

UNIVERSIDADE ESTADUAL PAULISTA

“JÚLIO DE MESQUITA FILHO”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Maria Cecília Perantoni Fuchs

Caracterização gênica para uma anomalia de *Eucalyptus* em
fase inicial de desenvolvimento.

Botucatu - SP
2014

Maria Cecília Perantoni Fuchs

Caracterização gênica para uma anomalia de *Eucalyptus* em
fase inicial de desenvolvimento.

Tese apresentada ao Instituto de Biociências,
Universidade Estadual Paulista “Júlio de
Mesquita Filho” – Campus de Botucatu, para a
obtenção do título de Doutor em Ciências
Biológicas – Área de Concentração Genética.

Orientador: Prof. Dr. Celso Luis Marino

Coorientador: Prof. Dr. Zengjian Jeffrey Chen

Botucatu - SP

2014

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE - CRB 8/5651

Fuchs, Maria Cecília Perantoni.

Caracterização gênica para uma anomalia de Eucalyptus em fase inicial de desenvolvimento / Maria Cecília Perantoni Fuchs. - Botucatu, 2014

Tese (doutorado) - Universidade Estadual Paulista, Instituto de Biociências de Botucatu

Orientador: Celso Luis Marino

Coorientador: Zengjian Jeffrey Chen

Capes: 20203004

1. Eucalipto - Melhoramento genético. 2. Expressão gênica. 3. Genética vegetal. 4. Anomalias (Plantas). 5. Ácido ribonucleico.

Palavras-chave: Anomalia genética vegetal; Eucalipto; Expressão gênica diferencial; RNAseq.

Aos meus pais Edécio e Maria Auxiliadora e avós maternos Mário e Luiza (*in memoriam*),
que sempre acreditaram e me apoiaram incondicionalmente.

Dedico.

AGRADECIMENTOS

A Deus, pai criador.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelas bolsas de doutorado no país e de doutorado sanduíche no exterior concedidas.

Ao Prof. Dr. Celso Luis Marino, pela orientação, oportunidade, confiança, paciência e ensinamentos.

Ao Prof. Dr. Jeffrey Chen e sua equipe da *The University of Texas at Austin*, pela recepção, colaboração, atenção e oportunidade de aprendizagem.

Ao Prof. Dr. Ney Lemke e Dr. Marcio Luis Acencio, pela colaboração e auxílio nas análises dos experimentos.

À empresa Suzano Papel e Celulose, especialmente ao Shinitiro, Esteban e Izabel, por fornecer o material vegetal estudado e colaboração no trabalho.

A todos os professores, funcionários e colegas do Departamento de Genética, pela disponibilidade em ajudar na condução deste trabalho.

Aos professores do Programa de Pós-Graduação em Ciências Biológicas (Genética), pela disponibilidade em ajudar e pela contribuição em minha formação acadêmica.

A todos os queridos amigos do laboratório CAGEM (Centro de Análises Genéticas e Moleculares), pelos incentivos e contribuições para a realização da pesquisa.

Aos funcionários da Seção Técnica da Pós-Graduação, por todo o apoio nos processos durante todo o doutorado.

À empresa Illumina Inc., pelo apoio técnico e concessão dos kits para os experimentos de sequenciamento de nova geração.

À minha avó Luiza (*in memoriam*), aos meus pais Edelcio e Maria Auxiliadora e irmã Raquel, por estarem sempre presentes nos momentos de alegria e tristeza e apoio incondicional.

Ao meu noivo Alexandre de Vicente Ferraz, por estar sempre ao meu lado me confortando, ajudando e apoiando.

Aos amigos do curso de pós-graduação, em especial à Ana Carolina T. Tavares, Karine Kettener, Lídia Carolina M. Arneiro, Marcela I. Rodrigues, Juliana Marino, Juliana Giusti, Júlio C. Otto, Vanusa do S. Leite, Leonardo C. Martins, Juliana Bravo e Alessandra Tenório pelo convívio, alegria, ajuda e paciência dedicados durante todo esse período.

Aos amigos Evandro V. Tambarussi, Frederico A de Jesus, Maíra Valle, Daniel M. Valverde e Amanda de Oliveira pelo apoio, lealdade, descontração e amizade dedicados desde sempre.

Enfim, a todos que contribuíram de alguma forma para que fosse possível a realização deste trabalho.

"Se quiseres seguir adiante com glória, tenha ciência de que às vezes é necessário mudar a estratégia do jogo radicalmente. Você poderá perder muitas peças e muitas batalhas no caminho. E não importa o quanto você sofra, o quanto você apanhe. Você precisa reunir suas forças e seguir em frente!"

Augusto Branco

RESUMO

Fuchs MCP. **Caracterização gênica para uma anomalia de *Eucalyptus* em fase inicial de desenvolvimento** [tese]. Botucatu: Instituto de Ciências Biológicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”; 2014.

O *Eucalyptus* é um dos gêneros mais importantes no setor florestal mundial. Por esse motivo, programas de melhoramento genético florestal são desenvolvidos no intuito de aumentar a produtividade, introduzir características desejáveis e reduzir impactos ambientais. No entanto, características deletérias podem ser incorporadas em determinados cruzamentos nos ciclos de melhoramento, seja por endogamia biparental ou segregação de *loci* heterozigóticos. As anomalias são um exemplo deste efeito e podem acarretar perdas na produção de mudas, na produtividade e atrasos no desenvolvimento de genótipos elite. Portanto, a identificação e caracterização de genes relacionados às anomalias são muito importantes nos programas de melhoramento. Ao realizar um cruzamento controlado entre dois indivíduos de *Eucalyptus grandis*, a empresa Suzano Papel e Celulose detectou uma anomalia com segregação mendeliana de 3 (normais) :1 (anormais) na progênie. As plântulas anômalas morrem em poucos meses e diferem significativamente em algumas características morfológicas, como altura, diâmetro da base do caule, dimensões e forma do limbo foliar e número de ramificações laterais. Estudos prévios permitiram identificar uma marca molecular ligada à anomalia que apresentou identidade com genes da superfamília Bet v1, família PR10 (*pathogenesis-related protein 10*). No entanto, não era claro o envolvimento desta família gênica no desenvolvimento da anomalia. Neste cenário, o presente trabalho teve como objetivo a identificação e caracterização dos genes envolvidos na anomalia detectada. Genes e vias metabólicas, diferencialmente expressos entre os fenótipos contrastantes, demonstraram elevada atividade em processos relacionados à resposta de defesa nas plantas anormais. Tais resultados sugerem que a anomalia é causada pela ativação inadequada do sistema imune da planta (resposta autoimune) associado à incompatibilidade genética. As famílias gênicas de proteínas similares à taumatina, proteínas Bet v1 e proteínas quitinase classe I foram consideradas candidatas na causa da incompatibilidade genética. Assim, foi iniciado o mapeamento fino dos genes diferencialmente expressos dessas famílias a fim de identificar o gene que determina a incompatibilidade genética.

Palavras-chave: Anomalia genética vegetal, Eucalipto, RNA-Seq, Expressão gênica diferencial.

ABSTRACT

Fuchs MCP. **Gene characterization of an *Eucalyptus* anomaly in the early stage of development** [thesis]. Botucatu: Instituto de Ciências Biológicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”; 2014.

Eucalyptus is one of the most important genus in worldwide forestry culture. For this reason forest improvement programs are developed aiming the increase of productivity, the introduction of desirable traits and the reduction of environmental impacts. However deleterious traits can be incorporated in certain crossings through the recurrent improvement cycles either by biparental inbreeding or by segregation of the heterozygous *loci*. Anomalies are an example of this deleterious effect and they can cause losses in seedling production, productivity and delays in the elite genotypes development. Thus, the identification and characterization of genes related to the anomalies are very important in forest improvement programs. In a controlled cross between two *Eucalyptus grandis* individuals, the Suzano Papel and Celulose SA company has detected an anomaly with Mendelian segregation ratio of 3 (normal) :1 (abnormal) in the progeny. Abnormal seedlings die in a few months and show significant difference in some morphological characteristics, such as height, stem-base diameter, dimensions and shape of leaf lamina, and number of branches. Previous studies allowed the identification of a molecular marker related to the anomaly that showed identity with Bet v1 superfamily genes, PR10 family (pathogenesis-related protein 10). However, it was not clear the involvement of this gene family in the anomaly. In this scenario, the present study aimed to identify and characterize the genes involved in the detected anomaly. Genes and metabolic pathways, differentially expressed between the contrasting phenotypes, showed high activity of process related to defense response in abnormal plants. These results suggest that the anomaly is caused by the inappropriate activation of the immune system of the plant (autoimmune response) associated with genetic incompatibility. The gene families of thaumatin-like proteins, Bet v1 proteins, and chitinase class I proteins were considered as candidates for as the genetic incompatibility. Thus, we initiated the fine mapping of differentially expressed genes of these families to identify the gene that determines the genetic incompatibility.

Keywords: Plant genetic anomaly, Eucalypt, RNA-Seq, Differential gene expression.

SUMÁRIO

1 INTRODUÇÃO	9
2 REVISÃO BIBLIOGRÁFICA	13
2.1 O gênero <i>Eucalyptus</i> e sua importância	13
2.2 Melhoramento genético florestal	14
2.3 Variabilidade genética: fonte de alelos deletérios e incompatibilidade genética	16
2.4 Resposta de defesa em plantas.....	17
3 OBJETIVOS	20
4 MATERIAL E MÉTODOS	21
4.1 Material vegetal	21
4.2 Análise morfológica	21
4.3 Análise in silico do marcador SCAR.....	21
4.4 Extração de RNA total e tratamento com DNase	22
4.5 RNA-Seq	22
4.6 Análise dos dados de RNA-Seq.....	23
4.7 RT-qPCR (real-time quantitative PCR).....	23
4.8 Análise funcional e metabólica	25
4.9 Extração de DNA.....	26
4.10 Mapeamento fino	26
5 RESULTADOS E DISCUSSÃO	28
5.1 O caráter anômalo.....	28
5.2 Região genômica da marca molecular	30
5.3 Caracterização gênica e vias metabólicas.....	31
5.4 Incompatibilidade genética.....	54
5.5 Mapeamento fino	56
6 CONCLUSÕES.....	58
7 REFERÊNCIAS BIBLIOGRÁFICAS	59
APÊNDICE A	70

1 INTRODUÇÃO

De ocorrência natural na Austrália, o gênero *Eucalyptus* possui uma ampla diversidade de espécies, variedades e híbridos (mais de 900), sendo a maioria arbóreos com altura variando de 30 a 50 metros (Brooker e Kleinig, 2004; Boland *et al.*, 2006). Por apresentar várias características vantajosas para o seu cultivo, como crescimento rápido, alta produtividade, grande capacidade de adaptação e utilização para diferentes finalidades, o gênero *Eucalyptus* tornou-se umas das essências florestais mais difundidas no Brasil e no mundo (Mora e Garcia, 2000; Grattapaglia e Kirst, 2008; Rockwood, 2012).

Técnicas de melhoramento genético têm sido utilizadas com o intuito de aumentar a produtividade, introduzir características desejáveis para o mercado e reduzir impactos ambientais (Golle *et al.*, 2009; Gonçalves *et al.*, 2013). Em espécies florestais, o melhoramento parte de uma população-base ou experimental onde serão realizadas seleções individuais ou seleções de procedências. Além disso, essas populações podem ter seu material genético recombinado em novos cruzamentos e utilizado em testes de progênies (Rockwood, 2012). O longo período necessário para que o gênero *Eucalyptus* atinja a maturidade reprodutiva e o fato de ser uma espécie alógama inviabiliza a produção de linhagens puras, como as existentes em outras culturas (*i.e.* soja e milho) (Tzfira *et al.*, 1998; Peña e Séguin, 2001). Dessa forma, há um elevado grau de locos em heterozigose neste gênero, mantido por cruzamentos entre os diferentes indivíduos (Quero-García *et al.*, 2009). No entanto, esses cruzamentos podem apresentar anomalias morfológicas e fisiológicas na progênie (Zobel e Talbert, 1984; Klekowski Jr, 1988). Essas anomalias podem ser relacionadas à presença de alelos recessivos deletérios presentes na carga genética da população e podem ser observadas em diferentes espécies, *e.g.* *Antirrhinum majus* (Huijser *et al.*, 1992), *Arabidopsis thaliana* (Rédei, 1975), *Oryza sativa* (Koh *et al.*, 1999) e *Cucumis sativus* (Marx, 1983).

Trabalhos que relatam os efeitos de alelos recessivos deletérios em *Eucalyptus*, em geral, restringem-se no estudo do tipo de cruzamento (aberto, controlado ou autopolinização) (Hardner e Potts, 1995; Burgess *et al.*, 1996; Hardner e Tibbits, 1998; McGowen *et al.*, 2004) ou no estudo de populações de tamanho reduzido (Young *et al.*, 1996; Kennington e James, 1997) e poucos explanam sobre os genes envolvidos e os fenótipos (Maschio *et al.*, 1996; Estopa, 2006).

Ao realizar um cruzamento controlado de *Eucalyptus grandis*, a empresa Suzano Papel e Celulose SA detectou uma anomalia com segregação mendeliana 3:1 (8360 normais e 2880 anômalas) na progênie (Tambarussi, 2006). As plântulas anômalas, que morriam em poucos

meses, foram caracterizadas por superbrotamento caulinar, redução de altura, redução drástica e alterações da forma do limbo foliar (Figura 1). Desde então, o Centro de Análises Genéticas e Moleculares da UNESP/Botucatu (CAGEM) vem estudando as causas genéticas de tal anomalia.

Após a constatação da segregação mendeliana de 3:1 na progênie, Tambarussi (2006) identificou uma marca molecular, pela técnica de RAPD (*Random Amplification of Polymorphic DNA*), relacionada com a anomalia (Figura 2). Segundo este mesmo autor, esta marca esteve presente em todos os 50 indivíduos anômalos e apenas em 31% dos 200 indivíduos normais avaliados.



Figura 1: Plântulas anômalas (à direita) e normais (à esquerda), com cerca de 4 meses de idade.

O RAPD possui algumas limitações, como a baixa reprodução das bandas por ser sensível a quaisquer alterações nas condições de reação e o fato de ser um marcador dominante, impossibilitando a diferenciação de indivíduos homocigotos e heterocigotos (Lacerda *et al.*, 2002; Agarwal *et al.*, 2008; Arif *et al.*, 2010). Dessa forma, Lourenção (2010) converteu-os em marcadores SCAR (*Sequence characterized amplified regions*), que são facilmente obtidos de fragmentos de DNA de 500 a 1500 pares de base (pb) (Negi *et al.*,

2000). Além disso, marcadores SCAR apresentam características vantajosas, tais como: codominância, o que permite a detecção de heterozigotos e, assim, torna-se um marcador mais informativo (Barret *et al.*, 1998; Negi *et al.*, 2000; Agarwal *et al.*, 2008); a reação de amplificação é mais rápida do que a do RAPD (Barret *et al.*, 1998); e permite reprodutibilidade (Hernández *et al.*, 1999).

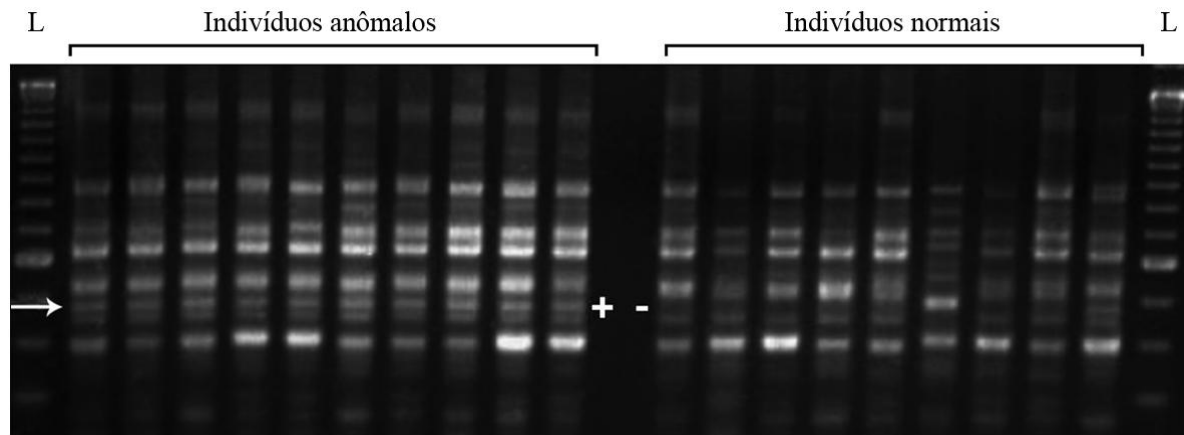


Figura 2: Padrão de bandas pela técnica RAPD. A marca polimórfica de, aproximadamente, 500 pb (flecha) presente em todos os indivíduos anômalos. L: marcador de DNA de 100 pb (Tambarussi, 2006).

O marcador SCAR confeccionado por Lourenção (2010) demonstrou eficiência em identificar (presença de banda) 100% dos indivíduos anômalos e em 22,2% dos indivíduos normais analisados, semelhante ao resultado apresentado por Tambarussi (2006) com RAPD. Ambos os trabalhos sugeriram que a característica seja controlada por um gene de efeito principal e que o marcador encontra-se ligado à este gene. Assim, o alelo que determina a anomalia seria recessivo e a porcentagem de plantas normais com a presença da marca molecular representariam os heterozigotos.

A sequência de 514 pb obtida a partir do marcador SCAR permitiu a identificação, através da análise *in silico* da cobertura de 4,5X do genoma de eucalipto, de uma região genômica que, provavelmente, está relacionada com a anomalia, além de detectar um polimorfismo de dois pares de bases nos indivíduos anômalos (Lourenção, 2010). A sequência SCAR apresentou alta identidade com consensos de sequências expressas (ESTs) de eucalipto e *Vitis vinifera* (84% e 67%, respectivamente) de proteínas PR10 (Pathogenesis-related protein 10) com domínio Bet v1 (Lourenção, 2010). É relatado que as proteínas com domínio Bet v1 (PR-10), estão associadas à resposta contra estresses bióticos e abióticos, assim como outros processos biológicos (Walter *et al.*, 1996; Hoffmann-Sommergruber *et*

al., 1997; Wen *et al.*, 1997; Poupard *et al.*, 2001; Liu e Ekramoddoullah, 2006). Apesar da marca molecular relacionada à anomalia possuir identidade com genes PR-10, não é claro seu envolvimento no desenvolvimento do caráter anômalo.

É evidente que anomalias advindas de alelos recessivos deletérios podem causar perdas na produção de mudas e atrasos em programas de melhoramento. Assim, estudos que buscam a identificação e elucidação da dinâmica dos genes envolvidos em anomalias contribuem para uma melhor compreensão do desenvolvimento vegetal e, conseqüentemente, para futuros planejamentos em programas de melhoramento.

2 REVISÃO BIBLIOGRÁFICA

2.1 O gênero *Eucalyptus* e sua importância

O gênero *Eucalyptus* pertence à família *Myrtaceae* (Angiospermas). A grande maioria das espécies desse gênero é originária da Austrália, mas existem duas espécies endêmicas em Nova Guiné (*E. deglupta*) e Timor (*E. urophylla*) (Pryor, 1985; Ladiges *et al.*, 2003; Brooker e Kleinig, 2004). Mesmo tendo sido previamente observado por exploradores e coletores europeus, não haviam registros botânicos do gênero até 1777. O primeiro exemplar foi coletado pelo botânico David Nelson na Tasmânia e nomeado pelo botânico L'Héritier como *Eucalyptus obliqua* (Brooker, 2002; Brooker e Kleinig, 2004). São reconhecidos mais de 900 espécies, subespécies, variedades e híbridos (Boland *et al.*, 2006; Grattapaglia e Kirst, 2008), que, em geral, possuem porte arbóreo (com altura de 30 a 50 metros), mas algumas espécies apresentam porte mediano e outras arbustivas (Brooker e Kleinig, 2004).

No final do século 18, o eucalipto já era utilizado na Austrália como fonte de madeira para construções civis, mourões para cercas, lenha para energia; assim como, para a produção de extratos medicinais (Turnbull e Booth, 2002). No mesmo período, o eucalipto foi rapidamente introduzido e difundido no Brasil (Grattapaglia e Kirst, 2008; Gonçalves *et al.*, 2013). Além das aplicações já citadas, as plantações de eucaliptos são destinadas à produção de papel e celulose, chapas de fibra, móveis (Wilcken *et al.*, 2008; Assis e Resende, 2011).

Atualmente, o Brasil é considerado como um dos principais países no setor florestal devido à grande extensão de áreas plantadas e desenvolvimento de tecnologias (Assis e Resende, 2011). Segundo a ABRAF (2013), existem mais de 5,1 milhões de hectares cultivados com eucaliptos, que correspondem à mais de 76% dos reflorestamentos comerciais estabelecidos no país. A maior parte desses plantios (mais de 53%) está concentrada nos estados de Minas Gerais, São Paulo, Bahia e Mato Grosso do Sul, com finalidade para produção de papel e celulose e energia para a metalurgia (Gonçalves *et al.*, 2013).

Sob perspectiva econômica e social, o setor florestal brasileiro contribui com uma parcela importante na geração de produtos, tributos, empregos e bem-estar. Em 2012, o valor bruto da produção (VBP) do setor correspondeu à R\$ 56,3 bilhões, os tributos arrecadados totalizaram R\$ 7,6 bilhões (0,5% da arrecadação nacional) e sua participação no superávit da balança comercial nacional foi de 28,1% (ABRAF, 2013). As atividades da cadeia produtiva do setor em 2012 contribuíram para a geração de 4,4 milhões de empregos diretos e indiretos,

além de investimentos de 149 milhões em programas de inclusão social, educação e meio ambiente (ABRAF, 2013).

2.2 Melhoramento genético florestal

Desde o início de sua domesticação, o eucalipto tem obtido ganhos genéticos através de seleção individual, estabelecimento de pomar de sementes e seleção de clones (Rockwood, 2012). Com o sucesso da introdução da cultura do eucalipto no Brasil, houve incentivos fiscais nos anos 60 para a expansão das áreas plantadas (Ferreira, 1992; Golle *et al.*, 2009; Gonçalves *et al.*, 2013). No entanto, as plantações de eucalipto estabelecidas em outras regiões não obtiveram o mesmo sucesso, manifestando problemas de adaptação à fatores bióticos e abióticos. Observações de campo demonstraram que esses problemas poderiam ser resolvidos utilizando clones de híbridos interespecíficos entre *E. urophylla* e *E. grandis*, iniciando assim, os trabalhos de melhoramento genético florestal no Brasil, baseados em seleção recorrente simples com polinização controlada (Gonçalves *et al.*, 2013).

As técnicas de melhoramento genético florestal vêm sendo utilizadas com o intuito de aumentar a produtividade, introduzir características desejáveis para o mercado, reduzir impactos ambientais e expansão da cultura em novos ambientes (Golle *et al.*, 2009; Harfouche *et al.*, 2012). Os programas de melhoramento genético possuem dois aspectos fundamentais: a seleção dos genitores e os mecanismos de herança dos caracteres a serem selecionados. Ambos os aspectos requerem um conhecimento detalhado da constituição genética das espécies a fim de se obter, eficientemente, ganhos genéticos no melhoramento (Bered *et al.*, 1997). Em espécies florestais, o melhoramento parte de uma população-base ou experimental, que podem ter seu material genético recombinado em novos cruzamentos, onde serão realizadas seleções individuais ou de progênies (Souza *et al.*, 2010; Potts *et al.*, 2011; Rockwood, 2012).

Programas de melhoramento genético, aliados à otimização do manejo florestal, geram ganhos significativos na produtividade (incremento médio anual - IMA) (Ferreira, 1992; Campinhos Jr., 1999; Gonçalves *et al.*, 2013). Plantios de eucalipto iniciados em 1967 com um IMA de 28 m³/ha/ano, através dos programas de melhoramento genético estabelecidos a partir de 1973, elevaram o IMA para 45 m³/ha/ano nos anos 90 (Ferreira, 1992; Campinhos Jr., 1999). Segundo a ABRAF (2013), as atividades de pesquisa e desenvolvimento em prol do melhoramento genético e otimização do manejo florestal, realizadas pelas empresas associadas individuais da ABRAF, aumentaram a produtividade média ponderada dos plantios de *Eucalyptus* de 39,4 m³/ha/ano (em 2006), para 40,7 m³/ha/ano em (2012).

Apesar dos grandes avanços, o longo período necessário para que as espécies florestais atinjam a maturidade reprodutiva permitiu que os programas de melhoramento genético florestal alcançassem apenas algumas gerações a partir da população original (base), e assim, são considerados menos avançados quando comparados com os programas de melhoramento de outras culturas (Tzfira *et al.*, 1998; Golle *et al.*, 2009; Grattapaglia e Kirst, 2008). Além disso, a produção de linhagens puras, que é muito comum em culturas como a soja e o milho, é inviável em eucalipto em consequência ao longo ciclo reprodutivo e reprodução alógama, havendo assim apenas indivíduos com a maioria dos locos em heterozigose (Tzfira *et al.*, 1998; Peña e Séguin, 2001).

Cruzamentos entre plantas diferentes mantêm o alto grau de variabilidade genética (heterozigose) na população (Quero-García *et al.*, 2009). No entanto, alguns cruzamentos demonstram alterações fisiológicas e morfológicas que podem resultar em: não formação de sementes; não germinação das sementes; presença de plântulas anormais, menores ou mais fracas que morrem prematuramente; e, raramente, plântulas melhores (Zobel e Talbert, 1984). Essas alterações podem ser decorrentes de efeitos negativos de alelos recessivos deletérios encobertos nos genitores heterozigotos (Klekowski Jr, 1988), que resultam em consideráveis perdas e atrasos nos programas de melhoramento em andamento.

Não há dúvidas de que informações do genoma de plantas podem contribuir para uma melhor caracterização do genoma, maior eficiência na seleção e na maximização dos ganhos genéticos obtidos pelos programas de melhoramento vegetal (Dasgupta *et al.*, 2010; Harfouche *et al.*, 2012; Gupta *et al.*, 2013). No entanto, apesar dos avanços tecnológicos, sua aplicabilidade ainda possui limitações (Grattapaglia e Kirst, 2008; Grattapaglia e Resende, 2011; Bolger *et al.*, 2014). Ferramentas moleculares têm sido utilizadas no melhoramento de diversas espécies cultivadas (Gupta *et al.*, 2013; Bolger *et al.*, 2014) e, em *Eucalyptus*, destaca-se o uso na manutenção da variabilidade genética, identificação de clones e paternidade (Grattapaglia e Kirst, 2008; Souza *et al.*, 2010). Neste cenário, o desenvolvimento de marcadores moleculares, tal como a compreensão da dinâmica dos genes em processos biológicos, podem ser ferramentas eficazes na identificação de alelos recessivos deletérios e, conseqüentemente, na eliminação de genótipos com alelos desfavoráveis nos programas de melhoramento, evitando inclusive, danos na produção e atrasos no desenvolvimento de genótipos elite.

2.3 Variabilidade genética: fonte de alelos deletérios e incompatibilidade genética

Alterações gênicas e cromossômicas são as principais fontes de variabilidade genética em uma população. Essas alterações são essenciais na geração de novos alelos e genes. O modo mais simples de geração de novos alelos é a mutação, no qual há a substituição de pares de bases, deleções ou inserções no DNA, que podem resultar na mudança da sequência de aminoácidos de uma proteína e, conseqüentemente, na sua função (Futuyma, 1997; Freeman e Herron, 2009). No entanto, recombinações cromossômicas via *crossing-over*, assim como a transposição de elementos genéticos móveis (transposons) também podem gerar novos alelos se estes ocorrerem dentro dos limites de um gene (Niklas, 1997; Stears e Hoekstra, 2003).

