	A1	A2	A3	A4
Word length	18	14	12	10
Index Word length	13	12	11	10
Mismatches	10%	15%	20%	20%
Ambiguities per read	4	4	16	16
Maximum % gaps per read	10%	15%	20%	20%
Gap size	1bp	2bp	5bp	5bp
Results of redundancy analysis				
# Unmatched singleton[a]	3,209	2,607	2,381	2,276
# Redundant sequences[b]	3,709	4,311	4,537	4,642
# Unique non-redundant sequences[c]	1,374	1,537	1,587	1,588
Total selected sequences[d]	4,583	4,144	3,968	3,864
Estimated rate of sequence redundancy	33.75%	40.0%	42.64%	44.14%

[a]Unique sequences not matching any of the other reads.
[b]Sequences that fall into multi-sequence clusters with more than two sequences per cluster.
[c]Unique sequences drawn from the redundant clusters.
[d]Sum of unmatched singleton and non-redundant sequences.
doi:10.1371/journal.pone.0044684.t002

were observed for the DArT probes (Table 2), an estimate consistent with the 4,189 DArT probes confidently aligned to a unique position in the genome. Nevertheless the BWA alignment to the genome assembly revealed that 2,252 additional probes mapped to exclusive positions so that, in total, 6,441 loci in the genome were sampled by the DArT array. This is the first time such an analysis has been carried out for a species for which a DArT array and a reference genome are available. It shows that redundancy estimates based on a simple assembly of DArT probe sequences tend to be conservative.

Framework Linkage Mapping Allows Reliable Estimates of the kbp/cM Ratio in the *Eucalyptus* Genome

Two versions of a linkage map were built in this study, each one with specific objectives. A Framework map was built as the "hypothesis" that best explained the segregation data observed, to provide more accurate information regarding marker order [59,60] and to be used for the estimation of the relationship between physical distance and recombination fraction (Figure 1). On the other hand, the Full map, that included all segregating DArT markers, was built to provide a preliminary position for all possible DArT markers and thus allow a more extensive assessment of the genome coverage and distribution of DArT markers relative to genes. Additionally, by including the largest number of markers (even if at a relaxed order), this map offered a better probability of assigning unanchored scaffolds to the assembled pseudochromosomes of the current *Eucalyptus* reference genome sequence. Marker order of the Framework map and the Full map were generally comparable and total map sizes were also close (Figure S2 and Table 1). However, inverted sets of markers were observed between these map versions as well as markers that dropped out when going from one map to the other and *vice versa*. These results substantiate the well-known fact that the resolution of marker order is not a trivial issue and more so when a large number of markers are mapped with a limited progeny size [61]. This observation also supports the fact that similar apparent inversions and non-colinearities reported in previous comparative linkage mapping studies across sexually compatible *Eucalyptus* species [29] are, by and large, ordering inconsistencies due to various sources of experimental error [56] and rarely should be taken as evidence of any relevant biological genomic occurrence unless independent validation data is available. With the availability of a reference genome for *Eucalyptus*, coupled to high throughput sequencing technologies and powerful assembly procedures, such validations might now become possible.

A simple visual inspection of the aligned Full and Framework maps (Figure S2) and the significant difference found between the distributions of map distances between consecutive markers in the two map versions ($p = 0.021$) (Figure S3), support the conclusion that framework map building largely removes highly clustered sets of DArT markers, leaving a sparser map with essentially equivalent genome coverage and improved statistical support for relative ordering. Moreover, when the linkage map of microsatellites and DArTs was compared with a microsatellite-only map, the total recombination distance did not change (data not shown). This additional result suggests that, while the microsatellites do provide adequate genome coverage, the DArT markers effectively cover the genome in previously unsampled genomic regions, thereby providing the necessary marker density for high-resolution mapping and genome-wide studies. The Framework map should therefore be taken as the most reliable map when it comes to comparative analysis with the reference genome or map-based efforts (such as looking for the co-localization of genes with potentially large effect QTLs).

The alignment and physical mapping of 869 framework mapped DArT and microsatellite markers to the 11 main scaffolds of the *Eucalyptus* genome sequence allowed an estimation of the relationship between physical distance and recombination fraction in each pseudochromosome and for the whole genome. This estimate varied considerably (357 to 736 kbp/cM) with a genome-wide average of 513 kbp/cM (Table 1). Interestingly, this estimate is not far from the coarse estimates of 395–559 kbp/cM reported early on, based on the first available linkage maps [34] and the first estimates of *Eucalyptus* genome size [62]. This time, however, by using the assembled genome sequence to which framework markers were mapped physically, an improved estimate was possible. Kullan et al. [30] using a different approach, based exclusively on a selected set of 153 pairs of flanking markers mapped at approximately 1 cM distance, estimated 633 kbp/cM. Besides the potential bias introduced by specifically selecting pairs of markers and, as a consequence, precluding the intrinsic variation in recombination *versus* physical distance along the genome, that estimate was based on markers ordered at a relaxed likelihood. We therefore consider the estimates presented in this work, both at the pseudochromosome level and whole-genome average to be better approximations for *Eucalyptus* (Table 1), although we acknowledge that recombination rates are expected to vary by orders of magnitude across a genome [63].

Linkage to Physical Mapping Suggests a Pan-genomic Feature of the DArT Array and Completeness of the *Eucalyptus* Genome Assembly

The overall consistency between the order of framework mapped DArT markers and their physical order in the genome sequence substantiate the quality of scaffold assembly in the current *Eucalyptus* genome sequence (Figure 2). While the linkage map reported by Kullan et al. [30] was used effectively to assist

scaffold ordering during genome assembly (J. Schmutz pers. comm.) the linkage map presented herein was not, and thus constitutes an independent validation of the current *Eucalyptus* genome. Furthermore, from the completeness standpoint, only 45 markers out of 2,274 linkage mapped ones could not be aligned to the 11 main scaffolds. Conversely, only 89 probes remained physically unmapped to the genome sequence. Although these unmapped markers could imply missing sections in the genome assembly, they could also correspond to sequences that do not exist in the *E. grandis* genome, recalling that the 7,680 probes on the array were developed from 18 genomic representations involving 64 different *Eucalyptus* species with a broad phylogenetic diversity. A scrutiny of the original source of these 89 unmapped DArT probes revealed that 65 of them came from genomic representation libraries built with DNA from species other than *E. grandis*, while 24 came from *E. grandis* [25]. Nevertheless we observed significantly more non-*E. grandis* DArT probes not mapping to the genome than what would be expected due to chance alone (Pearson Chi-square 5.03; p value = 0.0249). This result, together with the excellent performance demonstrated for diverse phylogenetic investigations in the genus [22], suggests a distinctive pan-genomic attribute of this *Eucalyptus* DArT array. While most probes correspond to core genomic features common to all individuals and *Eucalyptus* species, a few probes may be derived from the "dispensable genome" composed of partially shared and/or non-shared DNA sequence elements among species [64,65].

Completeness of the current assembly was also supported by the observation that, out of 85.4 Mbp of unanchored sequence in 4,941 small scaffolds, only 1.4 Mbp across 31 scaffolds was captured by 45 mapped markers and all, but a couple, were located in intermediate positions along the linkage groups and not at the extremes (Table S1). Were the genome assembly incomplete, one would expect to capture a considerably larger proportion of unanchored scaffolds and sequence. In fact, during the poplar genome assembly, Drost et al. [66], using a medium-density 608 marker map, were able to anchor 116 sequence scaffolds to unique genetic positions in linkage groups, thereby adding to the genome some 35.7 Mbp of sequence out of the 75 Mbp still unanchored at the time. These results, together with the fact that 86% of the 4,941 unanchored *Eucalyptus* genome scaffolds are less than 20 kbp in length, strongly suggest that the vast majority of unanchored scaffolds correspond to fragments of alternative haplotypes of already assembled pseudochromosomes, possibly derived from regions of high heterozygosity in the *Eucalyptus* genome and not to missing portions of the genome.

The *Eucalyptus* DArT Array Provides Uniform Genome-wide Coverage While Preferentially Targeting Gene-rich Regions

Results from BLAST hits and genome-wide analysis of the *Eucalyptus* DArT probe sequences (Table S2 and Figure 3) corroborated previous studies in other plant species reporting that *PstI*-based DArT markers are predominantly located in low-copy, gene-rich regions of the genome [23,54,67]. However, the opportunity to map the DArT probes to a fully annotated reference genome beyond a simple BLAST analysis against ESTs, revealed a highly significant relationship between the numbers of DArT markers and predicted gene models (Figure 4) with a small proportion of DArT probes located more than 10 kbp from the closest gene (Figure 5). This result is significant as it might help explain the unprecedented level of resolution that the DArT array has provided for population genetic and breeding studies across the full range of *Eucalyptus* species [22,28]. Based on the genomic characterization of the DArT probe sequences reported in this study, phylogenies or population genetic surveys based on DArT markers can now be further explored according to the gene proximity or gene content of particular markers sets. Alternatively, DArT markers from specific genes or selectively neutral regions can be selected *a priori* for targeted phylogenetic reconstructions. Moreover, the combination between genome-wide coverage and predominant association to the gene-space could also account for the good performance of the DArT array in providing markers for accurate genome-wide predictive models in recent Genomic Selection (GS) studies [28]. It might now be possible to correlate the genomic attributes of the DArT markers to their specific contributions to the predictive ability of GS models or to the resolution of specific phylogenies and hence look for specific markers or genomic segments of particular interest in subsequent studies.

Conclusion

The results of this work, following the recently published DArT-based genetic studies in *Eucalyptus* [22,28,29,30], further highlight the value of this genotyping platform for genetics, breeding and evolutionary genome-wide surveys in species of this genus. Given the commonality of the methods used in developing DArT arrays, the genomic properties of the markers described in this study are likely ubiquitous to most if not all angiosperm plant genomes. The DArT technology has now evolved by taking advantage of high-throughput short read sequencing [68]. By combining its long time established genome complexity reduction method, also adopted by recently described genotyping-by-sequencing (GbS) protocols [69,70], a considerable leap in genome-wide polymorphism detection has taken place. Nevertheless, the general genomic attributes of the GbS-derived markers as far as genome coverage and preferential targeting of gene-rich regions should remain essentially the same as those described in this study, although a much larger number of markers based on digital sequence counts rather than analog microarray signal are obtained, in addition to the scoring of co-dominant SNP markers. This advance might push down current costs of large scale high-throughput plant genotyping even further than the DArT and SNP platforms did in the last few years. However, the necessary informatics infrastructure required to handle, store and analyze the huge sequence files generated by GbS for several thousand samples will not be immediately available in the realm of most plant genetic resources and breeding operations. Microarray-based DArT genotyping with its standardized processing and analysis protocols shall therefore continue to be a useful tool for a number of applications in plant genetic analysis, particularly those that not necessarily require very high density genome-wide genotyping.

