

Autores

Danielle Assis de Faria Paiva
Doutoranda, PG
Ciências Genômicas e
Biotecnologia, UCB - CNPq

Rinaldo Wellerson Pereira
PG em Ciências
Genômicas e Biotecnologia,
UCB

Dário Grattapaglia
Embrapa Recursos Genéticos
e Biotecnologia

FREQÜÊNCIA DE SNPs E EXTENSÃO DO DESEQUILÍBRIO DE LIGAÇÃO AO LONGO DOS GENES CCR e CAD EM *E. grandis*, *E. globulus* e *E. urophylla*

Resumo

gênero *Eucalyptus* são de extrema importância para a indústria mundial tanto para a produção de papel e celulose, bem como para a produção de energia ou madeira sólida. O Brasil conta com mais de 3,5 milhões de hectares plantados e atualmente é o maior produtor mundial de celulose de fibra curta de eucalipto (cerca de 11 milhões de toneladas por ano) com a maior produtividade média mundial de floresta plantada de eucalipto, ao redor de 40m³ por hectare/ano. Ganhos significativos na produção de biomassa do eucalipto tem resultado dos programas de melhoramento genético. Estratégias de melhoramento tradicional que incorporam ferramentas genômicas tem sido o foco de pesquisas atuais. O conhecimento dos padrões de desequilíbrio de ligação (DL) ou associação não ao acaso entre alelos a genes ou marcadores moleculares ao longo de um determinado genoma é uma variável fundamental para o planejamento de estudos de associação com base na abordagem de varredura genômica (*whole genome scan*) ou de genotipagem de SNPs em genes candidatos. Genomas com valores significativos de DL por longas extensões de seqüência em kb apontam para a primeira possibilidade embora a resolução para o descobrimento de genes seja limitada. Extensões curtas de DL por outro lado apontam para a segunda possibilidade o que levaria à identificação dos SNPs causais ou fisicamente próximos a eles. Em humanos e organismos modelo esta variável já vem sendo estabelecida nos últimos cinco anos para diferentes regiões do genoma. Para espécies do gênero *Eucalyptus*, entretanto, ainda não existe informação sobre os padrões de DL no genoma que possam direcionar para a melhor abordagem a ser adotada em estudos de associação. O objetivo deste estudo foi a identificação de SNPs e o estabelecimento do padrão de DL nos genes Cinamoil-CoA Redutase (CCR) e Cinamil Álcool Dehidrogenase (CAD), dois genes chave no processo de lignificação, em populações naturais de *E. globulus*, *E. grandis* e *E. urophylla*. *Primers* para amplificação de fragmentos de aproximadamente 400 pb sobrepostos ao longo da região genômica dos dois genes foram desenhados utilizando seqüências depositadas em bancos de dados e seqüências genômicas geradas no âmbito do projeto GENOLYPTUS. Em média foi estimada uma diversidade nucleotídica de um SNP a cada 70 pb. Foram identificados SNPs polimórficos entre espécies e fixados dentro das diferentes espécies. Em termos de desequilíbrio de ligação os resultados mostram que para o gene Cinamoil-CoA Redutase (CCR) o DL não se estende por

mais de 200pb em uma região de 1.7kb. Nenhum bloco haplotípico característico foi identificado nessa região e como consequência o número de SNPs necessários para captar toda a diversidade haplotípica foi extremamente alto. Os resultados obtidos a partir do seqüenciamento de 2.7 Kb do gene CAD apontam para resultados similares aos encontrados para o gene CCR. Estes resultados corroboram a expectativa de uma elevada diversidade nucleotídica e uma extensão limitada de DL em espécies de *Eucalyptus*, tendo em vista seu hábito preferencialmente alógamo, histórico recente de domesticação e elevada heterogeneidade genética. Estes resultados sugerem ainda que dificilmente será possível adotar uma abordagem de “*whole genome scan*” em *Eucalyptus* dada a elevadíssima densidade de SNPs necessários, da ordem de milhões. Por outro lado reforçam a importância estratégica dos trabalhos voltados ao descobrimento de genes candidatos via estudos integrados de mapeamento de QTLs e expressão gênica.