Os efeitos de novos alelos sobre um fenótipo variam entre neutro, favorável e desfavorável, dependendo das condições ambientais. No entanto, alguns alelos, como os que causam a morte do indivíduo ou graves anomalias de desenvolvimento, são sempre deletérios (Futuyma, 1997). A maioria das plantas alógamas, como o eucalipto, são heterozigotos para os alelos recessivos deletérios, assim, permanecem encobertos na carga genética da população. Quando em homozigose, estes alelos causam danos drásticos na progênie (Zobel e Talbert, 1984; Klekowski Jr., 1988)

Da mesma forma que a mutação é o principal modo de geração de novos alelos, a duplicação gênica é a fonte mais importante de novos genes. Em geral, a duplicação gênica surge a partir de um processo chamado *crossing-over* desigual na meiose, resultando em um cromossomo com uma deleção e o outro com a duplicação dos genes em *tandem* (Zhang, 2003; Freeman e Herron, 2009). Após essa duplicação, podem ocorrer mutações que levam à divergências entre as cópias. Uma das cópias pode ser silenciada por mutações deletérias, tornando-se um pseudogene, ou ambas cópias podem permanecer e seguir um de três caminhos evolutivos: 1) podem preservar a similaridade das sequências, sendo assim cópias extras do gene original; 2) cada cópia pode manter parte das funções do gene ancestral (subfuncionalização), sendo complementares uma à outra e ambas necessárias para que o organismo funcione perfeitamente; e 3) uma das cópias permanece com a função original e a outra cópia adquire novas funções (neofuncionalização) (Force *et al.*, 1999; Gonzàlez-Duarte e Albalat, 2005).

Os genes duplicados representam grande parte dos genes totais nos organismos dos diferentes reinos, atingindo a proporção de até 65% em *Arabidopsis thaliana* (Zhang, 2003).

Em geral, duplicações recorrentes e divergências funcionais geram as denominadas famílias multigênicas (González- Duarte e Albalat, 2005).

Ambientes diferentes possuem pressões seletivas distintas, e estas podem ser contraditórias. Dessa forma, os genes de uma família gênica de um indivíduo, em um determinado ambiente, podem divergir evolutivamente dos genes de outro indivíduo em um ambiente completamente diferente (Bomblies e Weigel, 2007; Orr, 1996; Stears e Hoekstra, 2003). Os genes dos indivíduos adaptados aos ambientes diferentes, ao serem combinados, podem resultar em interações negativas por incompatibilidade genética (ou conflito genômico) (Stears e Hoekstra, 2003; Brideau *et al.*, 2006; Bomblies e Weigel, 2007; Bikard *et al.*, 2009;). A incompatibilidade genética, causa esterilidade, anomalias e morte na progênie, sendo considerada como um isolamento pós-zigótico no processo de especiação (Orr, 1996; Bomblies e Weigel, 2007; Ispolatov e Doebeli, 2009;). Em plantas, a incompatibilidade genética é, muitas vezes, associada à alterações no sistema de resposta de defesa (Bomblies *et al.*, 2007; Mizuno *et al.*, 2010).

2.4 Resposta de defesa em plantas

As plantas, apesar de apresentarem aparente vulnerabilidade, possuem inúmeras estratégias de defesa contra patógenos e predadores. Dentre as respostas de defesa, duas estratégias são consideradas principais:

- 1) A primeira estratégia envolve a percepção de padrões moleculares associados à micro-organismos, patógenos ou perigo (MAMPs, PAMPs e DAMPs, respectivamente) por proteínas R (proteínas de resposta) transmembrana, denominadas PRRs (receptores de reconhecimento de padrões). O reconhecimento pelas PRRs promove a ativação de uma cascata de sinalização, culminando na resposta de defesa (Coll *et al.*, 2011; Jones e Dangl, 2006; Tiffin e Moeller, 2006).
- 2) Denominada 'hipótese guarda', a segunda estratégia (mais complexa) ocorre no citoplasma e envolve a expressão basal de genes relacionados à resposta de defesa (genes R - 'guarda') (Jones e Dangl, 2006; Bomblies e Weigel, 2007). Uma vez que patógenos bem-sucedidos conseguem contornar a sistema de defesa transmembrana, estes liberam proteínas efetoras que contribuem para a patogenicidade (Jones e Dangl, 2006; Coll *et al.*, 2011). As proteínas efetoras são reconhecidas pelas proteínas NB-LRR específicas, que são os produtos dos genes R e que possuem domínio de ligação à nucleotídeos (*Nucleotide Binding* - NB) e domínio ricos em leucina (*Leucine Rich*

Repeat - LRR) (Jones e Dangl, 2006; Coll *et al.*, 2011). O reconhecimento pode ser de maneira direta, pela interação da proteína efetora com a NB-LRR, ou indireta, onde a proteína efetora interage com uma proteína alvo do hospedeiro, alterando a conformação desta que é reconhecida pela NB-LRR. Tanto o reconhecimento direto, como o indireto desencadeia uma cascata de sinalização de resposta de defesa (Heath, 2000; Bomblies e Weigel, 2007;).

Um dos primeiros passos na resposta de defesa envolve alterações na concentração de íons de cálcio (Ca^{2+}) na célula. Os canais de cálcio da membrana plasmática são ativados (abertos), permitindo a entrada de Ca^{2+} e consequente aumento da concentração deste elemento no citoplasma (Heath, 2000). Elevados níveis citoplasmáticos de Ca^{2+} possuem papel importante como ativador dos processos de explosão oxidativa e síntese de óxido nítrico (Delledonne *et al.*, 1998; Grant *et al.*, 2000; Romero-Puertas e Delledonne, 2005; Taiz e Zeiger, 2013). Na explosão oxidativa há a produção e acúmulo imediatos de espécies reativas de oxigênio, como o radical superóxido (O_2^-), radical hidroxila (OH) e o peróxido de hidrogênio (H_2O_2). O O_2^- é gerado a partir da adição de um elétron ao oxigênio molecular (O_2) e, ao passar por reações de oxirredução ou dismutação, gera peróxido de hidrogênio (Hammond-Kosack e Jones, 1996; Schwan-Estrada *et al.*, 2008). Segundo Zeier *et al.* (2004), o óxido nítrico possivelmente age inibindo a maquinaria antioxidante da célula, garantindo os níveis elevados de peróxido de hidrogênio por um tempo prolongado. O aumento nas concentrações tanto de óxido nítrico quanto de peróxido de hidrogênio são essenciais para a ativação da resposta de hipersensibilidade (Delledonne *et al.*, 1998; Heath, 2000; Taiz e Zeiger, 2013). A resposta hipersensitiva previne o crescimento e a proliferação do patógeno através da indução da morte celular no sítio de infecção (Dorey *et al.*, 1997; Heath, 2000; Coll *et al.*, 2011). Além do envolvimento na resposta hipersensitiva, o peróxido de hidrogênio atua na ativação de outros mecanismos de defesa, como por exemplo, a produção de lignina (Chen *et al.*, 1993; Hammond-Kosack e Jones, 1996; Durner *et al.*, 1997; Taiz e Zeiger, 2013).

A biossíntese do ácido salicílico (AS), hormônio vegetal, também é induzida pelo reconhecimento do patógeno pela planta (Ryals *et al.*, 1996; Sticher *et al.*, 1997; Durrant e Dong, 2004; Coll *et al.*, 2011). Tem sido relatado que o AS interage com enzimas catalase e peroxidase, resultando no aumento dos níveis de H_2O_2 pela inibição da atividade dessas enzimas (Chen *et al.*, 1993; Hammond-Kosack e Jones, 1996; Durner *et al.*, 1997; Durrant e Dong, 2004). Dessa forma, o AS intensifica a produção de espécies reativas de oxigênio e, consequentemente, a resposta de hipersensibilidade (Levine *et al.*, 1994; Durrant e Dong,

2004). Por outro lado, há evidências de que o H_2O_2 induz o acúmulo de AS através do aumento da atividade da enzima ácido benzóico-2 hidroxilase (AB2-H), também denominada citocromo P450 monoxigenase, que converte ácido benzóico em AS (Hammond-Kosack e Jones, 1996; Ryals *et al.*, 1996; Sticher *et al.*, 1997; Durrant e Dong, 2004).

O envolvimento do AS na sinalização e indução da resistência sistêmica adquirida (SAR) é demonstrado pela sua presença no floema após a infecção pelo patógeno (Ward *et al.*, 1991; Durner *et al.*, 1997; Durrant e Dong, 2004). A SAR confere aos tecidos não infectados resistência contra infecções secundárias pelo patógeno a partir de sinalizações decorrentes do sítio inicial de infecção (Durner *et al.*, 1997; Sticher *et al.*, 1997; Durrant e Dong, 2004). A SAR é caracterizada pela expressão de um grupo de genes, denominados SAR genes (Ward *et al.*, 1991; Ryals *et al.*, 1996). Os SAR genes são induzidos pelo tratamento com AS e pela transformação de genes que convertem corismato em AS (Ward *et al.*, 1991; Durner *et al.*, 1997; Verbene *et al.*, 2000), corroborando com a hipótese do AS como indutor da SAR. Os SAR genes consistem em pelo menos nove famílias gênicas, que incluem as proteínas PR (*pathogenesis-related*) PR-1, PR-2 (β -1,3-glucanase), PR-3 (quitinase classe II, também denominada PR-Q), PR-4 (*hevein-like protein*), PR-5 (*thaumatin-like protein*), assim com as formas básicas e ácidas da quitinase classe III, a forma básica da PR-1 e PR-Q extracelular (Ward *et al.*, 1991; Ryals *et al.*, 1996).

A ativação do sistema de defesa das plantas sem que ocorra ataque por patógenos (autoimunidade) está, frequentemente, associada à um tipo de incompatibilidade genética denominada necrose híbrida (Bomblies *et al.*, 2007; Mizuno *et al.*, 2010). Esta pode ser observada em cruzamentos inter ou intraespecíficos, nos quais parte da progênie apresenta fenótipo (fraqueza, nanismo e atrofiamento) similar à plantas atacadas por patógenos ou sob estresse abiótico (Knight, 1947; Bomblies e Weigel, 2007).

3 OBJETIVOS

O objetivo geral deste trabalho foi identificar e caracterizar os genes envolvidos na anomalia detectada em *Eucalyptus grandis* em fase inicial de desenvolvimento. Os objetivos específicos foram: caracterização morfológica do caráter anômalo; determinação do perfil transcricional de plantas normais e anômalas; análise e validação da expressão gênica diferencial entre os fenótipos contrastantes; análise funcional e metabólica dos genes diferencialmente expressos; e identificação das vias metabólicas relevantes comparando-se os dois fenótipos.

4 MATERIAL E MÉTODOS

4.1 Material vegetal

Foram utilizadas um total de 38 plântulas (19 normais e 19 anômalas) obtidas por cruzamento controlado de *Eucalyptus grandis* (indivíduos G07 e G26), cedidas pelo programa de melhoramento florestal da empresa Suzano Papel e Celulose SA. Todas as plântulas foram produzidas em tubete, via propagação seminal e condução em casa de vegetação. Aos 4 meses de idade, as 38 mudas de *E. grandis* foram submetidas às avaliações morfológicas e genéticas, conforme os itens seguintes.

4.2 Análise morfológica

Para a avaliação morfológica das plantas, foram quantificados: 1) a altura das plântulas, com auxílio de régua milimetrada; 2) o diâmetro da base do caule, por meio de paquímetro analógico com precisão de 0,1 mm; e 3) largura, comprimento e área do limbo foliar. Para as análises do limbo foliar, foram utilizadas 6 a 9 folhas de dois indivíduos de cada fenótipo e as mensuras realizadas pelo programa ImageJ software (Wayne Rasband, National Institutes of Health - NIH, Bethesda, MD). Além disso, foi realizada uma descrição visual do formato das folhas e contagem do número de ramificações laterais.

Após as avaliações, os resultados foram verificados quanto às pressuposições básicas estatísticas (ex. normalidade e homogeneidade da variância). Nos casos onde foram necessárias transformações dos dados, utilizou-se o método de potência máxima de Box-Cox. Após esta fase, os resultados foram submetidos à análise de variância (ANOVA) e, em seguida, comparados estatisticamente entre si através do teste de Tukey a 5% de nível de significância. Para as análises utilizou-se o programa estatístico SAS (Statistical Analysis System) v9.1 (SAS INSTITUTE, 2002).

4.3 Análise *in silico* do marcador SCAR

Com a atualização da cobertura do genoma de *Eucalyptus grandis* no banco de dados Phytozome v9.1 (<http://www.phytozome.net>), foi realizada nova análise *in silico* para complementar àquela feita anteriormente por Lourenção (2010). Utilizando a ferramenta BLAST (Altschul et al., 1997), foi comparada a sequência do marcador SCAR relacionado à anomalia com o genoma de *Eucalyptus grandis* referência do banco de dados Phytozome.

4.4 Extração de RNA total e tratamento com DNase

O RNA total foi extraído individualmente de três órgãos (raiz, caule e folha) com base na metodologia proposta por Chang et al. (1993), com algumas modificações. Foram utilizados, aproximadamente, 100 mg de material macerado, com posterior adição de 500 µl de solução de tampão de extração CTAB (2% CTAB; 2% PVP; Tris-HCl 100 mM, pH 8,0; EDTA 25 mM; NaCl 2 M; 2% β-mercaptoetanol) pré aquecido e incubação a 65°C por 15 min. Após incubação, foram adicionados 500 µl de solução de CIA (24 clorofórmio : 1 álcool isoamílico), que em seguida foi centrifugada a 10000 rpm por 10 min. O sobrenadante foi transferido para um novo tubo e uma nova lavagem com CIA foi realizada nas mesmas condições mencionadas anteriormente. Ao volume do sobrenadante coletado, foi adicionado um volume de cloreto de lítio (LiCl) 5M e o RNA total foi precipitado por 12 hs a 4°C. Após precipitação, centrifugou-se por 30 min, a 12000 rpm e a 4°C. O sobrenadante foi descartado e o precipitado eluído em 400 µl de TE-SDS (Tris-HCl 10 mM, pH 8,0; EDTA 1 mM, pH=8,0; SDS 1%), 200 µl de NaCl 5M e 600 µl de isopropanol. Este foi incubado por 30 min a -20°C para precipitação e centrifugado a 12000 rpm, por 20 min e a 4°C. Uma nova eluição e centrifugação foram realizadas, o sobrenadante descartado e o precipitado lavado duas vezes com 1 ml de etanol 70%. Após cada lavagem, as amostras foram centrifugadas a 10000 rpm, a 4°C, por 5 min e o precipitado eluído em 30 µl de água DEPC autoclavada com posterior armazenamento a -80°C. O RNA total obtido teve sua concentração determinada utilizando espectrofotômetro (Nanodrop ND-1000, Thermo Scientific). Três nanogramas de RNA total extraído de cada órgão foi tratado com DNase I Amplification Grade (Invitrogen), conforme o manual do fabricante e armazenado a -80°C.

4.5 RNA-Seq

Após o tratamento com DNase, determinação da concentração por espectrofotômetro (Nanodrop ND-1000, Thermo Scientific) e verificação da integridade do RNA pelo Agilent 2100 Bioanalyzer, utilizando a kit RNA 6000 Nano (Agilent), uma mesma quantidade de RNA de 5 indivíduos foram misturados, formando *pools* de RNA de plantas normais e anormais para cada órgão (raiz, caule e folha). Os *pools* foram utilizados na construção das bibliotecas de cDNA. As bibliotecas foram construídas utilizando o kit TruSeq RNA sample preparation (Illumina), de acordo com o protocolo Low-Throughput fornecido pelo fabricante e armazenadas a -20°C.

A quantificação das bibliotecas obtidas foi realizada por qPCR, utilizando o kit KAPA Library Quantification (KAPA Biosystems) e a qualidade verificada pelo Agilent 2100 Bioanalyzer, utilizando a kit Agilent High Sensitivity DNA (Agilent). As bibliotecas foram diluídas a 10 nM e construído um *pool* de bibliotecas com 10 µl de cada biblioteca. O pool de bibliotecas foi diluído em uma concentração final de 12,5 pM para o sequenciamento no equipamento MiSeq Personal Sequencing System (Illumina) de acordo com o guia do fabricante.

4.6 Análise dos dados de RNA-Seq

As sequências (*reads*) de cada biblioteca, obtidas pelo RNA-Seq, foram mapeadas utilizando o programa Bowtie (Langmead *et al.*, 2009), tendo como referência o genoma de *Eucalyptus grandis* (*Eucalyptus grandis* Genome Project 2010, <http://www.phytozome.net/eucalyptus>). Através da contagem do número de *reads* mapeados por gene, utilizando o programa BEDTools (Quinlan e Hall, 2010), foi obtido valores de expressão gênica digitais utilizados para a normalização e análise de expressão gênica diferencial no programa R (pacote DESeq) (Anders e Huber, 2010).

4.7 RT-qPCR (real-time quantitative PCR)

Oligonucleotídeos gene-específicos (Tabela 1) de transcritos representativos dos resultados do RNA-Seq de cada órgão foram desenhados através da ferramenta Primer Quest (IDT) no intuito de validar os resultados analisados do RNA-Seq. Como gene de controle interno foi utilizado a ubiquitina (Tabela 1), que demonstrou grande estabilidade em estudo de seleção de genes endógenos para estudos de expressão gênica em *Eucalyptus* (Boava *et al.*, 2010). Além da validação com os transcritos representativos, foi realizado um experimento com oligonucleotídeos específicos para quitinase, descrito por Boava *et al.* (2010) (Tabela 1).

Para as análises de RT-qPCR, três amostras de RNA total de cada órgão foram tratadas com DNase e submetidas à transcrição reversa utilizando o kit Omniscript Reverse Transcriptase (Qiagen), segundo as recomendações do fabricante. Todas as reações foram incubadas a 37°C por 1 hora, em um volume final de 20 µl contendo 726 ng de RNA, tampão de reação RT, dNTPs, RNase Out (Invitrogen), enzima transcriptase reversa e oligo(dT). O cDNA obtido de cada amostra foi diluído 10 vezes com água ultrapura autoclavada e armazenado a -20°C.

A verificação da qualidade do cDNA foi realizada por reação de PCR com o oligonucleotídeo do controle interno ubiquitina, nas seguintes condições: 1 µl de cDNA diluído; 1x Tampão PCR; 1,5 mM de MgCl₂; 0,08 mM de dNTPs; 0,5 µM de cada oligonucleotídeo primer e 1 unidade de Taq DNA polimerase, o volume foi completado para 15 µl com água ultrapura autoclavada. A amplificação foi realizada utilizando o seguinte programa: 2 minutos a 95°C, seguido de 25 ciclos de 30 segundos a 95°C, 30 segundos a 60°C, 30 segundos a 72°C e uma extensão final de 7 minutos a 72°C.

Tabela 1: Oligonucleotídeos utilizados para a análise por RT-qPCR.

Nome	Produto gênico	Sequência dos oligonucleotídeos senso/antissenso	Órgão
UBQ*	Ubiquitina	GAGGGACATCTATCTCTATGAC/ CAACAGTAAGCACACGAG	Controle interno - raiz, caule e folha
GRP	Proteínas ricas em glicina	GCTAACTACGGCCAAGGATATG/ GTGCTCGCCACCTTCATAC	Folha e caule
CHI I	Quitinase classe I	TCATTGCTCATGTACTIONGGTGAT/ GCTTCTGCAACTIONTTCATTGAGAT	Folha e caule
BARWIN	Proteínas com domínio Barwin	GCTTGTGCCAGAAAGAATAAG/ CGTAACGACACTIONTTCATTACATAACG	Folha e caule
THAU	Proteínas similares à taumatinas	TCTGAACTIONTGGATGGACCATIONT/ CCAGTGACGAGGGTACAAC	Folha e caule
BET vI	Proteínas com domínio Bet vI	CCACCGTGGAGTTGTTTAGT/ TACACACCCGACCATCAAATC	Folha
BET vI b	Proteínas com domínio Bet vI	CGCTGGTAGTGGATGTGTT/ CCTTCTTAATGTCTCTCTCTC	Caule
GLOB	Globinas	TTCAGATIONTGTAGCAGCAACAAG/ GAGATGTGGTGCCAGAAA	Raiz
PROT	Aspartil proteases	ATCGGCCTCAACAGCATATC/ AACAGTGTACGGAAGCACAG	Raiz
CUP	Cupinas	GAGGTAGTCAACCACCTTCTTG/ CCATTGCTAATGCTGTCTTTGG	Raiz
CHI I b	Quitinase classe I	GCAACCCGAACGTAGGATATTA/ GGGAGTTTAAGCCGTCGAAT	Raiz
QUI*	Quitinase	AAATCTTCAGCATCGCCA/ CCCAATAGACAGCAATIONTCTC	Raiz, caule e folha

* Oligonucleotídeos descritos por Boava *et al.* (2010).

As reações de RT-qPCR foram realizadas utilizando 1 µl de cDNA diluído, 5,75 µl FastStart SYBR Green Master (Roche) e 0,33 µM de cada oligonucleotídeo (senso e antissenso) para os genes específicos. Como plataforma para o RT-qPCR, foi utilizado o equipamento 7500 Real-Time PCR Systems (Applied Biosystems) com o seguinte programa: um ciclo de 50°C por 2 minutos; um ciclo de 95°C por 10 min; 40 ciclos de 95°C por 15 segundos e 60°C por 1 minuto.

Os níveis de expressão gênica revelados pelo RT-qPCR foram analisados pelo método comparativo de *Ct* (*threshold cycle*) (Schmittgen e Livak, 2008) e submetidos ao teste t de Student no programa estatístico SAS v9.1 (SAS INSTITUTE, 2002), com 5% de nível de significância.

4.8 Análise funcional e metabólica

As anotações funcionais (termos GO - *Gene Ontology*) dos genes diferencialmente expressos de cada órgão foram obtidas utilizando o programa Blast2GO (Conesa e Götzt, 2008), com os parâmetros *default*. O Blast2GO fornece as anotações funcionais para as categorias 'Processos biológicos' (P), 'Função celular' (F) e 'Componente celular' (C), através de buscas por similaridades e homologia entre as sequências pela ferramenta BLAST (Altschul et al., 1997). A análise de enriquecimento dos termos GO, foi realizada no programa BiNGO (*Biological Network Gene Ontology*) (Maere et al., 2005) pelo teste estatístico hipergeométrico com correção de Benjamini e Hochberg (FDR - *False Discovery Rate*), utilizando como referência as anotações obtidas pelo banco de dados Phytozome v9.1 (<http://www.phytozome.net>). Além dos termos GO, o Blast2GO fornece o código da enzima (EC - *Enzyme Commission number*) e o mapa KEGG (Enciclopédia de Genes e Genoma de Kioto), que permite a visualização das vias metabólicas nas quais as enzimas possuem participação (Conesa e Götzt, 2008).

Com base na anotação fornecida pelo banco de dados Phytozome v9.1, foram determinados os genes mapeados, nas análises de RNA-Seq, que codificam enzimas. As enzimas foram mapeadas em vias metabólicas, utilizando o banco de dados de vias metabólicas de plantas Gramene (<http://www.gramene.org/pathway/>). Para cada enzima foi calculada a variável $fold = \log_2(\text{expA}/\text{expB})$, onde expA e expB são as expressões de plantas anormais e normais, respectivamente, fornecidas pelos dados de RNA-Seq. Para cada via, foi calculada a média da variável *fold* das enzimas deste. Da mesma forma, para cada via foram selecionados valores aleatórios do total de variáveis *fold*, este processo foi repetido 100 vezes

e, então, foram calculados a média e o desvio padrão. O número de valores aleatórios é equivalente ao número de enzimas mapeadas na via. Com esses valores, foram calculados o *p-value* e o *Zscore*. Dessa forma, foram selecionadas as vias metabólicas subexpressas (*Zscore* negativo) e superexpressas (*Zscore* positivo) nas plantas anormais, com o critério de $pvalue < 0,05$.

4.9 Extração de DNA

O DNA genômico foi isolado de folhas, utilizando o método por CTAB, conforme metodologia proposta por Murray e Thompson (1980), com modificações. Cerca de 150 mg de material vegetal foi macerado em nitrogênio líquido. Em seguida, o macerado foi homogeneizado em 700 µl de solução de tampão de extração CTAB (2% CTAB; 2% PVP; 100 mM Tris-HCl pH=8.0; 25 mM EDTA; 2 M NaCl; 0.2% 2-mercaptoetanol) e aquecido por 45 min a 65 °C. Foram adicionados 600 µl de CIA (clorofórmio: álcool isoamílico 24:1) à cada amostra, que foram homogeneizadas e centrifugadas por 5 min a 12000 rpm. O sobrenadante foi transferido para um novo microtubo, onde foi adicionado 100 µl de solução de CTAB 5% (5% CTAB; 2% PVP; 100 mM Tris-HCl pH=8.0; 25 mM EDTA; 2 M NaCl) e nova extração com CIA foi realizada. O sobrenadante foi transferido para um novo microtubo, onde foram adicionados 400 µl de álcool isopropílico gelado e as amostras foram incubadas durante 12 hs a -20 °C. Após esse período, as amostras foram centrifugadas por 5 min a 12000 rpm. O sobrenadante foi descartado e o precipitado lavado com 500 µl de etanol 95%. O DNA foi eluído em 80 µl de solução de tampão TE (pH=8.0) contendo 10 ng/ µl de RNase. A concentração do DNA total foi determinada utilizando espectrofotômetro (Nanodrop ND-1000, Thermo Scientific) e a qualidade observada em gel de agarose 1% em 1X TBE (Tris-borato-EDTA), corado com brometo de etídio.

4.10 Mapeamento fino

Para o mapeamento fino, foram desenhados oligonucleotídeos (Tabela 2) nas regiões dos genes diferencialmente expressos candidatos, em um intervalo entre 3 kpb a montante e a jusante do gene. Os oligonucleotídeos foram testados em 4 indivíduos de cada grupo (normal e anormal) na procura de polimorfismo entre estes.

As reações de PCR foram realizadas nas seguintes condições: 30 ng de DNA; 1X Tampão PCR com 2 mM de MgCl₂; 0,2 mM de dNTPs; 0,4 µM de cada oligonucleotídeo (senso e antissenso) e 2 unidades de Taq DNA polimerase, o volume foi completado para 25

µl com água ultrapura autoclavada. A amplificação foi realizada utilizando o seguinte programa: 2 minutos a 95°C, seguido de 35 ciclos de 30 segundos a 95°C, 30 segundos na temperatura de anelamento do oligonucleotídeo, 30 segundos a 72°C e uma extensão final de 5 minutos a 72°C.

Tabela 2: Genes candidatos, oligonucleotídeos utilizados no mapeamento fino e temperatura de anelamento (Ta).