Supporting Information

Figure S1 Distributions of the number and percentages of DArT markers that passed the filtering thresholds adopted for reproducibility (≥95%), quality score (Q ≥65) and call rate (≥75%). A Venn diagram consolidates the information showing all possible classifications of the DArT markers according to the three filtering criteria adopted. Only markers that satisfied simultaneously all three criteria were used for linkage mapping.
(PDF)

Figure S2 Alignment of the Full map (yellow bars) to the Framework (Fmwk) map (green bars) for the eleven *Eucalyptus* pseudochromosomes built using JoinMap 3.0, showing the connections between the same loci on both maps. The Full map includes a total of 2,484 markers,

2,274 DArT and 210 microsatellites while the Framework map has 1,029 markers positioned with higher confidence for locus order, 861 DArT and 168 microsatellites. DArT markers in black and microsatellites in red; centiMorgan scale on the left.
(PDF)

Figure S3 Frequency distributions of Kosambi recombination distances between consecutive markers across the two linkage map versions. The distribution of map distances in the Framework map was significantly different from the one in the Full map ($p = 0.021$ of a non-parametric Komolgorov-Smirnov test), confirming the fact that a Framework map spreads out the retained markers with high support for ordering and reduces the proportion of inter-marker distances smaller than one centiMorgan from a total of 87% in the Full map to 65% in the Framework map.
(PDF)

Table S1 List of the 45 DArT markers linkage mapped to the eleven groups but aligning to small unanchored scaffolds of the current *Eucalyptus grandis* genome assembly (version 1.0 into Phytozome 6.0). These linkage mapped DArT markers allowed the assignment of 31 small scaffolds (1.4 Mbp of total sequence) to the 11 main pseudochromosomes.
(PDF)

Table S2 BLASTn hits of DArT marker probes (Genbank accession numbers HR865291-HR872186) searched against the complete NCBI EST database (August 12 2010 build); 3,703 (53.6%) returned with positive BLASTn hits (e value $<1e{-}5$).
(XLSX)

Acknowledgments

We acknowledge the help and technical support from many employees of Diversity Arrays Technology Pty. Ltd. where the genotyping work was performed.

Author Contributions

Conceived and designed the experiments: DG AK AAM REV DAS. Performed the experiments: CDP CPS JC. Analyzed the data: CDP CPS DG OBS GJP. Contributed reagents/materials/analysis tools: DG DAS REV AAM. Wrote the paper: DG CDP.

References

1. Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding - new tools for an old science. Bio-Technology 7: 257–264.
2. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. Nucleic Acids Research 18: 6531–6535.
3. Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, et al. (1995) AFLP - a new technique for DNA-fingerprinting. Nucleic Acids Research 23: 4407–4414.
4. Ganal MW, Altmann T, Roder MS (2009) SNP identification in crop plants. Current Opinion in Plant Biology 12: 211–217.
5. Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29: e25.
6. Kilian A (2009) Case 9: Diversity Arrays Technology Pty Ltd (DArT). Applying the open source philosophy in agriculture. In: Van Overwalle G, editor. Gene patents and collaborative licensing models: patent pools, clearinghouses, open source models and liability regimes: Cambridge University Press 204–213.
7. Mantovani P, Maccaferri M, Sanguineti MC, Tuberosa R, Catizone I, et al. (2008) An integrated DArT-SSR linkage map of durum wheat. Molecular Breeding 22: 629–648.
8. Xia L, Peng KM, Yang SY, Wenzl P, de Vicente MC, et al. (2005) DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. Theoretical and Applied Genetics 110: 1092–1098.
9. Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, et al. (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proceedings of the National Academy of Sciences of the United States of America 101: 9915–9920.
10. Supriya A, Senthilvel S, Nepolean T, Eshwar K, Rajaram V, et al. (2011) Development of a molecular linkage map of pearl millet integrating DArT and SSR markers. Theoretical and Applied Genetics 123: 239–250.
11. Howard EL, Whittock SP, Jakse J, Carling J, Matthews PD, et al. (2011) High-throughput genotyping of hop (*Humulus lupulus* L.) utilising diversity arrays technology (DArT). Theoretical and Applied Genetics 122: 1265–1280.
12. Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, et al. (2010) A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. Bmc Plant Biology 10.
13. Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, et al. (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. Bmc Genomics 10.
14. Bolibok-Bragoszewska H, Heller-Uszynska K, Wenzl P, Uszynski G, Kilian A, et al. (2009) DArT markers for the rye genome - genetic diversity and mapping. Bmc Genomics 10.
15. James KE, Schneider H, Ansell SW, Evers M, Robba L, et al. (2008) Diversity Arrays Technology (DArT) for pan-genomic evolutionary studies of non-model organisms. Plos One 3: e1682.
16. Van Schalkwyk A, Wenzl P, Smit S, Lopez-Cobollo R, Kilian A, et al. (2012) Bin mapping of tomato diversity array (DArT) markers to genomic regions of *Solanum lycopersicum x Solanum pennellii* introgression lines. Theoretical and Applied Genetics 124: 947–956.
17. Belaj A, Dominguez-Garcia MD, Atienza SG, Urdiroz NM, De la Rosa R, et al. (2012) Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. Tree Genetics & Genomes 8: 365–378.
18. Reddy UK, Rong JK, Nimmakayala P, Vajja G, Rahman MA, et al. (2011) Use of diversity arrays technology markers for integration into a cotton reference map and anchoring to a recombinant inbred line map. Genome 54: 349–359.
19. Milczarski P, Bolibok-Bragoszewska H, Myskow B, Stojalowski S, Heller-Uszynska K, et al. (2011) A high density consensus map of rye (*Secale cereale* l.) based on DArT markers. Plos One 6: e28495.
20. Bartos J, Sandve SR, Kolliker R, Kopecky D, Christelova P, et al. (2011) Genetic mapping of DArT markers in the Festuca-Lolium complex and their use in freezing tolerance association analysis. Theoretical and Applied Genetics 122: 1133–1147.
21. Yang SY, Saxena RK, Kulwal PL, Ash GJ, Dubey A, et al. (2011) The first genetic map of pigeon pea based on diversity arrays technology (DArT) markers. Journal of Genetics 90: 103–109.
22. Steane DA, Nicolle D, Sansaloni CP, Petroli CD, Carling J, et al. (2011) Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. Molecular Phylogenetics and Evolution 59: 206–224.
23. Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, et al. (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. Bmc Genomics 10: 39.
24. Schouten HJ, van de Weg WE, Carling J, Khan SA, McKay SJ, et al. (2012) Diversity arrays technology (DArT) markers in apple for genetic linkage maps. Molecular Breeding 29: 645–660.
25. Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, et al. (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. Plant Methods 6: 16.
26. Potts BM (2004) Genetic improvement of eucalypts. In: Burley J, Evans J, Youngquist JA, editors. Encyclopedia of Forest Science. Oxford: Elsevier Science. 1480–1490.
27. Grattapaglia D, Silva OB, Kirst M, de Lima BM, Faria DA, et al. (2011) High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. Bmc Plant Biology 11: 65.
28. Resende MD, Resende MF Jr, Sansaloni CP, Petroli CD, Missiaggia AA, et al. (2012) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol 194: 116–128.
29. Hudson C, Kullan A, Freeman J, Faria D, Grattapaglia D, et al. (2011) High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. Tree Genetics & Genomes: 1–14.
30. Kullan ARK, van Dyk MM, Jones N, Kanzler A, Bayley A, et al. (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 x E. urophylla*. Tree Genetics & Genomes 8: 163–175.
31. Jones N, Ougham H, Thomas H, Pasakinskiene I (2009) Markers and mapping revisited: finding your gene. New Phytologist 183: 935–966.
32. Grattapaglia D, Kirst M (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytologist 179: 911–929.
33. Grattapaglia D, Vaillancourt R, Shepherd M, Thumma B, Foley W, et al. (2012) Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. Tree Genetics & Genomes d.o.i. 10.1007/s11295-012-0491-x.

34. Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121–1137.
35. Verhaegen D, Plomion C (1996) Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. Genome 39: 1051–1061.
36. Marques CM, Araujo JA, Ferreira JG, Whetten R, O'malley DM, et al. (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. Theor Appl Genet 96: 727–737.
37. Bundock PC, Hayden M, Vaillancourt RE (2000) Linkage maps of *Eucalyptus globulus* using RAPD and microsatellite markers. Silvae Genetica 49: 223–232.
38. Myburg AA, Griffin AR, Sederoff RR, Whetten RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. Theor Appl Genet 107: 1028–1042.
39. Freeman JS, Potts BM, Shepherd M, Vaillancourt RE (2006) Parental and consensus linkage maps of *Eucalyptus globulus* using AFLP and microsatellite markers. Silvae Genetica 55: 202–217.
40. Byrne M, Murrell J, Allen B, Moran G (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theor Appl Genet 91: 869–875.
41. Thamarus KA, Groom K, Murrell J, Byrne M, Moran GF (2002) A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre, and floral traits. Theoretical and Applied Genetics 104: 379–387.
42. Neves LG, Mamani EMC, Alfenas AC, Kirst M, Grattapaglia D (2011) A high-density transcript linkage map with 1,845 expressed genes positioned by microarray-based Single Feature Polymorphisms (SFP) in *Eucalyptus*. BMC Genomics 12: 189.
43. Grattapaglia D (2004) Integrating genomics into *Eucalyptus* breeding. Genetics and Molecular Research 3: 369–379.
44. Brondani RP, Williams ER, Brondani C, 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 Biol 6: 20.
45. Faria DA, Mamani EMC, Pappas GJ, Grattapaglia D (2011) Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. Tree Genetics & Genomes 7: 63–77.
46. Faria DA, Mamani EMC, Pappas MR, Pappas GJ, Grattapaglia D (2010) A selected set of EST-derived microsatellites, polymorphic and transferable across 6 species of *Eucalyptus*. Journal of Heredity 101: 512–520.
47. Brondani RP, Grattapaglia D (2001) Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. Biotechniques 31: 793–795, 798, 800.
48. Van Ooijen J, Voorrips R (2001) JoinMap 3.0 software for the calculation of genetic linkage maps. Wageningen, the Netherlands: Plant Research International.
49. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, et al. (2012) Geneious v5.6. Available from http://www.geneious.com.
50. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589–595.
51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.
52. Li WZ, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658–1659.
53. Pevzner PA, Tang HX, Waterman MS (2001) An Eulerian path approach to DNA fragment assembly. Proceedings of the National Academy of Sciences of the United States of America 98: 9748–9753.
54. Wenzl P, Li HB, Carling J, Zhou MX, Raman H, et al. (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7: 206.
55. Myburg AA, Vogl C, Griffin AR, Sederoff RR, Whetten RW (2004) Genetics of postzygotic isolation in *Eucalyptus*: whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus*. Genetics 166: 1405–1418.
56. Hackett CA, 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.
57. Mace ES, Rami JF, Bouchet S, Klein PE, Klein RR, et al. (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biology 9: 13.
58. Wittenberg AHJ, van der Lee T, Cayla C, Kilian A, Visser RGF, et al. (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Molecular Genetics and Genomics 274: 30–39.
59. Keats BJB, Sherman SL, Morton NE, Robson EB, Buetow KH, et al. (1991) Guidelines for human linkage maps - an international system for human linkage maps (ISLM, 1990). Genomics 9: 557–560.
60. Vision TJ, Brown DG, Shmoys DB, Durrett RT, Tanksley SD (2000) Selective mapping: A strategy for optimizing the construction of high-density linkage maps. Genetics 155: 407–420.
61. Cheema J, Dicks J (2009) Computational approaches and software tools for genetic linkage map estimation in plants. Briefings in Bioinformatics 10: 595–608.
62. Grattapaglia D, Bradshaw HD (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. Canadian Journal of Forest Research 24: 1074–1078.
63. Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends in Genetics 20: 103–111.
64. Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10: 149–155.
65. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, et al. (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nature Genetics 43: 956–U960.
66. Drost DR, Novaes E, Boaventura-Novaes C, Benedict CI, Brown RS, et al. (2009) A microarray-based genotyping and genetic mapping approach for highly heterozygous outcrossing species enables localization of a large fraction of the unassembled *Populus trichocarpa* genome sequence. Plant Journal 58: 1054–1067.
67. Akbari M, Wenzl P, Caig V, Carling J, Xia L, et al. (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theoretical and Applied Genetics 113: 1409–1420.
68. Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, et al. (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. BMC Proceedings 5: P54.
69. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A robust, simple genotyping-by-sequencing (GbS) approach for high diversity species. Plos One 6: e19379.
70. Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7: e32253.

