

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
INSTITUTO DE BIOCÊNCIAS**

CAMPUS DE BOTUCATU

Leonardo Curi Martin

**Identificação de SNPs para uma anomalia de
desenvolvimento em *E. grandis* e seus efeitos
na estrutura proteica e interatoma**

**Botucatu-SP
2015**

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
INSTITUTO DE BIOCIÊNCIAS**

CAMPUS DE BOTUCATU

**Identificação de SNPs para uma anomalia de
desenvolvimento em *E. grandis* e seus efeitos
na estrutura proteica e interatoma**

Leonardo Curi Martin

**Tese apresentada ao Instituto de Biociências da
UNESP – Campus de Botucatu, para a obtenção do
título de Doutor em Ciências Biológicas – Genética.**

Orientador: Celso Luís Marino

Coorientador: Marcio Luis Acencio

Botucatu-SP

2015

Ficha catalográfica

A minha esposa, Priscila Keiko Nishi Martin, que sempre acreditou
e me apoiou em todos os momentos.

Dedico.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela bolsa de doutorado concedida.

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

Ao Dr. Marcio Luis Acencio, pela coorientação, paciência e ensinamentos.

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 e Departamento de Biofísica 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.

A minha tia Mirva Curi e meu irmão Gilberto Martin Filho por estarem sempre presentes nos momentos de alegria e tristeza e apoio incondicional.

A minha esposa Priscila Keiko Nishi Martin, por estar sempre ao meu lado me confortando, ajudando e apoiando.

Aos amigos do curso de pós-graduação, em especial à Pedro Gabriel Nachtigall, Karine Kettener, Maria Cecília Fuchs, Hernando Rivera-Jimenez, Lídia Carolina Arneiro, Agnes Takeda, Luiz Augusto Bovolenta, Júlio C. Otto, Vanusa do S. Leite, Natália Avelino pelo convívio, alegria, ajuda e paciência dedicados durante todo esse período.

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

Resumo

Martin LC. **Identificação de SNPs para uma anomalia de desenvolvimento em *E. Grandis* e seus efeitos na estrutura proteica e interatoma**[tese]. Botucatu: Instituto de Ciências Biológicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”; 2015.

Espécies de *Eucalyptus* têm sido utilizadas com sucesso para plantios florestais devido ao seu rápido crescimento, sua capacidade de adaptação às diversas condições climáticas e pelo seu potencial econômico na produção de energia, fibra e madeira sólida, reduzindo assim a pressão sobre as florestas tropicais e a biodiversidade associada. Investimentos em pesquisas e desenvolvimento estão sendo realizados pelo setor florestal, no intuito de aumentar os ganhos de produtividade através da pesquisa na área da biologia molecular. Visto que características anômalas nas progênes podem gerar perdas na produtividade, a identificação e caracterização dos genes relacionados a anomalias são muito importantes no auxílio em programas de melhoramento genético. Ao realizar um cruzamento controlado de *Eucalyptus grandis*, a empresa Suzano Papel e Celulose detectou uma anomalia com segregação mendeliana 3:1 (normais:anômalas) na progênie. As características eram bem significativas se comparadas com as plantas normais, tais como diferenças na altura, diâmetro da base do caule, tamanho das folhas e morte em poucos meses. Estudos anteriores identificaram uma marca molecular ligada à anomalia, e análises de comparação da sequência contra o banco de dados de eucalipto demonstrou identidade com proteínas da família Bet v1 PR10 (Pathogenesis-related protein 10), embora não estivesse claro o envolvimento deste gene no desenvolvimento da anomalia. Outros estudos traçaram o perfil transcricional dos fenótipos contrastantes pela técnica de RNAseq. Dentre os genes diferencialmente expressos, através do enriquecimento funcional, destacaram-se os pertencentes as famílias Bet v1, quitinase classe I e taumatina, todos relacionados à resposta contra patógenos, sugerindo que a anomalia seja causada pela ativação do sistema imune da planta (resposta autoimune). Com o intuito de se obter mais informações sobre a patogênese da anomalia genética sob investigação, adotamos 3 tipos de abordagens: (i) Identificação de SNPs entre os genes diferencialmente expressos das famílias Bet v1, quitinase e taumatina, (ii) identificação de miRNAs com potencial ligação aos mRNAs dos genes diferencialmente expressos e (iii) verificação na possível influência dos produtos dos genes diferencialmente expressos

sobre o interatoma. Utilizando os dados transcricionais de RNAseq, foram identificados, no total, de 97 SNPs, uma média de 1 SNP/142,7 pb, onde, mutações por transição e SNPs sinônimos prevaleceram. (i) A partir disso, marcadores moleculares foram desenvolvidos para validar os SNPs através da técnica de Sanger e poderão ser utilizados em progênes obtidas de cruzamentos controlados, pois foram desenvolvidos e validados em genes que possuem grande impacto à resposta da planta na defesa contra estresses bióticos e abióticos. (ii) Para dar suporte ao estudo na descoberta dos possíveis causadores da anomalia em questão, miRNAs em *Eucalyptus* foram identificados. Os transcritos preditos, foram alvos de 5 diferentes classes de miRNAs (euc-miR156, euc-miR157, euc-miR166, euc-miR167 e euc-miR390). Estes miRNAs foram considerados candidatos, atuando na regulação gênica causando prejuízo no desenvolvimento das ramificações laterais da raiz e no tamanho da planta anômala. (iii) Por fim, para verificar a influência dos genes/proteínas diferencialmente expressos sobre o interatoma, foi construído uma rede biológica utilizando como modelo a interação de proteínas de *Arabidopsis thaliana*. Nesta rede, foram introduzidos homólogos dos genes diferencialmente expressos (Eu2Ara) e, finalmente, análises de enriquecimento funcionais foram realizados para a descoberta dos processos sob maior influência das Eu2Ara na nessa rede. Nas análises de enriquecimento funcional, destacaram-se os processos biológicos enriquecidos significativamente relacionados a metabolismo energético, resposta à estímulos (resposta de defesa), RNAs não codificadores (incluindo miRNAs), replicação do DNA, sinalização e catabolismo de proteínas. Os resultados indicam que muitos processos biológicos que não estão envolvidos com os genes diferencialmente expressos, estão alterados em decorrência deles. Levando em consideração esses processos, concluímos que a morfofisiologia das plantas anormais parecem ser uma propriedade emergente das alterações desses múltiplos processos biológicos.

Palavras-chave: Anomalia genética vegetal; Eucalipto; Interatoma; SNP; miRNA.

Abstract

Martin LC. **SNPs discovery for a developmental anomaly in *Eucalyptus grandis* and its effects on protein structure and interactome**[thesis]. Botucatu: Instituto de Ciências Biológicas, Universidade Estadual Paulista “Júlio de Mesquita Filho”; 2015.

Eucalyptus species have been successfully used for forest plantations. Some intrinsic features such as rapid growth, adaptability to various weather conditions and economic potential in energy production, fiber and solid wood, help to reduce pressure on tropical forests and associated biodiversity. Investments in research and development on molecular biology are being carried out by forestry sector in order to increase productivity gains. Since anomalous characteristics of progeny may generate productivity loss, gene identification linked to abnormalities are crucial for breeding programs. During a controlled crossing with *Eucalyptus grandis*, the Suzano Papel e Celulose enterprise detected an atypical Mendelian segregation 3: 1 (normal: abnormal). Significant phenotypic differences concerning height, stem base diameter, and leaf size appeared, as well as an unusual early death, when compared to normal plants. Previous studies identified a molecular tag attached to this anomaly. Sequence comparison analysis of *eucalyptus* database showed identity with Bet v1 PR10 protein family (Pathogenesis-related protein 10). However, gene involvement in developing such aberration is undetermined. Other studies found a transcriptional profile of contrasting phenotypes by RNAseq technique. Among differentially expressed genes by functional enrichment highlighted those belonging to Bet v1 families, chitinases class I, and thaumatin, all related to pathogenic response. This fact suggests that aberration is caused by the plant immune system activation (autoimmune response). To investigate the pathogenesis of such genetic disorder, three approaches were adopted: (i) SNPs identification among the differentially expressed genes of Bet v1 families, chitinase, and thaumatin, (ii) miRNAs identification with potential binding to mRNAs differentially expressed genes, and (iii) check of possible influence of differentially expressed gene products on interactome. Using RNAseq transcriptional data, 97 SNPs were detected, 1 SNP/142.7 bp in average, being transition mutations and synonymous SNPs the most prevailing. (i) From this basis, molecular markers were developed to validate SNPs by Sanger technique. These markers may be used on progeny obtained from controlled crossings, since were created and

validated from genes with great impact on plant response in defense against biotic and abiotic stress. (ii) To endorse disclosing of possible cause of anomaly, miRNAs in *Eucalyptus* were detected. Predicted transcripts were targeted by 5 different miRNAs types (euc-miR156, euc-miR157, euc-miR166, euc-miR167 and euc-miR390). These miRNAs were considered candidates by acting on gene regulation and harming the evolution of lateral root branches and anomalous plant size. (iii) To verify influence of gene/protein differentially expressed on interatoma, a biological network was modeled from protein interactions of *Arabidopsis thaliana*. Gene homologues (Eu2Ara) from those differentially expressed were placed in this network, and analyzed by functional enrichment tests to determine processes under their higher influence. Biological enriched processes, significantly related to energy metabolism, such as response to stimuli (defense response), non-coding RNAs (including miRNAs), DNA replication, signaling and protein catabolism, peaked during the functional enrichment analysis. Findings indicate that many biological processes uninvolved with differentially expressed genes are altered because of them. As a conclusion, the morphophysiology of abnormal plant seems to be an emergent property of the changes in multiple biological processes.

Keywords: Plant genetic anomaly, Eucalypt, Interactome; SNP; miRNA.

Sumário

1 INTRODUÇÃO.....	12
2 REVISÃO DA LITERATURA.....	17
2.1 A biologia do gênero <i>Eucalyptus</i>	17
2.2 Importância econômica, ambiental e social do eucalipto no Brasil	18
2.3 Melhoramento genético de <i>Eucalyptus</i>	20
2.4 Mecanismos de defesa em plantas.....	22
2.5 Proteínas PRs (<i>pathogenesis-related</i>).....	26
2.6 Proteínas PR candidatas	28
2.6.1 Quitinases.....	28
2.6.2 Classificação da quitinase em plantas	29
2.6.3 Indução e Funções da quitinase de planta.....	30
2.6.4 Proteínas PR-10 e Bet v1	31
2.6.5 Função no estresse biótico e abiótico.....	32
2.6.6 Proteínas similares a taumatina - PR-5.....	33
2.7 Alelos deletérios e incompatibilidade genética.....	35
2.8 Marcadores Moleculares	37
2.9 SNPs (<i>Single-Nucleotide Polymorphisms</i>).....	40
2.10 Identificação de SNPs a partir de sequenciamento de nova geração (NGS)	42
2.11 MicroRNAs.....	43
2.12 Biogênese do miRNA em plantas	45
2.13 MiRNAs em espécies arbóreas	46
2.14 Redes biológicas.....	48
3 OBJETIVOS	51
4 MATERIAL E MÉTODOS	52
4.1 Material Vegetal	52
4.2 Extração de DNA.....	52
4.3 Seleção dos genes candidatos para análise <i>in silico</i>	54
4.4 Análise <i>in silico</i> e identificação dos SNPs.....	56
4.5 Seleção dos iniciadores específicos e condições de amplificação	57
4.6 Purificação dos produtos de PCR	58
4.7 Reação de sequenciamento	59
4.8 Precipitação das reações de sequenciamento	59
4.9 Alinhamento das Sequências para a procura dos SNPs	60
4.10 Classificação dos SNPs	60
4.11 Processamento e identificação dos miRNAs	61

4.12 Construção do interatoma da <i>Arabidopsis</i>	63
4.13 Construção do subgrafo dos diferencialmente expressos (Gdif)	64
4.14 Análise da estrutura da Gdif	64
4.15 Análise de enriquecimento funcional da Gdif e dos <i>hubs</i> e <i>bottlenecks</i>	65
5 RESULTADOS E DISCUSSÃO	67
5.1 Identificação dos SNPs entre os genes diferencialmente expressos das famílias Bet v1, quitinase e taumatina	67
5.1.2 Identificação <i>in silico</i> dos SNPs através dos dados obtidos por RNAseq e suas propriedades	67
5.1.3 Classificação dos SNPs	69
5.1.4 Desenho e amplificação dos iniciadores específicos	75
5.1.5 Sequenciamento pelo método de Sanger e validação dos SNPs	77
5.2 Identificação de miRNAs com potencial ligação aos mRNAs dos genes diferencialmente expressos	81
5.2.1 Processamento dos dados de sequenciamento para identificação dos prováveis miRNAs	81
5.2.2 Predição de alvos para miRNA em <i>Eucalyptus</i>	82
5.3 Verificação da possível influência dos produtos dos genes diferencialmente expressos sobre o interatoma.	89
5.3.1 Considerações iniciais	89
5.3.2 Determinação das proteínas em <i>A. thaliana</i> homologas às codificadas pelos genes diferencialmente expressos do <i>E. grandis</i> (Eu2Ara)	90
5.3.3 Construção do interatoma da <i>Arabidopsis thaliana</i> (Ara700)	94
5.3.4 Construção e análise da estrutura do subgrafo dos diferencialmente expressos (Gdif)	95
5.3.5 Análise de enriquecimento funcional	98
5.3.6 Análise de enriquecimento funcional da Gdif contra o proteoma de <i>A. thaliana</i>	98
5.3.7 Análise de enriquecimento funcional dos <i>hubs</i> da Gdif contra o proteoma de <i>A. thaliana</i>	110
5.3.8 Relação entre os processos biológicos possivelmente alterados e a morfofisiologia da planta anômala	114
6 CONCLUSÃO	117
7 REFERÊNCIAS BIBLIOGRÁFICAS	119

1 INTRODUÇÃO

O gênero *Eucalyptus* pertence ao grande grupo das angiospermas, família Myrtaceae. É nativo da Austrália, cobrindo 90% da área de todo o país, havendo cerca de 700 espécies identificadas com numerosas subespécies, variedades botânicas e híbridos. (Brooker, 2000; Poke, 2005). Suas características favoráveis para a produção de madeira (facilidade de propagação vegetativa, crescimento rápido, alta produtividade, boa forma de fuste e boa resposta a tratamentos culturais de manejo e melhoramento genético) fizeram com que o *Eucalyptus* se consolidasse como matéria-prima de base no setor florestal mundial e se tornasse o gênero florestal mais plantado no mundo (Mora & Garcia, 2000; Silva, 2005).

O melhoramento genético de espécies relacionadas à indústria florestal, vem centrando seus esforços, colaborando com o aumento da produtividade e qualidade dos produtos introduzidos no mercado (Golle et al., 2009; Harfouche et al., 2012). Existe uma série de obstáculos que dificultam a implementação de programas de melhoramento em espécies florestais, como fatores ligados à sobreposição de gerações, dificuldades de controle nos processos de polinização e fecundação além da complexidade na análise fenotípica dos descendentes. O grande período necessário para atingir o ciclo reprodutivo, a expressão de caracteres ao longo de várias idades, a necessidade de grandes áreas de plantio, dentre outros, também dificultam a execução de técnicas clássicas de melhoramento e, assim, inviabilizam a produção de linhagens puras (Tzfira et al., 1998; Resende, 2001).

Como consequência disso, 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, podendo elas estar relacionadas à presença de alelos recessivos deletérios presentes no material genético da população (Koh et al., 1999; Marx, 1983). Estopa (2006), comparando cruzamento de *Eucalyptus* em viveiros e no campo, relatou os efeitos dos alelos recessivos deletérios.