INTRODUÇÃO

Originário da Austrália e outras ilhas da Oceania, o eucalipto foi trazido para o Brasil na segunda metade do século XIX com o objetivo de ajudar na produção de dormentes para as linhas férreas que se instalavam no país. O gênero *Eucalyptus* é um membro da família *Myrtaceae* e é composto por cerca de 700 espécies. É um gênero importante na produção de madeira, energia e polpa para produção de papel. O eucalipto brasileiro se destina basicamente à produção de celulose e papel e ao carvão que abastece as siderúrgicas. Além da produção de celulose, papel e carvão para siderúrgicas, são produzidos diferentes tipos de tecidos sintéticos e cápsulas de remédios. A madeira é utilizada na produção de móveis, acabamentos refinados da construção civil, pisos, postes e mastros para barcos. Dele também se obtém o óleo essencial em produtos de limpeza, alimentícios, perfumes e remédios. As indústrias brasileiras que usam o eucalipto como matéria prima para a

produção de papel, celulose e demais derivados representam 4% do PIB brasileiro, 8% das exportações e geram aproximadamente 150 mil empregos (<http://www.celuloseonline.com.br>). Florestas de eucalipto de rápido crescimento suprem hoje, de modo racional e eficiente, a demanda por biomassa lenhosa com propriedades tecnológicas específicas para diversos setores industriais brasileiros, notadamente o de papel e celulose.

Eucalyptus grandis, *E. globulus*, *E. camaldulensis* e *E. urophylla* e seus híbridos estão entre as espécies mais utilizadas, compreendendo cerca de 80% da área plantada no mundo. Para produção de polpa e madeira *E. grandis*, *E. urophylla* e seus híbridos são comuns em áreas de clima tropical e *E. globulus* é a espécie predominante em áreas de clima temperado (POTTS & DUNGEY, 2004).

O eucalipto é uma espécie diplóide e tem 11 cromossomos.

GRATTAPAGLIA & BRADSHAW

(1994), estimaram que o genoma haplóide possui de 370, para *Corymbia citriodora*, gênero correlato, a 630 milhões de pares de bases (Mpb) para *E. grandis*, o que faz o genoma do *Eucalyptus* maior do que outras espécies que tiveram o seu genoma recentemente seqüenciado, tais como *Arabidopsis thaliana* (125 Mpb) (– THE ARABIDOPSIS GENOME INICIATIVE, 2000) e *Populus trichocarpa* (550 Mpb) (– TUKSCAN, et al., 2006), mas em comparação com outras gimnospermas, por exemplo *Pinus* (1C = 20 a 30 Gpb) o genoma do eucalipto é significativamente menor (BOGUNIC, 2003).

Grandes projetos de mapeamento de ligação e de QTL estão sendo desenvolvidos para a espécie. O projeto GENOLYPTUS, que é um consórcio entre 7 universidades brasileiras, 14 empresas de papel e celulose e a Embrapa e possui 24 famílias segregantes envolvendo diversas espécies de eucalipto e seus híbridos, as quais estão sendo utilizadas nos estudos de mapeamento. Outro objetivo do projeto é o estudo de genes candidatos, principalmente obtenção de SNPs para estudos de associação. Um SNP com associação significativa pode não ser aquele diretamente responsável pela variação fenotípica, mas sim que está associado, i.e. em desequilíbrio de ligação (DL) com o SNP causal. O DL tem sido estudado em espécies cultivadas como milho, porém em menor escala em *Eucalyptus* (THUMMA, et

al., 2005). Todavia, espera-se que o DL somente se estenderá a distâncias relativamente curtas para espécies florestais como o eucalipto (RAFALSKI, 2002).

Os genes candidatos mais comumente estudados em eucalipto são os que dão origem a enzimas que fazem parte da complexa via de lignificação, devido a sua importância biológica na formação da madeira juntamente com a celulose e hemicelulose. O conteúdo da lignina, especialmente a sua proporção siringil:guaiacil (S:G), influencia o processo de polpação, sendo a razão S:G diretamente proporcional à taxa de delignificação (PIQUEMAL et al., 1998). Neste estudo foram selecionados os genes que codificam para as enzimas cinamoil-CoA redutase (CCR), que está envolvida na regulação do fluxo de carbono na lignina e álcool cinamil dehidrogenase (CAD) que cataliza o último passo dos monolignóis, reduzindo os cinamaldeídos em álcool cinamil (BAUCHER et al., 2003).

Dentro da iniciativa do projeto GENOLYPTUS, este estudo procurou obter e estimar para esses genes, por meio de seqüenciamento total (CAD) e parcial (CCR), SNPs e DL ao longo dos mesmos nas espécies *E. globulus*, *E. urophylla* e *E. grandis*.