BMC Genomics

RESEARCH ARTICLE Open Access

A reference linkage map for *Eucalyptus*

Corey J Hudson[1*], Jules S Freeman[1,2], Anand RK Kullan[3], César D Petroli[4], Carolina P Sansaloni[4], Andrzej Kilian[5], Frank Detering[5], Dario Grattapaglia[6], Brad M Potts[1], Alexander A Myburg[3] and René E Vaillancourt[1]

Abstract

Background: Genetic linkage maps are invaluable resources in plant research. They provide a key tool for many genetic applications including: mapping quantitative trait loci (QTL); comparative mapping; identifying unlinked (*i.e.* independent) DNA markers for fingerprinting, population genetics and phylogenetics; assisting genome sequence assembly; relating physical and recombination distances along the genome and map-based cloning of genes. Eucalypts are the dominant tree species in most Australian ecosystems and of economic importance globally as plantation trees. The genome sequence of *E. grandis* has recently been released providing unprecedented opportunities for genetic and genomic research in the genus. A robust reference linkage map containing sequence-based molecular markers is needed to capitalise on this resource. Several high density linkage maps have recently been constructed for the main commercial forestry species in the genus (*E. grandis*, *E. urophylla* and *E. globulus*) using sequenced Diversity Arrays Technology (DArT) and microsatellite markers. To provide a single reference linkage map for eucalypts a composite map was produced through the integration of data from seven independent mapping experiments (1950 individuals) using a marker-merging method.

Results: The composite map totalled 1107 cM and contained 4101 markers; comprising 3880 DArT, 213 microsatellite and eight candidate genes. Eighty-one DArT markers were mapped to two or more linkage groups, resulting in the 4101 markers being mapped to 4191 map positions. Approximately 13% of DArT markers mapped to identical map positions, thus the composite map contained 3634 unique loci at an average interval of 0.31 cM.

Conclusion: The composite map represents the most saturated linkage map yet produced in *Eucalyptus*. As the majority of DArT markers contained on the map have been sequenced, the map provides a direct link to the *E. grandis* genome sequence and will serve as an important reference for progressing eucalypt research.

Background

Genetic linkage maps are valuable resources which can be used to provide a framework for many genomic analyses. Linkage maps can be used to investigate the organisation and evolution of genomes through comparative mapping [1-3] and serve as a basis for investigating phenotypic traits of ecological and economic importance through the localisation of quantitative trait loci [QTL; 4-6]. Subsequently, QTL results may be used to help guide the selection of candidate genes for association studies or be applied in marker-assisted breeding programmes [7,8]. Linkage maps can also be used to anchor physical maps and assist in the assembly of genome sequences [9-11]. The wide application of linkage maps in combination with their value to genetics research has led to numerous linkage mapping projects being undertaken in plants. Detailed linkage maps have been produced for all of the world's staple cereal species [12], and in forest trees, linkage maps have been produced for many of the most widely-planted species due to their commercial importance as wood and fibre crops [1,13,14].

Grattapaglia and Sederoff [15] published the first genetic linkage map in the forest tree genus *Eucalyptus* in 1994. Subsequently, many mapping pedigrees have been established for the purpose of linkage map construction and associated QTL analyses. More than 20 eucalypt genetic linkage maps have been reported with most being produced in the main commercially grown species, or their hybrids, from the *Eucalyptus* subgenus *Symphyomyrtus*. Thus, the majority of linkage mapping projects have focussed on *E. grandis*, *E. urophylla* and *E. globulus* [reviewed in 16], while a smaller number of maps have

* Correspondence: cjhudson@utas.edu.au
[1]School of Plant Science and CRC for Forestry, University of Tasmania, Private Bag 55 Hobart, Tasmania 7001, Australia
Full list of author information is available at the end of the article

also been produced for *E. nitens* [17], *E. teriticornis* [18,19], *E. camaldulensis* [20] and for species in the closely related genus *Corymbia* [21].

Many early eucalypt linkage maps were constructed using random amplification of polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) molecular markers [16,22]. However, the anonymous nature of these dominant markers has limited the transfer of linkage information between studies [16,23]. More informative, codominant markers such as isozyme and random fragment length polymorphism (RFLPs) have also been used in eucalypt linkage mapping, although, their low throughput, low inter-pedigree polymorphism and labour intensive genotyping requirements have limited their use [16,23]. The more recent development of highly polymorphic microsatellite markers made available a large potential suite of markers that are transferrable between species and polymorphic in multiple pedigrees. This enabled linkage group synteny to be established between maps containing common microsatellite markers and the positions and stability of QTL across multiple species to be examined [e.g. 24-27]. The ability to establish linkage group synteny has also enabled moderate-density comparative mapping studies [23,28].

Recent advances in molecular methods have led to high-throughput genotyping systems being developed [e.g. 29,30]. These have made it possible to quickly generate many hundreds of markers in single mapping pedigrees and have helped facilitate the construction of high density linkage maps [12]. Most recently in *Eucalyptus*, Diversity Arrays Technology [DArT; 31] has been used to generate large numbers of molecular markers for genetic linkage mapping in several mapping pedigrees [e.g. 11,32,33]. The eucalypt DArT markers are highly transferable across species from subgenus *Symphyomyrtus* [34] and the high-throughput array-based genotyping system provides wide genome coverage [35]. A key benefit of the *Eucalyptus* DArT markers is the public availability of the sequences of most of the 7680 markers contained on the genotyping array [GenBank accession numbers HR865291 - HR872186], thus making it possible to anchor DArT markers directly to the reference *E. grandis* genome sequence [v1.0 released January 2011; 36]. However, while the DArT technology offers many advantages, the DArT markers do suffer some limitations due to their dominant nature. For example, the incomplete segregation information provided by those DArT markers segregating in a 3:1 ratio (intercross) results in an exponential increase of marker-ordering calculations compared to fully-informative co-dominant markers [37]. Co-dominant markers also provide more complete information in QTL mapping studies [e.g. allowing estimation of additive and dominant allelic effects; 38] and are more useful in some genetic analyses, such as estimating population genetic parameters (e.g. inbreeding levels), relative to dominant marker types such as DArT. In addition, the DArT marker assay can be subject to cross-hybridization from duplicated loci in the genome, although most such artifacts can be excluded by preselecting markers exhibiting Mendelian segregation ratios in mapping pedigrees.

At present, DArT markers have been used to construct linkage maps in seven independent *E. globulus* and/or *E. grandis* × *E. urophylla* hybrid family mapping pedigrees [11,32,33]. All of these maps also contain a variable number of co-dominant microsatellite markers, which provide important links to many earlier eucalypt linkage maps. In the two largest mapping pedigrees (more than 500 individuals each), 1010 [32] and 2229 [33] DArT markers, were mapped at sub-centiMorgan marker densities and collectively more than 4000 DArT and microsatellite markers have been mapped in the seven pedigrees.

All DArT marker based linkage maps were constructed using the program JoinMap 4.0 [37]. This program is one of the most commonly used linkage mapping programs and appears to be the only software available for building linkage maps using the combined segregation data from multiple populations [39-41]. However, it is presently not feasible to combine the segregation data contained within the seven eucalypt mapping families describe above (collectively 1950 individuals), and successfully order such large numbers of markers within linkage groups (up to ~ 500) due to computational limitations (Van Ooijen *pers comm.*). To circumvent the limitations of traditional segregation-based methods of linkage map construction, alternative marker-merging strategies have been developed. A so-called 'composite map' can be produced in which markers from individual component maps are merged into a single map based on their position relative to common anchor loci. For example, the 'neighbours' marker-merging approach of Cone *et al.* [42] and the marker-merging method implemented in the PhenoMap program (GeneFlow Inc. USA) have been used to successfully construct high density composite maps containing several thousand markers in a number of plant species; including *Sorghum* [43], barley [41,44,45] and maize [42,46].

In this study, a marker-merging method was used to construct a high-density DArT and microsatellite marker composite linkage map from seven independently constructed maps. Recent comparative mapping analyses using 236 to 393 markers shared between three of the maps [see 32] showed that these linkage maps exhibited high synteny (> 93.4% markers occurring on the same linkage groups) and high colinearity (> 93.7% markers having the same order within linkage groups). This indicated that it would be possible to merge markers from

several component maps into a single high quality map featuring robust marker-order together with very high marker density. It is expected that this composite map will facilitate marker and map information exchange and serve as a valuable reference for species in the subgenus *Symphyomyrtus*.

Methods

The following terms are used to describe the various types of linkage maps reported in this paper; (1) sex-averaged map – a consensus of individually constructed male and female maps, built in a single family using segregation data from both parents, (2) consensus map – a consensus of multiple individually constructed male and female maps, built in multiple families (e.g. F_2 double-pseudo backcross) using segregation data from all of the families, and (3) composite map – an integrated map of multiple sex-averaged and/or consensus maps, built using a marker-merging method.

Component maps

The composite map was built using an *E. grandis* × *E. urophylla* F_2 double pseudo-backcross pedigree consensus linkage map [both species from section *Latoangulatae*; 33] plus one *E. grandis* × *E. urophylla* sex-averaged map constructed in a F_1 hybrid pedigree [11] and five pure-species *E. globulus* [section *Maidenaria*; 32] sex-averaged linkage maps constructed in either outcrossed F_2 or F_1 families (hereafter referred to as 'component' maps). Component map family sizes ranged from 172 (GLOB-F_2-1) to 547 (GU-SA) and collectively contained 1,950 individuals (Table 1). The component maps were constructed by different researchers. All used JoinMap 4.0 [37] with marker-ordering within linkage groups (LGs) estimated using the regression algorithm of Stam [47] combined with the Kosambi mapping function. All component maps comprised 11 linkage groups in accordance with the haploid chromosome number of *Eucalyptus* [48].

Before building the composite map, marker names were standardised across maps, homologous linkage groups were identified using common (anchor) loci and marker co-linearity between component maps was visually inspected in MapChart [49]. Map data was supplied for both framework (1032-marker) and comprehensive (2484-marker) maps built in the GU-Emb family [see 11]. Based on the level of marker-order agreement between linkage groups from these maps with other component maps, either GU-Emb framework (LG's 1, 3, 5, 7 and 9) or comprehensive (LG's 2, 4, 6, 8, 10 and 11) linkage groups were included in composite map construction. Five linkage groups from three of the smaller *E. globulus* mapping families (Table 1) were found to have substantial regions of non-colinearity (discordant marker-orders) with other component maps. Consequently, LG6 and LG10 from the GLOB-F_1-1 map, LG4 and LG9 from the GLOB-F_1-4 and LG4 from the GLOB-F_1-5 map were excluded from composite map construction.