Tendo em vista as perdas causadas por esses genes na produção de mudas para plantios comerciais, a empresa Suzano Papel e Celulose, realizando cruzamentos controlados com *Eucalyptus grandis*, observou uma anomalia, sugerindo uma segregação mendeliana 3:1 (8360 normais e 2880 anômalas) (Tambarussi, 2006). As características das plantas anômalas eram uma espécie de superbotamento no caule, redução de altura, alterações na forma e tamanho do limbo foliar, além de sua morte em poucos meses (Figura 1). Constatando essa segregação mendeliana de 3:1 na progênie, Tambarussi (2006) identificou uma marca molecular, utilizando a técnica de RAPD (Random Amplification of Polymorphic DNA), relacionada com a anomalia. Esta marca foi identificada em todos os 50 indivíduos anômalos e apenas em 31% dos 200 indivíduos normais avaliados, segundo este mesmo autor. Pelo fato dos marcadores RAPD serem dominantes e apresentarem baixa reprodutibilidade (Ferreira e Grattapaglia, 1998), estes, normalmente são convertidos em marcadores codominantes, chamados SCAR (Região amplificada de sequência caracterizada), com a finalidade de se aumentar a confiabilidade de resultados (Michelmore et al., 1993).



Figura 1: Plântula anômala (à direita) e normal (à esquerda), com cerca de 4 meses de idade.

Em vista disso, Lourenção (2010) converteu os marcadores RAPD em SCAR, que são facilmente obtidos de fragmentos de DNA de 500 a 1500 pares de base (pb) (Negi et al, 2000). Outras vantagens deste marcador são a codominância, o que permite a detecção de heterozigotos e, assim, torna-se 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); permite reprodutibilidade (Hernández et al., 1999). O marcador SCAR, confeccionado por Lourenção (2010), demonstrou eficiência em identificar a presença de bandas em 100% dos indivíduos anômalos e em 22,2% dos indivíduos normais analisados, semelhante ao resultado apresentado por Tambarussi (2006). Ambos os trabalhos sugeriram que a característica seja controlada por um gene de efeito principal e

que o marcador encontra-se ligado a 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.

Como resultado, foi realizada uma análise *in silico* da sequência de 514 pb obtida a partir do marcador SCAR, através do banco de dados do genoma de *Eucalyptus* (*Phytozome*). Esta sequência, apresentou um polimorfismo de 2 pares de base nos indivíduos anômalos e, provavelmente, ela esteja relacionada com a anomalia. A sequência SCAR apresentou alta identidade com ESTs de *Eucalyptus* e *Vitis vinífera* de proteínas PR10 (Pathogenesis related protein 10) com domínio Bet v1 (Lourenção, 2010). Segundo a literatura, essas proteínas estão relacionadas a respostas contra estresses bióticos e abióticos (Liu e Ekramoddoullah, 2006).

Embora esta marca molecular relacionada à anomalia possuir identidade com genes PR-10, não é claro seu envolvimento no desenvolvimento do caráter anômalo.

Desta maneira, Fuchs (2014) teve como objetivo caracterizar os possíveis genes envolvidos na anomalia em *Eucalyptus* em fase inicial de desenvolvimento, traçando o perfil transcricional dos fenótipos contrastantes pela técnica de RNAseq. O RNAseq 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). Como resultado, Fuchs (2014) pôde constatar que a grande maioria dos genes diferencialmente expressos apresentaram expressão diferenciada nas plantas anormais, e estas foram classificadas por categorias funcionais ao nível de 'Processos biológicos' conforme o termo GO - *Gene Ontology*. As famílias gênicas que foram mais representativas nesta categoria foram as quitinases classe I, proteínas similares à taumatina e proteínas da superfamília Bet v1. Segundo a literatura, 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). Levando em consideração estes genes diferencialmente expressos, os quais foram classificados nas categorias funcionais de processos relacionados à resposta contra patógenos e metabolismo, Fuchs (2014) sugeriu que a anomalia fosse causada pela ativação inadequada do sistema imune da planta (resposta autoimune) associado à incompatibilidade genética denominada necrose híbrida.

2 REVISÃO DA LITERATURA

2.1 A biologia do gênero *Eucalyptus*

O gênero *Eucalyptus* L'Herit pertence à família *Myrtaceae*. Possui mais de 700 espécies distribuídas em oito subgêneros, sendo o principal deles o *Symphyomyrtus*, com mais de 300 espécies, dentre as quais estão as seis espécies mais plantadas para fins comerciais *E.grandis*, *E. globulus*, *E. urophylla*, *E. camaldulensis*, *E. saligna* e *E. tereticornis* (Eldridge et al., 1994; Brooker, 2000; Poke et al., 2005). O gênero é endêmico da Austrália, com exceção de *E. urophylla* e *E. deglupta*, que são naturais do Timor e Papua Nova Guiné, respectivamente (Keane et al., 2000; Potts e Pederick, 2000). As extremas diferenças climáticas existentes de norte a sul deste vasto continente, com as variações de altitude e solo, tem resultado em uma imensa diversidade de habitats para as quais, o *Eucalyptus*, teve ótima adaptação. Este gênero mostra uma excepcional diferenciação e ampla variabilidade genética. Esta variação fornece a base para programas de melhoramento de espécies adaptadas a um amplo espectro de ambientes, inclusive no Brasil.

Florestas plantadas de *Eucalyptus* são conhecidas por seu rápido crescimento, forma reta, qualidade superior da madeira para múltiplas aplicações, ampla adaptabilidade a solos e climas e facilidade de manejo por plantio direto e rebrota (Eldridge et al., 1994; Potts, 2004). Na última década, constatou-se que mais de 90 países plantam *Eucalyptus*, onde são utilizadas para diversos fins, tais como, madeira sólida, postes, energia, celulose, carvão vegetal, óleos essenciais, mel e tanino, bem como para sombra em parques e jardins (Doughty, 2000).

Este gênero se reproduz predominantemente por alogamia, ou seja, por fecundação cruzada, em sua maioria com a ajuda de insetos, particularmente por abelhas

(Quintin, 2001). A autofecundação acaba sendo comum, ocorrendo em 10% a 30% neste gênero. Uma característica marcante do gênero é a presença de opérculo. Os frutos são capsulares e deiscentes, liberando sementes de cor preta a amarelada e de tamanho variável, de menos de 1 mm até 2 cm (Cavalcanti e Willadino, 1963). As folhas variam de acordo com a maturidade da planta, mas normalmente são coriáceas, lanceoladas, com elevada quantidade de cutina e ricas em esclerênquima. Normalmente as folhas são alternadas, com algumas espécies mostrando folhas opostas. Os troncos apresentam pouca ramificação e normalmente são lisos e com casca decídua. Possuem, em geral, alburno delgado e claro, e cerne de amarelo a avermelhado (Fao, 1981). Dentre as espécies de *Eucalyptus* há arbustivas e arbóreas, com algumas espécies alcançando até 100 metros de altura, e todas possuem troncos lenhosos.

Em termos genéticos, todas as espécies plantadas no mundo possuem $2n = 22$ cromossomos (Eldridge et al., 1993), e o tamanho do genoma varia desde cerca de 380 milhões de pares de base (Mpb) em *C. citriodora* até por volta de 650 Mpb em *E. grandis* (Grattapaglia e Bradshaw, 1994) e 644 Mpb em *E. globulus* (Pinto et al., 2004). A hibridização entre espécies de diferentes subgêneros em geral não ocorre. Porém, é possível entre espécies dentro de subgênero, especialmente se pertencem à mesma secção (Griffin et al., 1988; Assis, 2000).

2.2 Importância econômica, ambiental e social do eucalipto no Brasil

A partir do século 18, diversas espécies do gênero *Eucalyptus* foram introduzidas em países como Índia, França, Chile, Brasil, África do Sul e Portugal onde apresentaram excelente adaptação climática, sucesso reprodutivo e índices produtivos elevados (Potts, 2004).

Na década de 60, com o desenvolvimento de métodos industriais de processamento das fibras curtas do *Eucalyptus*, as plantações de espécies do gênero começaram a apresentar importância comercial crescente e os programas de melhoramento tiveram início em países como Estados Unidos (*E. grandis*) e Portugal (*E. globulus*) (Eldridge et al., 1994).

Em 1970, iniciaram-se as plantações clonais no Congo e no Brasil com destaque para a espécie *E. urophylla*, principalmente em função da maior facilidade de propagação vegetativa desta espécie.

No início dos anos 80, populações base de melhoramento foram formadas, envolvendo principalmente as espécies *E. grandis*, *E. tereticornis* e *E. viminalis* (Eldridge et al., 1994).

No começo da década de 90, o Brasil já era o país com maior acervo genético de *Eucalyptus*, atrás somente da Austrália e Indonésia (Ferreira, 1992).

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). Isto confirma o amplo interesse da indústria de base florestal pelos recursos oferecidos por esta árvore, envolvendo grandes esforços nos programas de melhoramento que permitam obter com eficiência produtos derivados da madeira para suprir a crescente demanda nacional e internacional.

Segundo a Indústria Brasileira de Árvores (IBÁ, 2014), existem mais de 7,6 milhões de hectares cultivados com espécies florestais, sendo o *Eucalyptus*, representando mais de 72%, 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. No ano de 2012, o valor bruto da produção (VBP) do setor correspondeu à R\$ 56,3 bilhões, sendo os tributos arrecadados totalizados em R\$ 7,6 bilhões (0,5% da arrecadação nacional) e sua participação no superávit da balança comercial nacional alcançando índices 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. O papel central das florestas plantadas na economia de base florestal é reconhecido mundialmente por sua grande importância (IBÁ, 2014).

2.3 Melhoramento genético de *Eucalyptus*

O melhoramento genético, vem se empenhando em concentrar seus esforços em espécies de interesse na indústria florestal. De maneira considerável, está havendo um aumento da produtividade e qualidade dos produtos florestais. Várias são as características que o melhoramento florestal pretende atingir, entre elas, incremento no crescimento e na produtividade, alterações das propriedades químicas e físicas da madeira, resistência a doenças, tolerância a estresse abiótico, melhoria da capacidade fotossintética, dos caracteres fisiológicos, produção de compostos farmacêuticos e alterações na arquitetura da árvore. Com esse fim, são utilizadas diversas estratégias de melhoramento florestal, tanto por meio do uso de técnicas clássicas, e, mais recentemente, com o uso da biotecnologia (Golle et al., 2009; Harfouche et al., 2012).

A possibilidade de prever ganhos é considerada uma das maiores contribuições para o melhoramento. Os materiais genéticos utilizados no setor florestal apresentam, de

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

As espécies perenes possuem características que comprometem o sucesso de programas de melhoramento genético que dependam de técnicas clássicas. Existe uma série de obstáculos que dificultam a implementação de programas de melhoramento em espécies florestais (Strauss et al., 1992). Fatores como sobreposição de gerações, dificuldades de controle nos processos de polinização e fecundação, e complexidade na análise fenotípica dos descendentes. O grande período necessário para atingir o ciclo reprodutivo, a reprodução sexuada e assexuada, a expressão de caracteres ao longo de várias idades, a necessidade de grandes áreas de plantio, dentre outros, também dificultam a execução de técnicas clássicas de melhoramento (Tzfira et al., 1998; Resende, 2001). A depressão endogâmica é também um fator importantíssimo, que torna impraticável a criação de linhagens endogâmicas. Além dos problemas técnico-práticos, é necessário considerar as questões de caráter financeiro.

Em programas de melhoramento florestal, independentemente da espécie, os cruzamentos controlados são geralmente caros, consomem tempo e exigem pessoal treinado e área para a realização dos mesmos.

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

2.4 Mecanismos de defesa em plantas

Inúmeros mecanismos de resposta de defesa ocorrem em plantas, que por sua vez, são específicos contra várias formas de estresse. Em função da falta de capacidade de locomoção destes organismos, estas características foram primordiais para a sobrevivência das espécies atuais. Analisando este argumento, podemos dizer que os vegetais possuem certa plasticidade, pois conseguem alterar o plano de desenvolvimento e contornar situações desfavoráveis, como os ataques de pragas ou patógenos, bem como fatores abióticos desfavoráveis (Agrios, 1997).

Portanto, a sobrevivência das plantas, depende da habilidade de adaptação ao estresse, que atua como pressão seletiva. A adaptação e a resistência traduzem-se por alterações no metabolismo da célula vegetal, entre elas a síntese de proteínas de defesa,

expressas por genes específicos. Vários papéis são exercidos por estas proteínas, sendo estes atuantes auxiliando na resistência e sobrevivência da planta, podendo agir de forma direta, combatendo o agente agressor, ou de forma indireta, mantendo a estrutura e as funções celulares (Jones e Dangl, 2006).

De maneira geral, os mecanismos de resistência são ativados perto da área infectada de modo a prevenir a difusão do patógeno ou deter a contínua predação por insetos. A velocidade com que a planta reconhece a presença do agressor determina o tempo de resposta à invasão, desencadeando uma ou mais reações de defesa (Ryan, 2000).

A defesa basal ou de não hospedeiro, é chamada de “PAMP-triggered immunity” (PTI), e acontece logo após o contato com o patógeno. Padrões moleculares associados a patógenos (PAMPs) ou a micróbios (MAMPs) são reconhecidos por receptores de reconhecimento específico de padrões (PRR) localizados nas membranas das células do hospedeiro. Este mecanismo é considerado a primeira linha de imunidade inata em planta (Zhang et al., 2010).

A resistência específica de hospedeiro, chamada de “effector-triggered immunity” (ETI), é baseada na interação altamente específica entre efetores do patógeno (proteínas Avr) e proteínas específicas da planta (proteínas R). De acordo com esta teoria, denominada de Teoria Gene-a-Gene (Cui et al. 2009), a resistência genética é controlada por um gene de resistência (R) na planta e um gene de avirulência (avr) correspondente no genoma bacteriano. Essa interação é denominada incompatível, e o hospedeiro tem seus mecanismos de defesa ativados e o patógeno não consegue infectar e desenvolver-se em seu interior, não havendo a indução de sintomas da doença (Hammond-Kosak e Jones, 1996; Daurelio et al., 2013). Essa interação, reconhecida como uma resposta hipersensibilidade (RH), é caracterizada por sintomas visíveis nas folhas da planta como morte celular localizada, lesões necróticas para limitar e conter o patógeno no

local de infecção (Oh et al., 2008; Mur et al., 2008), modificações estruturais, produções de espécies reativas de oxigênio (ROS), acúmulo de proteínas relacionadas à patogênese (PR-Proteínas) em espaços intracelulares e a expressão dos fatores de transcrição que ativam os genes das plantas que caracterizam a resposta do hospedeiro (Daurelio et al., 2013).

No entanto, se a planta é incapaz de reconhecer o patógeno ou não consegue ativar rapidamente as vias de defesa, como é o caso das plantas suscetíveis, a doença ocorre e esta interação é conhecida como interação compatível (Grewal et al., 2012). Nesse caso a interação Avr/R não ocorre (Febres et al., 2009; Domingues et al., 2011) e se dá a multiplicação do patógeno às custas do hospedeiro (Daurelio et al., 2013).

O fenômeno de reconhecimento é resultante da evolução conjunta do hospedeiro e do patógeno (Hammond-Kosack e Jones, 1996). A maioria dos genes R codifica proteínas citoplasmáticas contendo um sítio de ligação de nucleotídeo (nucleotide binding site-NBS) e um domínio carboxil rico em leucina (leucine rich repeats-LRR), geralmente denominados genes NBS-LRR (Tor et al., 2004). O desencadeamento das vias de defesa por genes R gera fluxo de íons celulares massivos, uma “explosão oxidativa”, levando ao acúmulo de superóxidos e peróxido de hidrogênio, óxido nítrico, resposta de hipersensibilidade (RH) e a produção de metabólitos antimicrobianos (Dangl e Jones, 2001; Taiz e Zeiger, 2013). A reação de hipersensibilidade resulta na indução da morte celular programada (MCP) no sítio de infecção, na ativação de diferentes sinais, os quais resultam na expressão de uma variedade de genes de defesa, no aparecimento de lesões necróticas no local da infecção e, conseqüentemente, no impedimento do desenvolvimento da infecção pelo patógeno. Os aspectos fisiológicos da RH incluem o aumento rápido e transitório de agentes oxidantes, a perda de íons potássio (K^+) e ganho de íons hidrogênio (H^+) pelas células, a destruição de compartimentos e o espessamento

das paredes celulares e da cutícula, além da síntese de toxinas (fitoalexinas) e proteínas relacionadas à defesa, denominadas proteínas PR (Linthorst, 1991; Nimchuk et al., 2003).