Gene candidato	Oligonucleotídeo	Sequências senso/antissenso	Ta	Amplicon (pb)
Eucgr.E00560	S5_1_1	CCTGTCGGAGTGTAGAGTG/ CGGTGTGGATTAGTGAATTGTG	59	643
	S5_1_3	CCCAAGGATGATGCTACCA/ TTTCCGTCTGTTCCAAACC	64	560
	S5_1_4	CACAAACAGAACAATGCGTC/ TATAGGGTTACTCCAAGCGG	59	787
Eucgr.G02175	S7_2_1	TTTCTTGAATAGGTTGGGCTGG/ CTTGTGATTGCTTGAGTGATGG	65	482
	S7_2_2	AATCAAGAACTCAAGCCCG/ CCCGCACCTTAATATCGTG	59	734
Eucgr.I01495	S9_3_1	TGGCGAAGATTGGTGAAAGAG/ GGACTTAGGATTTACGGGTTAGG	61	844
	S9_3_2	ATCTAACCGTGCCATCATACTC/ GTTCTCTGTATCAAATCTGCG	59	992
	S9_3_3	TGCCAATAGGTGAAGAATCG/ TCCAAACCAAGAAGATGCTC	58	744
Eucgr.J02518	S10_4_1	ATGAGATTGCTCTTCCCAACG/ GCAGCATTTCAGTAGATGTTACCC	62	794
Eucgr.J02519	S10_4_2	CCTTTACCATGTATGTGATCGCC/ CAAGTAGCAAGCTAAACGGGTC	61	432
Eucgr.H04012	S8_7_2	TCCTAAGATCACCATCCAACC/ GTTACAATGACTCACGAGCAG	63	479
Eucgr.E01382	S5_8_1	TCGATGGGCTTGACTACCG/ GGAGCGTTGAGGAACACTG	59	342
Eucgr.L00941	S86_12_1	GCTAAACGGGTCTGTCCA/ GAAATCCACAGCACTCCG	62	260
Eucgr.H00325	S8_13_1	TCAAGAGCACGAAGAATCTCAGG/ CCAATCTTACTAGGCTTAACGCTG	59	633
Eucgr.H00329	S8_14_1	CGATGAACACGAAAGGACTC/ GGCGACGGTAAGAATAATGG	59	234
	S8_14_2	ATATGGTCTTGCTCCTACTGTACG/ GGTTGTGCTTCTCCTTCG	62	176

5 RESULTADOS E DISCUSSÃO

5.1 O caráter anômalo

As plantas anômalas, detectadas no cruzamento controlado de *Eucalyptus grandis* realizado pela empresa Suzano Papel e Celulose SA, aparecem na progênie numa razão mendeliana de 3:1 (plantas normais: plantas anormais) (Tambarussi, 2006). Baseados na segregação mendeliana do fenótipo anômalo, Tambarussi (2006) e Lourenção (2010) sugerem que a anomalia seja controlada por um gene de efeito principal. O gene de efeito principal, com padrão de herança monogênica, apresenta alelos recessivos deletérios que, em homozigose, resultam na morte prematura das plantas.

A anomalia é caracterizada por um conjunto de alterações morfológicas: redução do volume da raiz (Figura 3A), presença de ramificações laterais, redução na altura da planta (Figura 3B), redução foliar e alterações na forma da folha (Figura 3C). A forte correlação entre as alterações destas diferentes características com a anomalia, sugerem uma ação pleiotrópica do gene de efeito principal. Este tipo de pleiotropia é denominada sindrômica por Ramalho *et al.* (2004), pois existe uma série de característica que aparecem em conjunto e constitui uma síndrome, que no caso desse estudo é a anomalia.

De acordo com a tabela 3, as plantas anormais foram 3 vezes menores em altura e duas vezes inferiores em diâmetro do colo, comparadas às plantas normais. Apesar da altura reduzida, as plantas anormais poderiam apresentar dificuldades de se manterem eretas após o seu plantio no campo e o seu tombamento, devido ao reduzido diâmetro do colo, poderia resultar em morte ou deformações que comprometeriam o valor silvicultural dos indivíduos arbóreos (Cunha *et al.*, 2005). Adicionalmente, a largura e o comprimento do limbo foliar nas plantas anormais foram 3 e 5 vezes menores, respectivamente, em comparação às plantas normais (Tabela 3). Por conseguinte, a área do limbo foliar nas plantas anormais também apresentou redução, chegando a ser 18 vezes menor que as plantas normais (Tabela 3). Segundo White e Scott (2006), as características morfológicas das folhas podem afetar a expansão total da copa e, conseqüentemente, afetar a interceptação da radiação luminosa e a eficiência fotossintética da planta.

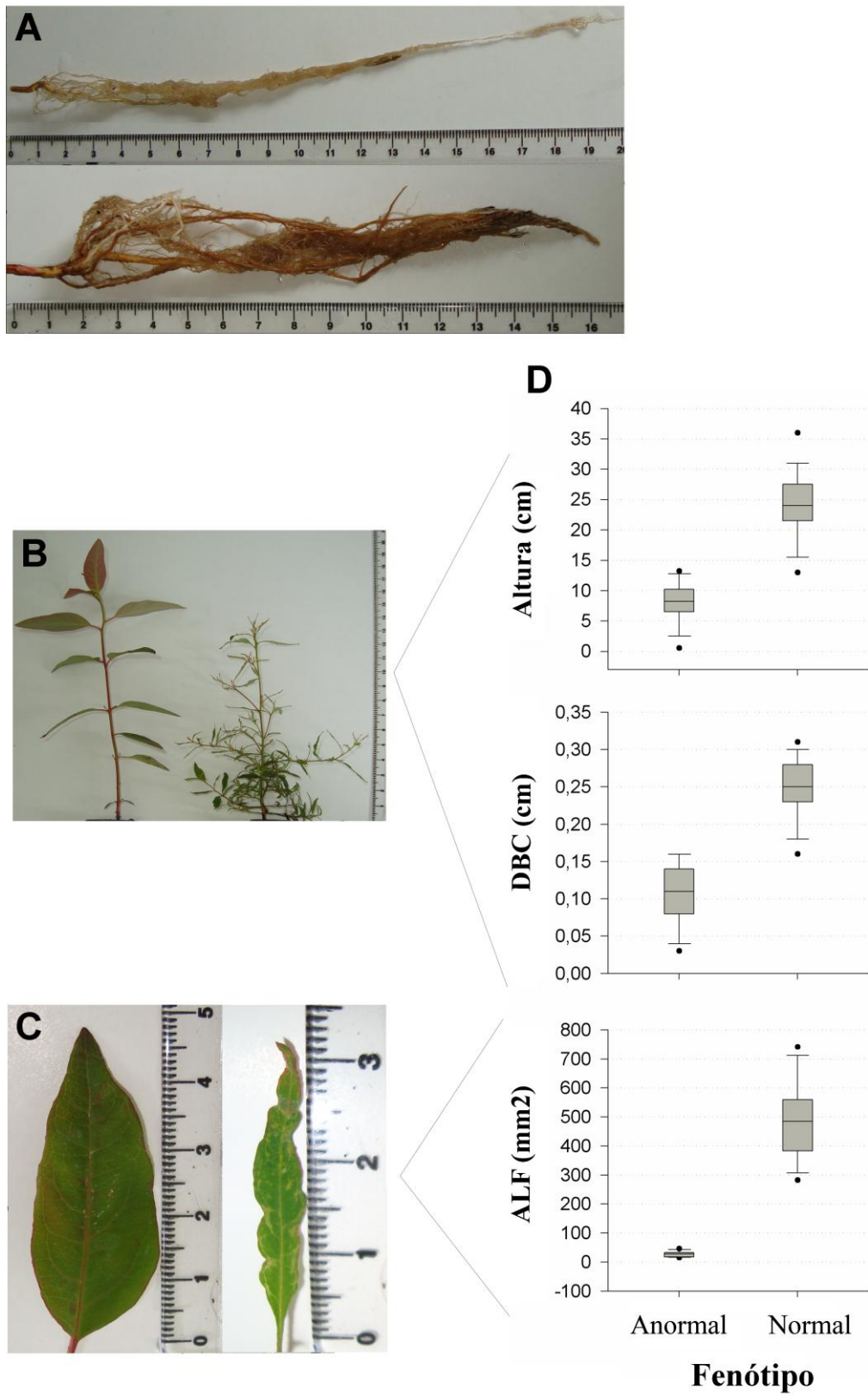


Figura 3: Morfologia das plantas normais e anormais. A: Raiz de planta normal (abaixo) e anormal (acima). B: Parte aérea de planta normal (esquerda) e anormal (direita). C: Folha normal (esquerda) e anormal (direita). D: Gráficos representativos da altura, diâmetro da base do caule (DBC) e área do limbo foliar (ALF) dos fenótipos anormal e normal.

Tabela 3: Médias das características morfológicas dos fenótipos Anormal e Normal.

Características	Anormal	Normal
Altura (cm)	8,21 b	24,07 a
Diâmetro da base do caule (cm)	0,11 b	0,25 a
Limbo Foliar: Área (mm ²)	27,49 b	484,78 a
Comprimento (mm)	15,87 b	48,4 a
Largura (mm)	3,68 b	16,48 a

Para cada característica morfológica (linha), médias seguidas de pelo menos uma mesma letra minúscula não diferem entre si pelo teste de Tukey ($P > 0,05$).

Além do limbo foliar apresentar dimensões reduzidas nas plantas anormais, este possui forma lanceolada, pontiaguda, com razão comprimento/largura maior de 8:1 e borda ondulada (Figura 3C). Morfologia diferente da observada em plantas normais juvenis de *Eucalyptus grandis*, que possui folhas ovaladas, mais amplas abaixo do meio da folha, pontiagudas no final, com razão comprimento/ largura de até 3:1 e borda lisa (Figura 3C) (Brooker e Kleinig, 2004).

Em geral, plântulas de eucalipto possuem um único caule após a germinação, assim como quando adultas (Brooker e Kleinig, 2004). Segundo esses mesmos autores, a ramificação lateral pode ocorrer em uma distância curta do nível do solo quando o eucalipto está localizado em ambientes de bosques. Como esperado, todas as plantas normais apresentaram um único caule após a germinação, no entanto, as plantas anormais possuíam ramificações laterais (Figura 3C) em um número variável de 1 a 13. Os resultados morfológicos demonstram uma clara distinção entre os fenótipos normal e anômalo detectados.

5.2 Região genômica da marca molecular

Uma vez que Tambarussi (2006) identificou uma marca molecular RAPD, de cerca de 500 pb co-segregante com o caráter anômalo (Figura 2), esta foi convertida em marcador SCAR e sequenciada por Lourenção (2010). Análises *in silico* da sequência do marcador SCAR, com base no genoma completo do eucalipto (cobertura de 8 vezes - www.phytozome.net), demonstrou que a marca está localizada na região intergênica de repetições em *tandem* de genes PR-10 (Pathogenesis-related protein 10) da superfamília Bet v1 (Figura 4). Esses dados complementam as análises *in silico* feitas previamente por Lourenção (2010), no qual a cobertura do genoma de eucalipto estava incompleta (cobertura

de 4,5 vezes) e, por isso apresentava algumas restrições. Mais detalhes sobre a sequência do marcador molecular e análises *in silico* encontram-se no Apêndice A.

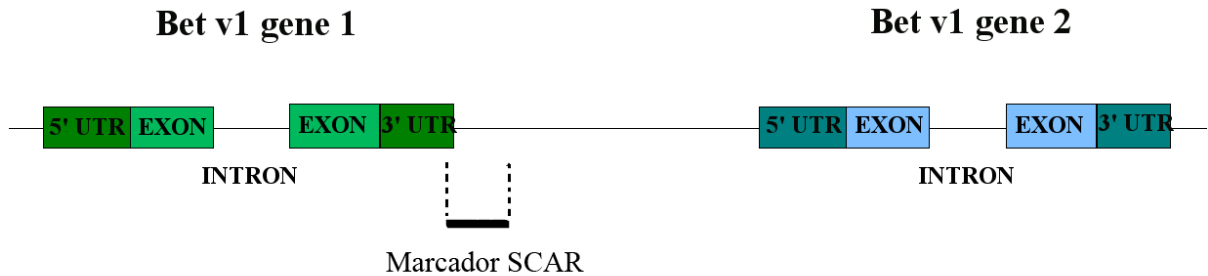


Figura 4: Esquema demonstrando o alinhamento da sequência SCAR com o genoma de *Eucalyptus grandis*.

De modo geral, as proteínas da superfamília 10 com domínio Bet v1 estão associadas à resposta de defesa contra estresses bióticos e abióticos (Hoffmann-Sommergruber *et al.*, 1997; Wen *et al.*, 1997; Liu e Ekramoddoullah, 2006), no entanto, sua função ainda não está totalmente esclarecida. Muitos estudos relatam homologia das proteínas PR-10 com ribonucleases, sugerindo que estas possam papel importante em processos biológicos, assim como no desenvolvimento (Walter *et al.*, 1996; Liu e Ekramoddoullah, 2006) e formação de raízes secundárias (Poupard *et al.*, 2001).

Apesar do marcador molecular relacionado à anomalia estar localizado em uma região de genes Bet v1 e possuir parte de sua sequência com identidade para estes genes, não era claro o envolvimento direto ou indireto destes genes com o desenvolvimento do caráter anômalo (Lourenção, 2010). Entretanto, estudos que envolvem fenótipos similares à anomalia, como o nanismo e ramificação abundante, apresentam associação com genes relacionados à resposta de defesa (Knight, 1947; Bomblies e Weigel, 2007; Bomblies *et al.*, 2007). Knight (1947) demonstra que o gene determinante para o caráter anormal (tamanho reduzido e numerosas ramificações), em algodão, está estritamente ligado ao gene de resistência à bactéria *Xanthomonas campestris* pv. *malvacearum*.

5.3 Caracterização gênica e vias metabólicas

Estudos de transcriptomas são fundamentais para a associação entre genótipo e fenótipo. Anteriormente ao sequenciamento de nova geração, os estudos de expressão gênica em alta-escala eram desenvolvidos por metodologias baseadas em hibridização (microarray) e em sequências (SAGE – *Serial Analysis of Gene Expression*, CAGE – *Cap Analysis of Gene*

Expression, e MPSS – *Massively Parallel Signature Sequencing*) (Marioni *et al.*, 2008; Mortazavi *et al.*, 2008; Wang *et al.*, 2009). No entanto, tais metodologias possuem algumas limitações como a necessidade do conhecimento prévio dos transcritos e *background* (Marioni *et al.*, 2008; Wang *et al.*, 2009).

O RNA-Seq é um método de análise de transcriptoma por sequenciamento em larga escala (Nagalakshmi *et al.*, 2008), no qual os fragmentos de cDNAs sequenciados são mapeados no genoma e analisados em programas estatísticos (Nagalakshmi *et al.*, 2008; Wang *et al.*, 2009). O mapeamento consiste no alinhamento das sequências, obtidas pelo RNA-Seq, ao genoma ou banco de dados de mRNAs ou ESTs (Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008; Severin *et al.*, 2010). Esta tecnologia oferece diversas vantagens: não é limitado, favorecendo a detecção de transcritos de menor abundância e *splicing* alternativo, e, assim, provê a análise total de transcritos ativos; possui pouco ou nenhum *background*, pois as sequências são mapeadas em uma única região do genoma; além de oferecer resolução de apenas uma base para anotações com custo menor que as técnicas anteriormente descritas (Wang *et al.*, 2009; Severin *et al.*, 2010).

A utilização da tecnologia de RNA-Seq na área florestal ainda é pouco empregada (Mizrachi *et al.*, 2010; Villar *et al.*, 2011), principalmente com relação à comparação de expressão gênica de fenótipos contrastantes. Há grande empenho de pesquisadores em desenvolver bancos de dados cada vez mais informativos de ESTs (FORESTS; NCBI; Genolyptus) e mRNA (NCBI; Eucspresso) de *Eucalyptus*. Estas informações, juntamente com o genoma completo de *Eucalyptus* (www.phytozome.net), provê ferramentas importantes na compreensão de processos biológicos como a anomalia em questão.

Para a identificação dos genes envolvidos no desenvolvimento de plantas anômalas, através da técnica de RNA-Seq, foram produzidas 6 bibliotecas de cDNA. Cada biblioteca é um *pool* de amostras composto por uma mesma quantidade de RNA total de 5 indivíduos de um mesmo fenótipo (anormal ou normal). Para cada fenótipo foram construídas três bibliotecas de cDNA correspondentes aos diferentes órgãos: folha, caule e raiz.

O sequenciamento das bibliotecas de cDNA no equipamento MiSeq Personal Sequencing System (Illumina) geraram um total de 4,87 milhões (0,63 a 1,09 milhões de *reads* de cada biblioteca) de *single-end reads* de 50 pb, compreendendo 237,4 Mpb de sequência do transcriptoma. Desse total, 94,75% dos *reads* apresentaram alta qualidade (Q score > Q30), no qual a probabilidade de ocorrer um erro no sequenciamento de uma base é 1 em 1000, ou seja, a precisão é de 99,9% (Illumina, 2011).

Os perfis transcricionais das 6 bibliotecas de cDNA foram traçados através de análises de expressão gênica digital. Esta análise consiste no mapeamento dos *reads* em um genoma de referência, contagem dos *reads* em uma anotação gênica de referência e normalização das bibliotecas. Análises de expressão gênica diferencial foram realizadas comparando cada órgão entre os fenótipos. Foram identificados genes diferencialmente expressos ($P < 0,05$) nos três órgãos: 26 em folha, 59 em caule e 11 em raiz (Figura 5). Dentre estes, foram selecionados 4 genes diferencialmente expressos em folha e caule: Eucgr.F00644 (GRP), Eucgr.L00937 (CHI I), Eucgr.B02124 (BARWIN) e Eucgr.E01382 (THAU); um apenas em folha: Eucgr.H04017 (BET vI); um apenas em caule: Eucgr.H04012 (BET vI b); e 4 genes somente em raiz: Eucgr.I01236 (GLOB), Eucgr.H00329 (CHI I b), Eucgr.A01578 (CUP) e Eucgr.L00605 (PROT) para a validação dos resultados via RT-qPCR. Com base nos dados de ΔCt , todos os genes apresentaram expressão gênica significativamente distintas pelo teste t de Student ($P < 0,05$). Estes resultados foram similares àqueles obtidos pela análise dos dados de RNA-Seq (Figura 6).

A grande maioria dos genes diferencialmente expressos apresentaram superexpressão nas plantas anormais (Figura 5). As famílias gênicas as quais estes genes pertencem foram verificadas no banco de dados Pfam 27.0 (Punta *et al.*, 2012) e, então, foram classificadas por categorias funcionais ao nível de 'Processos biológicos' conforme o termo GO - *Gene Ontology*. Uma vez que um gene pode estar em mais de uma categoria por apresentar mais de um termo GO, o número total de genes nas categorias funcionais ultrapassa o número total de genes superexpressos.

Em todos os órgãos, os genes superexpressos nas plantas anormais eram mais representativos nas categorias funcionais 'Resposta à estímulos' (46,15% - folha; 49,15% - caule; 50% - raiz) e 'Processos metabólicos' (46,15% - folha; 44,07% - caule; 72,72% - raiz) (Figura 7). Resultados similares foram encontrados no estudo de Mizuno *et al.* (2010), no qual analisa os fatores responsáveis pelo crescimento anormal (necrose do tipo III) em plantas triploides de trigo. Neste estudo, as categorias funcionais de 'Resposta à defesa' e 'Metabolismo', apresentaram 25,6% e 15,6%, respectivamente, do total de genes analisados.

De acordo com a tabela 4, as famílias gênicas mais representativas na categoria funcional 'Resposta à estímulos' são quitinase classe I, proteínas similares à taumatina e proteínas da superfamília Bet v1. As quitinases possuem papel importante na resistência contra patógenos, agindo na degradação da quitina, constituinte da parede celular dos fungos (Flach *et al.*, 1992; Kasprzewska, 2003). Por sua vez, as proteínas similares à taumatina (PR-

5; osmotina) possuem atividade na permeabilização da membrana plasmática dos patógenos (Ruiz-Medrano *et al.*, 1992; Kitajima e Sato, 1999).

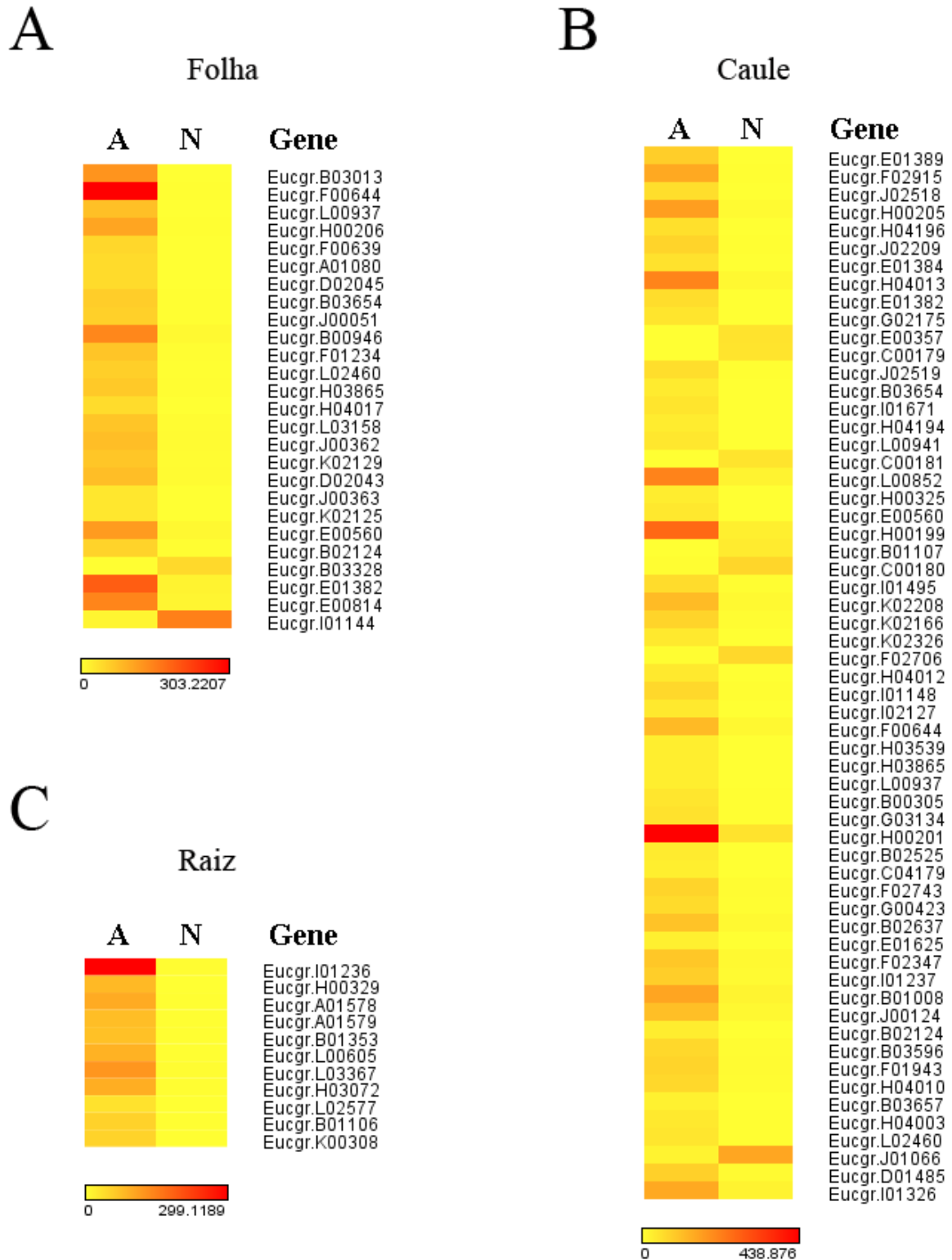


Figura 5: Agrupamento hierárquico dos genes diferencialmente expressos, com nível de significância de 5%, entre plantas normais e anormais em folha (A), caule (B) e raiz (C). A escala possui valores médios da contagem dos *reads* normalizados. A - biblioteca de plantas anormais; N - biblioteca de plantas normais.

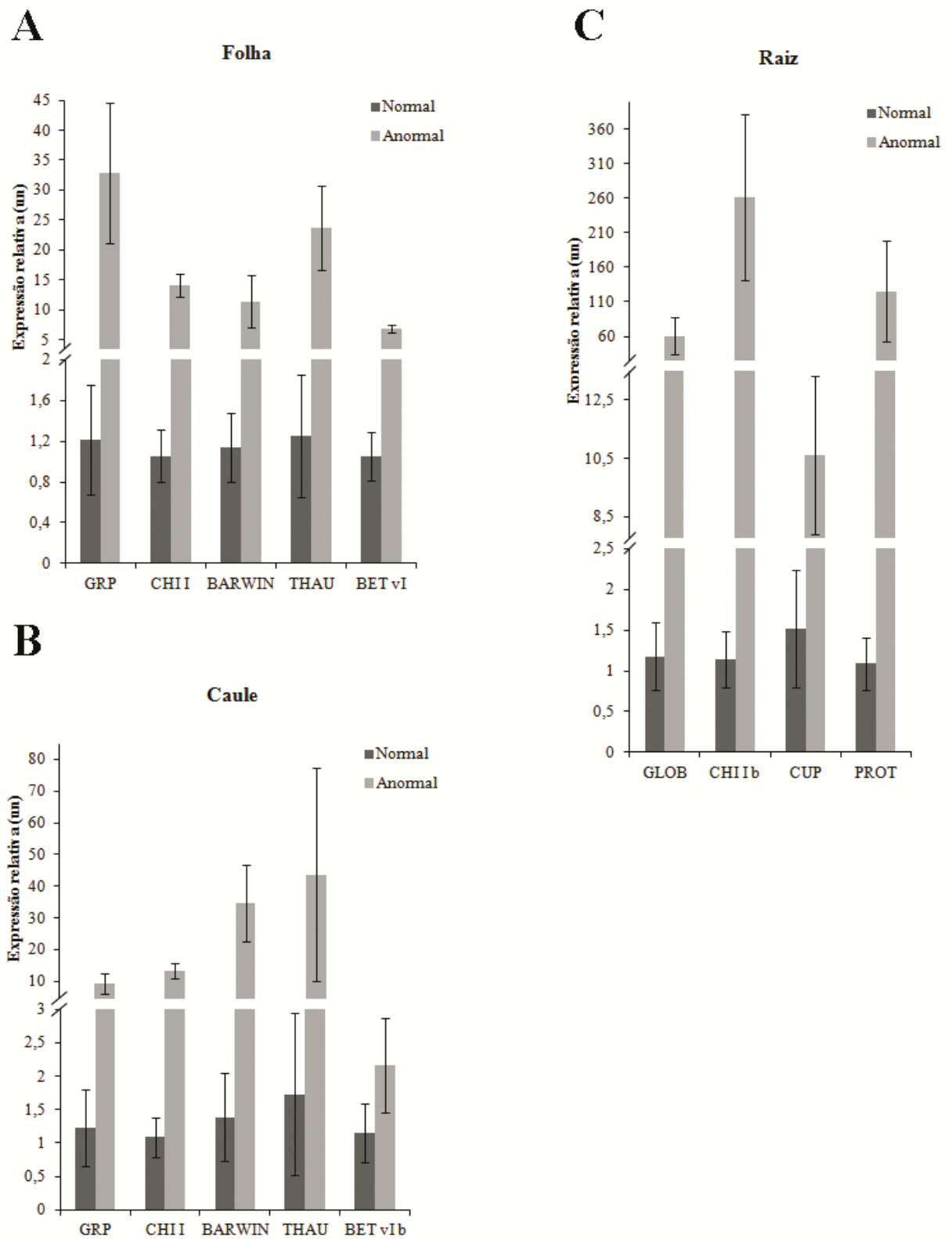


Figura 6: Gráficos de expressão relativa, com barra de erro padrão, dos genes selecionados para a validação com RT-qPCR, em folha (A), caule (B) e raiz (C).