The number of markers included for composite map construction ranged from 498 (GLOB-F_1-4) to 2290 (GU-SA; Table 1). In total, this consisted of 4350 individual markers, including: 4089 DArT, 253 microsatellites and eight mapped genes. Ninety-six markers (2.2% of the total number of markers; termed 'multicopy' markers) were mapped to two or more linkage groups across component maps. This resulted in the 4350 individual markers being mapped to 4457 positions. Of these 4457 positions, 1960 could be considered to be bridging loci, meaning that these markers had been mapped to syntenic linkage groups in two or more component maps and would serve as anchor loci during composite map construction. Conversely, 2497 marker positions were unique to single component maps.

Table 1 Component map details

Linkage map[a]	Map abbreviation	*n*	cM	MMI	Markers mapped (percentage of unique markers in pedigree)			
					DArT	SSR	Gene	Total
E. grandis × *E. urophylla* SA double pseudo-backcross F_2 [b]	GU-SA	547	1107	0.51	2229 (45%)	59 (46%)	2 (100%)	2290 (45%)
E. grandis × *E. urophylla* Embrapa F_1 [ce]	GU-Emb	177	1229	0.78	1617 (41%)	193 (77%)	0	1810 (44%)
E. globulus Lighthouse F_2 [d]	GLOB-LH	503	1151	1.21	1010 (27%)	50 (12%)	0	1060 (27%)
E. globulus FAM1 F_1 [d]	GLOB-F_1-1	184	1033	1.97	571 (14%)	4 (0%)	2 (0%)	577 (14%)
E. globulus FAM4 F_1 [d]	GLOB-F_1-4	184	1137	2.46	488 (10%)	6 (0%)	4 (25%)	498 (10%)
E. globulus FAM5 F_1 [d]	GLOB-F_1-5	183	1055	2.09	600 (22%)	4 (0%)	2 (0%)	606 (21%)
E. globulus FAM1 F_2 [d]	GLOB-F_2-1	172	1258	2.73	660 (18%)	30 (30%)	5 (40%)	695 (18%)

Summary of the component maps used to construct the composite map. For each map, progeny size (*n*), map length (cM; total for all 11 linkage groups), mean marker interval (MMI; average for all 11 linkage groups) and total number of mapped markers (using only those linkage groups included in composite map construction; see Methods) are given. For DArT, microsatellite (SSR) and gene markers mapped on each component map, the percentage of markers unique to that map (*i.e.* not mapped in any of the six other component maps) are given in parentheses. [a]Cross details and reference; [b]Kullan *et al.* [33], [c]Petroli *et al.* [11] and [d]Hudson *et al.* [32]. [e]Data for the *E. grandis* × *E. urophylla* Embrapa F_2 component map calculated using a combination of framework and comprehensive linkage groups (see Methods).

Composite map construction

The composite linkage map was constructed at Diversity Arrays Technology (DArT) Pty Ltd (Canberra, Australia) using specially developed R scripts which merged component map markers into the composite map based on their relative map positions. The *E. grandis* × *E. urophylla* SA F_2 (GU-SA) linkage map was used as the seed-map (*i.e.* the 'fixed backbone' to which markers from other component maps were added) due to it having the largest progeny size, the largest number of both mapped and unique markers (Table 1) and high overall marker colinearity to the 11 main superscaffolds of the assembled *E. grandis* genome sequence [33,36]. The procedure for building each composite map linkage group was as follows. Firstly, the number of common markers in each seed-map – component map linkage group comparison was identified. Spearman rank marker-order correlations were then estimated and a heuristic 'fit value' for each comparison was calculated as; Fit value = correlation × log (number of common markers); where the second term rewards for the number of common markers with a diminishing returns function. Following selection of the component map linkage group with the highest Fit value, unique markers (*i.e.* those not mapped on the seed linkage group, or the 'building' composite linkage group in following rounds) were added to the seed linkage group (or 'building' composite map linkage group) using linear regression. Here, the slope (m) and intersect (c) calculated from fitting the positions of common markers on the seed linkage group (pc) to their positions on the selected component map (pi) linkage group ($pc = m \times pi + c$) was used to calculate the positions of unique component map markers added to the seed linkage group. Once this first round was completed, the remaining component linkage groups were compared to this new 'building' composite map linkage group and the process was repeated. This continued until all unique markers had been added from remaining component maps which shared at least three common markers with the building composite map linkage group and had a marker-order correlation coefficient ≥ 0.50. This process was repeated for each linkage group to yield the final composite map of 11 linkage groups. Markers which mapped to the distal ends of composite linkage groups and which had relatively large inter-marker intervals (≥ 5 cM) and poor support (e.g. mapped in one component map only) were removed. The numbering and orientation of linkage groups followed the convention established in Brondani *et al.* [23]; this also corresponds to the numbering of pseudochromosome assemblies in the *E. grandis* genome sequence [36].

Composite map features

Following composite map construction, marker-order correlations between composite and component map linkage groups were calculated in SAS 9.2 (SAS Institute, Cary, USA) using the PROC CORR Spearman function. To test whether multicopy markers were distributed equally across linkage groups, a χ^2 test was used to compare the observed versus expected number of multicopy marker positions occurring on each linkage group. The expected number of multicopy markers per linkage group was calculated as; (total number of multicopy marker positions in the composite map/total number of DArT marker positions in the composite map) × number of DArT marker positions per linkage group for that linkage group. The BLAST server available at Phytozome [36] was used to search for DArT marker duplications. The bl2seq tool at NCBI [50] was used to examine DArT marker sequence similarity/redundancy. All graphical representations of linkage maps were drawn using MapChart [49].

Results

Composite map details

A total of 4101 individual markers, comprising 3880 DArT markers, eight gene-based markers and 213 microsatellite markers were included in the composite map. The composite map totalled 1107 cM which was within the range of component map lengths (1033–1258 cM; Table 1) and contained only eleven marker intervals ≥ 3 cM; with a maximum marker interval of 5.9 cM. The composite map contained 81 multicopy DArT markers (2.1% of total DArT markers) which were mapped to 171 map positions. Most multicopy markers occurred on two linkage groups only, however, one marker (ePt-574238) mapped to three linkage groups while four markers (ePt-503174, ePt-568818, ePt-637610, ePt-637861) mapped to four linkage groups. This resulted in the 4101 markers being mapped to 4191 positions (Table 2). Over half (2171 or 53%) of the markers mapped to these 4191 map positions had been mapped in a single component map only (*i.e.* were not shared among multiple component maps). Approximately 13% of DArT markers mapped to identical positions in the composite map. Therefore, the map contained 3634 unique map loci with an average interval of 0.31 cM.

The number of multicopy DArT marker positions on each linkage group ranged from 5 to 24 and represented 1.9-6.4% of the total number of DArT markers mapped per linkage group (Table 2). Although LG5 and LG7 contained a larger proportion of multicopy DArT marker positions (e.g. LG1 contained only 5 multicopy DArT marker positions, or 1.9% of the total number of DArT marker positions; Table 3), the proportion of multicopy DArT marker positions found on each linkage

Table 2 Composite map summary

LG	cM	Markers mapped				Average marker interval (cM)	MC DArT pos.[a]
		DArT	SSR	Genes	Total		
1	93.8	250	12	0	262	0.42	5
2	102.1	451	29	0	480	0.24	18
3	105.6	429	18	2	449	0.28	21
4	80.9	219	9	3	231	0.41	12
5	95.9	366	8	0	374	0.30	24
6	125.3	408	43	1	452	0.31	15
7	87.7	305	9	1	315	0.33	18
8	137.3	540	26	0	566	0.28	19
9	82.9	312	20	0	332	0.29	10
10	97.8	336	20	1	357	0.30	12
11	97.3	354	19	0	373	0.31	17
Total	1106.5	3970	213	8	4191	0.31	*171*[b]

Summary of the composite map: including the number of mapped markers, length and average marker intervals by linkage group (LG). [a]MC DArT pos. - indicates the number of multicopy (MC) DArT marker positions (pos.) occurring on each linkage group. [b]The 171 multicopy DArT marker positions represent 81 multicopy DArT markers (see Additional file 1).

group did not significantly differ from that expected by chance across all linkage groups ($\chi^2 = 12.99$, $P = 0.22$, $df = 10$). There was no trend within linkage groups for multicopy DArT markers to be clumped in either distal or central linkage group areas (data not shown). Composite map marker details, component map(s) marker origins and multicopy DArT marker information is presented in Additional file 1.

Composite – component map colinearity

Colinearity between component and composite map linkage groups can be viewed graphically in Figure 1 (for the GLOB-LH map) and in Additional file 2 (all component maps). Pair-wise linkage group marker-order correlations were generally high (greater than 0.90; Table 3) reflecting the high colinearity shown between common markers (Figure 1 and Additional file 2). However, a small degree of non-colinearity did occur between all component maps and the composite map. Eleven component map linkage groups had marker-order correlations of less than 0.90 (Table 3), however, these linkage groups were either, (1) identified as having poor marker colinearity with other component maps prior to composite map construction and excluded from analysis (five linkage groups with gray shading in Table 3), or (2) marker-order information from these linkage groups was not incorporated during composite map construction (correlation value without asterisk; six linkage groups Table 3) due to markers from these maps being previously added from other linkage groups having better fit values. Thus, these poorly correlated linkage groups did not adversely affect the composite map marker-order. For each linkage group, the average pair-wise marker-order correlation between the

Table 3 Composite – component map marker-order correlation coefficients

Composite map LG	Component map							
	GLOB-LH	GU-Emb	GLOB-F_2-1	GLOB-F_1-1	GLOB-F_1-4	GLOB-F_1-5	Average[a]	Average[b]
1	0.98*	0.56^{F}	0.99*	0.99*	0.98*	0.95*	0.91	0.98
2	0.98*	0.95^{C}*	0.93*	0.95*	0.98*	0.85	0.94	0.96
3	0.91*	0.99^{F}*	0.97*	0.99*	0.97*	0.99*	0.97	0.97
4	0.98*	0.74^{C}	0.96*	0.89	0.79^{ex}	$0.19^{ns,ex}$	0.76	0.97
5	0.99*	0.92^{F}	0.96*	0.96*	0.99*	0.96*	0.96	0.97
6	0.99*	0.99^{C}*	0.99*	0.63^{ex}	0.99*	0.86	0.91	0.99
7	0.98*	0.65^{F}	0.99*	0.98*	0.96*	0.91*	0.91	0.96
8	0.98*	0.99^{C}*	0.99*	0.66	0.94*	0.99*	0.93	0.98
9	0.99*	0.97^{F}*	0.98*	0.97*	0.65^{ex}	0.97	0.92	0.98
10	0.95*	0.98^{C}*	0.97*	$0.35^{ns,ex}$	0.92	0.97*	0.86	0.97
11	0.99*	0.99^{C}*	0.99*	0.97*	0.96	0.99*	0.98	0.99
Average[c]	0.98	0.88	0.97	0.85	0.92	0.87		
Average[d]	0.98	0.98	0.97	0.97	0.97	0.96		

Marker-order correlations between composite map and component map linkage groups; the GU-SA component map is not shown as this map was used as the seed-map (*i.e.* provided a fixed-order and all correlations were 1.0). For the GU-Emb component map, superscript letters indicates whether the framework (F) or comprehensive (C) linkage group was used in map construction. Component map linkage groups initially excluded from composite map construction are indicated by [ex] superscript. An asterisk following the correlation value indicates that marker-order information from the component map was incorporated during construction of the composite map linkage group. Apart from two correlations (indicated by [ns] superscript) all correlations were significant at $\alpha \leq 0.05$. Averages: [a]calculated using all six component maps, [b]calculated using only those linkage groups included in composite map construction (marked with an asterisk), [c]calculated using all 11 linkage groups, [d]calculated using only those linkage groups included in composite map construction (marked with an asterisk).