Sequências de genes de resistência a doenças de plantas mostram que, apesar de conferirem resistência a patógenos diferentes tais como vírus, bactérias e fungos, os mesmos contêm sequências similares e codificam peptídeos com semelhanças estruturais (Kanazin et al., 1996).

Há trabalhos que sugerem que várias vias metabólicas são estimuladas após a ativação do sistema imune das plantas. As vias do ácido salicílico (AS), ácido jasmônico (AJ) e do etileno (ET) resultam na ativação de fatores de transcrição de genes de defesa (De Souza et al., 2009, An e Mou, 2012).

Segundo pesquisadores, 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). Com isso, o AS intensifica a produção de espécies reativas de oxigênio (ROS) e, conseqüentemente, a resposta de hipersensibilidade (Levine et al., 1994; Durrant e Dong, 2004). Todavia, existem estudos que relatam que o H_2O_2 leva ao acúmulo de AS através do aumento da atividade da enzima ácido benzóico-2 hidroxilase, também denominada citocromo P450 monoxigenase, que converte ácido benzóico em AS (Hammond-Kosack e Jones, 1996; Ryals et al., 1996; Durrant e Dong, 2004).

A participação do AS na sinalização e indução da resistência sistêmica adquirida (SAR) é confirmado pela sua presença no floema após a infecção pelo patógeno (Ward et al., 1991; Durner et al., 1997; Durrant e Dong, 2004). A SAR atribui aos tecidos não infectados resistência contra infecções secundárias pelo patógeno a partir de sinalizações decorrentes do sítio inicial de infecção (Durner 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 genes SAR 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), confirmando 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 ocorrência da ativação do sistema de defesa das plantas sem que ocorra ataque por patógenos (autoimunidade) está, habitualmente, associada à um tipo de incompatibilidade genética, a qual se denomina necrose híbrida (Bomblies et al., 2007; Mizuno et al., 2010). Este tipo de incompatibilidade pode ser observado em cruzamentos inter ou intraespecíficos, nos quais parte da progênie apresenta fenótipo (fraqueza, nanismo e atrofiamento) semelhantes à plantas atacadas por patógenos ou sob estresse abiótico (Bomblies e Weigel, 2007).

2.5 Proteínas PRs (*pathogenesis-related*)

Proteínas relacionadas à patogênese (PR) foram definidas como proteínas codificadas pela planta hospedeira, mas induzida por vários tipos de agentes patogênicos, tais como vírus, bactérias, fungos, e também, pela aplicação de produtos químicos que “imitam” o efeito de uma infecção patogênica ou induzem algum tipo de estresse semelhante. A acumulação de proteínas PR são desencadeada por ataque de patógenos (biótico), estresse abiótico, reação de hipersensibilidade e SAR (Liu e Ekramoddoullah, 2006). As Proteínas PR foram classificadas em 17 famílias diferentes, variando de PR-1

a PR-17 à partir de sua estrutura primária, atividade sorológica, e atividade biológica em plantas mono e dicotiledôneas (Liu e Ekramoddoullah, 2006).

A Família PR-1 é induzida por AS e por patógeno e é comumente utilizado como um marcador para o SAR (Mitsuhara et al., 2008). As funções das proteínas da família PR-1 não são ainda conhecidas. A família PR-2 contém β -1,3-glucanase (glucano endo-1,3- β -glicosidases) e catalisa a hidrólise de β -1,3-glucanos. Eles agem na defesa dos fungos através da hidrólise da parede celular fúngica e também pela geração de elicitores (geram respostas de defesa). As proteínas PR-3, PR-4, PR-8 e PR-11 são quitinases vegetais. Membros da família PR-3 hidrolisa ligações β -1,4 que ligam os resíduos N-acetilglucosamina de quitina, que conferem resistência a doença com base na atividade antimicrobiana (Van Loon e Van Strein, 1999). Endoquitinases são proteínas denominadas PR-4 (Neuhaus et al., 1996) e possuem um Domínio C-terminal chamado Barwin e são induzíveis por ataque de patógenos. A família de proteínas PR-5 inclui permatinas, osmotinas, zeamatinas, e taumatinas (thaumatin-like protein - TLP). TLPs causam a ruptura osmótica dos poros da transmembrana nas membranas plasmáticas dos fungos. Proteínas PR-6 são referidas como inibidores de proteinases. Eles reduzem a capacidade do patógeno em usar suas enzimas líticas e completar o seu ciclo de replicação (Sels et al., 2008). PR-7 é uma proteinase e foi caracterizada em tomate (Vera e Conejero, 1988). A família de proteína PR-9 é mencionada como peroxidases, e são enzimas-chave nos processos de construção da parede celular. Elas catalisam deposição de lignina e reforçam a parede celular das plantas durante o ataque patogênico (Taheri e Tarighi, 2011). Proteínas PR-12 são referidas como as defensinas vegetais, e sua função exata não esta muito clara, mas existem estudos que a consideram com capacidade em inibir o crescimento de fungos e leveduras. Proteínas PR-13 são mencionadas como tioninas, estes são identificados com efeitos antimicrobianos causando a permeabilização das

membranas celulares. Tioninas são uma fonte de enxofre, encontrado em proteínas de reserva em sementes (Sels et al., 2008). Proteínas PR-14 são citadas como proteínas de transferência de lipídios (LTPs), e estão envolvidas na síntese da cutina, β -oxidação, sinalizando na defesa de planta (Sels et al., 2008). PR-15 e PR-16 são relatados à partir de cevada e são induzidos por *Erysiphe graminis*. A PR-17 tem PRp27 de tabaco como membro (Ebrahim et al., 2011).

2.6 Proteínas PR candidatas

2.6.1 Quitinases

Quitinases são catalizadores enzimáticos que clivam uma ligação entre C₁ e C₄ de dois monômeros consecutivos de quitina N-acetil-D-glucosamina. Elas são amplamente distribuídas na natureza, ocorrendo em bactérias, fungos, animais e plantas. A quitina é um componente comum das paredes das células fúngicas e do exoesqueleto de artrópodes (Bartnicki-Garcia, 1968). Quitinases de plantas são geralmente endoquitinases, capazes de degradar a quitina, um dos principais constituintes de certas paredes celulares de fungos, bem como inibir o crescimento dos fungos (Schlumbaum et al., 1986; Broekaert et al., 1988). Algumas quitinases classe I estão localizadas nos vacúolos celulares, enquanto outras enzimas, incluindo quitinases classe III, estão localizados fora da célula (Neuhaus et al., 1996). Quitinases extracelulares podem bloquear diretamente o crescimento das hifas, invadindo o espaço intercelular, possivelmente liberando os elicitores fúngicos, que, em seguida, induzem a biossíntese de quitinase e outras reações de defesa no hospedeiro (Barber et al., 1989; Mauch e Staehelin, 1988). Há fortes indícios de que quitinase, em conjunto com β -1, 3 glucanase participem no sistema de defesa da

planta contra patógenos fúngicos. As quitinases são encontradas nas folhas de diversas plantas, ao passo que não existe nenhuma função conhecida no crescimento e desenvolvimento da planta. Quitina e β -1, 3-glucanase são os principais componentes da parede celular de muitos fungos e há possibilidade da quitinase e β -1,3-glucanase distinguir os componentes da parede celular de fungos tendo assim, a função antifúngica (Abeles et al., 1970; Pegg, 1988).

2.6.2 Classificação da quitinase em plantas

Estas enzimas são classificadas em duas categorias, endoquitinases e exoquitinases. Endoquitinases clivam a quitina aleatoriamente em pontos internos do polímero, produzindo solúveis, multímeros de baixo peso molecular, como quitotrose, e quitotetraose e o dímero, di-acetilquitobios. Exoquitinases são divididas em duas subcategorias: quitobioses, catalisando liberação de di-acetilquitobioses de microfibrilas de quitina, e β -1,4-N-acetilglucoseaminidase, clivando produtos oligoméricos de endoquitinases quitobiosidases e com a produção de monômero de N-acetilglucosamina (Cohen-Kupiec e Chet, 1998). Com base nas suas estruturas primárias, quitinases de plantas têm sido classificados em sete classes, classe I a VII. Diferentes classes dessas enzimas não possuem nenhuma correlação aparente ao estar presente em uma espécie de planta em particular e órgãos da planta ou tecido. No entanto, certas isoformas de quitinases são por vezes induzidos por um eliciador em particular. Por exemplo, na batata, as quitinases básicas e ácidas, podem ser ativadas. Também apenas isoformas particulares têm atividades antifúngicas e algumas isoformas têm mostrado um outro papel como atividade anticongelante (Sela-Buurlage et al., 1993; Yeh, et al., 2000).

Quitinases Classe I são ricas em cisteína e têm um domínio de ligação N-terminal a quitina (CBD). CBD é separado do domínio catalítico por uma prolina e possui uma região espaçadora rica em glicina, variável, tanto em tamanho como em composição. Quitinase Classe I, no tabaco, devido à supressão do CBD e da região espaçadora, isolada ou em combinação, reduz a atividade hidrolítica em 50%, também a função antifúngica é reduzida em 80% (Suarez et al., 2001).

Quitinases classe III são únicos em estrutura e não têm relação com qualquer outra classe de quitinases vegetais. Estas enzimas pertencem à família PR-8 e da família 18 de glicosil-hidrolases. Quitinases classe III geralmente têm atividade lisozima e parecem estar mais intimamente relacionado com as quitinases bacterianas.

2.6.3 Indução e Funções da quitinase de planta

O efeito inibitório da quitinase na planta sobre o crescimento fúngico foi demonstrada por estudos *in vitro* sobre o crescimento de fungos que contêm quitina como um componente da sua parede celular (Mauch et al., 1988; Schlumbaum et al., 1986). A indução destas enzimas foi inicialmente apresentado nas plantas de ervilha infectadas com *Fusarium solani* (Mauch et al., 1988). O efeito inibidor da proteína extraída à partir de plantas de ervilha foi infectado por *Fusarium solani* e mostraram inibir o crescimento de 15 das 18 espécies de fungos testadas *in vitro*.

Quitinase purificada inibiu o crescimento de uma única espécie de fungo, enquanto que uma combinação destas enzimas e outra proteína PR, β -1, 3-glucanase, inibiu o crescimento de todos os fungos testados, mostrando assim, um sinergismo em atividades (Mauch et al., 1988). O número de estudos confirmou os resultados em tabaco

(Yun, et al., 1990) uvas (Derckel et al., 1998), grão de bico (Giri et al., 1998) arroz (Velazhahan et al., 2000) e outras plantas.

Vários estudos têm demonstrado que a expressão de quitinase contra sistemas de fitopatógenos é mais elevada e a indução é mais forte nas variedades resistentes em comparação com variedades susceptíveis na beterraba sacarina (Nielsen et al., 1993), trigo (Anguelova et al., 2001) e as variedades de tomateiro (Lawrence et al., 2000).

Transformação de genes quitinase foram realizadas em videiras, arroz e amendoim e maior resistência à doença foi obtido. Na uva, o gene da quitinase de arroz teve maior resistência das plantas contra o patógeno *Uncinula necator* (Yamamoto et al., 2000).

Quitinases que não eram de plantas também foram transformadas nas plantas modelos e apresentaram um aumento na resistência em variedades susceptíveis. Um gene da quitinase do fungo *Trichoderma harzianum* foi transformado em plantas de tabaco e batata. As plantas transgênicas aumentaram sua resistência em relação a vários fungos patogênicos, incluindo *Alternaria alternata*, *A. solani*, *Botrytis cinerea* e *Rhizoctonia solani* (Lorito et al., 1998).

2.6.4 Proteínas PR-10 e Bet v1

Proteínas PR-10, incluindo os alérgenos de pólen das árvores estão presentes como família multigênica em plantas que possuem sementes. Com base em semelhanças nas suas sequências de aminoácidos, localização subcelular e provável função, eles são classificados em dois grupos distintos: proteínas relacionadas com patogênese intracelulares (DPI) com homologia à ribonucleases e (S)-norcoclaurina sintases (NCS). Em plantas dicotiledôneas, proteínas PR-10 foram relatados em salsa, ervilha, batata,

feijão, soja, aipo, alfafa. Em monocotiledóneas, eles estão presentes em aspargos, arroz, lírio, sorgo. Entre as espécies de árvores, foram relatadas em *Pinus lambertiana*, *Pinus strobus*, *Pinus monticola*, *Pseudotsuga menziesii*, *Pinus pinaster*, *Picea glauca*, *Fagus sylvatica*, *Betula verrucosa*, *Jatropha curcas*. Há pelo menos cinco genes PR-10 em *Pisum sativum*, dezoito genes Mal d 1 em *Malus domestica*, dez genes Bet v1 em bétula, oito genes Fra um 1 em morango, seis genes PR-10 em *Solanum surattense*, oito em tremoço amarelo, cinco em arroz e dezessete genes PR-10 em *Vitis vinifera* (Lebel et al., 2010). Eventos de duplicação de genes ocorreram dentro do genoma, e assim, surgindo várias cópias de PR-10, que estão presentes sob a forma de grupos nos cromossomos (Liu e Ekramoddoullah, 2006). Isto é observado como uma consequência da evolução.

2.6.5 Função no estresse biótico e abiótico

A indução de genes PR relacionadas à patogênese em resposta a patógenos e parasitas que vão desde vírus microscópicos até insetos herbívoros e diferentes estímulos ambientais qualifica a sua implantação como gene importante para a tolerância a múltiplos estresses. Patógenos, como vírus, bactérias, fungos, bem como fatores abióticos como salinidade, frio, desidratação estão envolvidas na expressão induzível de proteínas PR-10. A maioria das proteínas PR-10 são induzidas pela ação dos compostos de sinalização como AS, AJ, ABA (ácido abscísico) ou ET, e possuem atividade antimicrobiana *in vitro* por meio de atividades hidrolíticas em paredes celulares, toxicidade por contato, e, portanto, estão envolvidos em múltiplas sinalizações sob stress.

A expressão do gene do arroz *JIOsPR10* foi induzida por moléculas de sinalização na presença de luz, por outro lado, no escuro, a expressão não foi observada, indicando que o sinal de luz pode estar envolvido na regulação da sua expressão (Jwa et al., 2001).

A Bet v1 é a representante mais conhecida da família PR-10. É constituída de diferentes isoformas que ocorrem naturalmente, codificados por inúmeros genes, que partilham mais do que 95% de identidade de sequência (Schenk et al., 2009).

A estrutura terciária da proteína Bet v 1 já foi determinada por difração de raio-X (Gajhede et al., 1996) e por espectroscopia NMR- Nuclear Magnetic Resonance (Faber et al., 1996; Gajhede et al., 1996). Swoboda et al. (1995) sugeriram que proteínas Bet v1 têm um papel importante durante a germinação do pólen no estigma, protegendo os tecidos reprodutivos femininos de infecção por patógenos. Apesar de ser, de longe, a proteína mais abundante no pólen de bétula, a função fisiológica de Bet v1 ainda permanece indefinida. Como o principal alérgeno de pólen de bétula branca, Bet v1 representa um dos melhores modelos de alérgeno caracterizado em imunologia. É o agente causador de diversas reações alérgicas tais como rinite alérgica e de polinose. Em todo o mundo, estima-se que 100 milhões de pessoas sejam sensíveis a Bet v 1, especialmente na América do Norte e Europa Ocidental. Em alguns países, mais de 90% de todos os pacientes sensíveis ao pólen de bétula exibiram reação ao soro de IgE dirigido a Bet v 1. (Durrant e Dong, 2004). Com base em sua capacidade de ligação a IgE, isoformas de Bet v 1 foram divididos em variedades de hiper e hipoalergênicos (Liu Ekramoddoullah, 2006).