MATERIAL E MÉTODOS

Material Genético

Foram utilizados indivíduos geneticamente não relacionados oriundos de sementes coletadas da Austrália e plantadas no Brasil, de cada uma das espécies: *E. grandis*, *E. globulus* e *E. urophylla*. Para o gene CAD e CCR respectivamente foram analisados 8 e 24 indivíduos. Os indivíduos de *E. grandis* são oriundos da procedência de Pine Creek na região de Atherton no nordeste da Austrália; os de *E. globulus* da procedência de Jeeralang no sul da Austrália e as de *E. urophylla* da região de Timor na Indonésia. Para extração do DNA total foram utilizados tecidos foliares. A extração foi feita pelo método de CTAB 2% (DOYLE & DOYLE, 1987).

Desenho de Primers e Sequenciamento

Para amplificação do gene CCR foram sintetizados *primers* descritos por Poke et al. (2003), além de *primers* já disponíveis no nosso laboratório. Já para a amplificação do gene CAD foi utilizada uma seqüência disponível no projeto Genolyptus produto de sequenciamento “*shot gun*” de um BAC de ~35 kpb contendo o gene. A partir dessa seqüência foram desenhados pares de *primers* a cada 500 pb em média, cobrindo todo o gene utilizando o software Primer3 (<http://www.genome.wi.mit.edu>), Rozen & Skaletsky (1998/2000). Na PCR do gene CAD, especificamente, foram usados três pares de *primers* para amplificação do gene e posteriormente as

reações de sequenciamento foram realizadas com *primers* internos com intervalos de 500 pares de bases. A eletroforese da reação de seqüenciamento foi realizada em plataforma 3100 de 16 capilares da Applied Biosystems.

Busca de SNPs e estimativas de desequilíbrio de ligação

Após a obtenção das seqüências as mesmas foram alinhadas pelo software SeqScape (Applied Biosystems) e foram identificados SNPs. O desequilíbrio de ligação entre cada par de SNPs foi calculado utilizando os parâmetros clássicos D' e r^2 , com seus respectivos valores de significância com $p < 0,05$, ou seus intervalos de confiança, utilizando o software Haploview (<http://broad.mit.edu/mpg/haploview/>). O valor D' é afetado somente pela recombinação e não pelas diferenças de freqüências alélicas entre os sítios polimórficos. Já r^2 é também afetado pela diferença de freqüências alélicas entre pares de sítios polimórficos. Somente locos com freqüência do alelo mais raro maior ou igual a 5% foram considerados na análise. Os valores de r^2 em relação à distância foram plotados em gráficos, com uma curva de tendência baseada em regressão logarítmica.

RESULTADOS E DISCUSSÃO

As amplificações dos fragmentos para o gene CAD geraram bandas específicas e de cerca de 1000 pares de base cada uma. O

seqüenciamento destas bandas a partir de *primers* internos desenhados especificamente para este trabalho gerou seqüências com ótima qualidade e sobrepostas cobrindo todos os *exons* e *introns* do gene. Já para o gene CCR não foi possível uma cobertura completa devido ao grande número de insucessos nas ampliações a partir de *primers* disponíveis na literatura o que nos levou a seqüenciar separadamente dois fragmentos de 500 pb cada com cerca de 100 pb entre eles. O número de amostras analisadas para os dois genes e para cada espécie foi variável devido principalmente à impossibilidade de seqüenciamento devido ao grande número de sítios de inserções/deleções (indels) (CCR) em heterozigose ou até mesmo a SNPs em amostras de espécies específicas na região 3' de anelamento dos *primers* (CAD) que foram desenhados a partir de seqüências de espécies correlatas disponíveis no *Genbank* mas não especificamente a partir de seqüências das espécies estudadas no trabalho. Por exemplo, para o gene CCR foram utilizados *primers* desenhados por Poke et al., (2003) que por sua vez se baseou numa seqüência de *E. gunnii* (*Genbank* X97433). Estes *primers* quando presentes em *introns* em sua maioria não resultaram em amplificação de produto para a espécie *E. globulus*, dificultando a obtenção de seqüências para um maior número de plantas desta espécie. Assim, apesar de terem

sido amplificadas 24 amostras de cada espécie para os dois fragmentos, o número efetivo de amostras para as quais foram obtidas seqüências foi inferior a 24. Especificamente, 23 para *E. grandis*, 9 para *E. globulus* e 7 para *E. urophylla*.