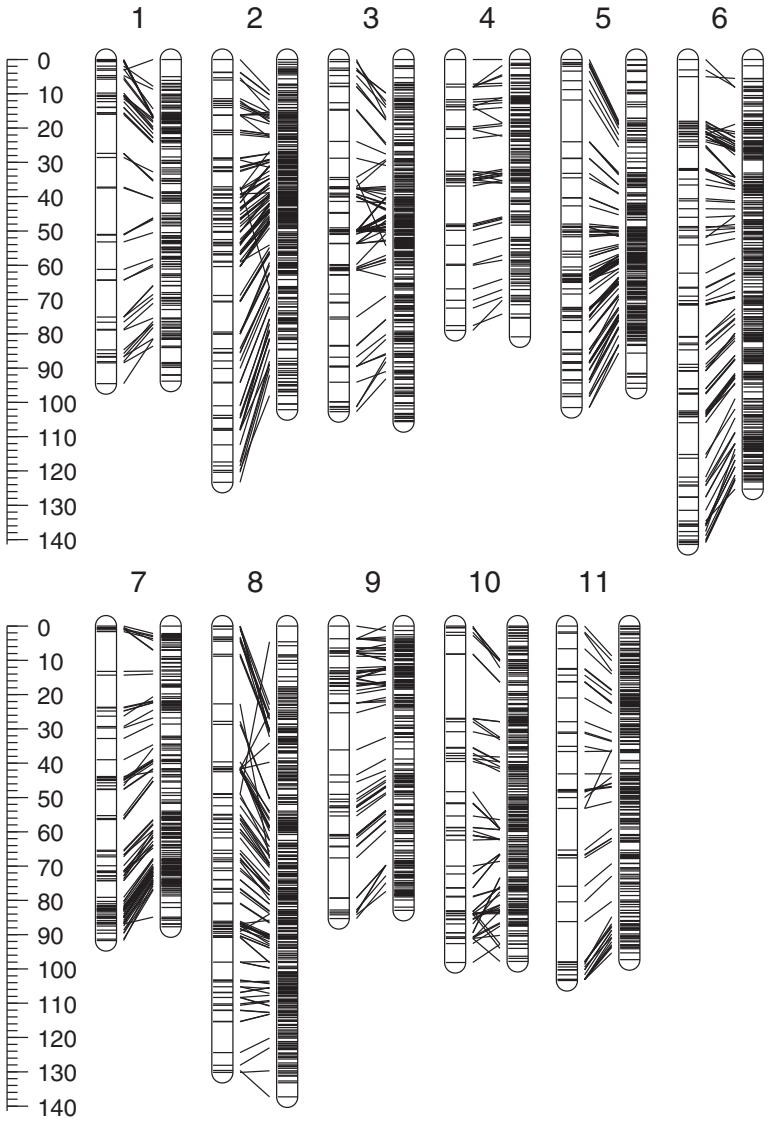


Figure 1 Marker colinearity between the GLOB-LH component map (left) and composite map (right). Lines between each homologous linkage group pair indicate the positions of common markers. The scale bar is in Kosambi's centiMorgans.

composite map and those component maps included in map construction ranged from 0.96 to 0.99 (Average[b] column; Table 3).

DArT marker duplications

Although not a main focus of this study, evidence for the occurrence of duplicated DArT marker loci within the assembled *E. grandis* genome sequence [36] was investigated for the five multicopy markers which had been mapped to three or more linkage groups. Two of these markers (ePt-637610 and ePt-637861; see Additional file 1) mapped to the same map position on each of four linkage groups (LGs 2, 3, 5 and 8) and were found to be redundant markers (*i.e.* identical sequences) based on their marker sequence similarity (bl2seq: 583/606 base-pair similarity, e-value: 0.0). For the four unique multicopy markers, three were detected to have loci duplications within the *E. grandis* genome sequence. In each case, the positions of duplicated loci detected in the *E. grandis* genome sequence corresponded to the linkage groups to which the marker was mapped.

Discussion

Composite map construction

Data from seven component maps were integrated into a single composite map which represents the highest density map yet produced in *Eucalyptus*. A major advantage of the marker-merger method used in this study was the substantial time and labour savings made when compared to the effort required to produce comparable

maps using traditional, segregation-based methods. For example, Li *et al.* [40] constructed a 2111 marker composite map from four barley mapping pedigrees and reported that it took '*several thousand hours*' of computing time. In a larger barley study, Wenzl *et al.* [41] produced a 2935 loci composite map from ten mapping populations using JoinMap 3.0 [51] in combination with specially built Perl scripts and reported that the project required several months of semi-manual data processing [41]. In contrast, the composite map produced in this study was built in a single day.

Utility of the composite map

As sequences are available for the majority of DArT markers on the map (91%; data not shown), the composite map provides a direct link to the *E. grandis* genome sequence [36]. We have made use of this link to search the *E. grandis* genome sequence for candidate genes associated with QTL locations and to facilitate the placement of candidate genes in the component linkage maps without the need for time consuming marker development and genotyping. Sequence-based linkage maps have also provided useful tools to aid in the assembly of genome sequences [e.g. 52,53] and can be particularly beneficial in taxa (such as eucalypts) which have a relatively small genome size. For example, during the assembly of the *E. grandis* genome sequence, a DArT linkage map was valuable in guiding contigs into the 11 main pseudochromosomes [16]. However, not all contigs could be aligned and approximately 12% of the 693 Mbp *E. grandis* genome sequence remains unassembled in more than 4900 small unlinked scaffolds [54]. With the composite map containing many more DArT markers (1600+) than the linkage map used to aid genome assembly, the composite map markers may provide further positional information and help to anchor some of the unlinked scaffolds and refine the current *E. grandis* genome sequence.

Over half (53%) of the markers placed in the composite map originated from a single component map (*i.e.* were not shared among multiple component maps). Therefore, the ability to determine the relative positions of markers mapped in different maps has been greatly enhanced through the integration of this data into a single map. This has already proven advantageous to our research group, with the composite map being used to quickly identify the linkage relationships of microsatellite markers used in population genetic studies. Although now a relatively simple task, it was previously necessary to consult multiple linkage maps and assess their colinearity to obtain this same information. Furthermore, any marker developed in eucalypts which has known sequence, can now potentially be found in the eucalypt genome sequence and then aligned against the reference map in order to estimate its distance to other markers in units of recombination (cM); which are evolutionary meaningful units compared to base pair distances. Additionally, it is also important to understand the relationship between physical map (*i.e.* genome) and genetic map distances as this can have implications for map-based cloning efforts and/or marker-assisted selection. For example, uneven recombination rates across a genome [12,55] may result in physically distant markers appearing to be genetically close to each other, or vice versa. In eucalypts, Kullan *et al.* [33] recently compared 153 linkage map intervals of approximately 1 cM against contigs of the *E. grandis* genome and found that the genetic map to physical distance relationship varied considerably; ranging from 100 kb to 2.4 Mbp per 1 cM. Therefore, the composite map will be useful to provide further insight into the relationship between physical and genetic map distance in addition to identifying hot (or cold) spots of recombination.

A key use of the composite map will be for comparison of QTL and candidate gene positions detected across variable genetic backgrounds and/or environments in different studies. This has previously been limited due to a lack of common markers being shared between maps [23]. For example, Thumma *et al.* [27] detected multiple co-locating growth-related QTL on LG5 in *E. nitens* but could not accurately compare the position of this QTL to similar growth-related trait QTL detected on this same linkage group in two other studies [24,56]. Although most of the markers contained on the composite map are DArT markers, which to date have only been mapped in the pedigrees included in this study, the map does contain several hundred microsatellite markers (213) which will enable synteny and colinearity to be established with many earlier linkage maps used for QTL detection; e.g. 13 out of 22 earlier studies have mapped a variable number of microsatellites [16]. This will enable QTL to be aligned against the composite map which may provide deeper insight into the genetic control of phenotypic traits in the genus. For example, following the construction of an integrated map for melon (*Cucumis melo*) which used data from eight independent mapping experiments, it was possible to align 370 QTL detected for 62 traits from 18 experiments [57]. Through this alignment, QTL detected in different studies for economically important traits were found to co-locate [57]; providing supporting evidence to substantiate the biological basis of the observed marker-trait association [7,8].

As in all linkage mapping studies, it is important to consider both the quality of the map produced and any specific map characteristics. In the alignment of 6480 DArT marker sequences against the *E. grandis* genome sequence [36], Petroli *et al.* [11] reported that although the majority of markers (4189) occurred at a single genome position with high support, many marker sequences (2291), albeit at lower confidence, also exhibited similarity to a second

genome position and that about half of these genome regions contained repeat elements. Furthermore, preliminary analysis of the *E. grandis* genome sequence suggests that (as has been observed in some *Rosid* genomes) a whole-genome duplication event has occurred in the lineage (*Myrtales*) subsequent to the ancient hexaploidy event shared by all rosids (Myburg *et al*., unpublished). Such whole-genome, as well as, segmental duplication events will affect thousands of marker loci, but most would be expected to diverge in sequence with evolutionary time yielding mostly unique marker loci. Thus, the presence of multicopy markers (representing putatively duplicated loci) in the composite map was not unexpected. It is worth noting that in the construction of each component map, only those markers which segregated as a single Mendelian locus were mapped. Therefore, in the event of a marker duplication being present within a pedigree, only one locus could be polymorphic in order for that marker to produce a single loci segregation ratios. Consequently, it is likely that only a subset of the duplicated loci present within the eucalypt genome have been identified in the composite map. Given that the *Pst*I enzyme used in the complexity reduction step of DArT marker development [35] preferentially produces markers located in hypomethylated, gene rich regions [55], and that many DArT markers contain protein coding sequences [33], it is possible that some of the multicopy markers identified may be associated with different gene family members and/or be part of larger duplicated regions. Further studies are required to examine the full extent and evolution of the duplicated loci. We also expected some marker redundancy (markers with the same sequence) among the 3808 composite map DArT markers; an issue which arises due to the process by which DArT markers are generated, resulting in the same amplified genomic fragment being represented more than once on the genotyping array [31,35]. Therefore, identical clones (e.g. the same DArT fragment, but with different DArT marker names) are expected to produce identical genotype scores and should map to identical (or near identical) map positions; as found for the markers ePt-637610 and ePt-637861 identified as identical clones in this study.

The marker-merging method used in this study took advantage of the fact that individual component maps were constructed using high marker-ordering stringency which resulted in linkage maps having robust marker-orders [32]. The comparison of the composite map marker-order against individual component maps gives an indication of the quality of the composite map. Marker-order correlations were mostly excellent with high pair-wise linkage group marker-order correlations found in most comparisons. For example, in 48 out of 66 pair-wise comparisons the marker-order correlation exceeded 0.95. Despite these high correlations, most component maps did exhibit some marker-order inconsistencies with the composite map. A number of (mostly) single marker-order inconsistencies did occur over large distances, but most marker-order disagreements occurred among tightly grouped markers in regions of less than 5 cM. Although it is possible that some of these marker-order differences could be real and represent local chromosomal rearrangements or marker duplications between the different mapping pedigrees and/or species, they are more likely to reflect marker-order inaccuracies within any of the component maps or simply be artefacts of the statistical uncertainty associated with ordering tightly linked markers [see 58]. While users of this map should be aware of these limitations and how they may affect marker ordering, overall, the generally high marker-order correlations observed and the exclusion of component map linkage groups having poor marker colinearity from initial composite map construction (and thus not adversely affecting composite map marker-order) suggests that the composite map is of a sufficiently high quality to facilitate the transfer of genetic information between studies.