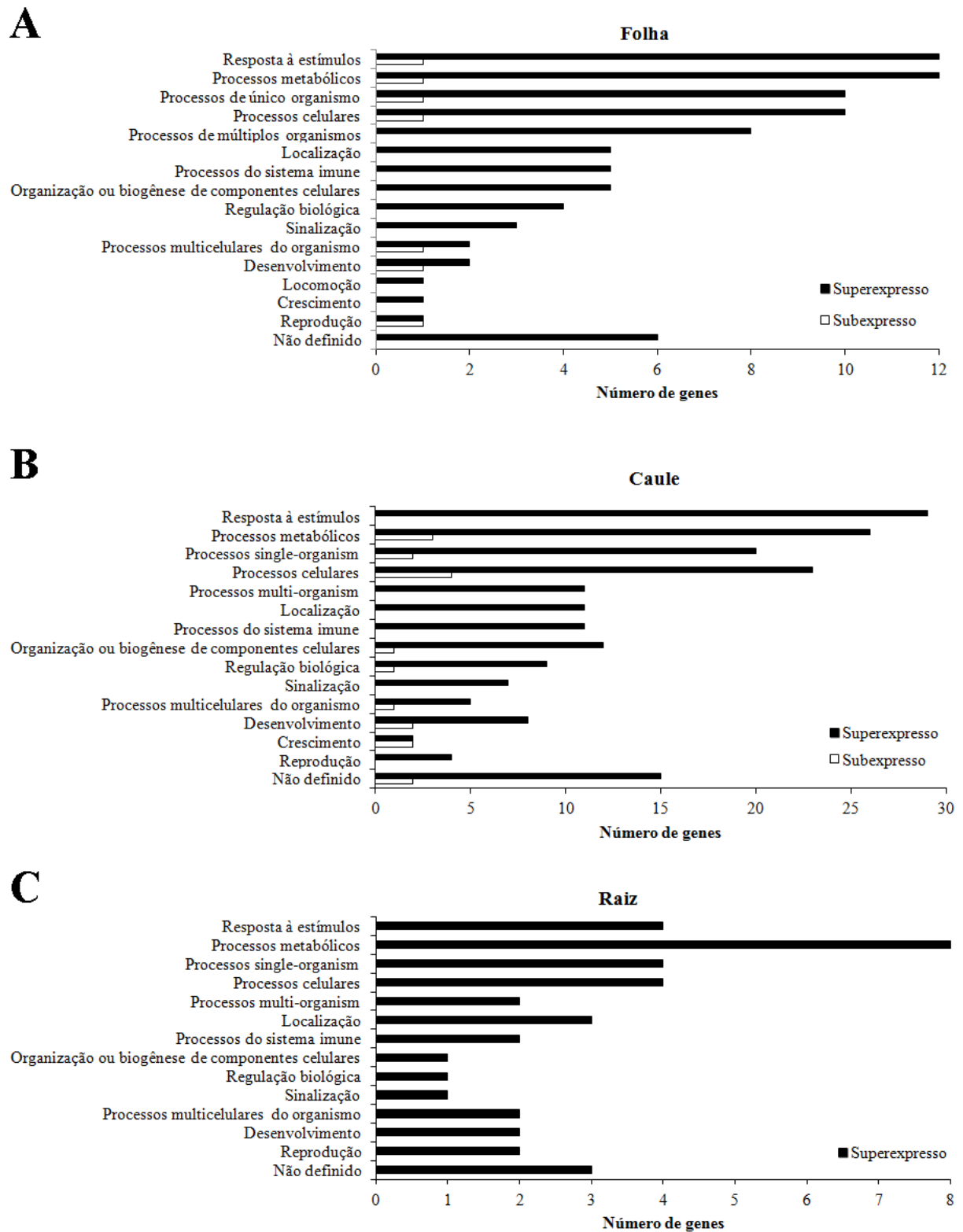


Figura 7: Distribuição dos genes diferencialmente expressos nas plantas anormais, de acordo com a categoria funcional e os órgãos da planta: folha (A), caule (B) e raiz (C).

Tabela 4: Genes diferencialmente expressos entre plantas normais e anormais, com nível de significância de 5%.

				(continua)
Gene	Descrição	n° Pfam	Categorias funcionais	Órgão
Eucgr.B03013	Ribulose-1,5-bifosfato subunidade menor/cadeia menor	PF12338; PF00101	Processos metabólicos, Resposta à estímulos, Processos Celulares e Organização ou biogênese de componentes celulares.	F
Eucgr.F00644	Família de proteínas ricas em glicina	PF07172	N/D	F e C
Eucgr.F00639	Família de proteínas ricas em glicina	PF07172	N/D	F
Eucgr.A01080	Família de proteínas ricas em glicina	PF07172	N/D	F
Eucgr.L00937	Endoquitinase básica; Quitinase classe I	PF00182	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos de múltiplos organismos, Organização ou biogênese de componentes celulares, Sinalização, Processos do sistema imune e Regulação biológica.	F e C
Eucgr.H00329	Endoquitinase PR4; Quitinase classe I	PF00182	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Processos de desenvolvimento, Organização ou biogênese de componentes celulares, Reprodução, Processos de múltiplos organismos, Localização e Processos multicelulares do organismo.	R
Eucgr.J02518	Endoquitinase básica; Quitinase classe I	PF00182	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Organização ou biogênese de componentes celulares, Regulação biológica, Processos de múltiplos organismos e Sinalização.	C
Eucgr.J02519	Endoquitinase básica; Quitinase classe I	PF00182	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Organização ou biogênese de componentes celulares, Regulação biológica, Processos de múltiplos organismos e Sinalização	C
Eucgr.L00941	Endoquitinase isoforma I; Quitinase classe I	PF00182	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo e Organização ou biogênese de componentes celulares.	C

Tabela 4: Genes diferencialmente expressos entre plantas normais e anormais, com nível de significância de 5%.

					(continua)
Gene	Descrição	n° Pfam	Categorias funcionais	Órgão	
Eucgr.H00325	Endoquitinase PR4; Quitinase classe I	PF00182	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Processos de desenvolvimento, Organização ou biogênese de componentes celulares, Reprodução, Processos de múltiplos organismos, Localização e Processos multicelulares do organismo.	C	
Eucgr.I01495	Quitinase classe I		Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Organização ou biogênese de componentes celulares, Regulação biológica, Processos de múltiplos organismos, Localização e Sinalização.	C	
Eucgr.K02166	Endoquitinase PR4; Quitinase classe I	PF00182	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Processos de desenvolvimento, Organização ou biogênese de componentes celulares, Reprodução, Localização e Processos multicelulares do organismo.	C	
Eucgr.H03865	Família de proteínas similares à taumatina	PF00314	Resposta à estímulos, Processos de múltiplos organismos e Processos do sistema imune.	F e C	
Eucgr.E00560	Família de proteínas similares à taumatina	PF00314	N/D	F e C	
Eucgr.E01382	Família de proteínas similares à taumatina	PF00314	N/D	F e C	
Eucgr.E01389	Família de proteínas similares à taumatina	PF00314	N/D	C	
Eucgr.E01384	Família de proteínas similares à taumatina	PF00314	N/D	C	
Eucgr.H04017	Proteína relacionada à patogenicidade Bet vI (sth-2)	PF00407	Resposta à estímulos.	F	
Eucgr.H04013	Proteína relacionada à patogenicidade Bet vI (Pru ar 1)	PF00407	Resposta à estímulos.	C	
Eucgr.G02175	Proteína relacionada à patogenicidade Bet vI (MLP 328)	PF00407	Resposta à estímulos.	C	
Eucgr.H04012	Proteína relacionada à patogenicidade Bet vI (Pru ar 1)	PF00407	Resposta à estímulos.	C	
Eucgr.H04010	Proteína relacionada à patogenicidade Bet vI (Pru ar 1)	PF00407	Resposta à estímulos.	C	
Eucgr.H04003	Proteína relacionada à patogenicidade Bet vI (Pru ar 1)	PF00407	Resposta à estímulos.	C	
Eucgr.H00206	Subunidade 28 do complexo mediador	PF11594	N/D	F	
Eucgr.L00852	Subunidade 28 do complexo mediador	PF11594	N/D	C	

Tabela 4: Genes diferencialmente expressos entre plantas normais e anormais, com nível de significância de 5%.

					(continua)
Gene	Descrição	n° Pfam	Categorias funcionais	Órgão	
Eucgr.H00199	Subunidade 28 do complexo mediador	PF11594	N/D	C	
Eucgr.H00201	Subunidade 28 do complexo mediador	PF11594	N/D	C	
Eucgr.D02045	Proteína rica em prolina de 14kda	PF00234	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Processos de desenvolvimento, Processos de múltiplos organismos, Localização e Processos multicelulares do organismo.	F	
Eucgr.D02043	Proteína rica em prolina de 14kda	PF00234	Resposta à estímulos, Processos de único organismo, Processos do sistema imune, Processos de múltiplos organismos e Localização.	F	
Eucgr.B03657	β - glucosidase 12; Família 1 das glicosil hidrolases	PF00232	Processos metabólicos.	C	
Eucgr.B03654	β - glucosidase 12; Família 1 das glicosil hidrolases	PF00232	Resposta à estímulos, Processos metabólicos, Processos de único organismo, Processos do sistema imune, Processos de múltiplos organismos e Localização.	F e C	
Eucgr.J00051	Aquaporina tipo 1-3	PF00230	Resposta à estímulos, Processos de múltiplos organismos, Processos celulares, Processos de único organismo, Localização e Regulação biológica .	F	
Eucgr.L02460	Peroxidase n1	PF00141	Resposta à estímulos, Processos metabólicos e Processos de múltiplos organismos.	F e C	
Eucgr.B00946	Fosfoetanolamina n-metiltransferase 1	PF08241	Resposta à estímulos, Processos metabólicos, Processos de múltiplos organismos, Processos celulares, Organização ou biogênese de componentes celulares, Processos de único organismo, Processos de desenvolvimento, Regulação biológica, Reprodução, Crescimento, Locomoção, Processos multicelulares do organismo e Sinalização.	F	
Eucgr.K02125	Quinona-oxidoreductase cloroplástica	PF08240; PF00107	Processos metabólicos	F	
Eucgr.F01234	Glicoproteína epidérmica específica	PF01453; PF08276	N/D	F	
Eucgr.L03158	Miraculina; Inibidor de tripsina e protease	PF00197	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo e Localização.	F	

Tabela 4: Genes diferencialmente expressos entre plantas normais e anormais, com nível de significância de 5%.

				(continua)
Gene	Descrição	n° Pfam	Categorias funcionais	Órgão
Eucgr.K02326	Miraculina; Inibidor de tripsina e protease	PF00197	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo e Localização.	C
Eucgr.B02124	Proteína relacionada à patogenicidade - PR4; Família <i>Barwin</i>	PF00967	Resposta à estímulos, Processos de múltiplos organismos, Processos de único organismo, Localização e Processos do sistema imune.	F e C
Eucgr.J00362	Omega-hidroxipalmitato O-feruloil transferase	PF02458	Processos metabólicos, Processos celulares, Organização ou biogênese de componentes celulares e Processos de único organismo.	F
Eucgr.J00363	Omega-hidroxipalmitato O-feruloil transferase	PF02458	Processos metabólicos, Processos celulares, Organização ou biogênese de componentes celulares e Processos de único organismo.	F
Eucgr.I01144	Oxidase de redução férrica cloroplástica	PF01794; PF08030; PF08022	Resposta à estímulos, Processos metabólicos e Processos celulares	F*
Eucgr.I01236	Globina	PF00042	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Regulação biológica, Processos de múltiplos organismos, Localização e Sinalização.	R
Eucgr.I01237	Globina	PF00042	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Regulação biológica, Processos de múltiplos organismos, Localização e Sinalização.	C
Eucgr.A01578	Cupina	PF00190	N/D	R
Eucgr.A01579	Cupina	PF00190	N/D	R
Eucgr.B01353	Cupina	PF00190	Processos metabólicos	R
Eucgr.L03367	Cupina	PF00190	Processos metabólicos	R
Eucgr.H04196	Cupina	PF00190	N/D	C
Eucgr.H04194	Cupina	PF00190	N/D	C
Eucgr.E00814	Proteína <i>Dirigent-like</i> ; Proteína de resposta à doenças	PF03018	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Regulação biológica e Sinalização.	F

Tabela 4: Genes diferencialmente expressos entre plantas normais e anormais, com nível de significância de 5%.

				(continua)
Gene	Descrição	n° Pfam	Categorias funcionais	Órgão
Eucgr.K00308	Endoquitinase ácida; Família 18 das glicosil hidrolases	PF00704	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos multicelulares do organismo, Processos de desenvolvimento e Reprodução.	R
Eucgr.L00605	Globulina básica 7s; Protease; peptidase	N/D	Processos metabólicos.	R
Eucgr.B01106	Globulina básica 7s; Protease; peptidase	N/D	Processos metabólicos.	R
Eucgr.F02915	Glutathiona S-transferase f11	PF00043; PF02798	Resposta à estímulos, Processos metabólicos, Processos celulares e Regulação biológica.	C
Eucgr.J02209	Formato desidrogenase mitocondrial	PF02826; PF00389	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos de desenvolvimento e Processos multicelulares do organismo.	C
Eucgr.L02577	Família EXS; Transportador de fosfato pho 1	PF03124	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo e Localização.	R
Eucgr.K02208	Família PLAC8;	PF04749	Resposta à estímulos.	C
Eucgr.E00357	Expansina	PF01357; PF03330	Processos celulares, Processos de único organismo, Processos de desenvolvimento, Crescimento e Organização ou biogênese de componentes celulares.	C*
Eucgr.G03134	Alfa-expansina 8	PF01357; PF03330	Processos celulares, Processos de único organismo, Processos de desenvolvimento, Crescimento e Organização ou biogênese de componentes celulares.	C
Eucgr.E01625	Precursor da beta-expansina	PF01357; PF03330	Processos celulares, Processos de único organismo, Processos de desenvolvimento, Crescimento, Organização ou biogênese de componentes celulares e reprodução.	C
Eucgr.C00179	xiloglucano endotransglicosilase-hidrolase; Família 16 das glicosil hidrolases	PF00722	Processos metabólicos e Processos celulares.	C*
Eucgr.C00181	xiloglucano endotransglicosilase-hidrolase; Família 16 das glicosil hidrolases	PF00722	Processos metabólicos e Processos celulares.	C*
Eucgr.C00180	xiloglucano endotransglicosilase-hidrolase; Família 16 das glicosil hidrolases	PF00722	Processos metabólicos e Processos celulares.	C*

Tabela 4: Genes diferencialmente expressos entre plantas normais e anormais, com nível de significância de 5%.

(continua)				
Gene	Descrição	n° Pfam	Categorias funcionais	Órgão
Eucgr.I01671	Transportador de lisina histidina 1	PF01490	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Organização ou biogênese de componentes celulares, Regulação biológica, Localização e Sinalização.	C
Eucgr.J01066	Transportador vacuolar de aminoácidos 1	PF01490	N/D	C*
Eucgr.F02706	Família transportadora <i>EamA-like</i> ; proteína induzida por auxina 5ng4	PF00892	N/D	C*
Eucgr.I01148	Ferredoxina-nitrito cloroplástica	PF03460; PF01077	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo e Localização.	C
Eucgr.H03539	Transportador de açúcares (e outros)	PF00083	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo e Localização.	C
Eucgr.B02637	Família Aldo-keto redutase	PF00248	Processos metabólicos.	C
Eucgr.B00305	Poli (ADP-ribose)	PF00644; PF12174	Resposta à estímulos, Processos metabólicos e Processos celulares.	C
Eucgr.F02347	Alfa-amilase	PF00128; PF07821	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo e Localização.	C
Eucgr.B03596	Citocromo P450	PF00067	Processos metabólicos.	C
Eucgr.K02129	Citocromo P450	PF00067	Processos metabólicos e Processos celulares.	F
Eucgr.J00124	Superóxido dismutase	PF00081; PF02777	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Localização, Processos de desenvolvimento, Reprodução, Organização ou biogênese de componentes celulares, Processos de múltiplos organismos e Processos multicelulares do organismo.	C
Eucgr.C04179	Proteína <i>Early nodulin 93 ENOD93</i>	PF03386	Processos de múltiplos organismos.	C
Eucgr.F02743	Ácido graxo dessaturase	PF03405	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos do sistema imune, Processos de desenvolvimento, Organização ou biogênese de componentes celulares, Regulação biológica, Processos de múltiplos organismos e Sinalização.	C

Tabela 4: Genes diferencialmente expressos entre plantas normais e anormais, com nível de significância de 5%.

				(conclusão)
Gene	Descrição	n° Pfam	Categorias funcionais	Órgão
Eucgr.B01008	Proteína ligante de clorofila A-B	PF00504	Resposta à estímulos, Processos metabólicos, Processos celulares, Processos de único organismo, Processos de desenvolvimento, Regulação biológica e Processos multicelulares do organismo.	C
Eucgr.D01485	Proteína de domínio MATH	PF00917	N/D	C
Eucgr.F01943	Receptor quinase associado à parede celular	PF08488; PF07714; PF07645	Processos metabólicos e Processos celulares	C
Eucgr.I01326	Frutose-bisfosfato aldolase classe I	PF00274	Resposta à estímulos, Processos metabólicos e Processos celulares	C
Eucgr.H03072	N/D	N/D	N/D	R
Eucgr.H00205	N/D	N/D	N/D	C
Eucgr.B01107	Proteína ancorada ao GPI	N/D	Processos de único organismo, Processos de desenvolvimento, Crescimento, Regulação biológica e Processos multicelulares do organismo.	C*
Eucgr.I02127	N/D	N/D	N/D	C
Eucgr.B02525	Proteína grip22 relacionada ao amadurecimento	N/D	N/D	C
Eucgr.G00423	N/D	N/D	N/D	C

* Superexpresso em plantas normais.

F - Folha; C - Caule; R - raiz; N/D - Não determinado.

Alguns genes, apesar de não pertencerem à categoria funcional de 'Resposta à estímulos', também podem estar envolvidos em processos de defesa. O citocromo P450 atua na conversão do ácido benzóico em ácido salicílico, importante componente da resposta de defesa (Hammond-Kosack e Jones, 1996). Assim como a peroxidase n1 que, através de seu código enzimático (EC 1.11.1.7), foi mapeado pela base KEGG como participante da via de biossíntese de fenilpropanoides.

Compostos fenilpropanoides são sintetizados a partir da fenilalanina, pela ação da peroxidase, e estão associados à formação da lignina, que é um dos componentes estruturais da parede celular secundária das plantas (Sato *et al.*, 1993; Croteau *et al.*, 2000; Taiz e Zeiger, 2013). A lignificação da parede celular é uma estratégia de defesa, tornando-a mais resistente ao ataque de patógenos (Hammond-Kosack e Jones, 1996; Durner *et al.*, 1997). É importante ressaltar que a inibição de peroxidases pelo ácido salicílico não se aplica às peroxidases relacionados à biossíntese de lignina (Hammond-Kosack e Jones, 1996). No entanto, a lignificação precoce da parede celular impede a expansão celular e, conseqüentemente, restringe o crescimento vegetal (Taiz e Zeiger, 2013), podendo ser responsável pela redução na altura e diâmetro da base do caule das plantas anormais. Outros genes nesta circunstâncias são: cupina (Dunwell *et al.*, 2004), superóxido dismutase (Scandalios, 1993), ácido graxo dessaturase (Kachroo *et al.*, 2001) e receptores quinase associado à parede celular (He *et al.*, 1998).

As análises de expressão gênica diferencial revelaram a superexpressão de genes da superfamília Bet v1 nas plantas anormais. Pelas análises *in silico* desses genes foi possível localizá-los nos *scaffolds* 7 (Eucgr.G02175) e 8 (Eucgr.H04017, Eucgr.H04013, Eucgr.H04012, Eucgr.H04010 e Eucgr.H04003). A família Bet v1 é representada por aproximadamente 129 genes em *Eucalyptus*, segundo o banco de dados *EucGenIE: The Eucalyptus Genome Integrative Explorer* (<http://anjie.bi.up.ac.za/eucGenIE/EucGenIE.html>), estando os mesmos distribuídos em diferentes *scaffolds* no genoma. Esse grande número de representantes de uma mesma família gênica pode ser atribuído à duplicação genômica ancestral em *Eucalyptus* (Myburg *et al.*, 2011), no qual a espécie seria um poliploide antigo ou residual que comporta-se atualmente como diploide, sendo denominada como paleopoliploide (Schifino-Wittmann, 2004; Paterson *et al.*, 2010; Levy *et al.*, 2013).

Uma das grandes dificuldades da genômica funcional atual é o entendimento da dinâmica de expressão de famílias gênicas (Guan *et al.*, 2014; Coate e Doyle, 2013). A marca molecular relacionada à anomalia está localizada no *scaffold* 1, flanqueada por genes Bet v1 dispostos em *tandem* (Figura 4). Contudo, as análises de expressão gênica diferencial

apresentaram apenas os genes para cupina (Eucgr.A01578 e Eucgr.A01579) e para proteínas ricas em glicina (Eucgr.A01080) com diferença de expressão significativa (Tabela 4) neste *scaffold*. Estudos mais detalhados deverão ser realizados para um maior entendimento da participação do *scaffold* 1 no fenótipo estudado. Uma das hipóteses sugeridas é que alguma região genômica próxima à marca molecular esteja envolvida no desenvolvimento da anomalia.

Uma vez que os três órgãos apresentaram superexpressão da família gênica quitinase classe I (Tabela 4), foi analisado os níveis de expressão de quitinase com oligonucleotídeos descritos por Boava *et al.* (2010) em cada órgão. Com esta análise, verificou-se que a expressão diferencial da quitinase foi significativa apenas na folha e no caule (Figura 8). Pela análise *in silico* do oligonucleotídeo, através da ferramenta BLAST (Altschul *et al.*, 1997) no banco de dados do Phytozome v9.1, foi verificado que a quitinase amplificada é pertencente à família glicosil hidrolase 18. Esta família consiste de quitinases das classes III e V, diferentes da quitinase classe I (Kasprzewska, 2003) encontrada na análise dos dados de RNA-Seq. Dessa forma, os diferentes padrões de expressão da quitinase (glicosil hidrolase 18) e a divergência de genes diferencialmente expressos nos três órgãos, indicam que cada parte da planta exhibe alterações fisiológicas e expressão de quitinases diferentes durante o desenvolvimento da anomalia.

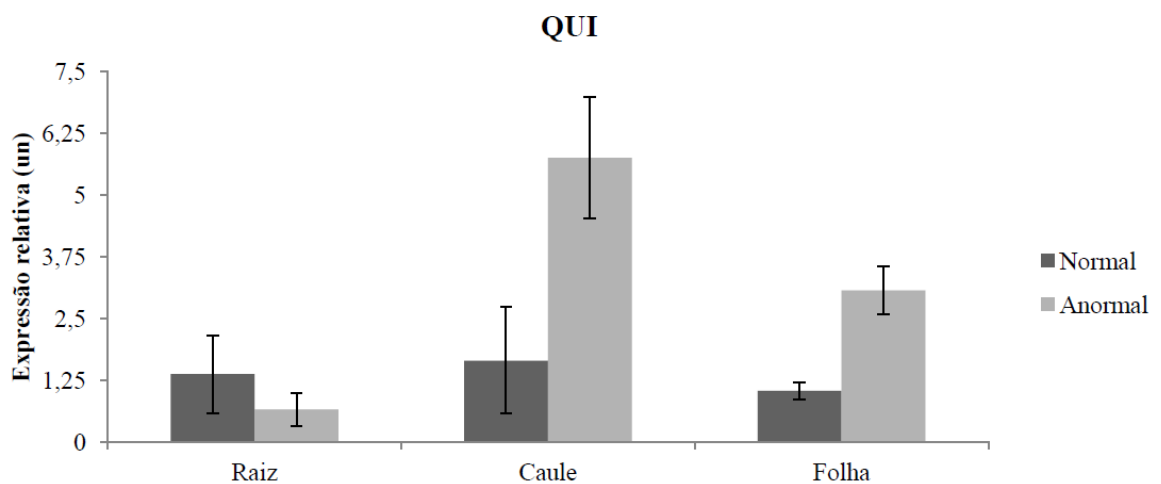


Figura 8: Expressão relativa, com barra de erro padrão, para o gene de quitinase com oligonucleotídeos descritos por Boava *et al.* (2010).

As análises de enriquecimento funcional consistem na comparação dos termos GO dos genes diferencialmente expressos com os termos GO de uma anotação de referência,

permitindo determinar os termos mais representativos (Tabela 5). Nesta análise, ficou evidente o predomínio de termos envolvidos com os mecanismos de defesa nos genes superexpressos das plantas anormais.

Tabela 5: Termos GO enriquecidos dos genes superexpressos em plantas anormais, nos diferentes órgãos avaliados.

GO ID	Termo GO	<i>p-value</i> corrigido	Órgão
GO:0004866	Atividade inibidora de peptidase	3,3546E-2	F
		3,5307E-2	C
GO: 0042742	Resposta de defesa contra bactéria	3,3546E-2	F
		4,2207E-2	C
GO: 0050832	Resposta de defesa contra fungo	3,3546E-2	F
		4,2207E-2	C
GO: 0006032	Processos catabólicos da quitina	6,9395E-13	C
		4,0370E-2	R
GO: 0004568	Atividade quitinase	6,9395E-13	C
		4,0370E-2	R
GO: 0045735	Atividade de reserva de nutrientes	4,2207E-2	C
		1,6721E-6	R
GO: 0016984	Atividade ribulose-bifosfato carboxilase	3,3546E-2	F
GO: 0015977	Fixação de carbono	3,3546E-2	F
GO: 0016747	Atividade acil-transferase	4,5042E-2	F
GO: 0016998	Processos catabólicos de macromoléculas da parede celular	5,7941E-10	C
GO: 0008061	Ligação à quitina	9,7289E-6	C
GO: 0009607	Resposta à estímulos bióticos	9,7289E-6	C
GO: 0006952	Resposta de defesa	9,9566E-6	C
GO: 0004556	Atividade alfa-amilase	3,5307E-2	C
GO: 0004784	Atividade superóxido dismutase	4,2005E-2	C
GO: 0004332	Atividade frutose-bifosfato aldolase	4,2005E-2	C
GO: 0003950	Atividade NAD ⁺ ADP-ribosiltransferase	4,2005E-2	C
GO: 0005509	Ligação à íons cálcio	4,2005E-2	C
GO: 0006801	Processos metabólicos com superóxido	4,5448E-2	C
GO: 0043169	Ligação à cátions	4,5448E-2	C
GO: 0004190	Atividade similar à endopeptidase aspártica	7,4105E-3	R

p-value ≤ 0,05; F - folha; C - caule; R - raiz.

Na folha (Figura 9), os termos GO enriquecidos implicam no acionamento de reações em resposta ao ataque de fungos e bactérias e no aumento da catálise da produção do composto 3-fosfo-D-glicerato. O 3-fosfo-D-glicerato participa do ciclo de Calvin-Benson, gerando eritrose-4-fosfato, que é precursor do corismato (Coruzzi e Last, 2000; Malkin e Nyogi, 2000; Taiz e Zeiger, 2013). Os compostos ácido salicílico e as fitoalexinas são derivados do corismato, e ambos participam na respostas de defesa em plantas (Weaver e Herrmann, 1997; Wildermuth *et al.*, 2001; Shah, 2003).