Especificamente para o re-seqüenciamento de genes candidatos em *Eucalyptus* necessita-se de um maior número de seqüências espécie-específicas que permita, por meio de alinhamentos múltiplos, a identificação de regiões conservadas onde possam ser desenhados *primers* com maior probabilidade de sucesso. Após o alinhamento múltiplo das seqüências foi identificado um grande número de SNPs entre as três espécies avaliadas e dentro de cada espécie. A diversidade nucleotídica para uma porção do gene CCR e CAD, respectivamente, nas três espécies, são demonstradas nas figuras 1 e 2. Uma das observações que podem ser feitas é que o número de SNPs para os dois genes é menor na espécie *E. globulus*, sendo as espécies *E. grandis* e *E. urophylla* mais polimórficas para vários sítios. É possível encontrar sítios polimórficos em *E. grandis* e monomórficos em *E. globulus* e *E. urophylla*, como por exemplo, os nucleotídeos 1654 e 1678. Há também sítios polimórficos em *E. grandis* e *E. urophylla* e monomórficos em *E. globulus*, como por exemplo, o nucleotídeo 1714. Foram identificados SNPs tanto em

exons como em *introns*, sendo que para o gene CCR foi identificado um maior número de SNPs dentro de introns, porém para o gene CAD essa proporção de SNPs *exons/introns* é semelhante. A análise também demonstrou a identificação de sítios específicos para cada espécie e sítios polimórficos para as três espécies. Foram também identificadas INDELS de um (CAD) a vários pares de bases (CCR). A análise demonstra ainda um grande número de sítios em heterozigose nas três espécies. Alguns sítios chegam a ter quatro formas alélicas quando comparadas as três espécies, o que indica a mutação recorrente nestes sítios.

As estimativas de desequilíbrio de ligação e seleção de tag SNPs foram calculadas para cada gene em cada espécie estudada (ZHANG, et al., 20045). Dados relativos à estimativa de r^2 mostram que o DL é baixo e se estende por cerca de 200 pb. Nenhum bloco haplotípico é identificado nesta região seguindo a metodologia de Gabriel (GABRIEL et al., 2002) para o gene CAD na espécie *E. globulus*. Como consequência, um grande número de tag SNPs é selecionado. Entre os 34 SNPs utilizados para a estimativa de haplótipos e cálculo do DL, 17 SNPs foram selecionados. Ou seja, baseados nestes dados, estudos de associação por DL em uma amostra desta população deveria genotipar pelo menos estes 17 SNPs. Para o gene CCR na espécie *E. globulus*, dos 19 SNPs detectados, 16 SNPs

foram selecionados como sendo capazes de capturar a informação dos demais SNPs. Ou seja, cada SNP está capturando quase que exclusivamente a informação que diz respeito a si próprio. Esses resultados indicam que o número de SNPs a serem genotipados para capturar a variabilidade da região do gene candidato é muito grande, o que sugere claramente que estudos de genética de associação baseados na abordagem de varredura genômica completa ("*whole genome scan*") constituiriam um enorme desafio técnico em *Eucalyptus* dada a grande quantidade de SNPs que teriam que ser genotipados (da ordem de centenas de milhares ou mesmo milhões).

A análise das seqüências do gene CAD, revelou ainda uma particularidade interessante, exclusivamente para a espécie *E. grandis*. Foi descoberto um bloco haplotípico com cerca de 300 nucleotídeos em forte desequilíbrio de ligação alto (Figura 3). Este bloco formado compreende uma região intrônica e exônica e pode corresponder a uma região genômica de forte pressão de seleção natural nessa espécie. Estudos mais detalhados serão conduzidos para confirmação desta hipótese e possível identificação de sua causa.

Os gráficos plotados relacionando o r^2 com a distância em pares de base mostram uma queda do DL com o aumento da distância para os dois genes nas três espécies avaliadas (Figura 3). Esses resultados

corroboram estimativas feitas para *E. globulus* (POKE et al., 2003) e estão na mesma linha do padrão de DL observado em outras plantas alógamas como o milho (REMINGTON, et al., 2001). Segundo Poke et al., (2003), o gênero *Eucalyptus* possivelmente possuiria duas cópias do gene CAD em seu genoma (CAD1 e CAD2), nos resultados desse estudo provavelmente está representada a cópia CAD2, devido à grande homologia entre a seqüência obtida e a seqüência de referência do CAD2. Nas análises de alinhamento foi utilizada a seqüência do BAC gerado no Genolyptus o que nos mostra também que essa seqüência, muito provavelmente corresponde à cópia CAD2.