2.6.6 Proteínas similares a taumatina - PR-5

As proteínas da família PR-5 também são conhecidas como *taumatina-like* proteínas (TLPs) por causa de sua sequência de aminoácidos e semelhanças estruturais à proteína da espécie *Thaumatococcus daniellii* (Edens et al., 1982). A família da proteína TLP (PR-5) também inclui as proteínas intimamente relacionadas permatina, osmotina e zeamatina. Várias isoformas de proteínas PR-5 foram identificados em uma ampla gama

de espécies de plantas, particularmente em espécies utilizadas em cultivo, tais como *Hordeum vulgare* (Reiss et al., 2001) *Oryza sativa* (Reimann e Dudler 1993) e *Avena sativa* (Lin et al., 1996). Muitas proteínas PR-5 foram purificadas e demonstraram possuir atividade antifúngica, em experiências *in vitro*. Uma proteína PR-5 de *Arabidopsis thaliana*, *ATLP3*, expressa em bactérias, mostrou atividade antifúngica considerável contra fungos patogênicos, tais como, *Verticillium alboatrum*, *Verticillium dahliae* e *Fusarium oxysporum* bem como fungos não patogênicos tipo *Trichoderma reesei* e *Candida albicans* (Xu et al., 1997). Proteínas PR-5 à partir de cevada mutante inibiu o crescimento de *Trichoderma viride* e *C. albicans* (Hejgaard et al., 1991). *WAS-3a*, que codifica a principal isoforma de taTLPs de trigo de inverno, inibiu o crescimento das hifas de *Microdochium nivale* quando em baixa temperatura (Kuwabara et al., 2002). Plantas de tabaco transgênicos com superexpressão da proteína PR-5 de arroz apresentaram uma tolerância aumentada para *Alternaria alternata*, o patógeno da mancha marrom (Velazhahan e Muthukrishnan, 2003). Plantas de batata transgênicas que expressam um gene PR-5 de tabaco aumentou significativamente a resistência ao *Phytophthora infestans* (Zhu et al., 1996).

Outros estudos têm mostrado que proteínas do tipo PR-5 também foram induzidos sob estresse ambiental. Por exemplo, as proteínas PR-5 acumuladas no apoplasto de centeio de inverno durante o frio exibiram atividade anticongelante (Antikainen et al., 1996) e as proteínas básicas PR-5 foram induzidas em tabaco, em resposta ao estresse osmótico (Singh et al., 1987).

Um subgrupo de proteínas PR-5 de monocotiledóneas foram identificados sendo substancialmente menores em tamanho (abaixo de 18 kD), com uma deleção em sua sequência de cerca de um quarto dos aminoácidos (Reimann e Dudler, 1993). WRTLTP2 exibiram atividades antifúngicas quando superexpressos em *Escherichia coli* (Chan et al.,

1999), indicando que as TLPs menores que carecem 6 dos 16 resíduos conservados de cisteína também estão envolvidos na defesa das plantas contra infecções fúngicas (Reimann e Dudler, 1993).

No entanto, o mecanismo pelo qual as proteínas PR-5 exercem sua atividade antifúngica ainda não é totalmente compreendido. Foi sugerido que a atividade antifúngica estaria associado a um aumento na permeabilidade das membranas dos fungos por mecanismos de formação de poros (Selitrennikoff, 2001). Foi demonstrado que todas as proteínas PR-5 tem uma fenda de superfície ácida. A superfície é polarizada eletrostaticamente considerando o sítio de ligação para um receptor situado na membrana plasmática de fungos e podem ser críticos para a atividade antifúngica deste grupo de proteínas (Batalia et al., 1996).

Os fitormônios AS, AJ e ET podem desempenhar papéis-chave na regulação de indução na resposta de defesa em plantas. Enquanto AS está acionando o caminho da resistência local adquirida (LAR) e resistência sistêmica adquirida (SAR), que está associado ao acúmulo de proteínas PR, AJ e ET estão ativando a via de resistência sistêmica induzida em *A. thaliana* (Pieterse et al., 1998).

Outros estudos revelaram que as proteínas PR-5 ao interagir especificamente com (1,3)- β -D-glucano da parede celular fúngica irá impedir a montagem adequada da parede celular dos fungos durante a extensão das hifas, resultando num aumento da permeabilidade da membrana plasmática (Bormann et al., 1999).

2.7 Alelos deletérios e incompatibilidade genética

As principais fontes de variabilidade genética em uma população são as alterações gênicas e cromossômicas.

Esse fenômeno se torna essencial na geração de novos alelos, sendo a mutação o modo mais simples, onde ocorre a substituição de pares de bases, deleções ou inserções no material genético. O resultado dessas alterações, podem acarretar na mudança de sequência de aminoácidos de uma proteína e, conseqüentemente, a perda da função (Freeman e Herron, 2009).

Entretanto, outras formas de mutação podem ser abordadas, tais como, as recombinações cromossômicas via crossing-over e a transposição de elementos genéticos móveis (transposons), podendo resultar na geração de novos alelos, caso isto ocorra dentro dos genes (Stears e Hoekstra, 2003).

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 maior parte das plantas que tem como característica a preferência pela polinização cruzada, como o eucalipto, são heterozigotos para os alelos recessivos deletérios, assim, permanecem encobertos na carga genética da população. Entretanto, o cruzamento entre indivíduos aparentados e a autofecundação podem reduzir a heterozigose e permitir que alelos recessivos de efeito desfavorável se manifestem em homozigose (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 (Freeman e Herron, 2009).

Após uma 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 outros caminhos evolutivos, se tornando em cópias do gene original, ou manter parte das funções do gene original e por último, adquirir novas funções. Em geral, duplicações recorrentes e divergências funcionais geram as denominadas famílias multigênicas (González- Duarte e Albalat, 2005).

Diferentes ambientes, em geral, apresentam pressões de seleção distintas. À vista disso, um indivíduo que possui certos genes de uma mesma família gênica, podem divergir evolutivamente dos genes de outro indivíduo que vive em um ambiente totalmente diferente (Bomblies e Weigel, 2007). Caso esses genes dos indivíduos adaptados aos ambientes diferentes sejam combinados, podem acabar resultando em interações negativas por incompatibilidade genética (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 (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).

2.8 Marcadores Moleculares

Os diferentes tipos de marcadores moleculares disponíveis variam conforme a tecnologia utilizada para revelar variabilidade em nível de DNA, habilidade para detectar diferenças entre indivíduos, custo, facilidade de uso, consistência e repetibilidade (Oliveira et al., 2007).

Nos últimos tempos, diversos tipos de marcadores moleculares foram desenvolvidos e, por padrão, são classificados como marcadores dominantes e codominantes.

Nos marcadores dominantes, a variação no DNA é detectada pela presença ou ausência de fragmentos, o que torna impossível a distinção entre locos em homozigose ou heterozigose, quando o fragmento está presente. Exemplos de marcadores dominantes são RAPD, Random Amplified Polymorphic DNA, (Williams et al., 1990) e AFLP, Amplified Fragment Length Polymorphism, (Zabeau, 1993). Os marcadores dessa classe não são dirigidos a uma região particular do genoma e revelam o polimorfismo presente em dezenas de locos simultaneamente (Souza, 2001).

A base do RAPD pode ser resumida em uma reação de PCR utilizando-se o DNA da amostra estudada e um *primer* único de sequência randômica. Dessa forma, mutações, deleções e ou inserções no sítio de iniciação podem causar a não complementariedade entre o DNA da amostra e o *primer*, impedindo o pareamento do iniciador e consequentemente a não amplificação (Williams et al., 1990; Souza, 2001). Assim a detecção do polimorfismo é baseada na presença ou ausência do fragmento.

A técnica de AFLP consiste na amplificação seletiva de um subconjunto de fragmentos do genoma obtidos após digestão do material genético realizada com auxílio de enzimas de restrição. Para isto, a mutação é revelada devido à presença ou ausência de um dado sítio de restrição entre dois genótipos (Ferreira e Grattapaglia, 1998).

Os marcadores codominantes apresentam os dois alelos de um loco, permitindo a distinção entre locos homozigotos de heterozigotos. Os mais conhecidos são os marcadores RFLP, Restriction Fragment Length Polymorphism, (Botstein et al., 1980); microssatélites (SSR) *Simple Sequence Repeat*, (Tautz, 1989); EST-SSR, *Expressed*

Sequence Tag contendo microssatélites, (Cato et al., 2001) e SNP (*Single Nucleotide Polymorphisms*), (Gupta, 2001).

A técnica do RFLP é fundamentada na fragmentação do DNA por meio de enzimas de restrição e hibridização com sonda específica marcada com radioatividade ou com quimioluminescência. Diante disso, o polimorfismo é comprovado devido a observação de fragmentos de restrição de diferentes tamanhos (Ferreira e Grattapaglia, 1998).

Microssatélites são sequências de DNA que contém poucos nucleotídeos, 1 a 6 pares de bases de comprimento, repetidas em tandem. Embora o número de repetições em tandem de um microssatélite em uma mesma região pode possuir variabilidade entre genótipos diferentes, as sequências ao microssatélite pode ser única no genoma e conservada. Com isso, se torna possível o desenho de *primers* específicos para as sequências adjacentes a um dado microssatélite. Os produtos dessa amplificação irão revelar um polimorfismo de acordo com o tamanho do produto de PCR amplificado. A visualização das bandas no gel pode ser feita diretamente por coloração ou através de autoradiografia (Ferreira e Grattapaglia, 1998).

O sequenciamento de DNA e marcadores moleculares começaram a convergir quando os dados acumulados passaram a mostrar variações entre sequências de indivíduos de uma mesma espécie. Logo se percebeu que Single-Nucleotide Polymorphisms (SNP) eram os marcadores mais abundantes nos genomas (Ganal et al., 2009).

2.9 SNPs (*Single-Nucleotide Polymorphisms*)

Um SNP, ou polimorfismo de um único nucleotídeo, é entendido como uma posição no DNA onde diferentes sequências alternativas (alelos) coexistem em indivíduos em uma população, e o alelo menos frequente possui frequência maior que 1% (Brookes, 1999). Ao considerar a frequência alélica na definição de SNPs, distingue-se esta classe de marcador de meras mutações ao acaso que ocorrem no genoma ou mesmo de erros de sequenciamento que, pela baixa frequência, dificultariam o uso da tecnologia em análise genética. Não estão incluídos como marcadores SNP pequenas inserções e deleções (indels), ainda que alguns autores mantenham esses dois tipos de polimorfismo (SNP e indels) na mesma categoria. Em teoria, é aceitável a existência de quatro alelos diferentes para cada nucleotídeo em um sítio SNP, uma vez que existem quatro bases nitrogenadas que compõem o DNA (A, C, T, G).

Todavia, na prática, observa-se em mais frequência duas possíveis variações, fato que pode ser explicado pela ocorrência desigual de substituições de base do tipo transição ($A \leftrightarrow G$, $T \leftrightarrow C$) e transversão ($A \leftrightarrow C$, $A \leftrightarrow T$, $G \leftrightarrow C$, $G \leftrightarrow T$). Assim, apesar do número possível de transversões ser duas vezes maior do que o de transições, o que se observa na prática é que a ocorrência de transições é cerca de duas vezes maior do que de transversões (Vignal, 2002).

Portanto, marcadores SNP normalmente possuem uma natureza bialélica e a tecnologia criada para identificar este polimorfismo é também baseada no bialelismo.

Estas mutações pontuais possuem comportamento codominante, isto é, no genótipo heterozigoto os dois alelos são identificados. Nesses últimos anos, o interesse e a procura por SNPs nos diferentes organismos aumentou muito. O sequenciamento completo do genoma de espécies modelo e o aumento do número de sequências

disponíveis em bancos de dados públicos vem estimulando o interesse da comunidade científica por este tipo de marcador (Ganal et al., 2009). A natureza bialélica dos SNPs, os torna menos informativos do ponto de vista genético quando comparados a marcadores microssatélites. Isto pode ser compensado pela abundância de SNPs, encontrados em larga escala ao longo do genoma (Kruglyak, 1997).

A estimativa do número de SNPs observados em uma espécie depende, naturalmente, das relações de vínculo genético entre as amostras de acessos utilizadas na análise. Se as amostras apresentam grande diversidade genética, a tendência é se observar maior número de SNPs a cada par de base (pb) analisado. A grande abundância de SNPs no genoma vem estimulando o seu emprego em pesquisa básica e aplicada de diferentes espécies. SNPs têm sido utilizados com sucesso, por exemplo, na construção de mapas genéticos de alta densidade, caracterizados por uma ocorrência no genoma muito além da capacidade dos marcadores tradicionais (Hyten et al., 2010). A elevada densidade de SNPs nos mapas genéticos, em teoria, aumenta a possibilidade de que estas mutações próximas a genes de interesse agrônômico sejam amostrados e, com isso, associações entre genótipo e fenótipo sejam mais prontamente encontradas, o que facilitaria a seleção assistida para características complexas. Deve ser frisado que a frequência de SNPs é heterogênea ao longo do genoma, diferindo entre região codificadora/região não-codificadora, por exemplo, ou entre sequências gênicas ou mesmo dentro de genes (éxon/intron). Em geral, os SNPs são menos abundantes em regiões gênicas (codificadoras) do que em outras regiões (Fladung e Buschbom, 2009).

Segundo Emara e Kim (2003), os SNP podem ser classificados em dois tipos quando localizados regiões codificadoras: não sinônima, que são aquelas que resultam na mudança de aminoácidos de uma proteína (Figura 3b), ou sinônimas que não resultam em mudança de aminoácido (Figura 3a).



Figura 3a. Mutação do tipo sinônima (ou silenciosa) – A troca de uma adenina por uma guanina não acarretou na mudança de aminoácido.

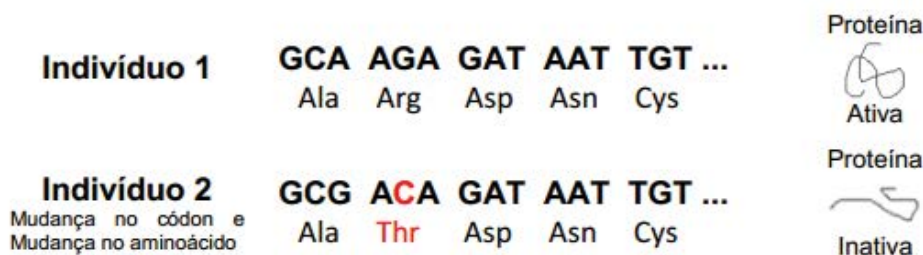


Figura 3b. Mutação do tipo não-sinônima – A troca de uma guanina por citosina acarretou na mudança de aminoácido.

2.10 Identificação de SNPs a partir de sequenciamento de nova geração (NGS)

O desenvolvimento de novas tecnologias de sequenciamento em larga escala revolucionaram o cenário da geração de sequência, com custos que baixam a cada ano, tendo aumentado exponencialmente a velocidade de geração de dados de sequências nos últimos anos. As chamadas tecnologias de sequenciamento de segunda geração ou NGS (Next Generation Sequencing), incluindo Roche 454 (Margulies et al., 2005); Solexa/Illumina (Bennett, 2004) e ABI Solid (Valouev et al., 2008), são capazes de produzir uma quantidade de enorme de dados de sequência, da ordem de bilhões de bases, em um curto intervalo de tempo, em sistemas altamente automatizados, e a um baixo custo por base sequenciada (Ganal et al., 2009). Com isso, essas tecnologias aumentaram expressivamente a velocidade na identificação de SNPs e, conseqüentemente,

estimularam a seu aproveitamento em diversas áreas da ciência, inclusive no melhoramento genético de plantas, como em cevada (Wicker et al., 2008), eucalipto (Novaes et al., 2008; Grattapaglia et al., 2011; Hendre et al., 2012; Thumma et al., 2012), pinus (Cronn et al., 2008) e milho (Emrich et al., 2007). A tecnologia de alta performance apresenta vantagens para a análise do genoma estrutural de plantas, mesmo sem o genoma de referência. Também pode ser aplicada para o estudo de uma variedade bastante ampla de questões biológicas, partindo de uma análise da estrutura do genoma ou de variações de níveis de expressão de transcriptomas, até a identificação de microssatélites e de SNPs, além de análise de interações DNA-proteína.

Entre todos, a plataforma Illumina é o preferido para descoberta de SNP devido à profundidade e natureza quantitativa de seus dados (Harismendy et al., 2009; Paszkiewicz e Studholme, 2012).

2.11 MicroRNAs

MicroRNAs (miRNAs) são uma classe de pequenos RNAs não codificadores endógenos que possuem de 20 a 22 nucleotídeos e que regulam a expressão gênica após ou durante a transcrição. Os miRNAs são codificados por genes MIR, os quais se encontram amplamente distribuídos no genoma de plantas e animais (Bartel et al., 2009).