The composite map will be most useful for studies involving species from subgenus *Symphyomyrtus* sections *Latoangulatae* and *Maidenaria*; due to the composite map being built from linkage maps constructed in species from these sections. However, due to the high level of genome synteny and colinearity detected between species from these relatively distant sections [28,32,34], information from the composite map should also be applicable to many other commercially important eucalypt species in closely related sections (e.g. *E. camaldulensis* from subgenus *Symphyomyrtus* section *Exsertaria*).

Future marker integration

A number of recent studies have focussed on the development of molecular markers for use in eucalypts. In addition to the DArT genotyping array developed for use in eucalypts [35], the feasibility of high-throughput SNP genotyping has been explored [59] and several tens of species-transferrable EST-based SSR markers have been recently reported [60,61]. Furthermore, DArT genotyping by sequencing (GBS), which combines the complexity reduction method of DArT [31] with next generation sequencing (NGS), and which can potentially deliver up to three-fold as many markers as conventional DArT genotyping [see 62] is becoming a cost-competitive genotyping option due to the recent plummeting costs of NGS sequencing. Therefore, to broaden the use of the composite map for comparative analyses and to optimise its' worth, it will be necessary to add new markers to the current version of the composite map in the future. Although beyond the scope of this study, it would also be valuable to compare the marker order of the composite map to maps built

using the same data with other marker-merging software (e.g. BioMercator [63], CarthaGene [64] or MergeMap [65]). The R scripts and map marker positions of the component maps used in this study can be made available upon request.

Conclusion

The integration of markers from seven individual genetic linkage pedigrees has resulted in a composite, reference map for eucalypts with 4101 DArT and microsatellite markers. Although some small marker-order inconsistencies exist between component maps and the composite map, there is a relatively high agreement of marker-order between component maps; which indicates that the composite map represents a good estimation of the true marker positions in most cases. However, at finer scales (sub-cM) marker-orders may differ between component and composite maps due to limited statistical power to order such tightly linked markers. Overall, the genome coverage and marker density of the composite map greatly exceeded that achieved in any of the single mapping populations. It is expected that this composite map will provide a valuable reference map for the world-wide *Eucalyptus* research community, facilitate the transfer of genetic information between different studies and allow for the integration of DArT marker information with other genomic resources.

Additional files

Additional file 1: *Eucalyptus* composite map details. Details of markers mapped in the *Eucalyptus* composite map. Includes, linkage group and position of mapped markers, marker type and which component map(s) markers were mapped. A '1' in the 'Multicopy marker' column indicates that the marker occurs on two or more linkage groups.

Additional file 2: Composite map – component map marker colinearity. Marker colinearity among all six component maps and the *Eucalyptus* composite map. For each linkage group, three linkage group 'triplets' show marker colinearity between two component maps (outside) and the composite map (centre). Horizontal lines on linkage group bars indicate marker positions and lines between linkage groups indicate the position of common markers. The scale bar shown is in Kosambi's centiMorgans. Component map names (abbreviations; see Table 1) are given above each linkage group. Linkage groups excluded from composite map construction are indicated in parentheses following the component map name. An asterisk indicates whether marker-order information from the component map was incorporated during composite map construction (see Methods). For the GU-Emb component map, superscript letters indicates whether the framework (f) or comprehensive (c) linkage group from this pedigree was used in composite map construction.

Competing interest

'The authors declare that they have no competing interests.

Authors' contributions

CJH built the GLOB-LH linkage map, coordinated the collection of map data, building of the composite map, performed analyses and wrote the manuscript. JSF built all other *E. globulus* linkage maps. AAM and ARKK, DG and CDP, built and contributed GU-SA and GU-Emb linkage maps, respectively. AK and FD constructed the composite map. REV and AAM conceived the study, and along with BMP and JSF, contributed to the design of the study. All authors read and approved the final manuscript.

Funding

Funding for this project was provided by the Australian Research Council (DP0770506 & DP110101621) as well as the Cooperative Research Centre for Forestry (Australia). Construction of the GU-SA map was supported by Sappi, Mondi, the Technology and Human Resources for Industry Program (THRIP), the National Research Foundation (NRF) and the Department of Science and Technology (DST) of South Africa.

Author details

[1]School of Plant Science and CRC for Forestry, University of Tasmania, Private Bag 55 Hobart, Tasmania 7001, Australia. [2]CRN Research Fellow, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Locked Bag 4, Maroochydore, QLD 4558, Australia. [3]Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa. [4]EMBRAPA Genetic Resources and Biotechnology - EPqB Final W5 Norte 70770–917 Brazilia DF and Dep. Cell Biology, Universidade de Brazilia – UnB, Brasilia, DF, Brazil. [5]Diversity Arrays Technology Pty Ltd, PO Box 7141, Yarralumla, ACT 2600, Australia. [6]EMBRAPA Genetic Resources and Biotechnology - Parque Estação Biológica - PqEB - Av. W5 Norte (final), Brasília, DF - Brazil - 70770–917, Universidade Catolica de Brasília- SGAN, 916 modulo B, 70790-160DF, Brasilia, Brazil.

Received: 23 March 2012 Accepted: 4 June 2012
Published: 15 June 2012

References

1. Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB: **Comparative mapping in the *Pinaceae*.** *Genetics* 2004, **168:**447–461.
2. Lefebvre-Pautigny F, Wu F, Philippot M, Rigoreau M, Zouine M, Frasse P, Bouzayen M, Broun P, Pétiard V, *et al*: **High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes.** *Tree Genetics & Genomes* 2009, **6:**565–577.
3. Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C-X, Katsar CS, Lan T-H, Lin Y-R, Ming R, Wright RJ: **Comparative genomics of plant chromosomes.** *Plant Cell* 2000, **12:**1523–1540.
4. Anderson JT, Lee C-R, Mitchell-Olds T: **Life-history QTLs and natural selection on flowering time in *Boechera stricta*, a perennial relative of *Arabidopsis*.** *Evolution* 2011, **65:**771–787.
5. Freeman JS, O'Reilly-Wapstra JM, Vaillancourt RE, Wiggins N, Potts BM: **Quantitative trait loci for key defensive compounds affecting herbivory of eucalypts in Australia.** *New Phytologist* 2008, **178:**846–851.
6. Kearsey MJ, Farquhar AGL: **QTL analysis in plants; where are we now?** *Heredity* 1998, **80:**137–142.
7. Brown GR, Bassoni DL, Gill GP, Fontana JR, Wheeler NC, Megraw RA, Davis MF, Sewell MM, Tuskan GA, Neale DB: **Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping.** *Genetics* 2003, **164:**1537–1546.
8. Wheeler NC, Jermstad KD, Krutovsky K, Aitken SN, Howe GT, Krakowski J, Neale DB: **Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping.** *Molecular Breeding* 2005, **15:**145–156.
9. Semagn K, Bjornstad A, Ndjiondjop MN: **Principles, requirements and prospects of genetic mapping in plants.** *African Journal of Biotechnology* 2006, **5:**2569–2587.
10. The Potato Genome Sequencing Consortium: **Genome sequence and analysis of the tuber crop potato.** *Nature* 2011, **475:**189–195.
11. Petroli C, Sansaloni C, Carling J, Mamani E, Steane D, Myburg A, Vaillancourt R, Kilian A, Pappas G, Bonfim da Silva O, Grattapaglia D: **Genomic characterization, high-density mapping and anchoring of DArT markers to the reference genome of *Eucalyptus*.** *BMC Proceedings* 2011, **5(Suppl 7):**P35.
12. Jones N, Ougham H, Thomas H, Pašakinskienė I: **Markers and mapping revisited: finding your gene.** *New Phytologist* 2009, **183:**935–966.
13. Grattapaglia D, Plomion C, Kirst M, Sederoff RR: **Genomics of growth traits in forest trees.** *Current Opinion in Plant Biology* 2009, **12:**148–156.
14. Poke F, Vaillancourt RE, Potts B, Reid J: **Genomic research in *Eucalyptus*.** *Genetica* 2005, **125:**79–101.