CONCLUSÕES

Os resultados deste estudo indicam que espécies de *Eucalyptus* apresentam uma elevada diversidade nucleotídica condizente com o seu hábito reprodutivo preferencialmente alógamo e histórico muito recente de domesticação. A freqüência de SNPs polimórficos é elevada, da ordem de 1 a cada 70 a 200 pb. Os resultados obtidos a partir do seqüenciamento de 2.7 Kb do gene CAD apontam para resultados similares ao encontrado para o gene CCR corroborando a extensão limitada de DL em espécies de *Eucalyptus* com distâncias menores do que 200 pb. Embora este estudo tenha se limitado a dois genes, é razoável supor que este mesmo

padrão de DL muito limitado seja encontrado no resto do genoma. Estes resultados sugerem que com a tecnologia atual de genotipagem de SNPs dificilmente será possível adotar uma abordagem de “*whole genome scan*” para estudos de associação em *Eucalyptus* dada a elevadíssima densidade de SNPs necessários, da ordem de milhões. Por outro lado, reforçam a importância estratégica dos trabalhos voltados ao descobrimento de genes candidatos via estudos integrados de mapeamento de QTLs e expressão gênica. Outro aspecto importante dos resultados encontrados neste trabalho é a grande diversidade encontrada, o que representa a presença no Brasil de um grande estoque para futuros trabalhos de melhoramento genético em *Eucalyptus*.

REFERÊNCIAS BIBLIOGRÁFICAS

- BAUCHER, M.; HALPIN, M.; PETIT-CONIL, M.; BOERJAN, W. (2003) Lignin: genetic engineering and impact on pulping. **Critical. Reviews in. Biochemistry. and Molecular. Biology**,. Boca Raton, USA, v. 38, n. (4), p. :305- – 350, . 2003.
- BOGUNIC, F.; MURATOVIC, E.; BROWN, S. C.; SILJAK-YAKOVLEV, S. (2003) Genome size and base composition of five *Pinus* species from Balkan region. **Plant. Cell. Reports**, Berlin, v. 22, n. (1), p. : 59 – 63,. 2003.

- DOYLE, J. J.; DOYLE, J. L. (1987). Isolation of plant DNA from fresh tissue. **Focus**, n. (12), p. : 13 – -15, .1987.
- GABRIEL, S. B.; SCHAFFNER, S. F.; NGUYEN, H.; *et al.* MOORE, J. M.; ROY, J.; BLUMENSTIEL, B.; HIGGINS, J.; DEFELICE, M.; LOCHNER, A.; FAGGART, M.; LIU-CORDERO, S. N.; ROTIMI, C.; ADEYEMO, A.; COOPER, R.; WARD, R.; LANDER. E. S.; DALY, M. J.; ALTSHULER, D. (2002) The structure of haplotype blocks in the Human Genome. **Science**, (296) v. 21296:, p. 2225-2229, .2002.
- GRATTAPAGLIA, D.; & BRADSHAW, H. D. (1994). Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. **Canadian. Journal. of Forest. Research**,. v. 24, n. 5, p. :11074- – 1078, . 1994.
- PIQUEMAL, J.; LAPIERRE, C.; MYTON, K.; O'CONNELL, A.; SCHUCH, W.; GRIMA-PETTENATI, J.; BOUDET, A. M. (1998) Down-regulation of cinnamoyl CoA-reductase induces significant changes of lignin profiles in transgenic tobacco plants. **Plant Journal**, . v. 13, n. (1, p.):71- – 83, .1998.
- POKE, F. S.; VAILLANCOURT, R. E.; ELLIOTT, R. C.; & REID, J. B. (2003). Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase 2 (CAD2). **Molecular Breeding**,. Dordrecht, Netherlands, v. 12, n. (2, p.): 107- – 118, .2003.
- POTTS, B. M.; & DUNGEY, H. S. (2004). Interspecific hybridization of eucalypts: key issues for breeders and geneticists. **New Forests**, Dordrecht, Netherlands,. v. (27, p.): 115- – 138, .2004.
- RAFALSKI, J. A. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches. **Plant Science**,. Oxford, v. 162, n. (3, p.): 329- – 333, .2002.
- REMINGTON, D. L.; THORNSBEERRY, J. M.; MATSUOKA, Y.; WILSON, L. M.; WHITT, S. R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M. M.; BUCKLER IV, E. S. *et al.* (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. **The Proceedings. of the national Academy of Sciences**, . Washington, v. 98, n. (20, p.): 11479- – 11484, 2001.
- ROZEN, S.; & SKALETSKY, H. J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: **Bioinformatics Methods and Protocols: Methods in Molecular Biology**. Totowa, NJ: Humana Press, Totowa, NJ,2000. pp. 365-386.
- THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome of the flowering plant *Arabidopsis thaliana*. **Nature**, v. 408, p. 796-815, 2000.
- THUMMA, B. L.; NOLAN, M. F.; EVANS, R.; MORAN, G. F. (2005) Polymorphisms in Cinnamoyl CoA Reductase

(CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. **Genetics**, v. 171, p. :11257-1265, .2005.