Em animais, os genes MIR estão localizados principalmente no interior de regiões intrônicas, enquanto que, em vegetais, estes são encontrados em regiões intergênicas, apresentando suas próprias unidades de transcrição, caracterizadas por regiões promotoras e terminadoras (Megraw et al., 2006; Voinnet, 2009). Os miRNAs regulam a expressão de genes alvo atuando principalmente ao nível pós-transcricional, por meio de interações RNA-RNA (Ha et al., 2008).

Tanto a clivagem quanto o mecanismo de repressão da tradução requerem uma complementaridade de pares de bases entre o miRNA e o seu RNA mensageiro (mRNA) alvo. A inibição traducional é encontrada principalmente em animais, nos quais ocorre maior número de “mismatches” (não complementaridade entre os nucleotídeos) entre o miRNA e sua sequência alvo. Em plantas, os miRNAs e seus alvos apresentam alto grau de complementariedade (Voinnet, 2009) e a inibição ocorre, de um modo geral, via clivagem do mRNA, ocorrendo usualmente no centro da complementariedade, no resíduo que pareia-se com o décimo ou décimo primeiro nucleotídeo do miRNA (Yang et al., 2007).

A existência de uma complementaridade quase perfeita entre os miRNAs de plantas e seus respectivos mRNAs alvos tem facilitado a identificação de genes -alvo para muitos miRNAs já descritos (Rhoades et al., 2002). Uma possível explicação para o alto grau de complementaridade entre os miRNAs e seus alvos em plantas é provavelmente a consequência da atuação de processos evolutivos nos genes MIR os quais ocorreram via a “hipótese de duplicação invertida”. Tal hipótese propõe que os genes de miRNAs surgiram a partir de duplicações invertidas do seus genes-alvos nos genomas vegetais (Allen et al., 2004). Ao longo do tempo, estas duplicações invertidas acumulam mutações que tornaram-se altamente divergentes do seu ancestral, exceto nas regiões correspondentes às sequências do miRNA maduro e da sua sequência complementar (Allen et al., 2004).

Desde que foram reportados em plantas em 2002, várias evidências sugerem que esses pequenos RNAs são reguladores cruciais em diversos aspectos do desenvolvimento vegetal (Ori et al., 2007; Wu et al., 2009; Aukerman e Sakai, 2003, Nogueira et al., 2007).

2.12 Biogênese do miRNA em plantas

A biogênese dos miRNAs é iniciada pela transcrição dos genes MIR pela RNA polimerase II (Pol II), gerando um transcrito primário ou precursor (pri-miRNA) cuja estrutura é semelhante a um mRNA fita-simples, possuindo um cap 7-metilguanosina na extremidade 5' e uma cauda poli-A na extremidade 3'. Além disso, o pri-miRNA forma uma estrutura secundária tipo "hairpin" (Lee et al., 2004), e é processado via a ação de diferentes enzimas, incluindo a enzima do tipo RNase III DICER-LIKE1 (DCL1), gerando um precursor intermediário ou pré-miRNA (Vaucheret, 2006). O pré-miRNA é novamente clivado pela enzima DCL1, gerando um duplex imperfeito de RNA (20-22 nt) que contém tanto o miRNA maduro quanto sua fita complementar (denominado miRNA*). Nesta fase, a DCL1 atua junto com duas proteínas nucleares: HYL1 (double stranded RNA binding protein1) e SERRATE (zinc finger protein) (Han et al., 2004; Lobbes et al., 2006). O duplex miRNA-miRNA* é metilado nas extremidades 3' por ação de uma metiltransferase nuclear (HEN1), conferindo-lhe maior estabilidade e evitando sua degradação (Yu et al., 2005). O duplex é subsequentemente transportado para o citoplasma por meio da enzima HASTY (Park et al., 2005). Já no citoplasma, a dupla fita é desfeita provavelmente pela ação de uma enzima helicase, ainda não identificada.

Finalmente, o miRNA maduro é incorporado ao complexo de ribonucleoproteínas denominado RISC que contém a enzima ARGONAUTA1 (AGO1), enquanto que o miRNA* é excluído do RISC e conseqüentemente exposto para degradação (Jones-Rhoades et al., 2006). A proteína AGO1, "guiada" pelo miRNA maduro, cliva mRNAs específicos ou promove a repressão da tradução de genes alvo (Voinnet, 2009). Tanto a

clivagem quanto o mecanismo de repressão da tradução requerem uma complementaridade de pares de bases entre o miRNA e o seu mRNA alvo.

Em plantas, os miRNAs e seus alvos apresentam alto grau de complementariedade (Voinnet, 2009) e a inibição ocorre, de um modo geral, via clivagem do RNA mensageiro e acontece usualmente no centro da complementariedade, no resíduo que parecia com o décimo ou décimo primeiro nucleotídeo do miRNA (Yang et al., 2007). A existência de uma quase perfeita complementaridade entre os miRNAs de plantas e seus mRNAs alvos tem facilitado a identificação de genes-alvo para muitos miRNAs já descritos (Rhoades et al., 2002).

2.13 MiRNAs em espécies arbóreas

Espécies arbóreas têm como principais diferenciais a formação de madeira, o tamanho e a característica perene que determina a necessidade de lidar com variações extremas de clima, de estresse biótico e abiótico, aspectos, sabidamente, controlados por miRNAs (Khraiweh, et al., 2011; Kruszka et al., 2012). A compreensão dos mecanismos relacionados a respostas a estresse se torna imprescindível, especialmente, em um momento de grandes alterações climáticas e do potencial risco de aumento da suscetibilidade a patógenos. Assim como grande parte dos estudos de função de miRNAs em plantas, vários experimentos com miRNAs em árvores foram desenvolvidos relacionados com desenvolvimento e condições de estresse (Osakabe et al., 2011; Sun et al., 2011).

Entre os representantes de espécies florestais no miRBase (Griffiths-Jones et al., 2008) estão *Hevea brasiliensis* (31 miRNAs maduros) (Gebelin et al., 2012), *Acacia mangium* e *A. auriculiformis* (3 e 7 miRNAs, respectivamente) (Wong et al., 2011; Ong e Wickneswari, 2012), as coníferas *Pinus taeda* (38 miRNAs) (Lu, et al., 2007), *Pinus*

densata (31 miRNAs) (Wan et al., 2012) e *Picea abies* (41 miRNAs) (Yakovlev et al., 2010). A espécie florestal mais bem representada é a espécie modelo para árvores *Populus trichocarpa*, que teve seu genoma completo publicado em 2006 (Tuskan et al., 2006), com cerca de 352 sequências de miRNAs maduros. A descoberta de miRNAs foi, recentemente, relatada pela primeira vez em uma espécie da família Myrtaceae, *Eugenia uniflora*, com a identificação de 25 miRNAs conservados e 17 novos (Guzman et al., 2012).

Experimentos com mutações de ganho de função em *A. thaliana* envolvendo genes de fatores de transcrição HD-Zip classe III (*homeo domain leucine-zipper*) mostraram que estes são importantes na formação do meristema e diferenciação do câmbio e que sua expressão é regulada por miRNAs. A partir destas informações, a expressão do gene *PtaHB1*, homólogo a um HD-Zip de *A. thaliana*, e a expressão do miRNA *Pta-miR166*, foram descritas como sendo inversamente correlacionadas, associadas ao desenvolvimento de xilema secundário e reguladas ao longo do desenvolvimento da planta e das estações do ano em um híbrido de *Populus tremula* x *Populus Alba* (Ko et al., 2006). Também em *Populus*, o uso de miRNA artificial e mutações no sítio de ligação ao miR166 de ortólogos de HD-Zip classe III mostraram lignificação anormal (Du et al., 2011), anormalidades no crescimento primário e secundário, na formação de câmbio e reversão da polaridade do tecido vascular (floema interior ao xilema) (Robischon et al., 2011).

Uma abordagem alternativa para a identificação de novos genes MIR para espécies com o genoma conhecido é a estratégia *in silico* (Allmer e Yousef, 2012). Mesmo com extensiva cobertura de sequenciamento, alguns genes MIR não são identificados devido a baixos níveis de expressão ou padrões particulares de expressão em tecidos ou estágio de desenvolvimento. Assim, metodologias experimentais não são

capazes de identificar todo o repertório de pequenos RNAs expressos em uma mesma amostra. A disponibilidade de sequências completas de genomas abre a possibilidade de se utilizar de uma varredura ampla buscando regiões com estrutura secundária compatíveis com precursores de miRNAs (Thakur et al., 2011; Xuan et al., 2011; Tempel e Tahi, 2012).

2.14 Redes biológicas

Os grafos ou redes são objetos matemáticos formados pelo conjunto de vértices ou nodos (componentes) e um conjunto de arestas ou interações que conectam cada dois nodos (Barabasi et al., 2004) (Figura 4). Uma rede biológica é, portanto, uma representação abstrata de um sistema biológico em forma de rede onde o conjunto de nodos é o conjunto de componentes biológicos (proteínas, genes, metabólitos, etc) e o conjunto de arestas é conjunto de interações de natureza biológica (interação física entre proteínas, interações metabólicas, interações de regulação transcricional, etc) que conectam cada dois componentes biológicos (Barabasi et al., 2004).

As redes podem ser não-direcionadas ou direcionadas. Redes não-direcionadas são aquelas nas quais as interações não têm uma direção definida e as direcionadas são aquelas nas quais as interações têm uma orientação bem definida (Figura 4). Como exemplo biológico de rede não-direcionada podemos citar as redes de interação física entre proteínas. Nessas redes, as interações entre as proteínas não têm uma orientação definida: se a proteína A se liga à Proteína B, então a proteína B também se liga à proteína A. As redes metabólicas e de regulação transcricional, por outro lado, são exemplos de redes direcionadas: nas redes metabólicas, cada interação representa a direção do fluxo material de um substrato para um produto em uma reação metabólica enquanto que nas

redes de regulação transcricional, cada interação representa a direção do fluxo de informação entre o fator de transcrição e seu gene alvo.

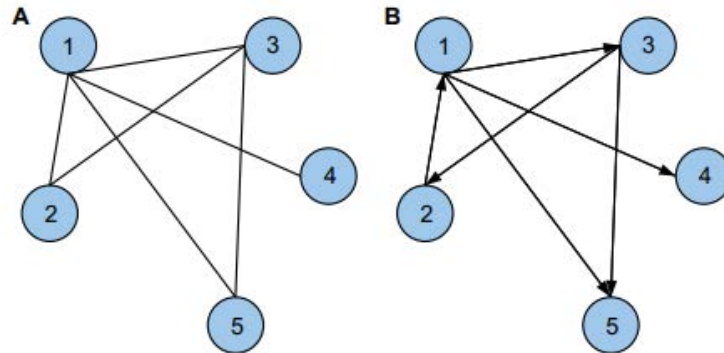


Figura 4: Exemplos de redes: a rede A é uma rede não-direcionada formado por um conjunto de cinco vértices ou nodos e um conjunto de seis arestas ou interações; a rede B é igual à rede A, mas suas interações são direcionadas.

A posição de um determinado nodo na rede em relação aos outros nodos pode ser determinada por funções chamadas de medidas de centralidade. As medidas de centralidade medem, como o próprio nome diz, a centralidade ou a importância de um nodo na rede. Em redes biológicas, as medidas de centralidade poderiam indicar, por exemplo, quais enzimas são as mais importantes em uma determinada rede metabólica ou quais fatores de transcrição são os reguladores globais em uma determinada rede de regulação transcricional. As medidas de centralidade mais utilizadas na análise de redes biológicas são:

- Grau de conectividade, $k(g)$: número de interações que determinado nodo g possui. Caso a rede seja direcionado, há um grau de entrada, $k_{in}(g)$, para o número de interações direcionadas para g , e um grau de saída, $k_{out}(g)$, para o número de interações que partem do nodo g .

- Grau de intermediação, *inbet* (*g*): relação entre o número total de caminhos geodésicos –sequencias de nodos adjacentes entre os nodos g_i e g_j contendo o menor número de nodos entre todas as sequencias existentes entre g_i e g_j – da rede e o número de caminhos geodésicos que passam por *g*, ou seja:

$$inbet(g) = \sum_{g_i \neq g \neq g_j} \frac{\sigma_{g_i g_j}(g)}{\sigma_{g_i g_j}}$$

sendo que $\sigma_{g_i g_j}$ é o número de caminhos geodésicos entre os nodos g_i e g_j e $\sigma_{g_i g_j}(g)$ é o número de caminhos geodésicos entre g_i e g_j que passam por *g* (Anthonisse, 1971; Freeman, 1977).

3 OBJETIVOS

-Identificar SNPs entre os genes diferencialmente expressos das famílias Bet v1, quitinase e taumatina;

-Identificar miRNAs com potencial ligação aos mRNAs dos genes diferencialmente expressos;

-Verificar a possível influência dos produtos dos genes diferencialmente expressos sobre o interatoma.

4 MATERIAL E MÉTODOS

4.1 Material Vegetal

Para validação dos SNPs, foram utilizadas, no total, folhas de 20 plântulas (10 normais e 10 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 mudas de *E. grandis* foram submetidas às avaliações morfológicas e, posteriormente, utilizadas para extração de DNA. Esta população utilizada no estudo, consiste em uma F₁ segregante para o caráter da anomalia.

4.2 Extração de DNA

Para a extração do DNA foi utilizado o protocolo descrito por Ferreira e Grattapaglia (1998) com adaptação na etapa de CTAB 10%, que foi substituída pela utilização de CTAB 5%. O CTAB 5% apresenta uma manipulação mais fácil, tendo menor viscosidade, sua utilização em trabalhos anteriores não mostrou alteração na qualidade do DNA final (Domingues, 2003).

Em um tubo foram macerados, aproximadamente, 150mg de folhas com nitrogênio líquido, em seguida, foram adicionados 700µl de solução de extração 29 com polivinilpinolidona (PVP) e 2-mercapto-etanol (2µl de mercapto-etanol para cada ml de solução de extração), aplicado em capela de exaustão e misturado até que todo o material macerado entrasse em contato com a solução. Esta mistura foi incubada por 45 a 60 minutos à temperatura de 65°C. Em seguida, foi realizada a primeira extração, adicionando-se 600µl de clorofórmio-álcool isoamílico (CIA 24:1). Esta solução foi

misturada com o auxílio de vortex e os tubos levados à centrífuga refrigerada a 10°C, numa rotação de 12000 rotações por minuto (rpm), durante 5 minutos.

Após a centrifugação, a solução apresentou três fases, a superior foi transferida para um novo tubo ($\pm 500\mu\text{l}$) e a esta solução foram adicionados 100 μl de CTAB 5%. Foi realizada então a segunda extração adicionando-se 600 μl de CIA 24:1. Essa solução foi homogeneizada em vortex e centrifugada nas mesmas condições anteriores. Após a centrifugação a solução apresentou duas fases separadas por uma película de proteínas. O sobrenadante ($\pm 500\mu\text{l}$) foi transferido para um novo tubo onde foi adicionado 400 μl de NaCl 5M e centrifugado nas mesmas condições anteriores. O sobrenadante ($\pm 500\mu\text{l}$) foi novamente transferido para um novo tubo onde foram adicionados 400 μl de álcool isopropílico gelado. O material foi mantido à temperatura de -20°C por um período de aproximadamente 10 horas. As amostras foram novamente centrifugadas nas mesmas condições anteriores.

Após a centrifugação pode-se visualizar um pellet no fundo do tubos, o sobrenadante foi descartado e o pellet lavado com 500 μl de etanol 95% por 2 minutos, o etanol foi descartado e o pellet após secagem em centrífuga a vácuo foi suspenso em tampão TE contendo 10ng/ μl de RNase (volume variou de 100 a 50 μl , dependendo do tamanho do “pellet”). As amostras foram incubadas a 3°C para a digestão do RNA por no mínimo 40 minutos ou até completa dissolução do “pellet”.

As amostras de DNA que após a digestão do seu RNA pela RNase, que ainda apresentaram resíduo de RNA, verificado após quantificação, foram tratadas com 1 μl de RNase a 10ng/ μl . As amostras, que não apresentaram completa dissolução do pellet após 2 horas incubadas a 37°C, foram deixadas over-night à temperatura ambiente e/ou foram levadas ao aquecedor a 65°C por 30 minutos.