15. Grattapaglia D, Sederoff R: **Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers.** *Genetics* 1994, **137:**1121–1137.
16. Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Kulheim C, Potts BM, Myburg AA: **Progress in *Myrtaceae* genomics: *Eucalyptus* as the pivotal genus.** *Tree Genetics & Genomes*, **8:**463–508. in press.
17. Byrne M, Murrell JC, Allen B, Moran GF: **An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers.** *Theoretical and Applied Genetics* 1995, **91:**869–875.
18. Gan S, Shi J, Li M, Wu K, Wu J, Bai J: **Moderate-density molecular maps of Eucalyptus urophylla (S. T. Blake) and E. tereticornis (Smith) genomes based on RAPD markers.** *Genetica* 2003, **118:**59–67.
19. Marques CM, Araújo JA, Ferreira JG, Whetten R, O'Malley DM, Liu BH, Sederoff R: **AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*.** *Theoretical and Applied Genetics* 1998, **96:**727–737.
20. Agrama HA, Salah SF: **Construction of a genome map for *Eucalyptus camaldulensis* DEHN.** *Silvae Genetica* 2002, **51:**201–206.
21. Shepherd M, Kasem S, Lee D, Henry R: **Construction of microsatellite linkage maps for *Corymbia*.** *Silvae Genetica* 2006, **55:**228–238.
22. Myburg AA, Potts BM, Marques CM, Kirst M, Gion J, Grattapaglia D, Grima-Pettenatti J: **Eucalypts**. In *Genome mapping and molecular breeding in plants. Volume 7*. Edited by Kole C. Berlin: Springer; 2007:115–160.
23. Brondani RPV, Williams ER, Brondani C, Grattapaglia D: **A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus.** *BMC Plant Biology* 2006, **6:**20.
24. Freeman J, Whittock S, Potts B, Vaillancourt R: **QTL influencing growth and wood properties in *Eucalyptus globulus*.** *Tree Genetics & Genomes* 2009, **5:**713–722.
25. Gion J-M, Carouche A, Deweer S, Bedon F, Pichavant F, Charpentier J-P, Bailleres H, Rozenberg P, Carocha V, Ognouabi N, *et al*: **Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*.** *BMC Genomics* 2011, **12:**301.
26. Marques C, Brondani R, Grattapaglia D, Sederoff R: **Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species.** *Theoretical and Applied Genetics* 2002, **105:**474–478.
27. Thumma B, Baltunis B, Bell J, Emebiri L, Moran G, Southerton S: **Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families.** *Tree Genetics & Genomes* 2010, **6:**877–889.
28. Myburg AA, Griffin AR, Sederoff RR, Whetten RW: **Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F_1 hybrid based on a double pseudo-backcross mapping approach.** *Theoretical and Applied Genetics* 2003, **107:**1028–1042.
29. Appleby N, Edwards D, Batley J: **New technologies for ultra-high throughput genotyping in plants**. In *Methods in Molecular Biology, Plant genomics. Volume 153*. Edited by Somers DJ, Langridge P, Gustafson JP. New York: Humana Press; 2009:19–39.
30. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML: **Genome-wide genetic marker discovery and genotyping using next-generation sequencing.** *Nat Rev Genet* 2011, **12:**499–510.
31. Jaccoud D, Peng K, Feinstein D, Kilian A: **Diversity arrays: a solid state technology for sequence information independent genotyping.** *Nucleic Acids Research* 2001, **29**(4):e25.
32. Hudson CJ, Kullan ARK, Freeman JS, Faria D, Grattapaglia D, Kilian A, Myburg A, Potts BM, Vaillancourt RE: **High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping.** *Tree Genetics & Genomes* 2012, **8:**339–352.
33. Kullan A, van Dyk M, Jones N, Kanzler A, Bayley A, Myburg A: **High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F_2 pseudo-backcross of *Eucalyptus grandis* x *E. urophylla*.** *Tree Genetics & Genomes* 2012, **8:**163–175.
34. Steane DA, Nicolle D, Sansaloni CP, Petroli CD, Carling J, Kilian A, Myburg AA, Grattapaglia D, Vaillancourt RE: **Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (*Myrtaceae*) using high-throughput, genome-wide genotyping.** *Molecular Phylogenetics and Evolution* 2011, **59:**206–224.
35. Sansaloni C, Petroli C, Carling J, Hudson C, Steane D, Myburg A, Grattapaglia D, Vaillancourt R, Kilian A: **A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*.** *Plant Methods* 2010, **6:**16.
36. *Eucalyptus grandis genome (JGI v1.0)*. http://www.phytozome.net/cgi-bin/gbrowse/eucalyptus/.
37. Van Ooijen J: *JoinMap 4, software for the calculation of genetic linkage maps in experimental populations*. Wageningen, Netherlands: Kyazma B.V; 2006.
38. Grattapaglia D: **Molecular breeding of *Eucalyptus***. In *In Molecular biology of woody plants Volume 1*. Edited by Jain S, Minocha S. Netherlands: Kluwer; 2000:451–474.
39. Cheema J, Dicks J: **Computational approaches and software tools for genetic linkage map estimation in plants.** *Briefings in Bioinformatics* 2009, **10:**595–608.
40. Li H, Kilian A, Zhou M, Wenzl P, Huttner E, Mendham N, McIntyre L, Vaillancourt R: **Construction of a high-density composite map and comparative mapping of segregation distortion regions (SDRs) in barley.** *Molecular Genetics and Genomics* 2010, **284:**319–331.
41. Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, *et al*: **A high-density consensus map of barley linking DArT markers to SSR. RFLP and STS loci and agricultural traits.** *BMC Genomics* 2006, **7:**206.
42. Cone KC, McMullen MD, Bi IV, Davis GL, Yim Y-S, Gardiner JM, Polacco ML, Sanchez-Villeda H, Fang Z, Schroeder SG, *et al*: **Genetic, physical, and informatics resources for *Maize*. On the road to an integrated map.** *Plant Physiology* 2002, **130:**1598–1605.
43. Mace E, Rami J-F, Bouchet S, Klein P, Klein R, Kilian A, Wenzl P, Xia L, Halloran K, Jordan D: **A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers.** *BMC Plant Biology* 2009, **9:**13.
44. Alsop B, Farre A, Wenzl P, Wang J, Zhou M, Romagosa I, Kilian A, Steffenson B: **Development of wild barley-derived DArT markers and their integration into a barley consensus map.** *Molecular Breeding* 2011, **27:**77–92.
45. Varshney R, Marcel T, Ramsay L, Russell J, Röder M, Stein N, Waugh R, Langridge P, Niks R, Graner A: **A high density barley microsatellite consensus map with 775 SSR loci.** *Theoretical and Applied Genetics* 2007, **114:**1091–1103.
46. Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S, *et al*: **Physical and genetic structure of the *Maize* genome reflects its complex evolutionary history.** *PLoS Genet* 2007, **3:**e123.
47. Stam P: **Construction of integrated genetic linkage maps by means of a new computer package: JoinMap.** *Plant J* 1993, **3:**739–744.
48. Bachir O, Abdellah B: **Chromosome numbers of the 59 species of *Eucalyptus* L'Herit (*Myrtaceae*).** *Caryologia* 2006, **59:**207–212.
49. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs.** *Journal of Heredity* 2002, **93:**77–78.
50. NCBI: *NCBI BLAST*. http://blast.ncbi.nlm.nih.gov/.
51. Van Ooijen J, Voorrips RE: **JoinMap 3.0, software for the calculation of genetic linkage maps**. In *Plant Research International*. Wageningen, Netherlands; 2001.
52. Wang Y, Sun S, Liu B, Wang H, Deng J, Liao Y, Wang Q, Cheng F, Wang X, Wu J: **A sequence-based genetic linkage map as a reference for *Brassica rapa* pseudochromosome assembly.** *BMC Genomics* 2011, **12:**239.
53. Hwang T-Y, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, Hisano H, Sasamoto S, Sato S, Tabata S, *et al*: **High-density integrated linkage map based on SSR markers in soybean.** *DNA Research* 2009, **16:**213–225.
54. *Eucagen: Early release of the E. grandis genome sequence*. http://web.up.ac.za/eucagen/.
55. van Os H, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, de Jong W, van Koert P, *et al*: **Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map.** *Genetics* 2006, **173:**1075–1087.
56. Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR: **Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers.** *Genetics* 1996, **144:**1205–1214.
57. Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, Staub J, Zalapa J, Cuevas H, Dace G, *et al*: **A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (Cucumis melo L.).** *BMC Plant Biology* 2011, **11:**111.
58. Collard B, Mace E, McPhail M, Wenzl P, Cakir M, Fox G, Poulsen D, Jordan D: **How accurate are the marker orders in crop linkage maps generated from large marker datasets?** *Crop and Pasture Science* 2009, **60:**362–372.

59. Grattapaglia D, Silva-Junior O, Kirst M, de Lima B, Faria D, Pappas G: **High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species.** *BMC Plant Biology* 2011, **11**:65.
60. Faria D, Mamani E, Pappas G, Grattapaglia D: **Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests.** *Tree Genetics & Genomes* 2011, **7**:63–77.
61. Acuña C, Fernandez P, Villalba P, García M, Hopp H, Marcucci Poltri S: **Discovery, validation, and in silico functional characterization of EST-SSR markers in *Eucalyptus globulus*.** *Tree Genetics & Genomes* 2012, **8**:289–301.
62. Sansaloni CP, Petroli CD, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A: **Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*.** *BMC Proceedings* 2011, **5**(Suppl 7):P54.
63. Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J: **BioMercator: integrating genetic maps and QTL towards discovery of candidate genes.** *Bioinformatics* 2004, **20**:2324–2326.
64. de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T: **CarthaGene: multipopulation integrated genetic and radiation hybrid mapping.** *Bioinformatics* 2005, **21**:1703–1704.
65. *MergeMap*. http://alumni.cs.ucr.edu/~yonghui/mgmap.html.

doi:10.1186/1471-2164-13-240
Cite this article as: Hudson *et al.*: A reference linkage map for *Eucalyptus*. *BMC Genomics* 2012 **13**:240.

Sansaloni et al. BMC Proceedings 2011, 5(Suppl 7):P37
http://www.biomedcentral.com/1753-6561/5/S7/P37

BMC Proceedings

POSTER PRESENTATION Open Access

How many genes might underlie QTLs for growth and wood quality traits in *Eucalyptus*?

Carolina Sansaloni[1*], César Petroli[1], Georgios Pappas[2], Orzenil Da Silva[2], Dario Grattapaglia[3]

From IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery
Arraial d'Ajuda, Bahia, Brazil. 26 June - 2 July 2011

Background

QTL mapping is an unbiased approach where the phenotype reveals the location of regulatory genes or genomic regions affecting the trait of interest. The development of transferable molecular markers and the increased use of multiple pedigrees for QTL mapping have allowed comparative analysis of QTLs across independent studies thus providing validation data. Such QTL positional information, together with the availability of annotated genome sequences, now promises to identify strong candidate genes for a number of traits [1]. As large pedigrees become available and higher resolution mapping with SNPs, DArT and genotype-by-sequencing technologies becomes routine in forest trees, QTL positional information could be an alternative to the current approaches that rely on tentative candidate genes for association genetics studies. Among the several traits for which QTLs have been mapped in forest trees those that display higher heritability, such as wood chemical composition, are more likely to involve candidate genes of stronger effect although recent association studies show that even such genes explain very small proportion of the variation [2]. In this study we used a high-resolution map with over 2,000 Diversity Arrays Technology (DArT) markers to carry out an initial assessment of the number of annotated gene models in the reference genome sequence of Eucalyptus that putatively co-locate with QTLs for growth and wood quality traits.

Methods

A QTL mapping study was carried out with a clonally replicated segregating population of 171 F1 individuals derived from an *E. grandis* x *E. urophylla* cross. Individuals were genotyped with the Eucalyptus DArT microarray described earlier [3]. The DArT marker data were combined with 222 microsatellites and a linkage map for each parent was constructed using JoinMap 3.0 [4]. Six traits were measured: height growth (HG), circumference at breast height (CBH); wood specific gravity (WSG); cellulose pulp yield (%PULP); Total Lignin (TL); syringyl/guaiacyl ratio (S/G). QTL mapping was carried out using QTL Cartographer [5] on the two parental maps separately at 1 cM intervals. Empirical threshold significance levels for QTL detection were determined by 1,000 permutations considering a significance level of 5%. All the segregating DArT and microsatellite markers were mapped onto the 11 pseudo-molecules of the *Eucalyptus grandis* draft genome sequence covering 609 Mbp.

Results

QTL analyses were carried out using framework genetic linkage maps with high likelihood support for order. The maternal map had 825 markers (684 DArTs + 141 SSR) and the paternal map 511 markers (410 DArTs + 101 SSR). A total of 16 QTLs in *E. grandis* and 14 in *E. urophylla* were detected influencing growth and wood quality traits. High and significant positive phenotypic correlations were found between CBH and HG, TL and S/G, and S/G and %PULP. In the maternal *E. grandis* map, five QTLs were identified for TL (Linkage groups (LG) 1, 3, 4, 5 and 8), two QTLs for %PULP (LG 4 and 5), three for S/G (LG 1, 5 and 8), WSG (LG 6, 8 and 10) and HG (LG 1, 2 and 6). In the *E. urophylla* paternal map we detected three QTLs for %PULP (LG 1, 4 and 9), three for HG (LG7, 8 and 10), three for TL (LG3, 4 and 8), two for CBH (LG7 and 10), two for S/G (LG 8 and 9), and one for WSG (LG 8). More than two QTLs were clustered on LG 4, 5 and 8 in *E. grandis* and

* Correspondence: carols@cenargen.embrapa.br
[1]EMBRAPA Genetic Resources and Biotechnology - EPqB Final W5 Norte 70770-917 Brazilia DF and Dep. Cell Biology, Universidade de Brazilia – UnB, Brazil
Full list of author information is available at the end of the article

on LG 4, 8 and 9 in *E. urophylla* suggesting interesting genomic regions to look for candidate genes to be tested in association mapping. Several of these QTLs were syntenic to QTLs found in other studies [6,7] providing some indirect support for their validity. The sequences of DArT and microsatellite markers bracketing QTLs were used to extract the gene models from the Eucalyptus reference genome. A total of 7,125 predicted gene models are found across all maternal QTLs, with an average of 445 genes per QTL. For the paternal QTLs 5,076 gene models exist, with an average of 362 genes per QTL.