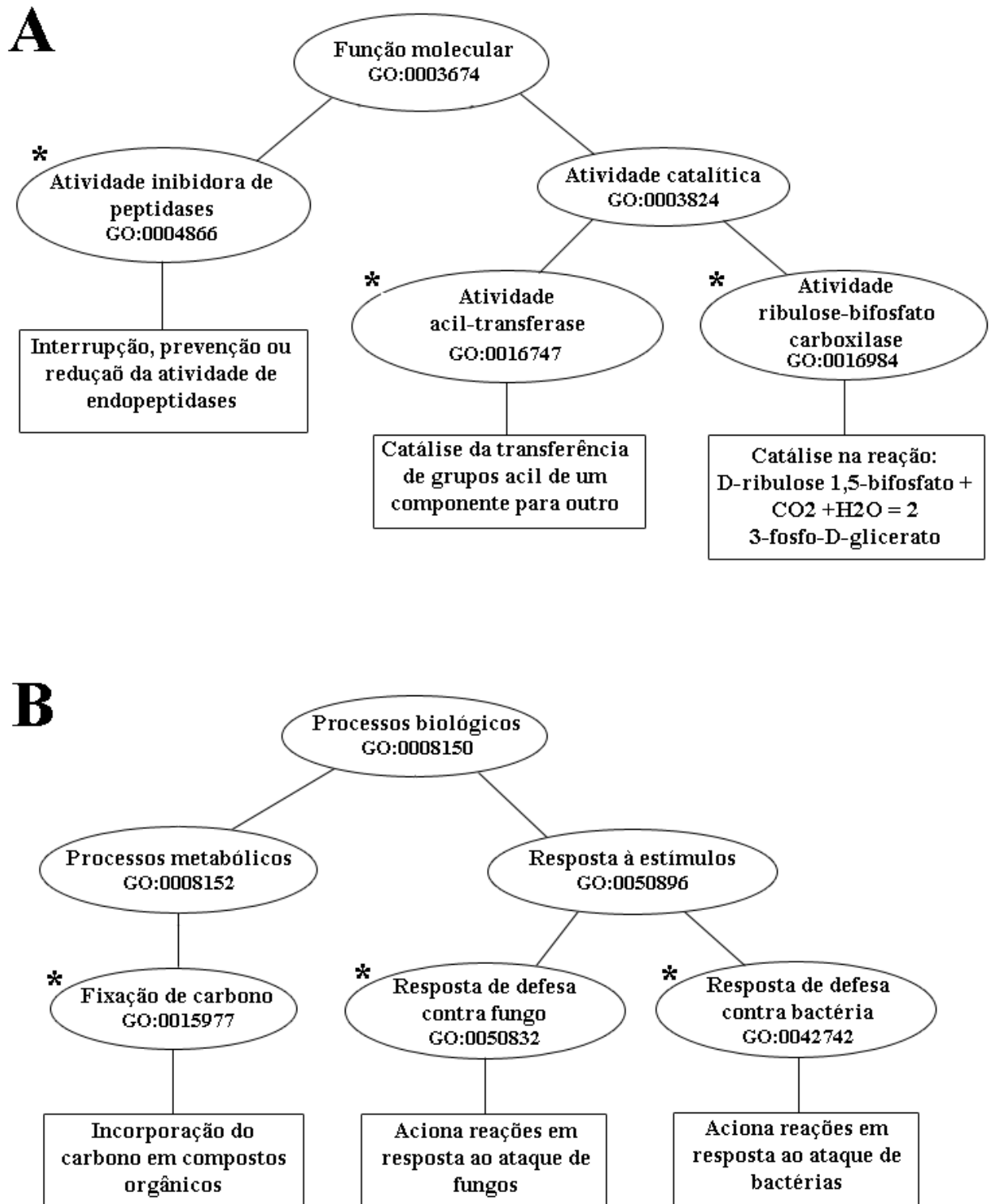


Figura 9: Diagrama demonstrando os termos GO enriquecidos (asteriscos) na folha, relacionados aos níveis de 'Função molecular' (A) e 'Processos biológicos' (B) dos genes superexpressos de plantas anormais.

Na raiz (Figura 10), destaca-se os termos GO relacionados à quebra de moléculas de quitina que são constituintes da parede celular dos fungos (Flach *et al.*, 1992; Kasprzewska, 2003). O caule foi o órgão com mais termos GO relevantes, possuindo alguns termos em comum com a raiz e a folha (Tabela 5). Assim como na folha e na raiz, o caule apresentou termos GO enriquecidos relacionados à estímulos e à degradação da quitina (Figuras 11 e 12), respectivamente. Em adição, estes termos também estão envolvidos em processos de ligações ao íon cálcio e em processos que envolvem o composto superóxido. O aumento na atividade superóxido dismutase eleva os níveis de peróxido de hidrogênio no organismo, componente importante na resposta hipersensitiva e na resistência sistêmica adquirida (Hammond-Kosack e Jones, 1996; Heath, 2000; Durrant e Dong, 2004).

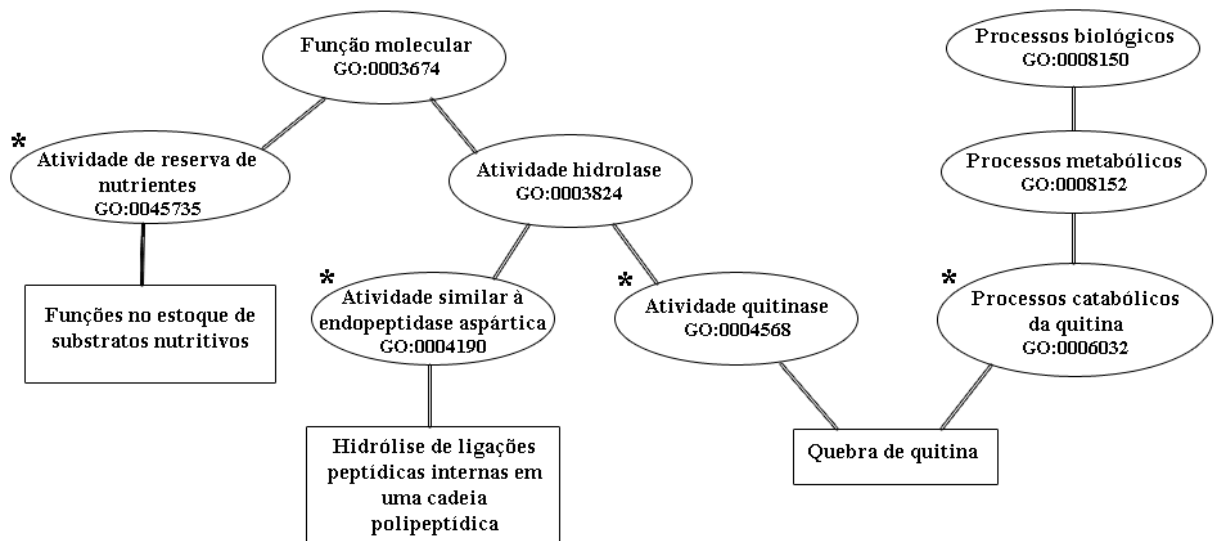


Figura 10: Diagrama demonstrando os termos GO enriquecidos (asteriscos) na raiz, relacionados aos níveis de 'Função molecular' e 'Processos biológicos' dos genes superexpressos de plantas anormais.

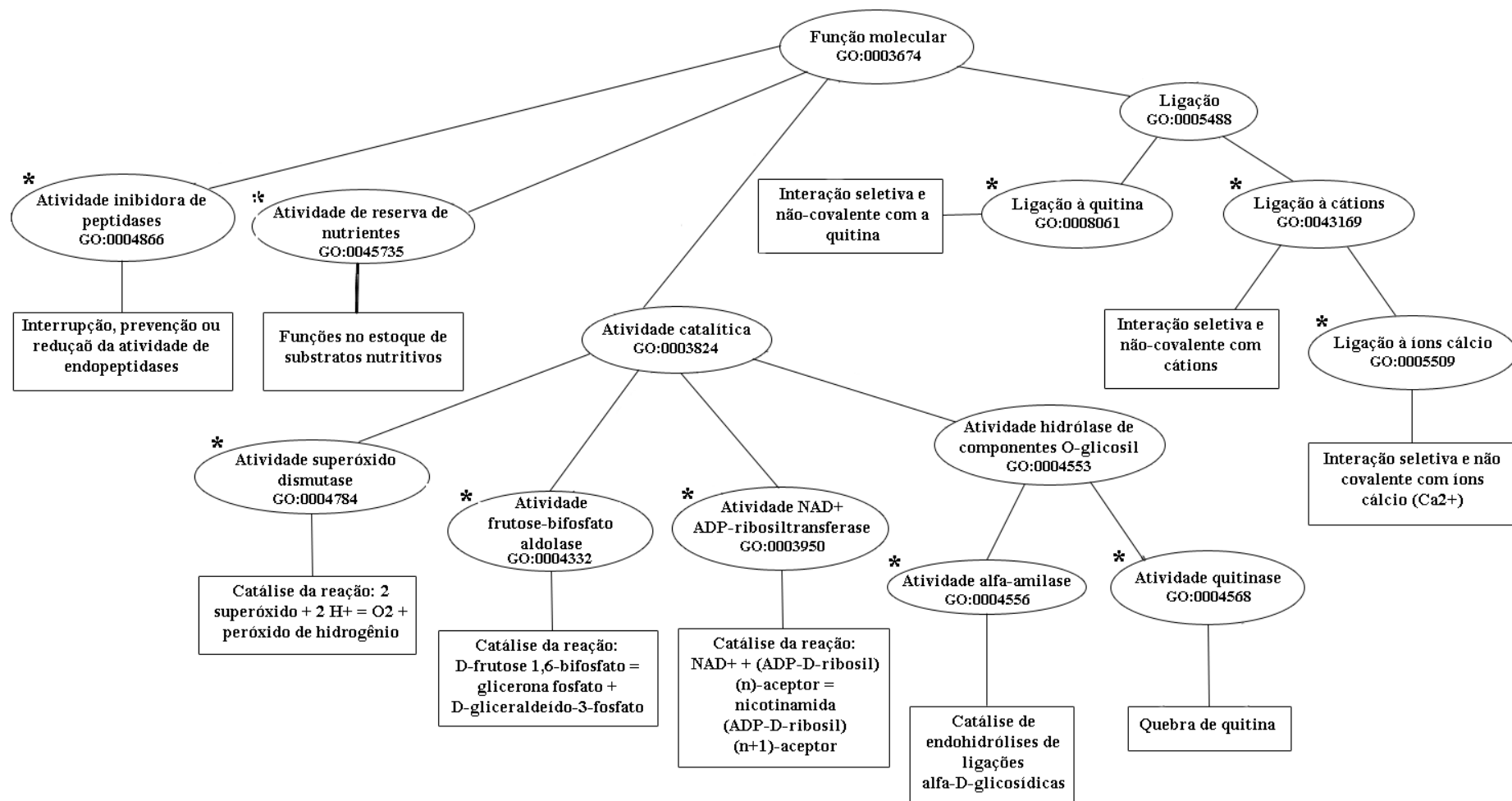


Figura 11: Diagrama demonstrando os termos GO enriquecidos (asteriscos) no caule, relacionados ao nível de 'Função molecular' dos genes superexpressos de plantas anormais.

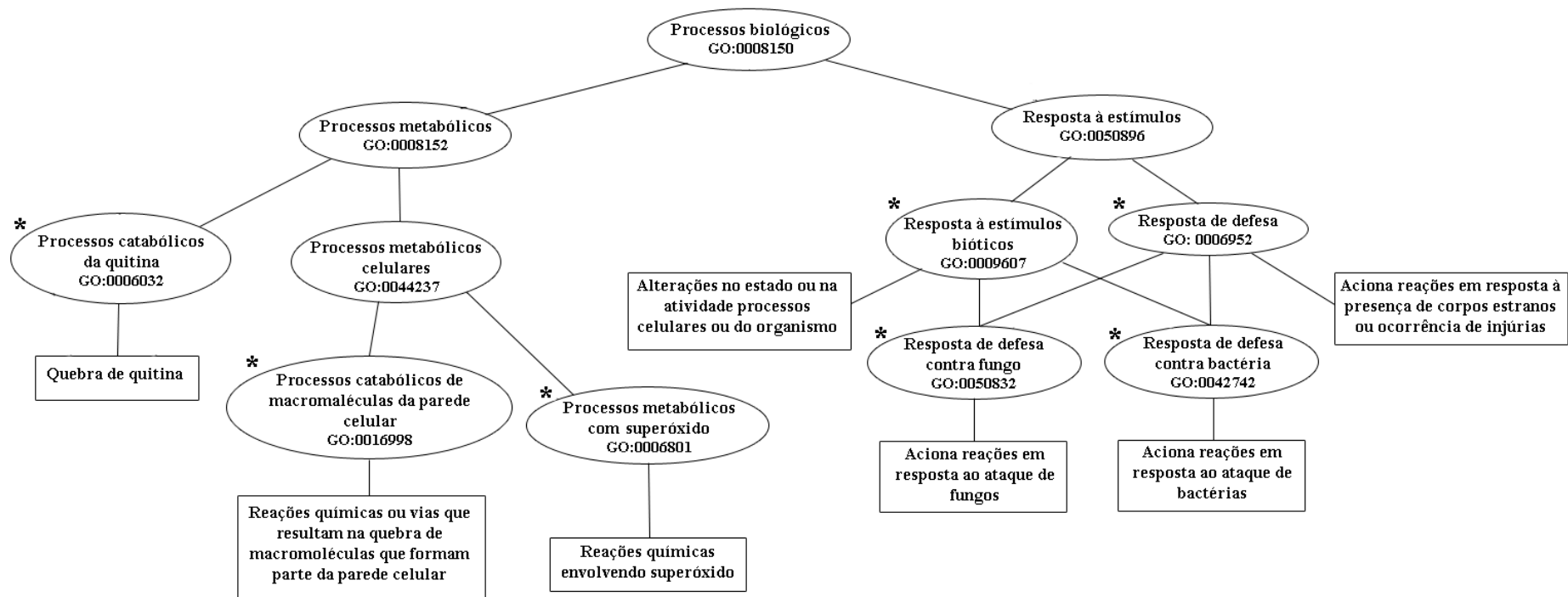


Figura 12: Diagrama demonstrando os termos GO enriquecidos (asteriscos) no caule, relacionados ao nível de 'Processos biológicos' dos genes superexpressos de plantas anormais.

Informações sobre as vias metabólicas relevantes nos três órgãos (Tabela 6) foram obtidas nas bases de dados MetaCyc (Caspi *et al.*, 2012) e PlanCyc (www.plantcyc.org). Apesar do caule apresentar mais genes expressos diferencialmente (Figura 5), o órgão que apresentou mais vias metabólicas relevantes entre plantas normais e anormais foi a folha. Algumas vias aparecem em comum nos três órgãos: degradação de lisina II, ciclo da ureia, metabolismo de citrulina e biossíntese de glicogênio II. Assim como nas outras análises (expressão gênica diferencial e termos GO enriquecidos), as plantas anormais apresentaram superexpressão de vias metabólicas relacionadas à resposta de defesa.

Tabela 6: Vias metabólicas expressas diferencialmente entre os fenótipos normal e anormal.

(continua)

Código	nome	Z-score	p-value	Órgão
ARGSYN-PWY	Biossíntese de arginina I	1,874	0,030	F
		1,773	0,038	C
ARO-PWY	Biossíntese de corismato I	1,889	0,029	F
		1,708	0,044	R
CALVIN-PWY	Ciclo de Calvin-Benson-Bassham	2,571	0,005	F
CHLOROPHYLL-SYN	Biossíntese de clorofila a I	4,123	0,00002	F
FAO-PWY	Oxidação de ácidos graxos I	-1,655	0,049	F
		-2,038	0,021	C
GLYCLEAV-PWY	Complexo de quebra de glicina	2,16	0,015	F
		1,938	0,026	R
LIPASYN-PWY	Fosfolipases	-1,763	0,039	F
LYSINE-DEG1-PWY	Degradação de lisina II	-2,475	0,007	F
		-2,032	0,021	R
LYSINE-DEG2-PWY	Degradação de lisina II	-2,299	0,011	F
		-2,091	0,018	C
		-1,888	0,029	R
PWY0-501	Biossíntese e incorporação de lipoato I	1,717	0,043	F
PWY-1187	Biossíntese de glicosinolatos a partir de homometionina	1,682	0,046	F
PWY-1422	Biossíntese de vitamina E	-2,833	0,002	F
		-2,838	0,002	R
PWY-181	Fotorespiração	1,835	0,033	F
PWY-2201	Transformação de folatos	1,886	0,03	F
		2,072	0,019	C
PWY-2821	Biossíntese de glicosinolatos a partir de fenilalanina	2,428	0,008	F
PWY-3841	Biossíntese de formil-THF II	1,951	0,025	F
PWY-4661	Biossíntese de fitato independente de lipídeo	-1,897	0,029	F
		-1,667	0,048	R
PWY-4984	Ciclo da ureia	1,945	0,026	F
		2,043	0,02	R

Tabela 6: Vias metabólicas expressas diferencialmente entre os fenótipos normal e anormal.

Código	nome	Z-score	p-value	(conclusão)
				Órgão
		1,913	0,028	C
PWY-5004	Metabolismo de citrulina	1,98	0,024	F
		1,747	0,04	R
		1,957	0,025	C
		-2,108	0,017	F
PWY-5067	Biossíntese de glicogênio II	-1,813	0,035	R
		-2,05	0,02	C
		-1,646	0,049	F
PWY-5080	Biossíntese de ácidos graxos de cadeia longa	-1,646	0,049	F
PWY-5136	β -oxidação de ácidos graxos II	-2,211	0,013	F
PWY-5188	Biossíntese de tetrapirrol I	2,591	0,005	F
		2,771	0,003	R
PWY-6124	Biossíntese de inosina-5'-fosfato II	-1,73	0,041	F
PWY-6163	Biossíntese de corismato a partir de 3-dehidroquinato	1,676	0,047	F
PWY-6435	Biossíntese de hidroxibenzoato V	-3,002	0,001	F
		-2,559	0,005	C
PWYQT-4470	Ciclo do γ -glutamil	-1,799	0,036	F
PWYQT-4471	Biossíntese de glicosinolatos a partir de dihomometionina	1,66	0,048	F
RIBOSYN2-PWY	Biossíntese de flavina	-2,051	0,02	F
		-1,851	0,032	R
THISYNARA-PWY	Biossíntese de tiamina II	2,486	0,006	F
		2,026	0,021	R
THISYN-PWY	Biossíntese de tiamina	2,6	0,005	F
		2,128	0,017	R
TYRFUMCAT-PWY	Degradação de tirosina I	-4,066	0,00002	F
GLUTORN-PWY	Biossíntese de ornitina	-1,725	0,042	R
PWY1F-FLAVSYN	Biossíntese de flavonoides	-2,583	0,005	R
PWY-6168	Biossíntese de flavina II	-1,721	0,043	R

Z-scores positivos e negativos representam, respectivamente, vias superexpressas e subexpressas nas plantas anormais; p-value $\leq 0,05$; F - folha; C - caule; R - raiz.

A conversão de L-arginina em L-citrulina, nas vias de metabolismo da citrulina e ciclo da ureia, envolvem a produção de óxido nítrico (Romero-Puertas e Delledonne, 2005). Assim como a produção de tiamina (vias THISYNARA-PWY e THISYN-PWY) e de glicosinolatos (vias PWY-1187, PWY-2821 e PWYQT-4471), a produção de óxido nítrico é intensificada na resposta de defesa em plantas (Delledonne *et al.*, 1998; Wang *et al.*, 2006; Clay *et al.*, 2009). O óxido nítrico, juntamente com o peróxido de hidrogênio, agem na ativação da resposta de hipersensibilidade (Delledonne *et al.*, 1998; Heath, 2000; Zeier *et al.*, 2004; Taiz e Zeiger 2013). A tiamina possui função na indução da expressão de proteínas

relacionadas à defesa (PRs) nas vias dependentes de ácido salicílico (Wang *et al.*, 2006; Goyer 2010). E os glicosinolatos são componentes requisitados na deposição de calose entre a membrana e a parede celular (Clay *et al.*, 2009), impedindo a difusão célula a célula de micro-organismos patogênicos (Hammond-Kosack e Jones, 1996).

A fotossíntese, em geral, é reduzida quando as plantas encontram-se em situações de defesa contra estresses bióticos ou abióticos (Scharte *et al.*, 2005; Chaves *et al.*, 2009; Bilgin *et al.*, 2010). No entanto, as plantas anormais apresentaram o gene para RuBisCo (Ribulose-1,5-bifosfato), termos GO (atividade ribulose-bifosfato carboxilase, fixação de carbono e atividade frutose-bifosfato aldolase) e vias metabólicas (ciclo de Calvin-Benson-Bassham, biossíntese de clorofila a, fotorespiração e biossíntese de tetrapirrol) superexpressos, que sugerem um aumento na atividade fotossintética no fenótipo anômalo. Ayres (1984 e 1991) e Ney *et al.* (2013) relatam que o aumento na taxa fotossintética pode ocorrer como um mecanismo de compensação. Assim, a redução da fotossíntese no local da infecção induz o aumento da atividade fotossintética dos outros órgão e tecidos não infectados, no intuito de contrabalancear os níveis dessa atividade funcional.

A biossíntese do corismato é realizada pela via do chiquimato, no qual o produto inicial é a eritrose-4-fosfato derivada da fotossíntese (Coruzzi e Last, 2000; Malkin e Nyogi, 2000; Mustafa e Verpoorte, 2005; Taiz e Zeiger, 2013), que se apresentou elevada nas plantas anormais. O corismato pode ser convertido em ácido salicílico por meio de duas vias distintas: 1) através da ação das enzimas isocorismato sintase e isocorismato piruvato liase, resultando na conversão do corismato em isocorismato e, em seguida, em ácido salicílico; e 2) através da ação de várias outras enzimas, o corismato é primeiramente convertido em fenilalanina que, por sua vez, é convertida em ácido benzóico; para então, ser convertido em ácido salicílico pela enzima citocromo P450 monoxigenase (Hammond-Kosack e Jones, 1996; Shah, 2003; Mustafa e Verpoorte, 2005). Vale ressaltar que a fenilalanina provinda do corismato pode gerar compostos fenilpropanoides, precursores da lignina (Sato *et al.*, 1993; Croteau *et al.*, 2000; Taiz e Zeiger, 2013).

Os resultados de superexpressão de genes, termos GO enriquecidos e vias metabólicas nas plantas anômalas, sugerem que a anomalia envolve alterações nos sistemas em resposta à estresses bióticos e abióticos, com ênfase nos mecanismos de defesa contra patógenos. Uma vez que as progênes estudadas não apresentavam quaisquer indícios de ataques patogênicos, a resposta de defesa das plantas anômalas pode ser decorrente à algum processo autoimune originado por incompatibilidade genética.

5.4 Incompatibilidade genética

A incompatibilidade genética associada à ativação inapropriada do sistema imune da planta (sistema autoimune) é denominada como necrose híbrida (ou fraqueza híbrida) (Bomblies e Weigel, 2007; Bomblies *et al.*, 2007; Mizuno *et al.*, 2010). Esta incompatibilidade pode ser verificada em cruzamentos interespecíficos (Burkart-Waco *et al.*, 2012; Jeuken *et al.*, 2009; Masuda *et al.*, 2007; Mizuno *et al.*, 2010) e intraespecíficos entre diferentes acessos (Bomblies *et al.*, 2007; Smith *et al.*, 2011), cultivares (Khanna-Chopra *et al.*, 1998; Reiber e Neuman, 1999), variedades (Knight, 1947; Yamamoto *et al.*, 2010) ou, ainda, entre indivíduos da mesma espécie provenientes de populações diferentes (Singh e Gutiérrez, 1984; Galloway e Etterson, 2005). E como resultado, parte da progênie possui características fenotípicas típicas àquelas associadas com respostas de defesa contra estresses bióticos e abióticos (Bomblies e Weigel, 2007; Bomblies *et al.*, 2007; Jeuken *et al.*, 2009; Mizuno *et al.*, 2010; Yamamoto *et al.*, 2010; Smith *et al.*, 2011). Além do sistema autoimune, outros mecanismos fisiológicos podem explicar a necrose híbrida, como infecção viral epissomal, adaptação à estresse ambiental e sinalização hormonal (Bomblies e Weigel, 2007).

Em *Arabidopsis thaliana*, dos 20 casos independentes de necrose híbrida identificados por Bomblies *et al.* (2007), quatro são causados pela ativação de genes associados à resposta contra patógenos. Experimentos de Mizuno *et al.* (2010), envolvendo uma progênie de híbridos triploides originada pelo cruzamento interespecífico entre *Triticum turgidum* (autotetraploide) e *Aegilops tauschii* (trigo selvagem diploide), demonstraram necrose híbrida em parte da progênie. Além disso, o estudo do transcriptoma destes híbridos com necrose revelou superexpressão de genes relacionados à defesa (Mizuno *et al.*, 2010). A analogia entre nossos resultados e os estudos de Bomblies *et al.* (2007) e Mizuno *et al.* (2010) corrobora com a hipótese do envolvimento de mecanismos autoimune na anomalia do eucalipto estudada. Eventos similares são descritos em alface (Jeuken *et al.*, 2009), *Nicotiana sp* (Masuda *et al.*, 2007) e arroz (Yamamoto *et al.*, 2010).

Assim como na fisiologia da necrose híbrida, a genética também envolve vários sistemas independentes (Bomblies e Weigel, 2007; Bomblies *et al.*, 2007). Por exemplo, Bomblies *et al.* (2007) relatam que existem pelo menos 5 sistemas genéticos independentes em seu estudo que envolve 20 casos de necrose híbrida. O sistema genético mais comum para a incompatibilidade híbrida é a interação entre *loci*, explicada pelo modelo de Bateson-Dobzhansky-Muller (BDM) (Orr, 1996; Bomblies e Weigel, 2007; Bikard *et al.*, 2009; Yamamoto *et al.*, 2010).

O modelo de BDM simplesmente explica o desenvolvimento da incompatibilidade genética pós-zigótica em híbridos devido a interações epistáticas negativas entre alelos de linhagens diferentes (Orr, 1996; Mizuno *et al.*, 2010; Burkart-Waco *et al.*, 2012). Linhagens diferentes podem se originar de uma população inicial e, devido à pressões seletivas distintas, dois ou mais genes podem acumular mutações e estas serem fixadas na população, gerando novos alelos (Galloway e Etterson, 2005; Bomblies e Weigel, 2007). Esses alelos, que eram vantajosos dentro de cada linhagem, ao serem combinados podem conferir à progênie uma incompatibilidade genética deletéria devido às interações negativas (Orr, 1996; Galloway e Etterson, 2005). O modelo de BDM mais comum envolve dois *loci* (Khanna-Chopra *et al.*, 1998; Bomblies e Weigel, 2007; Bomblies *et al.*, 2007; Bikard *et al.*, 2009; Jeuken *et al.*, 2009), no entanto, alguns exemplos de interações BDM envolvem múltiplos *loci* (Bomblies *et al.*, 2007; Burkart-Waco *et al.*, 2012).

A segregação mendeliana de 3:1 (3 plantas normais: 1 planta normal) observada por Tambarussi (2006) indica que a anomalia não é causada por interação epistática, como apresentado no modelo de BDM, mas por interação alélica no qual a homozigose recessiva determina o caráter anômalo. Interações alélicas simples como causa de incompatibilidade gênica é pouco comum. Mizuno *et al.* (2010) verificaram que a necrose híbrida do tipo III, em trigo triploide, estava associado à um *locus* gênico do genoma proveniente do parental diploide e que resultava na anomalia quando em homozigose. Estudos em *Arabidopsis thaliana* (Smith *et al.*, 2011) e arroz (Li *et al.*, 1997; Heuer e Miézan, 2003; Chen *et al.*, 2008) demonstram interação alélica subdominante, no qual heterozigotos são desfavorecidos, ou seja, manifestam a incompatibilidade genética.

Muitos genes relacionados à resposta de defesa (genes R) possuem uma tendência em gerar incompatibilidade genética devido à características, como por exemplo, a rápida evolução da sequência nucleotídica e o grande número de cópias no genoma (Bomblies e Weigel, 2007). Assim como em *Populus trichocarpa* (Tuskan *et al.*, 2006), *Malus domestica* (Chagne *et al.*, 2013) e *Vitis vinifera* (Jaillon *et al.*, 2007), sugere-se que tenha ocorrido eventos de duplicação genômica ancestral (paleopoliploidia) em eucalipto (Myburg *et al.*, 2011). Estes eventos compreendem ciclos recorrentes de poliploidização e diploidização, que contribuem com o aumento do número de cópias dos genes R no genoma (Schifino-Wittmann, 2004; Paterson *et al.*, 2010). A rápida evolução nas sequências dos genes R pode ser atribuída ao processo de coevolução patógeno-hospedeiro. Neste processo, ocorre pressão seletiva mútua, contínua e intensa entre patógeno e hospedeiro, resultando na variabilidade genética (McHale *et al.* 2006; Ridley, 2006; Bomblies e Weigel, 2007; Ispolatov e Doebeli,

2009). Essa variabilidade, juntamente com o grande número de cópias de genes R no genoma, é muito importante para a adaptação aos diferentes ambientes e à ação de uma ampla variedade de patógenos (Paterson *et al.*, 2010). Por outro lado, as duplicações influenciam no aumento da incompatibilidade genética (Galloway e Etterson, 2005). Dentre os genes diferencialmente expressos na anomalia de eucalipto, os genes das famílias de proteínas similares à taumatina, de proteínas com domínio Bet v1 e da quitinase classe I possuem as características descritas acima, sendo considerados genes candidatos para este caso de incompatibilidade genética.