TUKSCAN, G. A.; DIFAZIO, S.; JANSSON, S.; BOHLMANN, J.; GRIGORIEV, I.; HELLSTEN, U.; PUTNAM, N.; RALPH, S.; ROMBAUTS, S.; SALAMOV, A.; SCHEIN, J.; STERCK, L.; AERTS, A.; BHALERAO, R. R.; BHALERAO, R. P.; BLAUDEZ, D.; BOERJAN, W.; BRUN, A.; BRUNNER, A.; BUSOV, V.; CAMPBELL, M.; CARLSON, J.; CHALOT, M.; CHAPMAN, J.; CHEN, G.-L.; COOPER, D.; COUTINHO, P. M.; COUTURIER, J.; COVERT, S.; CRONK, Q.; CUNNINGHAM, R.; DAVIS, J.; DEGROEVE, S.; DÉJARDIN, A.; DEPAMPHILIS, C.; DETTER, J.; DIRKS, B.; DUBCHAK, I.; DUPLESSIS, S.; EHITING, J.; ELLIS, B.; GENDLER, K.; GOODSTEIN, D.; GRIBSKOV, M.; GRIMWOOD, J.; GROOVER, A.; GUNTER, L.; HAMBERGER, B.; HEINZE, B.; HELARIUTTA, Y.; HENRISSAT, B.; HILLIGAN, D.; HOLT, R.; HUANG, W.; ISLAM-FARIDI, N.; JONES, S.; JONES-RHOADES, M.; JORGENSEN, R.; JOSHI, C.; KANGASJARVI, J.; KARLSSON, J.; KELLEHER, C.; KIRKPATRICK, R.; KIRST, M.; KOHLER, A.; KALLURI, U.; LARIMER, F.; LEEBENS-MACK, J.; LEPLE, J.-C.; LOCASCIO, P.; LOU, Y.; LUCAS, S.; MARTIN, F.; MONTANINI, B.; NAPOLI, C.; NELSON, D. R.; NELSON, C.; NIEMINEM, K.; NILSSON, O.; PEREDA, V.; PETER, G.; PHILIPPE, R.; PILATE, G.; POLIAKOV, A.; RAZUMOVSKAYA, J.; RICHARDSON, P.; RINALDI, C.;

RITLAND, K.; ROUZE, P.; RYABOY, D.; SCHMUTZ, J.; SCHRADER, J.; SEGERMAN, B.; SHIN, H.; SIDDIQUI, A.; STERKY, F.; TERRY, A.; TSAI, C. -J.; UBERBACHER, E.; UNNEBERG, P.; VAHALA, J.; WALL, K.; WESSLER, S.; YANG, G.; YIN, T.; DOUGLAS, C.; MARRA, M.; SANDBERG, G.; VAN DE PEER, Y.; ROKHSAR, D. *et al.* (2006) The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). **Science**, Washington, v. 313, p. :11596-1604, .2006.

THE ARABIDOPSIS GENOME INITIATIVE (2000) Analysis of the genome of the flowering plant *Arabidopsis thaliana*. **Nature** 408:796 – 815.

ZHANG, K.; QIN, ZSUN, F.; CHEN, TZHAO, H.; *et al.* (2004). Hapblockre: a program for : haplotype block partitioning and tag SNP selection software using a set of dynamic programming algorithms reconstruction in general pedigrees without recombination. **Bioinformatics**, Oxford, v. 21, n. (1, p.):131- – 134,. 2005.

GLOSSÁRIO

Alogamia: Fertilização cruzada; numa população panmítica, é o transporte e fusão do gameta masculino de um indivíduo com o gameta feminino de outro indivíduo; tipo de reprodução sexual com mais de 40% de polinização cruzada.