A quantificação foi realizada em espectrofotômetro Nano Drop®- ND1000. Este equipamento, que além de avaliar o DNA com precisão, utiliza volumes pequenos (1µl) sem necessidade de diluição, e fornece parâmetros para avaliar sua pureza, quanto a proteínas e compostos fenólicos. O DNA utilizado nas reações de amplificação padrão foram diluídos à concentração de 15 ng/µl em água ultra pura autoclavada.

4.3 Seleção dos genes candidatos para análise *in silico*

A seleção dos genes candidatos foi obtida através dos resultados das análises de expressão gênica digital do trabalho de Fuchs (2014). Segundo esta mesma autora, este ensaio consistiu no mapeamento e contagem dos reads utilizando o genoma de *E. grandis* como referência (*Eucalyptus grandis* Genome Project 2010, <http://www.phytozome.net/eucalyptus>). Análises de expressão gênica diferencial foram realizadas comparando cada órgão (caule, folha e raiz) entre os fenótipos, sendo identificados genes diferencialmente expressos em todos eles ($P < 0,05$). O material genético utilizado foi um pool de amostras composto por uma mesma quantidade de RNA total de 5 indivíduos de um mesmo fenótipo (anormal ou normal).

Foram selecionadas para a identificação dos SNPs *in silico* as famílias gênicas mais representativas na categoria funcional “Resposta à estímulos”.

No total, 20 genes foram selecionados, 6 deles, representando a família Bet v1, 8 representando a família da quitinase classe I e 6 representando a família da taumatina. Na Tabela 1, estão listados os genes selecionados e seus acessos no banco de dados *Phytozome* v10.1, sua posição no banco de dados do genoma de eucalipto (scaffold), família da proteína a qual pertence, número de acesso no Pfam 27.0 (Punta et al., 2012), anotações funcionais e número de acesso no *Gene Ontology* (termos GO).

Tabela 1 – Anotação dos genes selecionados para a identificação *in silico* dos SNPs através dos dados de RNAseq, nome das proteínas e número de acesso das funções.

Locus name (Phytozome)	Posição no genoma (scaffolds)	PFAM	Gene Ontology
Eucgr.H04013 (Bet-1)	(scaffold_8: 58053532 – 58054388)	[00407] Proteína relacionada à patogenicidade Bet vI	[0009607] Resposta à estímulos [0006952] Resposta de defesa
Eucgr.H04012 (Bet-2)	(scaffold_8: 58046919 – 58047976)	[00407] Proteína relacionada à patogenicidade Bet vI	[0009607] Resposta à estímulos [0006952] Resposta de defesa
Eucgr.H04010 (Bet-4)	(scaffold_8: 58041631 – 58042706)	[00407] Proteína relacionada à patogenicidade Bet vI	[0009607] Resposta à estímulos [0006952] Resposta de defesa
Eucgr.H04003 (Bet-5)	(scaffold_8: 57989730 – 57990787)	[00407] Proteína relacionada à patogenicidade Bet vI	[0009607] Resposta à estímulos [0006952] Resposta de defesa
Eucgr.H04017 (Bet-6)	scaffold_8: 58079240 – 58080145	[00407] Proteína relacionada à patogenicidade Bet vI	[0009607] Resposta à estímulos [0006952] Resposta de defesa
Eucgr.G02175 (Bet-7)	scaffold_7: 39401452 – 39402318	[00407] Proteína relacionada à patogenicidade Bet vI	[0009607] Resposta à estímulos [0006952] Resposta de defesa
Eucgr.I01495 (Chiti-3)	(scaffold_9: 25149898 – 25151718)	[00182] Quitinase classe I [00187] Proteínas de reconhecimento	[0008061] Ligação a quitina [0004568] Atividade quitinase
Eucgr.K02166 (Chiti-4)	(scaffold_11: 28826193 – 28833622)	[00182] Quitinase classe I [00187] Proteínas de reconhecimento	[0008061] Ligação a quitina [0004568] Atividade quitinase
Eucgr.L00937 (Chiti-6)	(scaffold_86: 17615 – 19041)	[00182] Quitinase classe I	[0004568] Atividade quitinase
Eucgr.H00329 (Chiti-9)	(scaffold_8: 4561159 – 4561965)	[00182] Quitinase classe I	[0004568] Atividade quitinase
Eucgr.J02518 (Chiti-10)	scaffold_10: 31114177 – 31115438	[00182] Quitinase classe I	[0006032] Processos catabólicos da quitina [0004568] Atividade quitinase
Eucgr.J02519 (Chiti-11)	scaffold_10: 31120635 – 31122278	[00182] Quitinase classe I	[0006032] Processos catabólicos da quitina [0004568] Atividade quitinase
Eucgr.L00941 (Chiti-12)	scaffold_86: 88370 – 89503	[00182] Quitinase classe I	[0006032] Processos catabólicos da quitina [0004568] Atividade quitinase
Eucgr.H00325 (Chiti-13)	scaffold_8: 4522265 – 4523444	[00182] Quitinase classe I [00187] Proteínas de reconhecimento	[0006032] Processos catabólicos da quitina [0004568] Atividade quitinase [0008061] Ligação a quitina [0016998] Processos metabólicos de macromoléculas da parede celular
Eucgr.H03865 (Thau-2)	(scaffold_8: 56401990 – 56402899)	[00314] Família da taumatina	N/D
Eucgr.E00560 (Thau-3)	scaffold_5: 5359381 – 5360262	[00314] Família da taumatina	N/D
Eucgr.E01382 (Thau-4)	scaffold_5: 15488251 – 15496310	[00314] Família da taumatina	N/D
Eucgr.E01389 (Thau-5)	scaffold_5: 15590354 – 15591612	[00314] Família da taumatina	N/D
Eucgr.E01384 (Thau-6)	scaffold_5: 15549626 – 15550908	[00314] Família da taumatina	N/D
Eucgr.E00560 (Thau-7)	scaffold_5: 5359381 – 5360262	[00314] Família da taumatina	N/D

4.4 Análise *in silico* e identificação dos SNPs

Os dados obtidos por Fuchs (2014), que consistiram no sequenciamento das bibliotecas de cDNA através do equipamento MiSeq Personal Sequencing System (Illumina), gerando 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, serviram de base para a identificação dos SNPs *in silico*. Desse total, 94,75% dos reads apresentaram alta qualidade (Q score > Q30), sendo que a probabilidade de ocorrer um erro no sequenciamento de uma base é 1 em 1000 (Illumina, 2011). A partir destes dados, a mesma autora mapeou as sequências (reads) de cada biblioteca, utilizando o programa Bowtie (Langmead et al., 2009), tendo como referência o genoma de *Eucalyptus grandis* (Phytozome v10.1). Foram gerados arquivos no formato SAM e BAM. O formato SAM (Sequence Alignment/Mapping), é um formato texto criado para representar o resultado do alinhamento dos reads gerados pelo RNAseq contra um genoma de referência (Li et al., 2009). Já o arquivo BAM tem exatamente a mesma estrutura do arquivo SAM, porém as informações estão codificadas em formato binário para economizar espaço em disco e na memória. Além disso, o formato BAM possui uma compactação em blocos chamada BGZF, que permite o acesso aleatório ao arquivo mesmo ele estando compactado.

Para a identificação dos SNPs, foi empregada a ferramenta Integrative Genome Viewer (IGV) versão 2.3.34 (Robinson et al., 2011). IGV, é um programa de visualização dos alinhamentos gerados, que permite modificações em seu sistema de filtros para auxiliar na tomada de decisão sobre falsos positivos e/ou variantes mais prováveis para a posição variável identificada. Os parâmetros utilizados para esta análise foram os padrões do próprio programa. Os reads foram filtrados de acordo com uma qualidade de

mapeamento na escala Phred (Ewin et al, 1998) no valor de no mínimo 30, ou seja, 99% de probabilidade de um mapeamento correto. SNPs localizados na extremidades dos reads (últimas três bases) não foram considerados na análise devido à possibilidade de erros de sequenciamento.

4.5 Seleção dos iniciadores específicos e condições de amplificação

Para a validação dos SNPs identificados através do método de RNAseq, pares de *primers* foram desenhados pelo programa PerlPrimer, versão 1.1.14 com base nas famílias gênicas mais representativas do trabalho elaborado por (Fuchs, 2014). Tendo como orientação o número de acesso de cada gene no banco de dados *Phytozome* v10.1 (Tabela 1), as sequencias codificadoras de cada gene foram localizadas e utilizadas como base para o desenho dos *primers*. Esses iniciadores foram confeccionados somente para amplificar fragmentos das regiões codificadoras.

A ressuspensão destes *primers* foi feita com tampão Tris-EDTA (T.E.) a uma concentração final de 100 μ M, sendo os mesmos diluídos a uma concentração final de 10 μ M para serem utilizados nas reações de PCR. A melhor temperatura de pareamento para amplificação de cada *primer* foi determinada via gradiente de temperatura utilizando-se um termociclador Eppendorf Mastercycler EP S (MJ Research, Inc., Watertown, MA, USA). As reações foram realizadas em cadeia da polimerase (PCR) em 32 ciclos de 1 minuto a 95 °C, 1 minuto a X °C e 1 minutos a 72 °C, onde X (temperatura de anelamento de cada *primer*). Cada reação conteve um volume final de 10 μ l, onde foram utilizados 0,3 μ M de cada *primer forward* e *reverse*, 15 ng de DNA genômico e PCR Master Mix, da Promega Corporation®, o qual contém 50 unidades/ml de Taq DNA polimerase,

400 μ M dATP, 400 μ M dGTP, 400 μ M dCTP, 400 μ M dTTP, 3mM MgCl₂, estocados em tampão para reação com pH 8.5.

A não amplificação ou amplificação de bandas fracas indicaram a necessidade de reduzir a temperatura de anelamento do *primer*. Por outro lado, quando houve a presença de várias bandas, indicou-se a amplificação de produtos inespecíficos, tornando-se necessário o aumento da temperatura de anelamento para aumentar a especificidade da reação. Todos estes procedimentos, de aumento ou diminuição da temperatura, foram utilizados de acordo com as necessidades de cada *primer*.

Os fragmentos resultantes foram visualizados em gel de agarose a 1%, corado com brometo de etídeo e comparado com um padrão de peso molecular conhecido (Ladder – Fermentas®) de 100 pb e 1Kb para confirmação da especificidade.

4.6 Purificação dos produtos de PCR

Os fragmentos de PCR de cada indivíduo foram purificados com tratamento da enzima ExoSAP IT (GE Healthcare). Este reagente de purificação, é formado por duas enzimas hidrolíticas – Exonuclease I e a Fosfatase Alcalina de Camarão (SAP). A enzima exonuclease I tem como função remover os resíduos de *primers* e alguns DNAs de fita simples “estranhos” produzidos durante a reação de PCR com terminal 3'. Já a função da enzima SAP, é a de remover os dNTPs não incorporados na reação de PCR. Foi adicionado 1 μ l da enzima ExoSAP IT à μ l do produto de PCR e, posteriormente, incubar a solução a 37°C durante 15 minutos (ocorrer a degradação do excesso de *primers* e nucleotídeos) e a 80 °C por 15 minutos (inativação da enzima).

4.7 Reação de sequenciamento

As sequências de DNA foram determinadas em sequenciador automático ABI 3100 (Applied Biosystems®) utilizando-se 4 µl de 2,5x Save Money (400mM Tris-HCl pH 9,0, 10mM MgCl₂), 4µl BigDye™ Terminator Cycle Sequencing Ready Reaction Kit versão 3,1 (Applied Biosystems, Foster City, CA), 3,2 pmol dos oligonucleotídeos forward e reverse, 4 pL do produto de PCR purificado e H₂O q.s.p.. As reações foram realizadas em termociclador Eppendorf MasterCycler Gradient (Eppendorf, Hamburg, Germany) com os ciclos de temperatura programados para: 25 ciclos de 95°C por 10 seg, 50°C por 5 seg, 60°C por 4 seg, com rampa de 1°C/seg, como recomendado pelo fabricante. Após a amplificação as amostras serão mantidas a 4°C até a precipitação. Para cada amostra serão realizadas duas reações, sendo uma para o *primer forward* e outra para o *reverse*.

4.8 Precipitação das reações de sequenciamento

Nas reações de sequenciamento foram adicionados 80µl de isopropanol 65%, incubando a temperatura ambiente (T.A.) por 20 min. Em seguida, as amostras foram centrifugadas a velocidade de 10.000xg por 25 min a T.A. O isopropanol foi removido invertendo os tubos e, em seguida, foram adicionados 200µl de etanol 60% e centrifugado a velocidade de 10.000xg por 5 min a T.A. Todo o etanol foi removido com o auxílio de uma micropipeta. As amostras foram secas em T.A. e o produto ressuspenso em 2µl de tampão de amostra contendo Formamida Hi-Di™ (Applied Biosystems) + Loading Buffer (25mM EDTA pH 8,0 contendo 50 mg/ml de Blue Dextran) (5:1). No momento

da aplicação em sequenciador automático ABI PRISM® 3100 (Applied Biosystems, USA) as amostras foram aquecidas a 95 °C por 5 min e rapidamente transferidas para o gelo.

4.9 Alinhamento das Sequências para a procura dos SNPs

Os eletroferogramas das sequências amplificadas pelos *primers forward* e *reverse*, gerados automaticamente, foram editados e inspecionados visualmente através do programa ChromasPro v1.7.6 (Technelysium Pty Ltd). Os consensos formados pelas sequências foram verificados através da ferramenta BLAST (Altschul et al., 1997) no banco de dados *Phytozome* v10.1, afim de se obter a confirmação de que os fragmentos amplificados eram referentes aos genes selecionados para este trabalho. A identificação visual dos SNPs se torna possível através do alinhamento das sequências e subsequente comparação entre elas.

4.10 Classificação dos SNPs

As sequências dos genes foram usadas para anotações e traduções através da ferramenta de predição de genes FgenesH (<http://www.softberry.com/>), tomando como referência genes da planta modelo *Arabidopsis thaliana*. As previsões de FgenesH foram consideradas apropriadas para confirmar se as regiões estudadas eram somente éxons, ou se possuíam, também, íntrons e regiões UTRs (Untranslated Region). Após esta etapa, os SNPs foram classificados visualmente por mutações de transição e transversão. Por fim, utilizou-se a ferramenta de tradução Expasy (<http://web.expasy.org/translate>) (Gasteiger et al., 2003) afim de se obter a posição de cada mutação por códon para inspeção visual da classificação das mutações em sinônimas ou não-sinônimas.

Obtenção das sequências de pequenos RNAs

Para obtenção de miRNAs de eucalipto para serem utilizados como referência na identificação de possíveis miRNAs envolvidos na baixa expressão de 9 genes da planta anômala, utilizou-se o arquivo SRA (Sequence Read Archive) com os reads de todos os pequenos RNAs extraídos de eucalipto obtidos em GEO NCBI (Gene Expression Omnibus) com o código de acesso SRR1373439. Estes dados foram desenvolvidos por Levy et al., 2014, onde o procedimento utilizado para a obtenção do material genético, consistiu na germinação das sementes e, após 14 dias, extração de RNA, das raízes, caules e folhas. O critério utilizado para a escolha destes dados para serem utilizados como ferramenta para a identificação de miRNAs deveu-se ao fato de ser o único trabalho publicado, até então, com dados brutos de pequenos RNAs de eucalipto.

4.11 Processamento e identificação dos miRNAs

Um protocolo computacional foi desenvolvido especificamente para o processamento desses dados de sequenciamento em larga escala. A etapa de pré-processamento realizada pelo conjunto de ferramentas Fastx (Gordon, 2009) foi utilizada para filtrar as sequências, além de executar a remoção das sequências dos adaptadores e checagem de contaminantes e a análise da qualidade (Figura 5) foi realizada pelo conjunto de ferramentas FastQC (Simon, 2010).

As sequências que passaram por esta etapa foram, então, agrupadas de acordo com o número de bases, quantificadas e usadas para criar um grupo de sequências não redundantes. Como esperado, a distribuição, no gráfico, das sequências apresenta um único pico em 20 bases (Figura 6).