5.5 Mapeamento fino

O mapeamento fino permite verificar genes R (Bombliès *et al.*, 2007; Mizuno *et al.*, 2010; Smith *et al.*, 2011; Yamamoto *et al.*, 2012) ou QTLs (*Quantitative Trait Loci*) (Jeuken *et al.*, 2009; Burkart-Waco *et al.*, 2012) envolvidos na incompatibilidade genética. Visto que a incompatibilidade genética associada à anomalia de *E. grandis* apresenta três famílias gênicas de genes candidatos (famílias de proteínas similares à taumatina, de proteínas com domínio Bet v1 e da quitinase classe I), foi iniciado o mapeamento fino dos genes diferencialmente expressos dessas famílias. No intervalo entre 3 kb a montante e a jusante dos genes candidatos, foram desenhados oligonucleotídeos (Tabela 2) e estes testados em plantas normais e anormais. No entanto, os oligonucleotídeos desenhados e testados até o momento não apresentaram polimorfismo entre os fenótipos (Figura 13), sendo necessário o desenvolvimento de novos oligonucleotídeos e, assim, a continuidade do mapeamento.

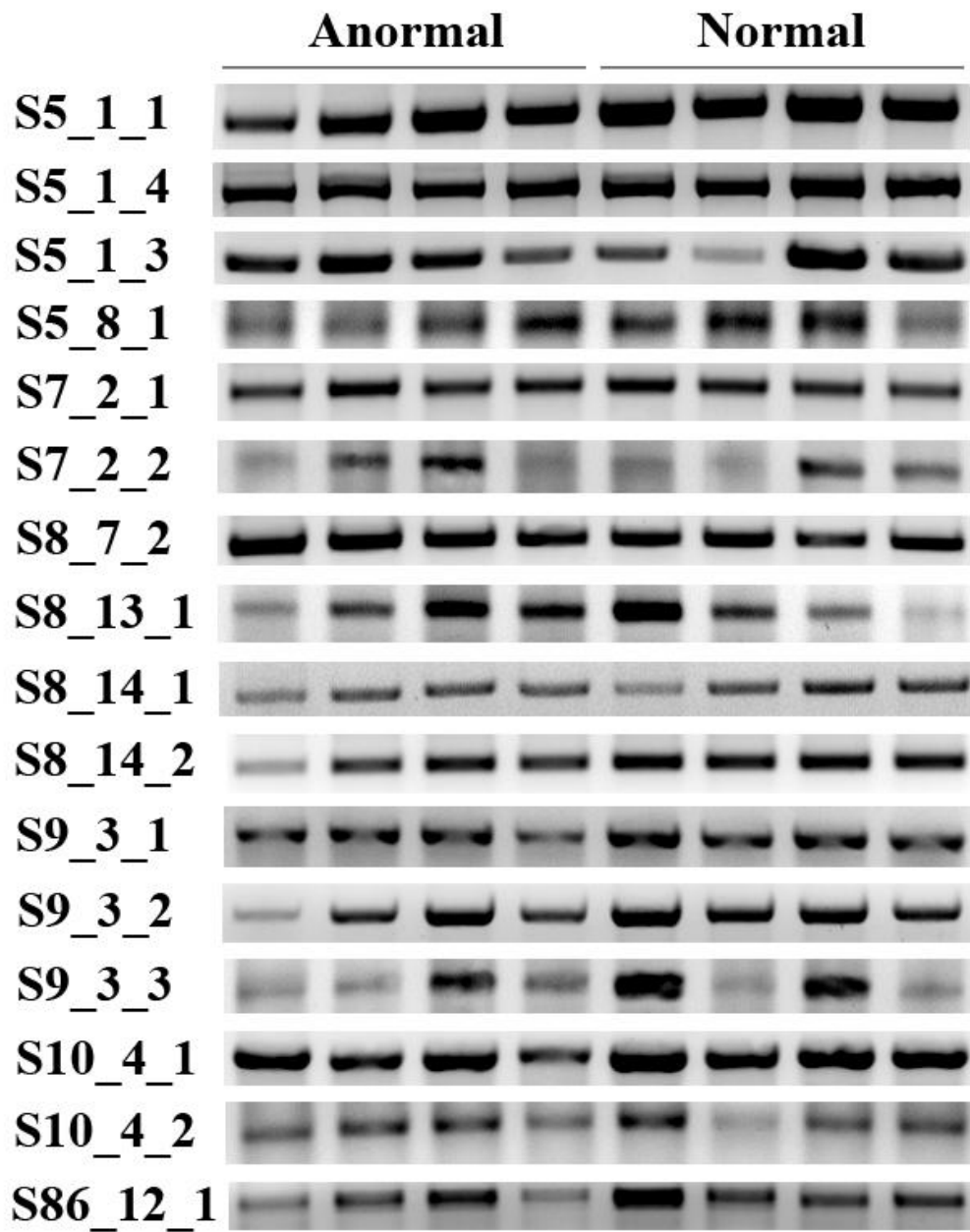


Figura 13: Amplificação com os oligonucleotídeos (código à esquerda) desenvolvidos para o mapeamento fino nos grupos anormal e normal.

6 CONCLUSÕES

As conclusões obtidas com o trabalho foram as seguintes:

- A anomalia altera significativamente a altura, diâmetro da base do caule, número de ramificações, dimensões e forma do limbo foliar das plântulas de *Eucalyptus grandis*.
- O fenótipo anômalo apresentou superexpressão de genes, termos GO e vias metabólicas relacionados à resposta de defesa.
- Os três órgãos estudados (folha, caule e raiz) apresentaram divergências quanto aos genes, termos GO e vias metabólicas diferencialmente expressos, assim como diferenças na expressão da quitinase (glicosil hidrolase 18).
- Os genes da superfamília gênica Bet v1 apresentaram superexpressão no fenótipo anormal, confirmando sua participação no desenvolvimento da anomalia.
- As famílias gênicas de proteínas similares à taumatina, proteínas com domínio Bet v1 e quitinase classe I foram consideradas candidatas na causa da incompatibilidade genética.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAF. Anuário estatístico ABRAF - ano base 2012. Brasília: ABRAF; 2013.
- Agarwal M, Shrivastava N, Padh H. Advances in molecular marker techniques and their applications in plant science. *Plant Cell Rep.* 2008; 27:617-631.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 1997; 25:3389-3402.
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010; 11: R106.
- Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Bahkali AH, Al Sadoon M, Shobrak M. A brief review of molecular techniques to assess plant diversity. *International Journal of Molecular Sciences.* 2010; 11:2079-2096.
- Assis TF, Resende MDV. Genetic improvement of forest tree species. *Crop Breeding and Applied Biotechnology.* 2011; S1: 44-49.
- Ayres PG. The interaction between environmental stress injury and biotic disease physiology. *Ann. Rev. Phytopathol.* 1984; 22:53-75.
- Ayres PG. Growth responses induced by pathogens and other stresses. In: Mooney HA, Winner WE, Pell EJ, editors. *Response of Plants to Multiple Stresses.* New York: Academic Press; 1991. p. 227-246.
- Barret P, Delourme R, Foisset N, Renard M. Development of a SCAR (sequence characterized amplified region) marker for molecular tagging of the dwarf BREIZH (Bzh) gene in *Brassica napus* L. *Theor. Appl. Genet.* 1998; 97:828-833.
- Bered F, Barbosa Neto JF, de Carvalho FIF. Marcadores moleculares e sua aplicação no melhoramento genético de plantas. *Ciência Rural, Santa Maria.* 1997; 27(3): 513-520.
- Bikard D, Patel D, Le Mette C, Giorgi V, Camilleri C, Bennett MJ, Loudet O. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science.* 2009; 323 623–626.
- Bilgin DD, Zavala JA, ZHU J, Clough SJ, Ort DR, DeLucia EH. Biotic stress globally downregulates photosynthesis genes. *Plant, Cell and Environment.* 2010; 33: 1597–1613.
- Boava L.P., Laia M.L, Jacob T.R, Dabbas K.M, Gonçalves J.F, Ferro J.A, Ferro M.I.T, Furtado E.L. Selection of endogenous genes for gene expression studies in Eucalyptus under biotic (*Puccinia psidii*) and abiotic (acibenzolar-S-methyl) stresses using RT-qPCR. *BMC Research Notes.* 2010; 3: 43-51.
- Boland DJ, Brooker MIH, Chippendale GM, Hall N, Hyland BPM, Johnson RD, Kleining DA, McDonald MW, Turner JD. *Forest trees of Australia.* Melbourne: CSIRO; 2006.

- Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer KFX. Plant genome sequencing — applications for crop improvement. *Current Opinion in Biotechnology*. 2014; 26:31–37.
- Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol*. 2007; 5:1962–1972.
- Bomblies K, Weigel D. Hybrid necrosis: Autoimmunity as a common barrier to gene flow in plants. *Nat Rev Genet*. 2007; 8: 382–393.
- Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*. 2006; 314:1292–1295.
- Brooker I. Botany of the eucalypts. In: Coppen JJW, editor. *Eucalyptus: The genus Eucalyptus*. New York: Taylor & Francis Inc.; 2002. p.3-35.
- Brooker MIH, Kleinig DA. *Field guide to Eucalyptus*. Vol. 3, 2nd ed. Melbourne: Bloomings Books; 2004.
- Burgess IP, Williams ER, Bell JC, Harwood CE, Owen JC. The effect of outcrossing rate on the growth of selected families of *Eucalyptus grandis*. *Silvae Genetica*. 1996; 45(2-3):97-100.
- Burkart-Waco D, Josefsson C, Dilkes B, Kozloff N, Torjek O, Meyer R, Altmann T, Comai L. Hybrid incompatibility in *Arabidopsis* is determined by a multiple-locus genetic network. *Plant Physiol*. 2012; 158: 801–812.
- Campinhos Jr. E. Sustainable plantations of high-yield *Eucalyptus* trees for production of fiber: the Aracruz case. *New Forests*. 1999; 17: 129–143.
- Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*. 2012; 40: D742–D753.
- Chagne D, Lin-Wang K, Espley RV, Marra M, Sandberg G, Rokhsar D. An ancient duplication of apple MYB factors is responsible for novel red fruitflesh phenotypes. *Plant Physiology*. 2013; 161(1): 225-239.
- Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter*. 1993; 11(2):113-116.
- Chaves MM, Flexas J, Pinheiro C. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Annals of Botany*. 2009; 103: 551–560.
- Chen JJ, Ding JH, Ouyang YD, Du HY, Yang JY, Cheng K, Zhao J, Qiu SQ, Zhang XL, Yao JL, Liu KD, Wang L, Xu CG, Li XH, Xue YB, Xia M, Ji Q, Lu JF, Xu ML, Zhang Q. A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility

- of indica–japonica hybrids in rice. *Proc. Natl. Acad. Sci. USA.* 2008; 105:11436–11441.
- Chen Z, Silva H, Klessig DE. Active Oxygen Species in the Induction of Plant Systemic Acquired Resistance by Salicylic Acid. *Science.* 1993; 262: 1883-1886.
- Clay NK, Adio AM, Denoux C, Jander G, Ausubel FM. Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response. *Science.* 2009; 323: 95-101.
- Coate JE, Doyle JJ. Genomics and transcriptomics of photosynthesis in polyploids. In: Chen ZJ, Birchler JA, editors. *Polyploid and hybrid genomics.* Oxford: Wiley-Blackwell; 2013.
- Coll NS, Epple P, Dangl JL. Programmed cell death in the plant immune system. *Cell Death Differ.* 2011; 18:1247–1256.
- Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *International Journal of Plant Genomics.* 2008; 2008: Article ID 619832.
- Coruzzi G, Last R. Amino Acids. In: Buchanan BB, Grissem W, Jones RL, editors. *Biochemistry & Molecular Biology of Plants.* Rockville: American Society of Plant Physiology; 2000.
- Croteau R, Kutchan TM, Lewis NG. Natural Products (Secondary Metabolites). In: Buchanan BB, Grissem W, Jones RL, editors. *Biochemistry & Molecular Biology of Plants.* Rockville: American Society of Plant Physiology; 2000.
- Cunha AO, Andrade LA, Bruno RLA, Silva JAL, Souza VC. Efeitos de substratos e das dimensões dos recipientes na qualidade das mudas de *Tabebuia impetiginosa* (Mart. Ex D.C.) Standl. *Revista Árvore.* 2005; 29(4): 507-516.
- Dasgupta MG, Chezian P, Yasodha R. Identification of putative trait based markers for Genetic Improvement of *Eucalyptus tereticornis*. *Electronic Journal of Plant Breeding.* 2010; 1(4): 1213-1226.
- Delledonne M, Xia Y, Dixon RA, Lamb C. Nitric oxide functions as a signal in plant disease resistance. *Nature.* 1998; 394: 585-588.
- Dorey S, Baillieul F, Pierrel MA, Saindrenan P, Fritig B, Kauffmann S. Spatial and temporal induction of cell death, defense genes, and accumulation of salicylic acid in tobacco leaves reacting hypersensitively to a fungal glycoprotein elicitor. *Mol. Plant—Microbe Interact.* 1997; 10: 646–655.
- Dunwell JM, Purvis A, Khuri S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry.* 2004; 65:7–17.
- Durner J, Shah J, Klessig DF. Salicylic acid and disease resistance in plants. *Trends Plant Sci.* 1997; 2: 266 274.
- Durrant WE, Dong X. Systemic acquired resistance. *Annu. Rev. Phytopathol.* 2004; 42:185–209.

- Estopa RA. Comparação do desempenho no viveiro e no campo dos descendentes de clones de *Eucalyptus* spp. autofecundados e cruzados [dissertação]. Lavras: Universidade Federal de Lavras; 2006.
- Ferreira M. Melhoramento e a silvicultura intensiva clonal. IPEF. 1992; 45: 22-30.
- Flach J, Pilet PE, Jollès P. What's new in chitinase research? *Experientia*. 1992; 48: 701-716.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999; 151: 1531–1545.
- Freeman S, Herron JC. *Análise Evolutiva*. Porto Alegre: Artmed; 2009.
- Futuyma DJ. *Evolutionary Biology*. 3rd ed. Sunderland, MA: Sinauer; 1997.
- Galloway LF, Etterson JR. Population differentiation and hybrid success in *Campanula americana*: geography and genome size. *J. Evol. Biol.* 2005; 18: 81–89.
- Golle DP, Reiniger LRS, Curti AR, Bevilacqua CB. Forestry improvement: emphasis on biotechnology application. *Ciência Rural, Santa Maria*. 2009; 39(5):1606-1613.
- Gonçalves JLM, Alvares CA, Higa AR, Silva LD, Alfenas AC, Stahl J, Ferraz SFB, Lima WP, Brancalion PHS, Hubner A, Bouillet JPD, Laclau JP, Nouvellon Y, Epron D. Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. *Forest Ecology and Management*. 2013; 301: 6–27.
- González-Duarte R, Albalat R. Merging protein, gene and genomic data: the evolution of the MDR-ADH family. *Heredity*. 2005; 95:184–197.
- Goyer A. Thiamine in plants: Aspects of its metabolism and functions. *Phytochemistry*. 2010; 71: 1615–1624.
- Grant M, Brown I, Adams S, Knight M, Ainslie A, Mansfield J. The RPM1 plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. *The Plant Journal*. 2000; 23(4): 441-450.
- Grattapaglia D, Kirst M. *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytologist*. 2008; 179:911-929.
- Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. *Tree Genetics & Genomes*. 2011; 7:241–255.
- Guan X, Pang M, Nah G, Shi X, Ye W, Stelly DM, Chen ZJ. miR828 and miR858 regulate homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre development. *Nature Communications*. 2014; 5: 3050.
- Gupta PK, Balyan HS, Varshney RK, Gill KS. Development and use of molecular markers for crop improvement. *Plant Breeding*. 2013; 132: 431–432.
- Hammond-Kosack KE, Jones JDG. Resistance Gene-Dependent Plant Defense Responses. *The Plant Cell*. 1996; 8: 1773-1791.

- Hardner CM, Potts BM. Inbreeding depression and changes in variation after selfing in *Eucalyptus globulus* ssp. *globulus*. 1995. *Silvae Genetica*. 1995; 44(1):46-54.
- Hardner C, Tibbits W. Inbreeding depression for growth, wood and fecundity traits in *Eucalyptus nitens*. *Forest Genetics*. 1998; 5(1):11-20.
- Harfouche A, Meilan R, Kirst M, Morgante M, Boerjan W, Sabatti M, Mugnozza GS. Accelerating the domestication of forest trees in a changing world. *Trends in Plant Science*. 2012; 17(2): 64-72.
- He Z-H, He D, Kohorn BD. Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. *Plant J*. 1998; 14: 55–63.
- Heath MC. Hypersensitive response-related death. *Plant Molecular Biology*. 2000; 44: 321–334.
- Hernández P, Martín A, Dorado G. Development of SCARs by direct sequencing of RAPD products: a practical tool for the introgression and marker-assisted selection of wheat. *Molecular Breeding*. 1999; 5:245-253.
- Heuer S, Miézan KM. Assessing hybrid sterility in *Oryza glaberrima* x *O. sativa* hybrid progenies by PCR marker analysis and crossing with wide compatibility varieties. *Theoretical and Applied Genetics*. 2003; 107: 902–909.
- Hoffmann-Sommergruber K, Vanek-Krebitz M, Radauer C, Wen J, Ferreira F, Scheiner O, Breiteneder H. Genomic characterization of members of the Bet v 1 family: genes coding for allergens and pathogenesis-related proteins share intron positions. *Gene*. 1997; 197:91-100.
- Huijser P, Klein J, Lönnig WE, Meijer H, Saedler H, Sommer H. Bracteomania. An inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. *EMBO J*. 1992; 11(4):1239–1249.
- Illumina, Inc. 2011. Quality scores for next-generation sequencing. Technical Note: Sequencing (http://res.illumina.com/documents/products/technotes/technote_q-scores.pdf).
- Ispolatov I, Doebeli M. Speciation due to hybrid necrosis in plant-pathogen models. *Evolution*. 2009; 63:3076–3084.
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 2007; 449:463-467.
- Jeuken MJ, Zhang NW, McHale LK, Pelgrom K, den Boer E, Lindhout P, Michelmore RW, Visser RGF, Niks RE. Rin4 causes hybrid necrosis and race-specific resistance in an interspecific lettuce hybrid. *Plant Cell*. 2009; 21:3368-3378.
- Jones JD, Dangl JL. The plant immune system. *Nature*. 2006; 444: 323–329.

- Kachroo P, Shanklin J, Shah J, Whittle EJ, Klessig DF. A fatty acid desaturase modulates the activation of defense signaling pathways in plants. *Proc. Natl. Acad. Sci. USA.* 2001; 98:9448-9453.
- Kasprzewska A. Plant chitinases - Regulation and function. *Cellular & Molecular Biology Letters.* 2003; 8: 809 – 824.
- Kennington WJ, James SH. The effect of small population size on the mating system of a rare clonal mallee, *Eucalyptus argutifolia* (Myrtaceae). *Heredity.* 1997; 78:252-260.
- Khanna-Chopra R, Dalal M, Kumar GP, Laloraya M. A genetic system involving superoxide causes F1 necrosis in wheat (*T. aestivum* L.). *Biochem. Biophys. Res. Commun.* 1998; 248:712–715.
- Kitajima S, Sato F. Plant Pathogenesis-Related Proteins: Molecular Mechanisms of Gene Expression and Protein Function. *J. Biochem.* 1999; 125: 1-8.
- Klekowski Jr EJ. Genetic load and its causes in long-lived plants. *Trees.* 1988; 2:195-203.
- Knight RL. The genetics of blackarm resistance v. dwarf-bunched and its relationship to B₁. *Journal of Genetics.* 1947; 48(1): 43-51.
- Koh HJ, Son YH, Heu MH, Lee HS, McCouch SR. Molecular mapping of a new genic male-sterility gene causing chalky endosperm in rice (*Oryza sativa* L.). *Euphytica* 1999; 106:57-62.
- Lacerda DR, Acedo MDP, Lemos Filho JP de, Lovato MB. A técnica de RAPD: uma ferramenta molecular em estudos de conservação de plantas. *Lundiana.* 2002; 3(2):87-92.
- Ladiges PY, Udovicic F, Nelson G. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography.* 2003; 30: 989-998.
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25.
- Levine A, Tenhaken R, Dixon R, Lamb C. H2O₂ from the Oxidative Burst Orchestrates the Plant Hypersensitive Disease Resistance Response. *Cell.* 1994; 79: 583-593.
- Levy AA, Tirosh T, Reikhav S, Bloch Y, Barkai N. Yeast hybrids and polyploids as models in evolutionary studies. In: Chen ZJ, Birchler JA, editors. *Polyploid and hybrid genomics.* Oxford: Wiley-Blackwell; 2013.
- Li HB, Wang J, Liu AM, Liu KD, Zhang QF, Zou JS. Genetic basis of low-temperature-sensitive sterility in indica-japonica hybrids of rice as determined by RFLP analysis. *Theor. Appl. Genet.* 1997; 95:1092–1097.
- Liu J-J, Ekramoddoullah AKM. The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. *Physiological and Molecular Plant Pathology.* 2006; 68:3-13.

- Lourenção JC. Caracterização de uma região genômica relacionada a uma anomalia de viveiro em *Eucalyptus* [dissertação]. Botucatu: Instituto de Ciências Biológicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”; 2010.
- Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics Applications Note*. 2005; 21(16): 3448–3449.
- Malkin R, Niyogi K. Photosynthesis. In: Buchanan BB, Gruissem W, Jones RL, editors. *Biochemistry & Molecular Biology of Plants*. Rockville: American Society of Plant Physiology; 2000.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*. 2008; 18:1509-1517.
- Maschio LMA, Ferreira CA, Auer CG, Grigoletti Jr. A, Wiecheteck MRS, Nardelli AMB, Bernardi CA. Descrição de uma anomalia ocorrida em *Eucalyptus grandis* na região de Arapoti-PR, Brasil. *Boletim de Pesquisa Florestal*. 1996; 32-33:85-87.
- Masuda Y, Yamada T, Kuboyama T, Marubashi W. Identification and characterization of genes involved in hybrid lethality in hybrid tobacco cells (*Nicotiana suaveolens* × *N. tabacum*) using suppression subtractive hybridization. *Plant cell Rep*. 2007; 26(9): 1595-1604.
- Marx GA. Developmental mutants in some annual seed plants. *Ann. Rev. Plant Physiol*. 1983; 34:389-417.
- McGowen M, Potts B, Vaillancourt R, Gore P, Williams D, Pilbeam D. The genetic control of sexual reproduction in *Eucalyptus globulus*. In: Borralho N, *et al.*, editors. *Eucalyptus in a changing world*. Proc. of IUFRO Conference. Aveiro; 2004. p. 11-15.
- McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol*. 2006; 7:212.
- Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics*. 2010; 11: 681-692.
- Mizuno N, Hosogi N, Park P, Takumi S. Hypersensitive response-like reaction is associated with hybrid necrosis in interspecific crosses between tetraploid wheat and *Aegilops tauschii* Coss. *PLoS One*. 2010; 5:e11326.
- Mora AL, Garcia CH. A cultura do eucalipto no Brasil. São Paulo: Sociedade Brasileira de Silvicultura; 2000.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5: 621–628.
- Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Research*. 1980; 8:4321-4325.

- Mustafa NR, Verpoorte R. Chorismate derived C6C1 compounds in plants. *Planta*. 2005; 222: 1–5.
- Myburg A, Grattapaglia D, Tuskan G, Jenkins J, Schumutz J, Mizrachi E, Hefer C, Sterck L, Hayes R, Rokhsar D. The *Eucalyptus grandis* genome project: genome and transcriptome resources for comparative analysis of woody plant biology. *BMC Proceedings*. 2011; 5(Suppl 7):I20.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 2008; 320:1344–1349.
- Negi MS, Devic M, Delseny M, Lakshmikumaran M. Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection. *Theor. Appl. Genet.* 2000; 101:146-152.
- Ney B, Bancal MO, Bancal P, Bingham IJ, Foulkes J, Gouache D, Paveley N, Smith J. Crop architecture and crop tolerance to fungal diseases and insect herbivory. Mechanisms to limit crop losses. *Eur J Plant Pathol*. 2013; 135:561–580.
- Niklas KJ. *The evolutionary biology of plants*. Chicago: University of Chicago Press; 1997.
- Orr HA. Dobzhansky, Bateson, and the genetics of speciation. *Genetics*. 1996; 144: 1331-1335.
- Paterson AH, Freeling M, Tang H, Wang X. Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology*. 2010; 61:349-372.
- Peña L, Séguin A. Recent advances in the genetic transformation of trees. *Trends Biotechnol.* 2001; 19(12):500-506.
- Potts B, Hamilton M, Blackburn D. Genetics of eucalypts: Traps and opportunities. In: Walker J, editor. *Developing a eucalypt resource : learning from Australia and elsewhere*. New Zealand: Wood Technology Research Centre, University of Canterbury; 2011. pp. 1-27.
- Poupard P, Brunel N, Leduc N, Viéumont J-D, Strullu D-G, Simoneau P. Expression of a Bet v 1 homologue gene encoding a PR 10 protein in birch root: induction by auxin and localization of the transcripts by *in situ* hybridization. *Aust. J. Plant Physiol.* 2001; 28:57-63.
- Pryor L. *Eucalyptus*. In: Halevy, A, editor. *CRC Handbook of Flowering*. Vol. 2, 1st ed. Boca Raton: Franklin Book Company; 1985. p. 476-482.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. The Pfam protein families database. *Nucleic Acids Res.* 2012; 40:D290–301.
- Quero-García J, Letourmy P, Ivancic A, Feldmann P, Courtois B, Noyer JL, Lebot V. Hybrid performance in taro (*Colocasia esculenta*) in relation to genetic dissimilarity of parents. *Theor Appl Genet.* 2009; 119:213–221.

- Quinlan A, Hall I. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841-842.
- Ramalho MAP, Santos JB, Pinto CABP. *Genética na Agropecuária*. 3ª ed. Lavras: UFLA; 2004.
- Rédei GP. *Arabidopsis* as a genetic tool. *Annu. Rev. Genet.* 1975; 9:111-127.
- Reiber JM, Neuman DS. Hybrid Weakness in *Phaseolus vulgaris* L. II. Disruption of Root-Shoot Integration. *J Plant Growth Regul.* 1999; 18:107-112.
- Ridley M. *Evolução*. 3ª ed. São Paulo: Artmed; 2006.
- Rockwood DL. History and Status of Eucalyptus Improvement in Florida. *International Journal of Forestry Research*. 2012; 2012: Article ID 607879.
- Romero-Puertas MC, Delledonne M. Nitric oxide signaling through the hypersensitive disease resistance response. In: Magalhaes JR, Singh RP, Passos L, editors. *Nitric oxide signaling in higher plants*. Houston: Studium Press, LLC; 2005. p. 199-218. *Plant Molecular Biology* 20: 1199-1202, 1992.
- Ruiz-Medrano R, Jimenez-Moraila B, Herrera-Estrella L, Rivera-Bustamante RF. Nucleotide sequence of an osmotin-like cDNA induced in tomato during viroid infection. *Plant Molecular Biology*. 1992; 20: 1199-1202.
- Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner H, Hunt MD. Systemic Acquired Resistance. *The Plant Cell*. 1996; 8: 1809-1819.
- SAS INSTITUTE. *SAS user's guide: statistics*. Cary: SAS INSTITUTE; 2002.
- Sato Y, Sugiyama M, Górecki RJ, Fukuda H, Komamine A. Interrelationship between lignin deposition and the activities of peroxidase isoenzymes in differentiating tracheary elements of *Zinnia*. *Planta*. 1993; 189: 584-589.
- Scandalios JG. Oxygen stress and superoxide dismutases. *Plant Physiol.* 1993;101:7-12.
- Scharte J, Schön H, Weis E. Photosynthesis and carbohydrate metabolism in tobacco leaves during an incompatible interaction with *Phytophthora nicotianae*. *Plant, Cell and Environment*. 2005; 28: 1421-1435.
- Schifino-Wittmann, MT. Poliploidia e seu impacto na origem e evolução das plantas silvestres e cultivadas. *R. bras. Agrociência*. 2004; 10(2): 151-157.
- Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protocols*. 2008; 3:1101-8.
- Schwan-Estrada, KRF, Stangarlin JR, Pascholati SF. Mecanismos bioquímicos de defesa vegetal. In: Pascholati SF, Leite B, Stangarlin JR, CIA P, editores. *Interação planta-patógeno: fisiologia, bioquímica e biologia molecular*. Piracicaba: Fealq; 2008. p. 227-248.