BAC (Bacterial Artificial Chromosome): Um plasmídeo recombinante, similar a um cromossomo, contruído através de engenharia genética. Atua como um vetor de clonagem podendo abrigar grandes insertos (cerca de 500 Kb) de DNA em um hospedeiro bacteriano (em geral *E. coli*).

Biblioteca Genômica: Coleção de fragmentos de DNA clonados que, idealmente, representam todas as seqüências de um genoma.

Deleção: Perda de um segmento cromossômico; pode envolver um ou mais genes, assim como pode envolver uma ou mais bases nucleotídicas do DNA.

Desequilíbrio de Ligação (DL): Associação não aleatória entre alelos a locos distintos; em geral consider-se esta ocorrência entre genes ligados no mesmo cromossomo. Dois alelos estão em desequilíbrio de ligação quando ocorrem simultaneamente com freqüência significativamente maior do que a esperada pela freqüência individual dos alelos.

DNA (Ácido Desoxirribonucléico): Molécula que codifica a informação genética. É uma molécula de dupla hélice de cadeias antiparalelas unidas por pontes de hidrogênio entre os pares de bases; os nucleotídeos são unidos por ligações fosfodiésteres. Na natureza, os pares de bases são formados

entre A (adenina) e T (timina) e entre G (guanina) e C (citosina); portanto a seqüência de bases de cada fita simples é complementar e pode ser deduzida uma da outra.

Éxon: Seqüência genômica codificadora de um gene.

Genoma: A seqüência de DNA completa de um organismo.

Genótipo: Contituição alélica para um determinado loco ou o conjunto gênico de um indivíduo ou espécie.

Genes candidatos: Genes com ação biológica conhecida que estão envolvidos com características de desenvolvimento ou fisiologia do organismo.

Haplótipo: Uma combinação de alelos e/ou polimorfismos de DNA encontrados em locos ligados em um único cromossomo e que tendem a ser herdados em conjunto.

INDELS: Variações no DNA caracterizadas pela presença ou ausência de um ou mais nucleotídeos.

Íntron: Seqüências de DNA que ocorrem dentro dos genes, mas que não codificam, a princípio, para proteínas. Essas seqüências são transcritas juntamente com os éxons em uma molécula de RNA, mas são removidas do RNA mensageiro durante o seu processamento antes da tradução e não são encontradas no mRNA funcional.

Kb (kilobase): Abreviação para pares de quilobases (1.000 bp).

Pares de base (bp): As duas fitas separadas de um ácido nucléico em dupla hélice são mantidas juntas por meio de pontes de hidrogênio específicas entre uma base purina e uma pirimidina, uma em cada fita. A base A está pareada com T no DNA (com U no RNA), enquanto G está pareada com C em ambos DNA e RNA. O tamanho de uma molécula de ácido nucléico frequentemente é dada em termos do número de pares de bases que ela contém.

Polimorfismo genético: Ocorrência simultânea, na população, de genomas que apresentam variações nos alelos de um mesmo loco, resultando em diferentes genótipos.

Primer: Iniciador; molécula de DNA sintética de seqüência arbitrária ou específica, em geral composta de 5 a 40 bases de comprimento, utilizada para iniciar a extensão de um novo polímero de DNA a partir de seu terminal 3' -OH livre.

QTL (*Quantitative Trait Loci*): Loco controlador de característica quantitativa, isto é, característica determinada por muitos genes (multifatorial) que apresenta uma distribuição fenotípica contínua.

SNP (*Single Nucleotide Polymorphism*): Polimorfismo

de sítio único; uma posição de base na qual dois nucleotídeos alternativos ocorrem a uma freqüência significativa (>1%) na população. Os SNPs são responsáveis pela maior parte da variação genética entre os indivíduos.

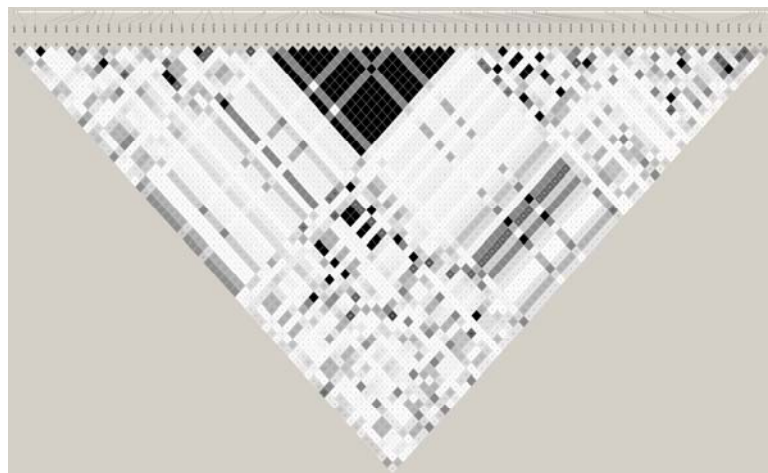
Shotgun: Método de clonagem no qual o Dna genômico é aleatoriamente clivado, mecânica ou enzimaticamente (com endonuclease de restrição), e os fragmentos são então ligados a um vetor de clonagem apropriado. Usado na preparação de bibliotecas genômicas.