Conclusions

As in many other QTL mapping studies in Eucalyptus [6-8] we have identified several QTLs that control a modest proportion of the phenotypic variation for a number of economically important traits. In this first assessment of putative candidate genes co-locating with these QTLs we found thousands of annotated gene models, hundreds of which could be tentatively suggested as being involved in trait variation. Notwithstanding the low mapping resolution provided by the small progeny, this preliminary study shows that tens or hundreds of genes will likely be always found underlying QTLs for such complex traits. Testing and validation of such large numbers of genes will require a gigantic effort. Furthermore a large proportion of the phenotypic variation remains unexplained by the few QTLs mapped. It is therefore questionable from the applied stand point how much useful information this approach will effectively provide for the advancement of association genetics and, for that matter, of breeding practice.

Acknowledgments

This work was supported by the Brazilian Ministry of Science and Technology through CNPq grant 577047/2008-6 and FAP-DF Grant NEXTREE 193.000.570/2009 and EMBRAPA Macroprogram 2 project grant 02.07.01.004

Author details

[1]EMBRAPA Genetic Resources and Biotechnology - EPqB Final W5 Norte 70770-917 Brazilia DF and Dep. Cell Biology, Universidade de Brazilia – UnB, Brazil. [2]EMBRAPA Genetic Resources and Biotechnology - EPqB Final W5 Norte 70770-917 Brazilia DF, Brazil. [3]EMBRAPA Genetic Resources and Biotechnology – Estação Parque Biológico, 70770-910, Brazilia, DF and Genomic Sciences Program - Universidade Católica de Brasília - Brazilia, Brazil.

Published: 13 September 2011

References

1. Price AH: **Believe it or not, QTLs are accurate!** *Trends Plant Sci* 2006, **11(5)**:213-216.
2. Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai CJ, Neale DB: **Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem.** *New Phytol* 2010, **188(2)**:515-532.
3. Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA, Grattapaglia D, Vaillancourt RE, Kilian A: **A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus.** *Plant Methods* 2010, **6**:16.
4. Stam P: **Construction of Integrated Genetic-Linkage Maps by Means of a New Computer Package - Joinmap.** *Plant Journal* 1993, **3(5)**:739-744.
5. Wang S, Basten CJ, Zeng Z-B: **Windows QTL Cartographer 2.5.** North Carolina State University, Raleigh, NC: Department of Statistics; 2011.
6. Thumma BR, Southerton SG, Bell JC, Owen JV, Henery ML, Moran GF: **Quantitative trait locus (QTL) analysis of wood quality traits in.** *Eucalyptus nitens. Tree Genetics & Genomes* 2010, **6(2)**:305-317.
7. Freeman JS, Whittock SP, Potts BM, Vaillancourt RE: **QTL influencing growth and wood properties in Eucalyptus globulus.** *Tree Genetics & Genomes* 2009, **5(4)**:713-722.
8. Grattapaglia D, Kirst M: **Eucalyptus applied genomics: from gene sequences to breeding tools.** *New Phytologist* 2008, **179(4)**:911-929.

doi:10.1186/1753-6561-5-S7-P37
Cite this article as: Sansaloni *et al.*: How many genes might underlie QTLs for growth and wood quality traits in *Eucalyptus*? *BMC Proceedings* 2011 **5**(Suppl 7):P37.

García et al. BMC Proceedings 2011, 5(Suppl 7):P26
http://www.biomedcentral.com/1753-6561/5/S7/P26

BMC Proceedings

POSTER PRESENTATION Open Access

A genetic linkage map for a Full sib population of *Eucalyptus grandis* using SSR, DArT, CG-SSR and EST-SSR markers

Martín García[1*], Pamela Villalba[1], Cintia Acuña[1], Javier Oberschelp[2], Leonel Harrand[2], Mauro Surenciski[2], María Martínez[1], César Petroli[3], Carolina Sansaloni[3], Danielle Faria[3], Dario Grattapaglia[3], Susana Marcucci Poltri[1]

From IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery
Arraial d'Ajuda, Bahia, Brazil. 26 June - 2 July 2011

Background

Eucalypts are the most widely planted hardwood trees in the world, occupying globally more than 18 million hectares, as an important source of carbon neutral renewable energy and raw material for pulp, paper and solid wood. Intensive planting programs of *Eucalyptus grandis* have been carried out in the Argentinian Mesopotamia.

Linkage maps are useful tools for quantitative trait loci (QTL) analyses and detection. Several maps for QTL analyses of growth and wood quality have been developed in this genus [1,2] and most of the *E. grandis* maps have been carried out in interspecific crosses. Improved marker density in genetic maps, high-throughput techniques and transferability across species are key aspects to increase resolution and speed for a variety of genomic applications in *Eucalyptus*.In this context, an important issue in association studies is the selection of appropriate mapped candidate genes that co-localize with QTL of interest.

As part of the Biotech MERCOSUR project (Marcucci et al., this journal), we here report the construction of a genetic linkage map for *E. grandis* in the context of a QTL study of this specie in an effort to understand the molecular basis for quantitative trait variation in wood quality. This map includes Diversity Arrays Technology (DArT) [3], microsatellite (SSR) markers [4], Candidate Genes-SSR (CG-SSR) for wood quality traits and stress responses functions and Expressed Sequence Tag-SSR (EST-SSR) for putative function related to stress responses and other functions (Acuña et al., this journal). These CG-SSR and EST-SSR were not mapped in *Eucalyptus* previously.

Material and methods

Plant material

E. grandis x *E. grandis* (EG-INTA-161 x EG-INTA-152) F1 population of 130 full-sib progeny cloned (3 ramets) and planted in 2007 in Entre Ríos, Argentina, was analyzed.

Genotyping

The parents of the mapping cross were initially screened with: 55 SSR, 12 CG-SSR and 37 EST-SSR markers; these last two classes of markers derive from a broad study (for details see Acuña et. al., this journal). Capillary electrophoresis and fluorescent detection were carried out on an ABI 3130xlGenetic Analyzer. A DArT Microarray of 7,860 clones was screened for useful polymorphic markers.

Linkage and bioinformatic analysis

All loci were tested for goodness of fit to expected Mendelian segregation ratios using Chi-square goodness of fit tests. The assignment of DArT sequence function was performed using the Blast2GO software [http://www.blast2go.org/]. A consensus genetic linkage map was constructed with JoinMap v3.0 [5]. Linkage parameters were set as 10 minimum LOD and 0.4 maximum recombination fractions.

Results and discussion

In this intraspecific *E. grandis* population, 78% of the SSR markers tested could be mapped. Most mapped

* Correspondence: mgarcia@cnia.inta.gov.ar
[1]Instituto Nacional de Tecnología Agropecuaria. Instituto de Biotecnología. De Los Reseros y Dr. Nicolás Repetto s/n°, CP 1686, Hurlingham, Buenos Aires, Argentina
Full list of author information is available at the end of the article

SSR loci were fully informative, segregating in approximate ratios of 1:1:1:1 (either heterozygous in both parents four alleles in total). Seven SSR loci that followed approximate segregation ratios of 1:1 (heterozygous in only one parent) were EMBRA 47,51,101,179,676, 1244,2010.

From the DArT Microarray of 7,860 clones, 31% of the marker were selected because of their high call rate (>0.80) and polymorphism between parents.

A large proportion (1,503/2,381=63%) of the DArT markers displayed a Mendelian behavior indicating that they sample single copy regions and provide markers that can be used for genetic analyses (65% segregating in 1:1 ratio and 35% in 3:1 ratio).

The map was assembled with 1032 markers, including 976 DArT, 43 SSR loci (2-6 per linkage group), seven EST-SSR and six CG-SSR. The resulting integrated map featured the expected 11 major linkage groups, yielding a genome coverage of 1358.4 cM, and an average consecutive intermarker distance of 1.3 cM in accordance to other reports [1,4]. Linkage groups were numbered following the standardized nomenclature for Eucalyptus proposed by Brondani et al [4].

The six Candidate Genes and seven ESTs include enzymes involved in lignin and cell-wall polysaccharide biosynthesis and stress responses genes, while 267 DArT (29.7%) were assigned to a gene ontology (GO) categories and 296 loci (32.9%) had significant matches with the nonredundant protein database using BLASTX. Thus, 25 enzymes in 56 metabolic pathways were represented by at least one sequence with its corresponding EC number.

The inclusion of common previously mapped SSR markers in several different eucalypt species within the subgenus *Symphyomyrtus* (*E. globulus*, *E. camaldulensis*, *E. dunnii*, *E. tereticornis*, and predominantly *E. grandis* and *E. urophylla*) allowed comparison of linkage groups among this *E. grandis* population and other species of the genus. Linkage orders previously reported in *E. globulus*, *E. grandis* and *E. urophylla* were also observed in this intraspecific *E. grandis* population, supporting the developed map.

Conclusions

In this work, 13 new functional SSR and 976 DArT (296 with assigned functions) markers are mapped in an intraspecific *E. grandis* F1 population. This map will be used to locate QTL for wood quality and growth traits in the specie. Also, this map can help to identify candidate genes and regions in the *Eucalyptus* genome useful for fine scale analysis with association studies that are being developed by our group (see Cappa et al., this volume).The putative functional approach combined with the genetic linkage mapping provides an advantage tool for future analysis in locating genes of interest in *Eucalyptus*.

Author details

[1]Instituto Nacional de Tecnología Agropecuaria. Instituto de Biotecnología. De Los Reseros y Dr. Nicolás Repetto s/n°, CP 1686, Hurlingham, Buenos Aires, Argentina. [2]EEA INTA Concordia. Grupo de Mejoramiento Genético Forestal. Estación Yuquerí s/n, CP 3200, Concordia, Entre Ríos, Argentina. [3]Embrapa Recursos Genéticos e Biotecnologia - Parque Estação Biológica - PqEB - Av. W5 Norte (final). Caixa Postal 02372, 70770-970 - Brasília, DF – Brazil.

Published: 13 September 2011

References

1. Petroli CD, Sansaloni CP, Kilian A, Steane DA, Myburg AA, Pappas GJ, Faria DA, Vaillancourt RE, Grattapaglia D: **A high-density sub-centiMorgan integrated DArT/microsatellite genetic linkage map for species of *Eucalyptus* based on 2,980 markers.** *Resumos do 56° Congresso Brazileiro de Genética* 2010 [http://web2.sbg.org.br/congress/sbg2008/pdfs2010/GP159-34030.pdf].
2. Thumma BR, Southerton SG, Bell JC, Owen JV, Henery ML, Moran GF: **Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*.** *Tree Genet Genom* 2010, **6**:305-317.
3. Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA, Grattapaglia D, Vaillancourt RE, Kilian A: **A high high-density Diversity Arrays Technology (DART) microarray for genome genome-wide genotyping in *Eucalyptus*.** *Plant Methods* 2010, **6**:16.
4. Brondani RP, Williams ER, Brondani C, Grattapaglia D: **A microsatellite-consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus.** *BMC Plant Biology* 2006, **6**:20.
5. Stam P: **Construction of integrated genetic linkage maps by means of a new computer package: JoinMap.** *Plant J* 1992, **3(5)**:739-744.

doi:10.1186/1753-6561-5-S7-P26
Cite this article as: García *et al.*: A genetic linkage map for a Full sib population of *Eucalyptus grandis* using SSR, DArT, CG-SSR and EST-SSR markers. *BMC Proceedings* 2011 **5**(Suppl 7):P26.