- Severin AJ, Woody JL, Bolon Y, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham A, Cannon SB, May GD, Vance CP, Shoemaker RC. RNA-Seq atlas of glycine max: a guide to the soybean transcriptome. *BMC Plant Biology*. 2010; 10:160.
- Shah J. The salicylic acid loop in plant defense. *Current Opinion in Plant Biology*. 2003; 6:365–371.
- Singh SP, Gutiérrez JA. Geographical distribution of the DL₁ and DL₂ genes causing hybrid dwarfism in *Phaseolus vulgaris* L., their association with seed size, and their significance to breeding. *Euphytica*. 1984; 33: 337-345.
- Smith LM, Bomblies K, Weigel D. Complex evolutionary events at a tandem cluster of *Arabidopsis thaliana* genes resulting in a single-locus genetic incompatibility. *PLoS Genetics*. 2011; 7:e1002164.
- Souza HG, Doria KMABVS, Basseto MA, Rosa DD, Furtado EL, Marino CL. Diversidade genética em populações-núcleo de *Eucalyptus grandis*. *Acta Scientiarum, Agronomy*. 2010; 32(4): 621-625.
- Sticher L, Mauch-Mani B, Métraux. Systemic acquired resistance. *Annu. Rev. Phytopathol.* 1997; 35:235–70.
- Stearns SC, Hoekstra RF. *Evolução: uma introdução*. São Paulo: editora Atheneu; 2003.
- Taiz L, Zeiger E. *Fisiologia vegetal*. 5ª ed. Porto Alegre: Artmed; 2013.
- Tambarussi EV. Associação de marcador RAPD e desenvolvimento de marcador SCAR para um tipo de anomalia de viveiro em *Eucalyptus grandis* [trabalho de conclusão de curso] Botucatu: Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”; 2006.
- Tiffin P, Moeller DA. Molecular evolution of plant immune system genes. *Trends Genet.* 2006; 22: 662–670.
- Turnbull JW, Booth TH. *Eucalypts in cultivation: an overview*. Coppen JJW, editor. *Eucalyptus: The genus Eucalyptus*. New York: Taylor & Francis Inc.; 2002. p. 52-74.
- Tuskan GA, Difazio S, Jansson S. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 2006; 313(5793): 1596-1604.
- Tzfira T, Zuker A, Altman A. Forest-tree biotechnology: genetic transformation and its application to future forests. *Trends Biotechnol.* 1998; 16: 439-446.
- Verberne MC, Verpoorte R, Bol JF, Mercado-Blanco J, Linthorst HJM. Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. *Nature Biotechnology*. 2000; 18: 779-783.
- Villar E, Klopp C, Noirot C, Novaes E, Kirst M, Plomion C, Gion JM. RNA-Seq reveals genotype-specific molecular responses to water deficit in eucalyptus. *BMC Genomics*. 2011; 12:538-556.

- Walter MH, Liu J-W, Wünn J, Hess D. Bean ribonuclease-like pathogenesis-related protein genes (Ypr10) display complex patterns of developmental, dark-induced and exogenous-stimulus-dependent expression. *Eur. J. Biochem.* 1996; 239:281-293.
- Wang₁ G, Ding₁ X, Yuan M, Qiu D, Li X, Xu C, Wang S. Dual function of rice OsDR8 gene in disease resistance and thiamine accumulation. *Plant Molecular Biology.* 2006; 60:437-449.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics.* 2009; 10:57-63.
- Ward ER, Uknes SJ, Williams SC, Dincher SS, Wiederhold DL, Alexander DC, Ahl-Goy P, Métraux J, Ryals JA. Coordinate Gene Activity in Response to Agents That Induce Systemic Acquired Resistance. *The Plant Cell.* 1991; 3: 1085-1094.
- Weaver M, Herrmann KM. Dynamics of the shikimate pathway in plants. *Trends im Plant Science.* 1997; 2(9): 346-351.
- Wen J, Vanek-Krebitz M, Hoffmann-Sommergruber K, Scheiner O, Breiteneder H. The Potential of Bet v 1 Homologues, a Nuclear Multigene Family, as Phylogenetic Markers in Flowering Plants. *Molecular Phylogenetics and Evolution.* 1997; 8(3):317-333.
- White JD, Scott NA. Specific leaf area and nitrogen distribution in New Zealand forests: Species independently respond to intercepted light. *Forest Ecology and Management.* 2006; 226(1-3): 319-329.
- Wilcken CF, Lima ACV, Dias TKR, Masson MV, Filho PJF, Pogetto MHFAD. Guia prático de manejo de plantações de eucaliptos. Botucatu: FEPAF; 2008.
- Wildermuth MC, Dewdney J, Wu G, Ausubel FM. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature.* 2001; 414: 562-571.
- Yamamoto E, Takashi T, Morinaka Y, Lin S, Wu J, Matsumoto T, Kitano H, Matsuoka M, Ashikari M. Gain of deleterious function causes an autoimmune response and Bateson–Dobzhansky–Muller incompatibility in rice. *Mol Genet Genomics.* 2010; 283:305–315.
- Young A, Boyle T, Brown T. The population genetic consequences of habitat fragmentation for plants. *Tree.* 1996; 11(10):413-418.
- Zeier J, Delledonne M, Mishina T, Severi E, Sonoda M, Lamb C. Genetic Elucidation of Nitric Oxide Signaling in Incompatible Plant-Pathogen Interactions. *Plant Physiology.* 2004; 136: 2875–2886.
- Zhang J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 2003; 18:292–298.
- Zobel B, Talbert J. Applied forest tree improvement. New York: John Wiley and Sons; 1984.

APÊNDICE A - Artigo submetido para a revista *Tree Genetics & Genomes* (em fase de revisão).

MOLECULAR CHARACTERIZATION OF A GENOMIC REGION ASSOCIATED WITH A NATURAL MUTANT OF *Eucalyptus grandis*

Maria C P Fuchs¹, Juliana C Lourenção¹, Evandro V Tambarussi², Leticia M Nogueira³, Tânia M Bortoloto¹, Esteban R González⁴, Shinitiro Oda⁴, Celso L Marino¹

¹ Departamento de Genética, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Distrito de Rubião Jr s/n, Botucatu, SP CEP 18618-970, Brazil.

² Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ), Universidade de São Paulo (USP), Avenida Pádua Dias 11, Piracicaba, SP CEP 13418-900, Brazil.

³ Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Dr Room 6E208 MSC 7248, Bethesda, MD 20892, USA.

⁴ Empresa Suzano Papel e Celulose SA, Av. Dr. José Lembo 1010, Itapetininga, SP CEP 18207-780, Brazil.

Correspondence author: PhD CL Marino, Departamento de Genética, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Distrito de Rubião Jr s/n, Botucatu, SP CEP 18618-970, Brazil. E-mail: clmarino@ibb.unesp.br. Telephone: +55 14 3880-0365.

ABSTRACT

Eucalyptus is one of the most important genus in worldwide forestry culture. It exhibits mixed mating systems with predominance of outcrossing, what contribute in the large quantity of heterozygosity in natural populations. High outcrossing rates can generate half-sib and full-sib progenies with a low frequency of anomalous individuals, as observed in a controlled cross of *Eucalyptus grandis*. Anomalous individuals were characterized by branching, reduced height and individual leaf area, and asymmetrical leaf shape. The anomaly presents Mendelian segregation with a ratio of 3 normal to 1 abnormal seedlings, suggesting that the anomalous phenotype has a polygenic character with a major gene with a recessive deleterious allele effect. We developed a molecular marker related with the anomalous character. Analysis of the sequence marker revealed a mutation of two base pairs between normal and abnormal individuals. A marker was localized in an intergenic region, but it is unclear whether this mutation is related with the anomaly by affecting nearby genes or others genomic regions. A Bet v1-like gene is near to the mutation and its 3'UTR contains identity with the molecular marker developed. Bet v1-homologous proteins participate in several biological processes, suggesting a role in anomalous character development. Thus, the polymorphism detected allows the development of functional markers to identify anomaly alleles in *Eucalyptus* genetic improvement programs.

Keywords: Eucalypt anomaly, molecular marker, Bet v1-like genes, deleterious recessive alleles.

INTRODUCTION

The genus *Eucalyptus* is the most extensively utilized forestry culture in the world due to its high adaptability, superior growth, and multipurpose applications, which confer advantages for its introduction and maintenance (Gattapaglia and Kirst 2008). This genus possesses a wide diversity of species, varieties and hybrids (more than 900) (Boland et al. 2006; Brooker and Kleinig 2004; Gattapaglia and Kirst 2008). The importance of *Eucalyptus* forestry culture in Brazil is reflected in the quantity of clonal and seed orchards with high eucalypt diversity in São Paulo and Espírito Santo States (Brune and Zobel 1981; Campinhos et al. 1998; Junghans et al. 1998). Actually, in Brazil about 76,6% of the forestry areas (5.10 million ha) are planted with this genus (ABRAF 2013).

Eucalypt presents mixed mating systems ($0.05 \leq t \leq 0.95$) but *E. grandis* is predominant of outcrossing ($t \geq 0.95$) (Chaix et al. 2003), what contribute in the large quantity of heterozygosis in the genome (Golle et al. 2009; Zobel and Talbert 1984). Outcrossing populations exhibit a low frequency of individuals with physiological and morphological anomalies (Klekowski Jr 1988). These anomalies can be generated by allele mutations and can be observed in different plant species, e.g. *Antirrhinum majus* (Huijser et al. 1992), *Arabidopsis thaliana* (Rédei 1975), *Oryza sativa* (Koh et al. 1999) and *Cucumis sativus* (Marx 1983). According to Klekowski Jr (1988), recessive deleterious alleles are hidden in heterozygotes of outcrossing populations. When homozygous, the recessive deleterious alleles results in several phenotypic manifestations that are not restricted to embryonic stage (Klekowski Jr 1988; Zobel and Talbert 1984), but can also involve seedling stage where the growth, morphology and physiology are altered (Klekowski Jr 1988; Zobel and Talbert 1984). Negative effects of deleterious recessive alleles cause considerable losses in seedling production and may cause delays in improvement programs. In this context, the identification of deleterious recessive alleles by molecular markers could provide a powerful tool to eliminate genotypes with unfavorable alleles in breeding programs and avoid damage to production.

In a full-sib offspring from a controlled cross of *Eucalyptus grandis*, we observed an anomaly in one fourth of the seedlings, which died within a few months. The principal characteristics of this anomaly were branching, reduced height, reduced individual leaf area, and modification of leaf shape. We hypothesize the anomaly observed is due to the presence of deleterious recessive alleles in homozygosis. Therefore, the purpose of this study was to analyze the anomalous character, as well as to characterize the genomic region and genetic causes of the observed phenotype in *Eucalyptus grandis*.

MATERIALS AND METHODS

Plant material and character evaluation

A full-sib population with 11,240 seedlings was obtained from a controlled cross between two *Eucalyptus grandis* (codes G026 and G07) from Suzano Papel and Celulose SA company. For statistical analysis of normal and abnormal seedlings frequencies, chi-squared (χ^2) test was used to determine the segregation of normal and abnormal seedlings in a nursery trial (200 normal and 50 abnormal seedlings). These seedlings were selected for identification of molecular markers linked to the anomaly.

For the character evaluation, we studied four traits from 19 plants of each phenotype. The studied traits were: *i*) seedling height (SH); *ii*) seedling stem-base diameter (SD); *iii*) leaf shape; and *iv*) number of lateral branches. For the individual leaf area (LA) evaluation we used 6 to 9 leaves from two plants of each phenotype

and the measurements performed by ImageJ software (Wayne Rasband, National Institutes of Health - NIH, Bethesda, MD). Data from SH, SD and LA traits were subjected to Tukey's test by Statistical Analysis System (SAS) software (SAS 1986), with a statistical significance determined at the $p < 0.05$.

Genomic DNA extraction, RAPD (Random Amplified Polymorphic DNA) marker identification, and SCAR (Sequence Characterized Amplified Region) marker development

Genomic DNA was isolated from leaves using the CTAB method proposed by Murray and Thompson (1980) with modifications. About 150 mg of plant material were ground in liquid nitrogen to a fine powder. The powder was homogenized in 700 μ l of CTAB extraction buffer (2% CTAB; 2% PVP; 100 mM Tris-HCl pH=8.0; 25 mM EDTA; 2 M NaCl; 0.2% 2-mercaptoethanol) and heated for 45 min at 65 °C. 600 μ l of chloroform-isoamyl alcohol (CIA 24:1) was then added and the mixture was centrifuged for 5 min at 12,000 rpm. The upper phase was transferred to a new tube, 100 μ l of 5% CTAB solution (5% CTAB; 2% PVP; 100 mM Tris-HCl pH=8.0; 25 mM EDTA; 2 M NaCl) was added, and a new extraction with CIA was carried out. The upper phase was transferred to a new tube and 400 μ l of chilled isopropyl alcohol was added and the mixture was incubated overnight at -20 °C. The samples were centrifuged for 5 min at 12,000 rpm, the supernatant discarded, and the pellet washed with 95% ethanol. The DNA was dissolved in 80 μ l TE buffer (pH=8.0) containing 10 ng/ μ l of RNase. Total DNA concentration was quantified by spectrophotometer (Thermo Scientific NanoDrop™ 1000), and visualized after electrophoresis in 1% agarose gel in 1X TBE (Tris–borate-EDTA) stained with ethidium bromide.

For RAPD marker identification, we used the Bulk Segregant Analysis (BSA) technique (Michelmore et al. 1991). DNA from ten normal and abnormal seedlings, extracted individually, was mixed in equal quantities, creating two bulks (normal bulk and abnormal bulk). The two DNA bulks were screened for polymorphic markers using random primers (kits Operon Technologies, Inc.). The reaction consisted of 15ng DNA, 1.30 μ l 10X PCR Buffer (Invitrogen), $MgCl_2$ 1 mM, 1.04 μ l BSA 10mg/mL (Bovine Serine Albumin - Invitrogen), 1 mM dNTP mix, 0.3 μ M of primers, 1U of *Taq* DNA polymerase (Invitrogen), and water up to final volume of 13 μ l. PCR were performed under following conditions: initial denaturation at 94°C for 3 min, followed by 41 cycles of 94°C for 1 min, 35°C for 1 min, 72°C for 2 min, and final extension 72°C for 5 min. Amplified products were visualized after electrophoresis in 1% agarose gel in 1X TBE (Tris–borate-EDTA) stained with ethidium bromide.

We selected the polymorphic marker band that co-segregates with the abnormal character and converted in SCAR marker. Based on the sequence information of the cloned polymorphic fragments of the marker, SCAR primers were designed (Barret et al. 1998; Hernández et al. 1999; Negi et al. 2000). SCAR reaction consisted of 20 ng DNA, 2 μ l 10X PCR Buffer (Invitrogen), $MgCl_2$ 1.5 mM, 0.5 mM dNTP mix, 0.5 μ M of primers, 1U of *Taq* DNA polymerase (Invitrogen), and water up to final volume of 20 μ l. PCR were performed under following conditions: initial denaturation at 94°C for 5 min, followed by 34 cycles of 94°C for 45 s, 66°C for 30 s, 72°C for 1 min 35 s, and final extension 72°C for 10 min. Amplified products were visualized after electrophoresis in 1% agarose gel in 1X TBE (Tris–borate-EDTA) stained with ethidium bromide.

3' RACE (Rapid Amplification of cDNA Ends)

The seedlings were quickly frozen in liquid nitrogen and used to extract the total RNA by Chang et al. (1993) protocol with modifications. About 200 mg of plant material were ground in liquid nitrogen to a fine powder. The powder was homogenized in 1 ml of CTAB extraction buffer (2% CTAB; 2% PVP; 100 mM Tris-HCl pH=8.0; 25 mM EDTA; 2 M NaCl) and heated for 15 min at 65 °C. 1 ml of chloroform-isoamyl alcohol (CIA 24:1) was then added and the mixture was centrifuged for 10 min at 10,000 rpm. The upper phase was transferred to a new tube, and a new extraction with CIA was carried out. The upper phase was transferred to a new tube and one volume of lithium chloride was added and the mixture was incubated overnight at 4 °C. The samples were centrifuged at 12,000 rpm for 30 min at 4°C, the supernatant discarded, and the pellet washed with 70% ethanol. Total RNA was dissolved in 30 µl DEPC-treated water, concentration quantified by spectrophotometer (Thermo Scientific NanoDrop™ 1000), and visualized after electrophoresis in 1% agarose gel in 1X TBE (Tris–borate-EDTA) stained with ethidium bromide.

The total RNA was used for cDNA synthesis and 3'RACE procedure, utilizing the 3' RACE System for Rapid Amplification of cDNA Ends kit (Invitrogen), following the manufacturer's guidelines. For nested amplification, gene-specific internal and external primers (GSA1 and GSA2, respectively) were designed by OligoAnalyser 3.1 tool (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx>) based on the sequences of SCAR marker products.

Cloning and sequencing

The RAPD polymorphic band, PCR products of SCAR marker, and 3'RACE products were extracted from agarose gel, and DNA fragments were purified using the illustra GFX PCR DNA (GE Healthcare). Purified fragments were cloned into the pGEM-T Easy Vector (Promega) and inserted into UltraMAX DH5α-FT competent *Escherichia coli* cells (Life Technologies), following the manufacturers protocol.

White bacterial colonies with insert of RAPD and 3' RACE products were selected, and plasmid DNA was extracted by plasmid miniprep. In the plasmid miniprep, the selected bacterial colonies were grown overnight at 37°C in 3 ml liquid medium (CircleGrow – QBiogene). 1.4 ml of culture was then transferred to a new tube and centrifuged at 12,000 rpm for 1 min. The supernatant was discarded and 1.4 ml of culture was added and centrifuged at 12,000 rpm for 1 min. The supernatant was discarded and the pellet eluted in 300 µl of P1 solution (Tris-HCl 50 mM pH 8.0, EDTA 10 mM pH 8.0 and RNase 100 µgml⁻¹). 300 µl of P2 solution was added (NaOH 200 mM and SDS 1%) followed by incubation at room temperature for 5 min, and then 300 µl of P3 solution was added (potassium acetate 3 M pH 5.5) to neutralize and then incubated on ice for 20 min. Subsequently, the samples were centrifuged at 14,000 rpm for 30 min and the supernatant was transferred to a new tube with 400 µl of absolute isopropanol. The samples were incubated at -20°C for 1 h and centrifuged at 12,000 rpm for 30 min. The supernatant was discarded, ethanol 70% was added to rinse, and then the pellet was dried. Next, the DNA was eluted in deionized autoclaved water. Plasmid DNA was quantified by spectrophotometer (NanoDrop - ND1000). White bacterial colonies with insert of SCAR products were selected and amplified with SCAR primers. The amplicons were extracted from agarose gel, and DNA fragments purified using the illustra GFX PCR DNA (GE Healthcare).

Purified plasmid DNA with RAPD, 3'RACE products, and SCAR amplicons were bi-directionally sequenced. The sequencing was performed using ABI 3100 Genetic Analyzer (Applied Biosystems) using

BigDye Terminator Cycle Sequencing kit v. 3.1 (Applied Biosystems) according to the manufacturer's instructions.

Southern blot

Total DNA (20 µg) from the two genitors was digested with restriction enzymes *NdeI* and *XbaI*, electrophoresed on a 0.7% agarose gel, and transferred to Hybond-N+ nylon membrane (Amersham). For probe construction, SCAR marker was amplified, electrophoresed on a 0.7% agarose gel and the fragment purified using PureLink Quick Gel Extraction Kit (Invitrogen). The fragments were labeled using AlkPhos Direct Labeling and Detection Systems Kit (Amersham, UK). Hybridization and detection were performed as recommended by instruction manual (Amersham Pharmacia, UK).

In silico analysis

Sequencing SCAR and 3'RACE data were analyzed with ChromasPro 1.5 (Technelysium Pty Ltd) and ClustalX 2.0.11 (Tompson et al. 1997) software to form a consensus sequence. These were compared with GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>) and Phytozome "egrandis1.0" gene set database ("*Eucalyptus grandis* Genome Project 2010 - <https://phytozome.net/eucalyptus>) utilizing BLASTN tool (Altschul et al. 1997). The found ESTs with high identity were translated by ExPasy translate tool (<http://expasy.org>) (Gasteiger et al. 2003) and analyzed by FGENESH tool (Softberry, Inc. – www.softberry.com) to predict the protein. The results were analyzed by PFAM database (<http://pfam.sanger.ac.uk>) (Finn et al. 2008) for domain identification.

SCAR and 3'RACE sequences were compared with eucalypt genome in Phytozome v7.0 database (<http://www.phytozome.net>) utilizing BLAST tool (Altschul et al. 1997). The genomic regions with high identity were analyzed by FGENESH tool (Softberry, Inc. – www.softberry.com) to find predicted genes in the region. Predicted genes find were translated by ExPASy translate tool (<http://expasy.org>) (Gasteiger et al. 2003), and the protein analyzed by PFAM database (<http://pfam.sanger.ac.uk>) (Finn et al. 2008) for domain identification.

RESULTS AND DISCUSSION

Two phenotypic classes, abnormal and normal, were identified in full-sib population of a total of 11,240 seedlings. The 2,880 abnormal seedlings were detected based on some traits: reduction of root volume (Fig. 1a), shoot branching, height reduction (Fig. 1b), individual leaf area reduction, changes in leaf shape (Fig. 1c), and death in a 3 to 4 months. Because of these features, we called the abnormal seedlings as AnoDel (Anomaly and Deleterious). Measurement of seedling height, seedling stem-base diameter and individual leaf area showed significant difference ($P \leq 0.05$) (Fig. 1d) between wild and AnoDel phenotypes. The large differences detected indicate that these traits can be used for detection of AnoDel plants.

In *E. grandis* the shape of seedling (juvenile) leaves possess flat border and they are ovate: broader below the middle, pointed at the end, and ratio between length and breadth up to 3:1; and flat leaf border (Brooker and Kleinig 2004). These leaf characters were observed in all normal seedlings, but not in abnormal seedlings (Fig. 1c). The abnormal seedlings showed lanceolate leaves: lance-shaped, finally pointed, and ratio between length and breadth up to 8:1 (Brooker and Kleinig 2004); with undulate border (Fig. 1c).

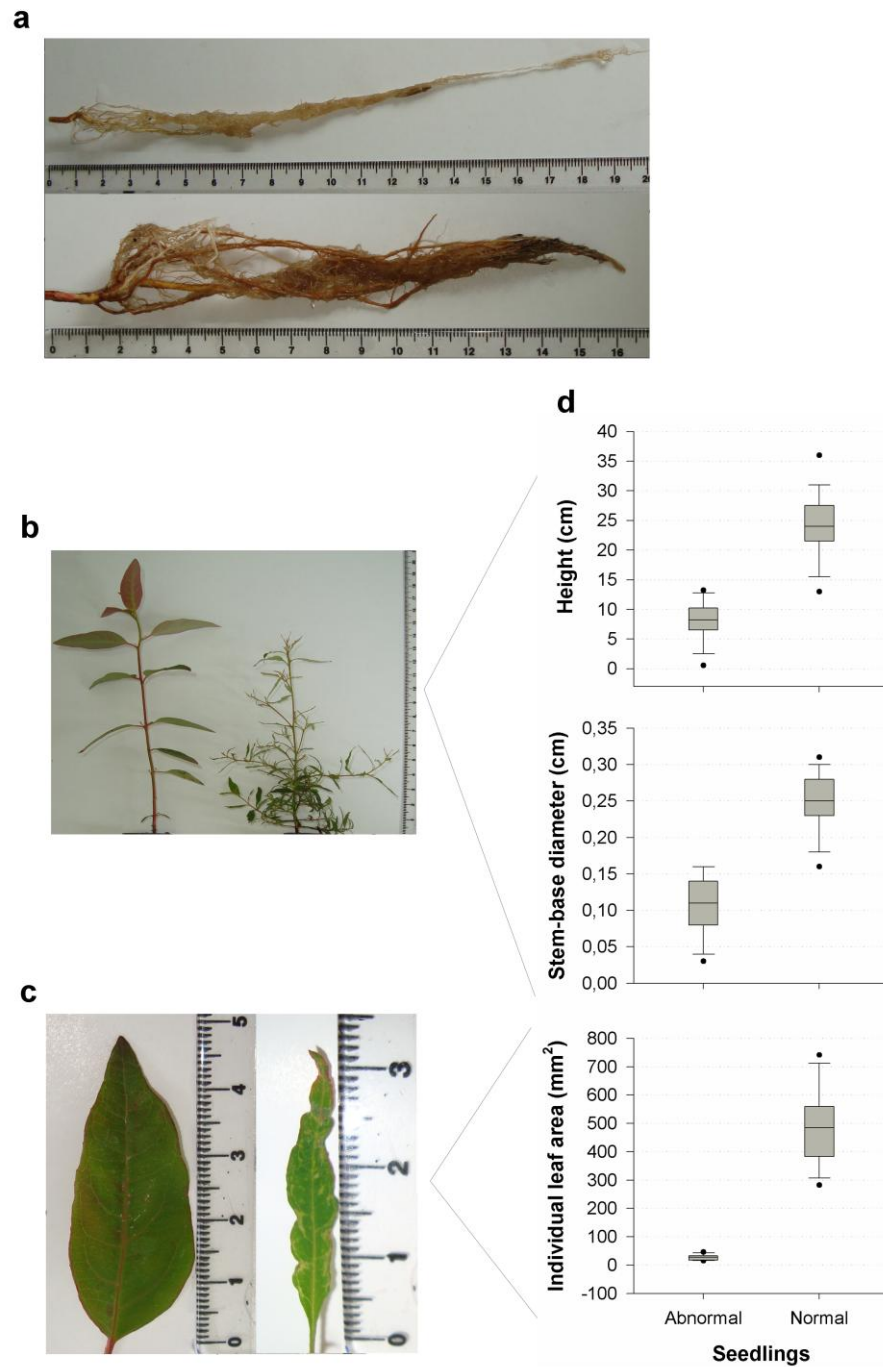


Fig. 1

a Root of normal (below) and abnormal (above) seedlings; **b** Shoot of normal (left) and abnormal (right) seedlings; **c** Leaf of normal (left) and abnormal (right) seedlings; **d** Graphic representation of height, stem-base diameter and individual leaf area of abnormal and normal seedlings

Nurserymen have related that eucalypt seedlings possess a single stem after germination and, in general, eucalypts trees are single-stemmed (Brooker and Kleinig 2004). The branching can occur in a short distance above ground level in woodland habitat (Brooker and Kleinig 2004). We observed that all normal seedlings were single-stemmed, whereas the abnormal seedlings showed lateral branches (Fig. 1b) in a variable number of 1 to 13.