Tag SNPs: Conjunto de SNPs que em função do desequilíbrio de ligação capturam a maior parte da diversidade haplotípica de uma determinada região do genoma. Em estudos de associação a utilização de Tag SNPs permite a redução no número de SNPs a serem genotipados.

Whole Genome Scan: Inspeção de genoma completo; termo utilizado para análise rápida de genomas completos pela tipagem de polimorfismos de centenas de marcadores de DNA, tipicamente microssatélites ou SNPs.

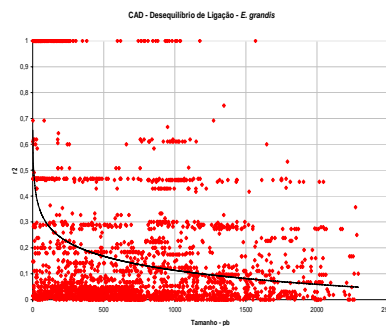
Figura 3: A – Representação gráfica dos níveis de Desequilíbrio de Ligação (DL) para comparações par a par entre sítios polimórficos que apresentaram freqüência do alelo menos freqüente de pelo menos 0,05 e que apresentavam resultados em todas amostras analisadas, obtido pelo software Haploview para o gene CAD na espécie *E. grandis*. **B** – Gráficos de dispersão dos valores de DL estimados via r^2 como função da distância física em pares de bases. Em preto é mostrada uma curva de tendência da dissipação do DL com o aumento da distância física baseada em regressão logarítmica. **1b** – Gene CAD em *E. grandis*. **2b** – Gene CCR em *E. globulus*. **3b** – Gene CAD em *E. urophylla*.

A

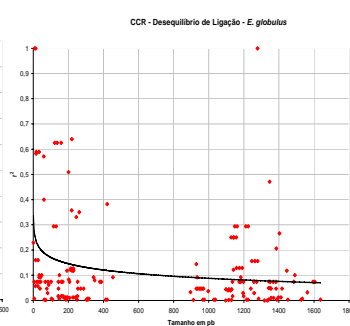


B

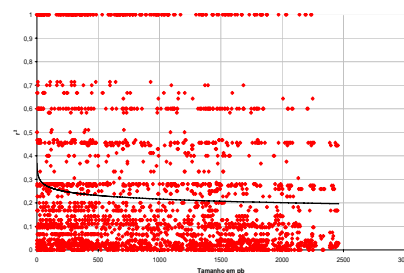
1b



2b



3b



<p>Circular Técnica, 47</p> <p>Ministério da Agricultura, Pecuária e Abastecimento</p>	<p>Exemplares desta edição podem ser adquiridos na Embrapa Recursos Genéticos e Biotecnologia Serviço de Atendimento ao Cidadão Parque Estação Biológica, Av. W/5 Norte (Final) – Brasília, DF CEP 70770-900 – Caixa Postal 02372 PABX: (61) 3448-4763 Fax: (61) 3340-3624 http://www.cenargen.embrapa.br e.mail:sac@cenargen.embrapa.br</p> <p>1ª edição 1ª impressão (2006):</p>	<p>Comitê de Publicações</p> <p>Expediente</p>	<p>Presidente: <i>Sergio Mauro Folle</i> Secretário-Executivo: <i>Maria da Graça Simões Pires Negrão</i> Membros: Arthur da Silva Mariante Maria da Graça S. P. Negrão Maria de Fátima Batista Maurício Machain Franco Regina Maria Dechechi Carneiro Sueli Correa Marques de Mello Vera Tavares de Campos Carneiro</p> <p>Supervisor editorial: <i>Maria da Graça Simões Pires Negrão</i> Normalização Bibliográfica: <i>Ligia Sardinha Fortes</i> Editoração eletrônica: <i>Maria da Graça Simões Pires Negrão</i></p>
-----------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------