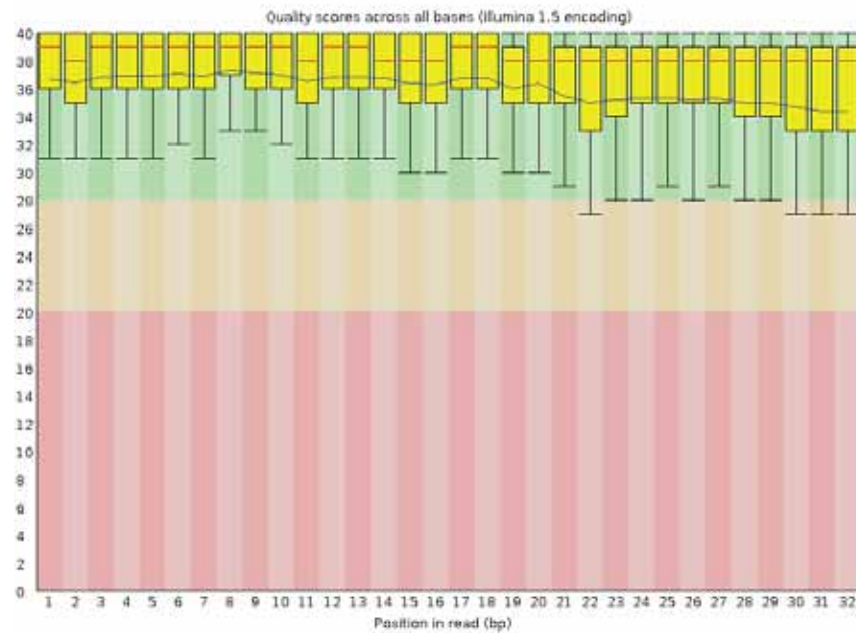


Figura 5: Demonstração do Phred Score, apresentando a ótima qualidade do sequenciamento obtido por Levy et al., 2014. O valor ideal de qualidade é 33, e todos os reads apresentaram resultados acima deste valor, confirmando a boa qualidade das sequências.

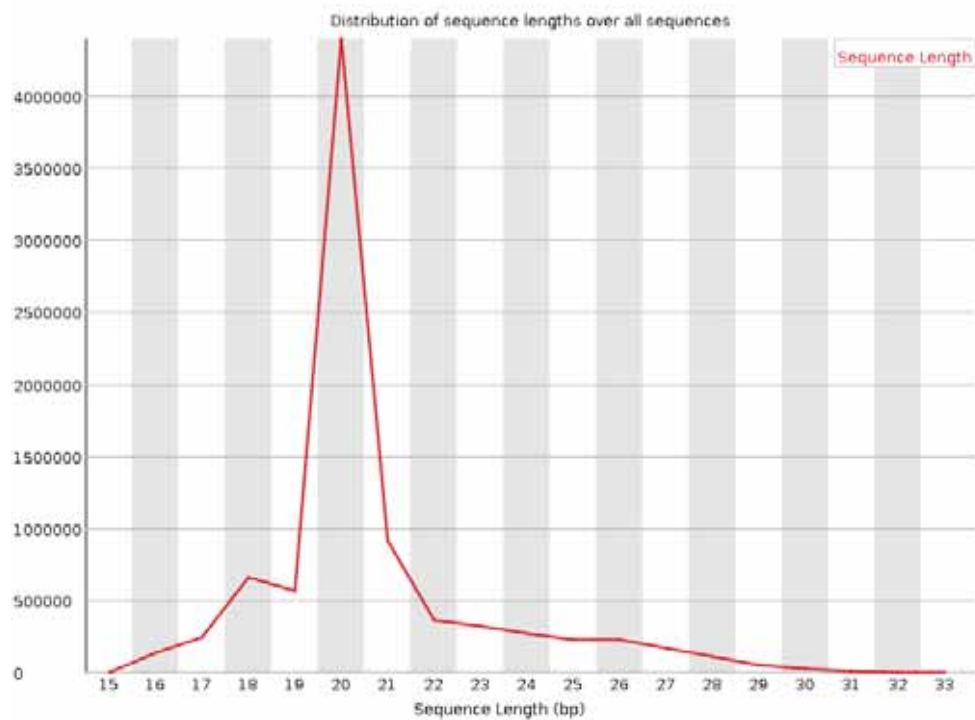


Figura 6: Distribuição do tamanho das sequências sobre todas as sequências processadas de pequenos RNAs.

O programa MIRPROF (Moxon et al., 2008) foi utilizado para mapear estas sequências contra o banco de dados de miRNA, Mirbase versão xxx (Griffiths-Jones et al., 2008), afim de se identificar sequencias de miRNAs conservadas e a busca por similaridades.

A busca por mRNAs alvo dos miRNAs foi realizada pelo programa psRNATarget (Dai e Zhao, 2011) que utilizou o banco de transcritos de *E. grandis* (phytozome 10.2), para identificação das interações com os genes alvo.

4.12 Construção do interatoma da Arabidopsis

Foram coletadas as interações proteína-proteína com escore de confiança igual ou maior a 0,7 do banco de dados STRING v10.0 (Search Tool for the Retrieval of Interacting Genes/Proteins). Este valor indica interações de alta confiança, segundo o String. Essas interações foram reunidas em uma rede que chamamos de rede Ara700.

O STRING (<http://string-db.org/>) é um banco de dados que tem como finalidade consultar informações sobre proteínas. É uma ferramenta para realizar buscas com o intuito de se obter redes de interação de proteínas (Jensen et al., 2009). O STRING disponibiliza um serviço que fornece informações sobre associações funcionais entre proteínas em um único lugar, facilitando a consulta e navegação nos dados em interfaces interativas e intuitivas (Szklarczyk et al, 2011). Um dos métodos usados para trabalhar com o banco de dados STRING é a Biologia de Sistemas. Esta é uma área interdisciplinar com o objetivo de compreender como funcionam as interações entre populações de moléculas, células e organismos devido ao aumento da complexidade de processos biológicos (Karsenti, 2012). As interações obtidas, neste banco de dados, podem ser experimentalmente verificadas ou interações computacionalmente preditas.

4.13 Construção do subgrafo dos diferencialmente expressos (Gdif)

Os dados com os resultados dos genes diferencialmente expressos, das amostras normais e anômalas de *E. grandis*, utilizados para esta etapa do trabalho, foram obtidos por Fuchs (2014).

As sequências proteicas dos genes diferencialmente expressos foram extraídas no *Phytozome* v10.2.

Utilizando a ferramenta blastp, com parâmetros padrões do próprio banco de dados *Phytozome*, foram encontrados os homólogos dessas sequências em *A. thaliana*. A partir daqui chamaremos essas proteínas homólogas de Eu2Ara. As Eu2Ara foram mapeadas na rede Ara700.

O subgrafo dos diferencialmente expressos foi finalmente construído através da extração a partir da rede Ara700 do conjunto de nodos contendo as Eu2Ara e todas as outras proteínas localizadas nos caminhos geodésicos entre as Eu2Ara. Para isso foi utilizado o pacote NetworkX (HAGBERG; SCHULT; SWART, 2008) para Python.

4.14 Análise da estrutura da Gdif

Foram analisados duas medidas de centralidade da Gdif: grau de conectividade e grau de intermediação. Em adição a isto, determinaram-se as distribuições de conectividades (como está distribuído o numero de conexões nas proteínas na rede).

Os *hubs* e *bottlenecks* foram determinados. Segundo Zotenko et al (2008), os *hubs* são aquelas proteínas entre as 20% mais conectadas em uma rede. Os *bottlenecks*, por sua vez, foram consideradas aquelas com grau de intermediação ≥ 0.01 .

4.15 Análise de enriquecimento funcional da Gdif e dos *hubs* e *bottlenecks*

Anotações de "*Gene Ontology*" (GO) foram determinados para os genes diferencialmente expressos de acordo com os dados das anotação funcionais obtidas a partir do *Phytozome* v10.2.

Análise de enriquecimento funcional baseado em GO foi realizada para estes genes diferencialmente expressos. Para cada tipo de termos GO, ou seja, processo biológico, função molecular e componente celular, foram comparadas as frequências de termos GO mapeados para os genes diferencialmente expressos com os de todo o conjunto de genes de eucalipto.

Para este fim, foi utilizado o serviço web Genetrail (BACKES et al., 2007) com os parâmetros padrões, sendo esta, uma ferramenta para determinar qual termos GO são significativamente representados em um conjunto de genes. O teste estatístico (teste hipergeométrico), os vários ajustamentos de teste (correção de Benjamini e Hochberg) e o nível de confiança ($p < 0,05$) foram os valores padrão. O conjunto de referência, a ontologia e os arquivos de anotação do organismo foram todos preparados de uma forma personalizada. Para cada tipo de ontologia, o conjunto de teste consistiu no conjunto de transcritos diferencialmente expressos e o conjunto de referência consistia em todo o conjunto de genes de eucalipto.

As anotações funcionais (termos GO - *Gene Ontology*) dos genes diferencialmente expressos da planta em estudo foram obtidas utilizando o banco de dados *Phytozome* v10.2. Utilizou-se somente a categoria 'Processos biológicos', 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 Genetrail (BACKES et al., 2007) 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* v10.2.

5 RESULTADOS E DISCUSSÃO

5.1 Identificação dos SNPs entre os genes diferencialmente expressos das famílias Bet v1, quitinase e taumatina

5.1.2 Identificação *in silico* dos SNPs através dos dados obtidos por RNAseq e suas propriedades

Utilizando os dados do sequenciamento, gerados por Fuchs (2014), das bibliotecas de cDNA geradas pelo equipamento MiSeq Personal Sequencing System (Illumina), foram analisadas as variações genéticas geradas pela população F₁ segregante para uma anomalia de desenvolvimento. Esta técnica de sequenciamento em larga escala, que consiste na identificação de SNPs em populações de plantas para avaliação de genes candidatos, já foram relatadas em arroz (Kharabian-Masouleh et al., 2011), *A. thaliana* (Schneeberger e Weigel, 2011), pinheiro (Gonzalez-Martinez et al., 2007), cana-de-açúcar (Bundock et al., 2009) e eucalipto em trabalhos comparando a síntese de metabólitos secundários (Kulheim, et al., 2009), genes relacionados ao crescimento da planta (Hendre et al., 2012), procura de marcadores relacionados à defesa contra patógenos (Thumma, et al., 2012a) e genes relacionados ao déficit hídrico (Thumma, et al., 2012b).

Através do programa IGV (Integrative Genome Viewer) (Figura 7), alinharam-se as sequências geradas pelos dados de RNAseq contra o genoma de referência de *E. grandis* Phytozome v10.1, seguindo as condições e parâmetros de qualidade que o programa determina. Com isto, diminuiu-se a probabilidade do aparecimento de falsos SNPs, nos 20 genes analisados. As variações alélicas foram identificadas por comparação entre pools das plantas normais e anômalas (Figura 7).



Figura 7 - Identificação de uma mutação na posição 4.561.181 do scaffold 8 (Chitinase classe I). Neste exemplo, a base mutada esta marcada com (G), sendo (A), a normal.

No total, foram avaliados “single-end reads” de 50 pb, compreendendo 13.849 pb de sequência do transcriptoma localizados em uma região genômica de 35.803 pb.

As análises mostraram que, nos genes, a frequência total de SNPs, variaram significativamente, sendo a média de 1 SNP para cada 142,7 pb. Esses dados, podem ser comparados aos relatados em espécies altamente heterozigotas, como em pinheiro (1 SNP por 102,6 pb, Dantec et al., 2004), videira (1 cada 64 pb, Lijavetzky et al., 2007), milho (1 a cada 60 pb, Ching et al., 2002) e centeio (1 a cada 52 pb, Li et al., 2011). Em comparação com outras espécies de plantas, como o arroz (1 SNP por 588 pb, Feltus et al, 2004; 1 SNP por 232 pb, Kharabian-Masouleh et al, 2011), sorgo (1 SNP por 357 bp, Feltus et al, 2004), soja (1 por 273 bp, Zhu et al., 2003), soja selvagem (1 por 425 bp, Hyten et al., 2010) e trigo (1 por 335 bp, Ravel et al., 2006), a frequência dessas mutações pontuais, quando levado em consideração o genoma de *E. grandis*, foi muito alta. Quando a comparação foi com o mesmo gênero, no caso a espécie *E.camaldulensis*, a frequência

foi de 1 SNP por 83,9 pb (Hendre et al., 2012), mas Kulheim et al. (2009) registraram em seus estudos, com esta mesma espécie, uma alta frequência considerada como uma das maiores entre as espécies vegetais com a ocorrência de 1 SNP para cada 16 bases. A diferença dos dados de Kulheim et al. (2009) com este trabalho, pode ser atribuída aos genes utilizados para a identificação dos SNPs. Aqui, foi levada em consideração somente as regiões codificadoras dos genes, que, segundo a literatura, possuem uma menor probabilidade na ocorrência de mutações. Assim sendo, esta explicação pode ser confirmada por diversos autores, que constataram que o maior número de variações ocorrem em regiões não codificadoras (introns e regiões UTR), onde existe pouca ou nenhuma pressão seletiva (Ching et al., 2002; Fladung e Buschbom, 2009). Desta forma, a seleção de genes candidatos, regiões genômicas, definição de um SNP, densidade do pool de DNA e a biologia do sistema (polinização cruzada têm maior frequência de variação do que a autopolinização) podem influenciar nessas estimativas (Lepoittevin et al., 2010, Grattapaglia et al., 2011). Outra razão importante para a alta diversidade de espécies florestais é a sua recente domesticação, que é em grande parte influenciada pela erosão genética vivida por vegetais domesticados (Doebley et al., 2006). Assim, espécies florestais em geral e *Eucalyptus ssp.*, em específico, pode ser considerado como um gênero com uma enorme riqueza alélica existente, mas inexplorada e não utilizada. Tal frequência alta de SNPs também faz com que as análises para identificação dessas mutações de ponto sejam complexas, como observado por Grattapaglia et al. (2011) na investigação de várias espécies de *Eucalyptus*.

5.1.3 Classificação dos SNPs

Foram identificados, no total, 97 SNPs. Somente o gene Bet-7 (Eucgr.G02175) não apresentou qualquer tipo de polimorfismo e, divergindo desse resultado, foram identificadas 12 mutações pontuais no gene Bet-5 (Eucgr.H04003), uma média de 1 SNP

para cada 40 pb (Tabela 2). Segundo Fladung e Buschbom (2009), a frequência de SNPs é heterogênea ao longo do genoma, diferindo entre região codificadora e região não-codificadora, por exemplo, ou entre sequências gênicas ou mesmo dentro de genes.

Uma mutação ocorrida no gene Bet-1 (Eucgr.H04013) foi considerada trialélica e em todos os outros genes investigados neste trabalho, bialélica. Assim, na prática, SNPs são considerados marcadores bialélicos, podendo ocorrer trialélicos em uma proporção menor (Gupta et al., 2001; Oraguzie et al., 2007).

Tabela 2 – SNPs obtidos a partir dos dados de RNAseq. A comparação dos polimorfismos foi feita entre uma população F₁ segregante para uma anomalia de desenvolvimento. Para a posição de cada mutação, foi usado como referência a sequência codificadora de cada gene depositada no banco de dados *Phytozome* v10.1.