The Chi-square (χ^2) test showed that the population segregated in the ratio 3 (normal): 1 (abnormal), confirming a Mendelian segregate of the trait ($\chi^2=2.32$ and $p<3.84$). Based on the segregation ratio, we hypothesized that the abnormal class (AnoDel), with undesirable features and premature death, has a polygenic character with a major gene effect, because the anomaly appears at different levels. The major gene would have a monogenic inheritance pattern with recessive deleterious allele effect (Byers and Waller 1999; Charlesworth et al. 1990; Klekowski Jr 1988; Young et al. 1996). Given the mixed mating system and open-pollinated progenies in *Eucalyptus*, that confers a high level of heterozygous loci (Golle et al. 2009; Klekowski Jr 1988; Zobel and Talbert 1984) and the unrelated genitors, these recessive deleterious alleles are allozygous, with the identical state of the alleles having independent origins through convergent evolution (Keller et al. 2011). Thus, we suggest that there is a pleiotropic gene with major effect that, when in homozygous recessive alleles, determines the early death and the anomalies at different levels in the seedlings.

To identify markers linked to abnormal traits, a combination of bulked segregant analysis (BSA) and molecular marker techniques were utilized. This combination has been shown to be an efficient method to identify markers linked to objectives genes or traits (Barret et al. 1998; Hyten et al. 2009; Lan et al. 2010; Liu et al. 2011; Michelmore et al. 1991). Since the complete genome sequencing of some species, such as *Eucalyptus* (Phytozome v7.0 database - <http://www.phytozome.net>), these combined techniques are an effective tool in localization of genomic regions of interest or molecular characterization (Hyten et al. 2009; Schneeberger and Weigel 2011).

A total of 100 random primers were tested on normal and abnormal bulks. Of these, 13 primers (AD01, AD02, AD06, AD11, AD17, AD18, AF03, AF05, AF12, AF14, AF17, AF18 and AF20) revealed polymorphisms between the bulks. However, only AD17 marker co-segregated with AnoDel phenotype (Fig. 2a) (Tambarussi et al. 2008). Among the 252 individuals screened (the genitors, 200 normal and 50 abnormal seedlings of progeny), the marker (band at 500 bp) was present in one genitor, in all abnormal, and in 31% of normal individuals; and absent in one genitor and in 69% of normal individuals. The co-segregation marker observed was similar with the expected (marker presence in all abnormal individuals and in 33% of normal individuals), indicating that it is linked in repulsion to the gene related with the AnoDel trait of one genitor without recombination. The results indicate that the linked gene possesses a pleiotropic character that, when in homozygous recessive alleles determinate the early death of seedlings, acting as deleterious recessive alleles, and causes several abnormalities in the seedlings at different levels. Thus, this pleiotropic gene had effect in a qualitative and monogenic character (early death), and a major effect in a quantitative and polygenic trait (several abnormalities at different levels).

The marker linked to abnormal trait was named AnoDel (Anomaly and Deleterious), the same as the mutant name, and cloned with the objective of converting these into a SCAR marker. SCAR markers are more advantageous than RAPD, because homozygous and heterozygous can be detected (it is a codominant marker) making it more informative (Agarwal et al. 2008; Negi et al. 2000; Barret et al. 1998); SCAR amplification reaction is faster than RAPD reaction (Barret et al. 1998); and have high reproducibility (Hernández et al. 1999). SCAR markers are easily obtained from RAPD fragments due to their size (500 to 1500 bp) (Negi et al. 2000). Based on the sequence data of the polymorphic marker detected, primers were designed for direct amplification of the corresponding loci from genomic DNA.

AnoDel marker (SCAR) genotyping of normal and abnormal individuals showed the same size (500bp) and pattern of the polymorphic marker detected (Fig. 2d). This result confirms the linkage of the AnoDel to abnormal trait. The knowledge generated in genomic studies allows the development of functional markers (FMs), derived from polymorphic sites within genes involved phenotypic trait variation, that can be used in marker-assisted selection (MAS) techniques (Andersen and Lübberstedt 2003). Thereby, the AnoDel marker developed will be used to screen elite plant material involved in eucalypt improvement programs. Since this marker is specific, future validation in other populations will allow for the detection of these recessive deleterious alleles in eucalypt species in general, so troubles to breeding improvement in eucalypt production can be avoided.

Sequence analysis detected a polymorphism of two base pairs between wild (Brasuz genome from Phytozome v7.0 database – <http://www.phytozome.net>) and mutant (abnormal) individuals. The polymorphism corresponds to the last two bases (CC) from the 3' end of AnoDel sequence of abnormal individuals (513-514bp) (Fig. 2c) that coincide with first two bases from the 5' end of RAPD primer, but have no similarity to the corresponding bases (GA) in the wild-type genome. This indicates that non-amplification of RAPD primer in normal individuals is due to non-pairing of these two bases of the primer with wild-type genome. This genomic difference can allow for the development of a FM to identify the anomaly in *Eucalyptus* genetic improvement programs. MAS has been shown to increase selection efficiency in genetic improvement programs, although in quantitative traits it is effective when selecting genes with large effects (Andersen and Lübberstedt 2003; Bernardo 2001; Lande and Thompson 1990). Utilizing FMs identified major effect genes related with phenotypic variation will reduce the risk of linkage drags (with undesirable effects), and the time and costs of the breeding programs (Andersen and Lübberstedt 2003).

Analysis of AnoDel (514bp) and 3' RACE (164bp and 174bp) sequences by BLASTN tool (Altschul et al. 1997) from Phytozome database and GenBank (est_others) showed high similarity (identity of 84%) with 3'UTR sequence of an eucalypt EST encoding a putative protein and with an EST of *Eucalyptus gunnii* sequence, respectively (Fig. 2f). Phytozome EST and *Eucalyptus gunnii* EST, when translated by ExPASy translate tool (Gasteiger et al. 2003), result in a predicted protein of 159 and 92 amino acids, respectively. Analysis by PFAM database (Finn et al. 2008) demonstrated that these putative proteins possess a Bet v1-type domain. Similar results were obtained when protein sequences were predicted with FGESH tool (<http://www.softberry.com>). Analysis of ESTs by BLASTX tool (Altschul et al. 1997) showed significant homology (68% of identity) with a sequence in the *Vitis vinifera* protein database in GenBank (nr). The *Vitis vinifera* hypothetical protein found had 159 amino acids and showed a Bet v1-type domain by PFAM database (Finn et al. 2008) analysis, exactly like FORETs and *Eucalyptus gunnii* ESTs analysis.

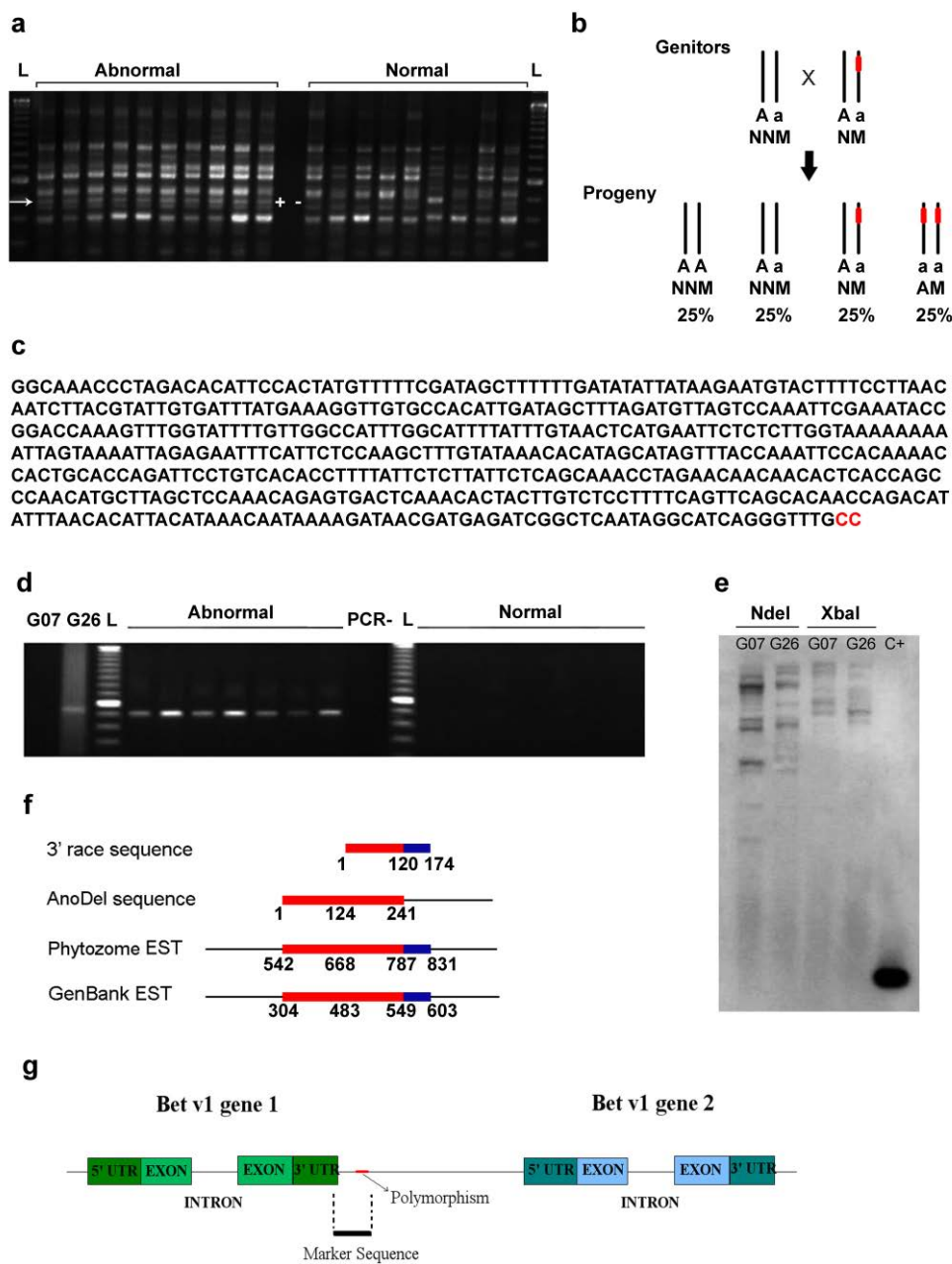


Fig. 2

a Polymorphic marker (arrow) obtained with primer AD17. The marker is shown in all abnormal individuals (left) and in some normal individuals (right). L: ladder 1Kb (Invitrogen); **b** Scheme of the marker presence in the genitors and full-sib progeny presents three classes NNM (normal without marker), NM (normal with marker), and AM (abnormal with maker); **c** SCAR (AnoDel) sequence shows polymorphism (red); **d** SCAR marker validation. G07 and G26 are the genitors. L: ladder 1Kb (Invitrogen); **e** Southern blot analysis; **f** Scheme showing alignment between AnoDel, 3' RACE, Phytozome EST; and GenBank EST sequences analysis. Identity between the four sequences (red) and between 3' RACE, Phytozome EST and GenBank EST sequences (blue); **g** Scheme of marker sequence alignment with the eucalypt genome

BLAST tool (Altschul et al. 1997) analysis in eucalypt genome database (<http://www.phytozome.net>) enable us localize the genome region of AnoDel marker. The AnoDel sequence showed similarity with 19 regions in the eucalypt genome. The region with the most similarity to AnoDel sequence plus the surrounding 2kb of genomic sequence on either side was analyzed by FGENESH tool (<http://www.softberry.com>) and showed that AnoDel sequence is located in an intergenic region between two protein-coding regions. In other

words, it is localized between the 3'UTR of the predicted protein with similarity with AnDel sequence and other protein-coding region (Fig. 2g). Both predicted proteins, by PFAM database (Finn et al. 2008) analysis, have a Bet v1-type domain.

Bet v1 is a multigene family of allergen peptides which are constituents of the White birch (*Betula verrucosa*) pollen, responsible for IgE binding in more than 95% of patients allergic to birch pollen (Breiteneder et al. 1989; Swoboda et al. 1995). They are homologous to major allergens of *Betulaceae* family (Breiteneder et al. 1989; Breiteneder et al. 1992; Valenta et al. 1991) and are related with allergens of *Rosaceae* and *Apiaceae* families (Breiteneder et al. 1995; Gao et al. 2005; Hoffmann-Sommergruber et al. 1999; Karlsson et al. 2004; Scheurer et al. 1997; Schoning et al. 1996; Vanek-Kreibitz et al. 1995). Bet v1-like genes consist of two exons interrupted by an intron of 76 to 359bp, localized between the first and second nucleotide of codon 62, which is highly conserved (Hoffmann-Sommergruber et al. 1997; Handschuh et al. 2007; Schenk et al. 2009; Wen et al. 1997). This structure corresponds with Phytozome EST and the two protein-coding regions through structure gene prediction by FGENESH tool (<http://www.softberry.com>).

Since the discovery of the Bet v1 sequence by Breiteneder et al. (1989), this protein has been compared with other proteins to identify its homologues and, consequently, its function (Breiteneder et al. 1989; Swoboda et al. 1995; Walter et al. 1996; Wen et al. 1997). Bet v1-homologous proteins form the PR10 family (Pathogenesis-related protein 10). In general, PR10 family proteins are associated with plant defense functions in response to biotic and abiotic stresses (Hoffmann-Sommergruber et al. 1997; Hoffmann-Sommergruber 2002; Liu e Ekramoddoullah 2006; Wen et al. 1997), but the exact function of Bet v1-like protein still unclear. PR10 proteins are reported to have important roles in biological processes, such as plant development, because they have high homology to ribonucleases (Liu e Ekramoddoullah 2006; Walter et al. 1996) and localization of PR protein isoforms in specific organs at certain stages of development has been observed (Hoffmann-Sommergruber 2002). In addition, Poupard et al. (2001) reported the presence of a Bet v1-like protein at sites of emerging secondary birch (*Betula pendula*) primordial root after induction by indole-3-acetic acid, suggesting that *Bet v1* could be involved in the processes of lateral root initiation.

The Bet v1 multigene family possesses a variable number of members in different species (Breiteneder 1989; Hoffmann-Sommergruber et al. 1997; Wen et al. 1997). Among the 19 regions with AnDel sequence similarity, 14 regions showed *Bet v1* proteins downstream. All regions are near, suggesting its organization is in tandem as well as other multigenic families (Capesius 1997; Passardia et al. 2004; Valério et al. 2004). Southern blot analysis showed several bands (Fig. 2e), confirming the different members of multigenic family *Bet v1* in *Eucalyptus grandis*. The larger number of multigenic family members in eucalypt can be explained by ancestral genomic duplication (Myburg et al. 2011).

The polymorphism of two base pairs detected is localized in an intergenic region, thus it is not possible to determine if the polymorphism is the cause of the anomaly. However, the polymorphism differentiates normal and abnormal individuals, suggesting its direct or indirect role in the anomaly phenotype. In other words, the polymorphism can influence directly or indirectly the gene involved in anomaly character, whether it be the Bet v1-like gene nearby or other genes situated in another region of genome.

CONCLUSIONS

The molecular marker developed is related with the anomaly character, because there is no recombination between it and the morphological marker (abnormal seedlings). The polymorphism detected in SCAR sequence suggests that the mutation is caused by a two base transversion. Because of the mutation is in an intergenic region it is unclear at this point whether these base substitutions are affecting nearby genes (such as identified Bet v1-like genes) or unknown genomic regions. However, the close proximity of Bet v1-like genes to the mutation and its identity with SCAR sequence, as well as the role of Bet v1-homologous proteins in several biological processes; suggests it is involved in anomaly character.

ACKNOWLEDGMENTS

We wish to thank the Suzano Papel and Celulose SA company for material research. We acknowledge Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp grant 08/52266-6), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support and fellowships. Finally, we thank Carlos Rodnei Nogueira for logistics assistance.

DATA ARCHIVING STATEMENT

Genotyping and phenotyping data have been deposited at TreeGenes Data Repository accession TGDR020. Sequence data from this article have been deposited with the DNA Data Bank of Japan under the accession nos. AB889687 and AB889688.

REFERENCES

- ABRAF (2013) Yearbook statistical ABRAF 2013 - base year 2012. ABRAF - Brazilian Association of Forest Plantation Producers, Brasília.
- Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant science. *Plant Cell Rep* 27:617-631. doi: 10.1007/s00299-008-0507-z.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389-3402.
- Andersen JR, Lübberstedt T (2003) Functional markers in plants. *Trends in Plant Science* 8:554-560. doi:10.1016/j.tplants.2003.09.010.
- Barret P, Delourme R, Foisset N, Renard M (1998) Development of a SCAR (sequence characterized amplified region) marker for molecular tagging of the dwarf BREIZH (Bzh) gene in *Brassica napus* L. *Theor Appl Genet* 97:828-833.
- Bernardo R (2001) What if we knew all the genes for a quantitative trait in hybrid crops? *Crop Sci* 41:1-4.
- Boland DJ, Broker MIH, Chippendale GM, Hall N, Hyland BPM, Johnson RD, Kleining DA, McDonald MW, Turner JD (2006) *Forest trees of Australia*, 5th edn. CSIRO Publishing, Collingwood.
- Breiteneder H, Ferreira F, Reikerstorfer A, Duchon M, Valenta R, Hoffmann-Sommergruber K, Ebner C, Breitenbach M, Kraft D, Scheiner, O (1992) Complementary DNA cloning and expression in *Escherichia coli* of Aln g 1, the major allergen in pollen of alder (*Alnus glutinosa*). *Journal of Allergy and Clinical Immunology* 90:909-917.
- Breiteneder H, Hoffmann-Sommergruber K, O'Riordain G, Susani M, Ahorn H, Ebner C, Kraft D, Scheiner O (1995) Molecular characterization of Api g 1, the major allergen of celery (*Apium graveolens*) and its

immunological and structural relationships to a group of 17kDa tree pollen allergens. *European Journal of Biochemistry* 233:484-489.

Breiteneder H, Pettenburger K, Bito A, Valenta R, Kraft D, Rumpold H, Scheiner O, Breitenbach M (1989) The gene coding for the major birch pollen allergen Betv1, is highly homologous to a pea disease resistance response gene. *The EMBO Journal* 8:1935-1938.

Brooker MIH, Kleinig DA. 2004. Field guide to *Eucalyptus*, 2nd edn, vol 3. Bloomings Books, Melbourne.

Brune A, Zobel B (1981) Genetic base populations, gene pools and breeding populations for *Eucalyptus* in Brazil. *Silvae Genetica* 30:146-149.

Byers DL, Waller DM (1999) Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Annual Review of Ecology and Systematics* 30:479-513.

Campinhos EN, Peters-Robinson I, Bertolucci FL, Alfenas AC (1998) Interspecific hybridization and inbreeding effect in seed from a *Eucalyptus grandis* x *E. urophylla* clonal orchard in Brazil. *Genet Mol Biol* 21:369-374.

Capesius I (1997) Analysis of the ribosomal RNA gene repeat from the moss *Funaria hygrometrica*. *Plant Molecular Biology* 33:559-564.

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* 11:113-116.

Chaix G, Gerber S, Razafimaharo V, Vigneron P, Verhaegen D, Hamon S (2003) Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis* *Theor Appl Genet*, 107:705-712.

Charlesworth D, Morgan MT, Charlesworth B (1990) Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* 44:1469-1489.

Finn RD, Tate J, Mistry J, Cogill PC, Sammut JS, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) The Pfam protein families database. *Nucleic Acids Research* 36:281-288. doi:10.1093/nar/gkm960.

Gao ZS, Van der Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, Van der Meer IM, Van der Geest AH, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen LJ (2005) Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Gen* 111:171-183. doi: 10.1007/s00122-005-2018-4.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucl Acids Res* 31:3784-3788. doi: 10.1093/nar/gkg563.

Golle DP, Reiniger LRS, Curti AR, Bevilacqua CB (2009) Forestry improvement: emphasis on biotechnology application. *Ciência Rural, Santa Maria* 39:1606-1613.

Grattapaglia D, Kirst M (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytologist* 179:911-929. doi: 10.1111/j.1469-8137.2008.02503.x.

Handsuh L, Femiak I, Kasperska A, Figlerowicz M, Sikorski MM (2007) Structural and functional characteristics of two novel members of pathogenesis-related multigene family of class 10 from yellow lupine. *Acta Biochimica Polonica* 54:783-796.

Hernández P, Martín A, Dorado G (1999) Development of SCARs by direct sequencing of RAPD products: a practical tool for the introgression and marker-assisted selection of wheat. *Molecular Breeding*, 5:245-253.

Hoffmann-Sommergruber K (2002) Pathogenesis-related (PR)-proteins identified as allergens. *Biochemical Society Transactions* 30:930-935.

Hoffmann-Sommergruber K, O'Riordain G, Ahorn H, Ebner C, Machado MLC, Puhlinger H, Scheiner O, Breiteneder H (1999) Molecular characterization of Dau c 1, the Bet v 1 homologous protein from carrot and its cross-reactivity with Bet v 1 and Api g 1. *Clinical and Experimental Allergy* 29:840-847.

- Hoffmann-Sommergruber K, Vanek-Krebitz M, Radauer C, Wen J, Ferreira F, Scheiner O, Breiteneder H (1997) Genomic characterization of members of the Bet v 1 family: genes coding for allergens and pathogenesis-related proteins share intron positions. *Gene* 197:91-100.
- Huijser P, Klein J, Lönnig WE, Meijer H, Saedler H, Sommer H. Bracteomania (1992) An inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. *EMBO J* 11:1239–1249.
- Hyten DL, Smith JR, Frederick RD, Tucker ML, Song Q, Cregan PB (2009) Bulk segregant analysis using the GoldenGate assay to locate the *Rpp3* locus that confers resistance to soybean rust in soybean. *Crop Sci.* 49:265–271. doi: 10.2135/cropsci2008.08.0511.
- Junghans TG, Peters-Robinson I, Bertolucci FL, Alfenas AC (1998) The use of self-incompatibility in the production of hybrid eucalyptus seed by 'Aracruz Celulose' in Brazil. *Genet. Mol. Biol.* 21:375–379.
- Karlsson AL, Alm R, Ekstrand B, Fjellkner-Modig S, Schiott A, Bengtsson U, Bjork L, Hjerno K, Roepstorff P, Emanuelsson CS (2004) Bet v 1 homologues in strawberry identified as IgE-binding proteins and presumptive allergens. *Allergy* 59:1277-1284. doi: 10.1111/j.1398-9995.2004.00585.x.
- Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189:237-249. doi: 10.1534/genetics.111.130922.
- Klekowski Jr EJ (1988) Genetic load and its causes in long-lived plants. *Trees* 2:195-203.
- Koh HJ, Son YH, Heu MH, Lee HS, McCouch SR (1999) Molecular mapping of a new genic male-sterility gene causing chalky endosperm in rice (*Oryza sativa* L.). *Euphytica* 106:57-62.
- Lan C, Liang S, Zhou X, Zhou G, Lu Q, Xia X, He Z (2010) Identification of genomic regions controlling adult-plant stripe rust resistance in Chinese landrace Pingyuan 50 through bulk segregant analysis. *Phytopathology* 100:313-318. doi:10.1094/PHYTO-100-4-0313.
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743-756.
- Liu F, Yao J, Wang X, Hu Z, Duan D (2011) Identification of SCAR marker linking to longer frond length of *Saccharina japonica* (Laminariales, Phaeophyta) using bulked-segregant analysis. *J Appl Phycol* 23:709–713. doi: 10.1007/s10811-010-9567-x.
- Liu J-J, Ekramoddoullah AKM (2006) The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. 68:3-13. doi:10.1016/j.pmpp.2006.06.004.
- Marx GA. 1983. Developmental mutants in some annual seed plants. *Ann. Rev. Plant Physiol.* 34:389-417.
- Michelmore RW, Paran I, Kesseli RV. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA* 88: 9828-9832.
- Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Research* 8:4321-4325.
- Myburg A, Grattapaglia D, Tuskan G, Jenkins J, Schumutz J, Mizrachi E, Hefer C, Sterck L, Hayes R, Rokhsar D. 2011. The *Eucalyptus grandis* genome project: genome and transcriptome resources for comparative analysis of woody plant biology. *BMC Proceedings* 5:120.
- Negi MS, Devic M, Delseny M, Lakshmikumaran M. 2000. Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection. *Theor. Appl. Genet.*, 101:146-152.
- Passardia F, Longeb D, Penela C, Dunand C. 2004. The class III peroxidase multigenic family in rice and its evolution in land plants. *Phytochemistry* 65:1879–1893.

- Poupard P, Brunel N, Leduc N, Viéumont J-D, Strullu D-G, Simoneau P. 2001. Expression of a Bet v 1 homologue gene encoding a PR 10 protein in birch root: induction by auxin and localization of the transcripts by *in situ* hybridization. *Aust. J. Plant Physiol.*, 28:57-63.
- Rédei GP. 1975. *Arabidopsis* as a genetic tool. *Annu. Rev. Genet.* 9:111-127.
- Schenk M F, Cordewener JHG, America AHP, Westende WPCV, Smulders MJM, Gilissen LJWJ. 2009. Characterization of PR-10 genes from eight *Betula* species and detection of Bet v 1 isoforms in birch pollen. *BMC Plant Biology* 9:24.
- Scheurer S, Metzner K, Hausteiner D, Vieths S. 1997. Molecular cloning, expression and characterization of Pru a 1, the major cherry allergen, *Molecular Immunology*, 34:619-629.
- Schoning B, Ziegler WH, Vieths S, Baltes W. 1996. Apple Allergy: The cDNA Sequence of the Major Allergen of Apple, Determined by Performing - PCR with a Primer Based on the N-Terminal Amino Acid Sequence, is Highly Homologous to the Sequence of the Major Birch Pollen Allergen. *Journal of the Science of Food and Agriculture*, 71:475-482.
- Schneeberger K, Weigel D. 2011. Fast-forward genetics enabled by new sequencing technologies. *Trends in Plant Science* 16:282-288.
- Swoboda I, Dang TCH, Heberle-Bors E, Vicente O. 1995. Expression of Bet v 1, the major birch pollen allergen, during anther development. An *in situ* hybridization study. *Protoplasma*, 187:103-110.
- Tambarussi EV, Lourenção JC, Bortoloto TM, Sasaki FT, Oda S, Marino CL (2008) Desenvolvimento de marcador SCAR (Sequence Characterized Amplified Region) relacionado a uma anomalia de viveiro em *Eucalyptus grandis*. In: 54º Congresso Brasileiro de Genética. Sociedade Brasileira de Genética, Salvador - BA, pp 17.
- Tompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG(1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.
- Valenta R, Breiteneder H, Petternburger K, Breitenbach M, Rumpold H, Kraft D, Scheiner O. 1991. Homology of the major birch-pollen allergen, Bet v I, with the major pollen allergens of alder, hazel, and hornbeam at the nucleic acid level as determined by cross-hybridization. *Journal of Allergy and Clinical Immunology*, 87(3):677-682.
- Valério L, De Meyer M, Penel C, Dunand C. 2004. Expression analysis of the *Arabidopsis* peroxidase multigenic family. *Phytochemistry* 65:1331–1342.
- Vanek-Krebitz M, Hoffmann-Sommergruber K, Machado MLD, Susani M, Ebner C, Kraft D, Scheiner O, Breiteneder H. 1995. Cloning and sequencing of Mal d 1, the major allergen from apple (*Malus domestica*), and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochemical and Biophysical Research Communications*, 214(2):538-551.
- Young A, Boyle T, Brown T. 1996. The population genetic consequences of habitat fragmentation for plants. *Tree* 11(10):413-418.
- Walter MH, Liu J-W, Wünn J, Hess D. 1996. Bean ribonuclease-like pathogenesis-related protein genes (Ypr10) display complex patterns of developmental, dark-induced and exogenous-stimulus-dependent expression. *Eur. J. Biochem.*, 239:281-293.
- Wen J, Vanek-Krebitz M, Hoffmann-Sommergruber K, Scheiner O, Breiteneder H. 1997. The Potential of Bet v 1 Homologues, a Nuclear Multigene Family, as Phylogenetic Markers in Flowering Plants. *Molecular Phylogenetics and Evolution*, 8(3):317-333.
- Zobel B, Talbert J. 1984. Applied forest tree improvement. John Wiley and Sons, New York, NY. 505 p.