SNP	Gene	Posição	Códon 1	Códon 2 (mutação)	Tipo de Mutação
01	BET-1	286	TG(T) - CISTEINA	TG(C) - CISTEINA	SINÔNIMA
02	BET-1	367	G(T)G - VALINA	G(A)G - AC. GLUTÂMICO	NÃO - SINÔNIMA
03	BET-1	369	(G)AT - AC. ASPÁRTICO	(A)AT - ASPARAGINA	NÃO - SINÔNIMA
04	BET-1	605	GT(G) - VALINA	GT(A) - VALINA	SINÔNIMA
05	BET-1	522	(T)GG - TRIPTOFANO	(A)GG - ARGININA	NÃO - SINÔNIMA
06	BET-1	520	G(A)G - AC. GLUTÂMICO	G(T)G - VALINA	NÃO - SINÔNIMA
07	BET-1	401	AA(G) - LISINA	AA(A) - LISINA	SINÔNIMA
08	BET-1	320	GC(G) - ALANINA	GC(A) - ALANINA	SINÔNIMA
09	BET-1	315	(G)TC - VALINA	(A)TC - METIONINA	NÃO - SINÔNIMA
10	BET-1	310	A(A)A - LISINA	A(G)A - ARGININA	NÃO - SINÔNIMA
11	BET-1	95	CC(C) - PROLINA	CC(G) - PROLINA	SINÔNIMA
				CC(A) - PROLINA	
12	BET-1	387	(G)AC - AC. GLUTÂMICO	(A)AG - LISINA	NÃO - SINÔNIMA
13	BET-1	383	CT(C) - LEUCINA	CT(G) - LEUCINA	SINÔNIMA
14	BET-2	320	GC(C) - ALANINA	GC(G) - ALANINA	SINÔNIMA
15	BET-2	110	GG(A) - GLICINA	GG(G) - GLICINA	SINÔNIMA
16	BET-2	215	AT(C) - ISOLEUCINA	AT(T) - ISOLEUCINA	SINÔNIMA
17	BET-2	83	CA(C) - HISTIDINA	CA(T) - HISTIDINA	SINÔNIMA
18	BET-2	287	AT(C) - ISOLEUCINA	AT(T) - ISOLEUCINA	SINÔNIMA
19	BET-2	47	CC(A) - PROLINA	CC(G) - PROLINA	SINÔNIMA
20	BET-4	37	G(T)C - VALINA	G(C)C - ALANINA	NÃO - SINÔNIMA
21	BET-5	95	CC(G) - PROLINA	CC(A) - PROLINA	SINÔNIMA
22	BET-5	108	CC(G) - PROLINA	CC(A) - PROLINA	SINÔNIMA
23	BET-5	146	GG(C) - GLICINA	GG(A) - GLICINA	SINÔNIMA
24	BET-5	119	AA(A) - LISINA	AA(G) - LISINA	SINÔNIMA
25	BET-5	128	GA(G) - AC. GLUTÂMICO	GA(A) - AC. GLUTÂMICO	SINÔNIMA
26	BET-5	278	TT(T) - FENILALANINA	TT(C) - FENILALANINA	SINÔNIMA
27	BET-5	299	AC(G) - TREONINA	AC(A) - TREONINA	SINÔNIMA
28	BET-5	363	(A)TT - ISOLEUCINA	(G)TT - VALINA	NÃO - SINÔNIMA
29	BET-5	383	CT(G) - LEUCINA	CT(A) - LEUCINA	SINÔNIMA
30	BET-5	386	AA(G) - LISINA	AA(A) - LISINA	SINÔNIMA
31	BET-5	407	GG(A) - GLICINA	GG(T) - GLICINA	SINÔNIMA
32	BET-5	435	(G)CT - ALANINA	(A)CT - TREONINA	NÃO - SINÔNIMA
33	CHITI-3	548	AT(C) - ISOLEUCINA	AT(A) - ISOLEUCINA	SINÔNIMA
34	CHITI-3	509	AT(T) - ISOLEUCINA	AT(G) - METIONINA	NÃO - SINÔNIMA
35	CHITI-3	626	GT(C) - VALINA	GT(G) - VALINA	SINÔNIMA
36	CHITI-3	188	AA(C) - ASPARAGINA	AA(A) - LISINA	NÃO - SINÔNIMA
37	CHITI-3	185	GG(T) - GLICINA	GG(C) - GLICINA	SINÔNIMA
38	CHITI-3	107	GG(A) - GLICINA	GG(C) - GLICINA	SINÔNIMA
39	CHITI-3	182	AC(C) - TREONINA	AC(G) - TREONINA	SINÔNIMA
40	CHITI-4	186	(G)CC - ALANINA	(C)CC - PROLINA	NÃO - SINÔNIMA
41	CHITI-4	189	(A)CC - TREONINA	(C)CC - PROLINA	NÃO - SINÔNIMA

Tabela 2 – SNPs obtidos a partir dos dados de RNAseq. A comparação dos polimorfismos foi feita entre uma população F₁ segregante para uma anomalia de desenvolvimento. Para a posição de cada mutação, foi usado como referência a sequência codificadora de cada gene depositada no banco de dados *Phytozome* v10.1. (Continua)

SNP	Gene	Posição	Códon 1	Códon 2 (mutação)	Tipo de Mutação
42	CHITI-4	194	CC(A) - PROLINA	CC(C) - PROLINA	SINÔNIMA
43	CHITI-4	386	GC(T) - ALANINA	GC(G) - ALANINA	SINÔNIMA
44	CHITI-4	755	GG(G) - GLICINA	GG(A) - GLICINA	SINÔNIMA
45	CHITI-6	18	(A)TA - ISOLEUCINA	(G)TA - VALINA	NÃO – SINÔNIMA
46	CHITI-6	209	AC(C) - TREONINA	AC(G) - TREONINA	SINÔNIMA
47	CHITI-6	185	GC(G) - ALANINA	GC(C) - ALANINA	SINÔNIMA
48	CHITI-6	203	TT(C) - FENILALANINA	TT(T) - FENILALANINA	SINÔNIMA
49	CHITI-6	32	CT(C) - LEUCINA	CT(T) - LEUCINA	SINÔNIMA
50	CHITI-6	497	AC(G) - TREONINA	AC(A) - TREONINA	SINÔNIMA
51	CHITI-6	503	GC(A) - ALANINA	GC(G) - ALANINA	SINÔNIMA
52	CHITI-6	754	G(T)G - VALINA	G(C)G - ALANINA	NÃO – SINÔNIMA
53	CHITI-9	43	C(C)A - PROLINA	C(G)A - ARGININA	NÃO – SINÔNIMA
54	CHITI-9	268	G(G)A - GLICINA	G(C)A - ALANINA	NÃO – SINÔNIMA
55	CHITI-9	270	(C)AC - HISTIDINA	(G)AC - AC. ASPÁRTICO	NÃO – SINÔNIMA
56	CHITI-9	20	AC(C) - TREONINA	AC(T) - TREONINA	SINÔNIMA
57	CHITI-9	74	AG(T) - SERINA	AG(G) - ARGININA	NÃO – SINÔNIMA
58	CHITI-9	431	TT(C) - FENILALANINA	TT(T) - FENILALANINA	SINÔNIMA
59	CHITI-9	598	G(C)T - ALANINA	G(G)T - GLICINA	NÃO – SINÔNIMA
60	THAU-2	175	G(T)G - VALINA	G(G)G - GLICINA	NÃO – SINÔNIMA
61	THAU-2	296	GG(C) - GLICINA	GG(T) - GLICINA	SINÔNIMA
62	THAU-2	182	CC(C) - PROLINA	CC(T) - PROLINA	SINÔNIMA
63	THAU-2	257	AC(T) - TREONINA	AC(C) - TREONINA	SINÔNIMA
64	BET-6	326	GA(T) - AC. ASPÁRTICO	GA(C) - AC. ASPÁRTICO	SINÔNIMA
65	BET-6	320	GC(G) - ALANINA	GC(A) - ALANINA	SINÔNIMA
66	BET-6	233	GA(C) - AC. ASPÁRTICO	GA(T) - AC. GLUTÂMICO	SINÔNIMA
67	BET-6	212	AA(A) - LISINA	AA(G) - LISINA	SINÔNIMA
68	CHITI-10	776	TG(C) - CISTEINA	TG(T) - CISTEINA	SINÔNIMA
69	CHITI-10	684	(A)AG - LISINA	(G)AG - AC. GLUTÂMICO	NÃO – SINÔNIMA
70	CHITI-11	727	T(A)C - TIROSINA	T(T)C - FENILALANINA	NÃO – SINÔNIMA
71	CHITI-11	419	CC(T) - PROLINA	CC(C) - PROLINA	SINÔNIMA
72	CHITI-11	402	(T)AC - TIROSINA	(C)AC - HISTIDINA	NÃO – SINÔNIMA
73	CHITI-11	333	(C)GA - ARGININA	(T)GA – STOP CÓDON	NÃO – SINÔNIMA
74	CHITI-12	351	(G)TC - VALINA	(A)TC - ISOLEUCINA	NÃO – SINÔNIMA
75	CHITI-12	368	AC(A) - TREONINA	AC(G) - TREONINA	SINÔNIMA
76	CHITI-12	610	C(T)G - LEUCINA	C(C)G - PROLINA	NÃO – SINÔNIMA
77	CHITI-13	146	GG(A) - GLICINA	GG(T) - GLICINA	SINÔNIMA
78	CHITI-13	525	(T)AC - TIROSINA	(C)AC - HISTIDINA	NÃO – SINÔNIMA
79	CHITI-13	636	(A)CC - TREONINA	(G)CC - ALANINA	NÃO – SINÔNIMA
80	CHITI-13	641	GT(G) - VALINA	GT(A) - VALINA	SINÔNIMA
81	CHITI-13	729	(A)TC - ISOLEUCINA	(G)TC - VALINA	NÃO – SINÔNIMA
82	THAU-3	182	CA(C) - HISTIDINA	CA(A) - GLUTAMINA	NÃO – SINÔNIMA
83	THAU-3	207	(C)CC - PROLINA	(T)CC - SERINA	NÃO – SINÔNIMA
84	THAU-3	210	(G)CT - ALANINA	(A)CT - TREONINA	NÃO – SINÔNIMA
85	THAU-3	309	(C)AG - GLUTAMINA	(G)AG - AC. GLUTÂMICO	NÃO – SINÔNIMA
86	THAU-3	341	AC(G) - TREONINA	AC(T) - TREONINA	SINÔNIMA
87	THAU-4	716	GG(T) - GLICINA	GG(C) - GLICINA	SINÔNIMA
88	THAU-5	395	GA(C) - AC. ASPÁRTICO	GA(T) - AC. ASPÁRTICO	SINÔNIMA
89	THAU-5	743	TG(T) - CISTEINA	TG(C) - CISTEINA	SINÔNIMA
90	THAU-6	421	C(T)G - LEUCINA	C(C)G - PROLINA	NÃO – SINÔNIMA
91	THAU-6	449	AG(C) - SERINA	AG(T) - SERINA	SINÔNIMA
92	THAU-6	650	AA(G) - LISINA	AA(C) - ASPARAGINA	NÃO – SINÔNIMA
93	THAU-7	182	A(C)A - TREONINA	A(A)A - LISINA	NÃO – SINÔNIMA
94	THAU-7	207	TC(C) - SERINA	TC(T) - SERINA	SINÔNIMA
95	THAU-7	210	CC(G) - PROLINA	CC(A) - PROLINA	SINÔNIMA
96	THAU-7	309	TC(C) - SERINA	TC(G) - SERINA	SINÔNIMA
97	THAU-7	341	C(G)C - ARGININA	C(T)C - LEUCINA	NÃO – SINÔNIMA

Os SNPs foram avaliados de acordo com a variação de base nucleotídica em transição (Tr) ou transversão (Tv). Dos 97 SNPs analisados, 60 foram mutações por transição (61,85%) e 37 por transversão (38,15%) (Tabela 3). A proporção (Tr/ Tv) encontrada aqui foi de 1,62, semelhante à relatada para o arroz (1,89, Feltus et al., 2004), pinheiro (1,40, Dantec et al., 2004), videira (1,46, Lijavetzky et al., 2007) e batata (1,5, Simko et al., 2006), mas superior ao de soja (0,92, Zhu et al., 2003) e gengibre (0,9, Chandrasekar et al., 2009). Já em *E. camaldulensis*, a média de Tr/Tv foi de 2,73 (Hendre et al., 2012). A hiper-mutabilidade das ilhas CpG é considerada como a principal razão para maior proporção de transições do que transversões presentes em éxons em planta, bem como genomas animais (Duncan e Miller, 1980; Jiang et al., 2008). Além disso, observou-se também que as transições, além de ocorrerem em maiores proporções, resultaram em mais mutações sinônimas, como o esperado (Dagan et al., 2002) (Tabela 2). Ainda assim, alguns genes tiveram frequências mais altas de mutações por transversão em relação à transição, como por exemplo, 3 genes da família quitinase Chiti-3 (Eucgr.I01495), Chiti-4 (Eucgr.K02166), Chiti-9 (Eucgr.H00329) com proporções de 0,16 ,0,25 e 0,40 respectivamente e 2 genes da família da taumatina Thau-3 (Eucgr.E00560) Thau-7 (Eucgr.E00560) com proporções de 0,66 e 0,66 respectivamente. Embora as variações de transição ocorram numa frequência mais alta que as transversões em genomas eucariotos (Li e Graur, 1991), possivelmente como resultado de mecanismos moleculares pelos quais são geradas, existem evidências que isso não é universal (Keller et al., 2007). Moriyama e Powell (1996) detectaram 54% de transversões em *Drosophila*, valor bem parecido com o encontrado em soja (52%) por Nasu et al. (2002).

Os SNPs também foram classificados considerando a sua posição no códon e o seu efeito sobre o aminoácido codificado, em mutações sinônimas ou mutações não-sinônimas (Tabela 2). As mudanças não-sinônimas podem determinar o fenótipo

alternativo e mostrará 100% de associação com a característica e dessa forma será de grande importância para a seleção assistida por marcadores (MAS) ou para o isolamento do gene. Conforme Chen et al. (2007), do total de SNP encontrados em um genoma, 1 a 2% estão localizados em regiões codificadoras de modo que os SNPs não-sinônimos podem ser considerados como candidatos para mudanças de funções gênicas. Ao contrário, os SNPs sinônimos são aqueles que têm efeito pequeno na expressão gênica, entretanto, tanto os SNP sinônimos quanto os não sinônimos são de grande importância como marcadores genéticos.

Tabela 3 – Diferentes tipos de classificação atribuídas aos SNPs identificados nos genes de *E. grandis*. São mostrados na tabela abaixo as mutações por transição (Tr) e transversão (Tv). Frequência dos SNPs e tipos de mutação, sinônimas e não-sinônimas e suas respectivas proporções.

Gene	Tr + Tv	Total	Tr/Tv	Freq. SNP	SNP sinônimo (S)	SNP Não-sinônimo (N)	S/N
BET-1	8 + 5	13	1,60	58,46	6	7	0,85
BET-2	5 + 1	6	5,00	80	6	0	-
BET-4	1 + 0	1	-	480	0	1	0
BET-5	10 + 2	12	5,00	40	10	2	5,0
CHITI-3	1 + 6	7	0,16	136,28	5	2	2,5
CHITI-4	1 + 4	5	0,25	167,4	3	2	1,5
CHITI-6	6 + 2	8	3,00	100,5	6	2	3
CHITI-9	2 + 5	7	0,40	97,3	2	5	0,4
THAU-2	3 + 1	4	3,00	170	3	1	3
BET-6	2 + 2	4	1,00	120	3	1	3
BET-7	-	-	-	-	-	-	-
CHITI-10	2 + 0	2	-	400,5	1	1	1
CHITI-11	3 + 1	4	3,00	201	1	3	0,33
CHITI-12	3 + 0	3	-	214	1	2	0,5
CHITI-13	4 + 1	5	4,00	169,2	2	3	0,66
THAU-3	2 + 3	5	0,66	141,6	1	4	0,25
THAU-4	1 + 0	1	-	750	1	0	-
THAU-5	2 + 0	2	-	375	2	0	-
THAU-6	2 + 1	3	2,00	250	1	2	0,5
THAU-7	2 + 3	5	0,66	141,6	3	2	1,5
TOTAL	60 + 37	97	1,62	142,7	58	39	1,48

Das 97 substituições de nucleotídeos, 20 ocorreram na primeira posição do códon (20,61%), 14 na segunda posição (14,45%) e 63 na terceira posição (64,94%). É sabido que o pareamento do anticódon com a terceira base no códon é menos rígido se comparado as duas primeiras. Uma vez que código genético degenerativo envolve, na maioria dos casos, somente a terceira base do códon, o pareamento oscilante (*wobble base pairing*) explica as modificações sinônimas encontradas em vários organismos. Em adição a estes dados, é possível relacionar a alta incidência de modificações na primeira e segunda base do códon com a elevada frequência de substituições não sinônimas (Alberts et al., 2010), como as encontradas neste trabalho.

Foram detectadas 58 (59,79%) substituições sinônimas e 39 (40,21%) substituições não sinônimas (Figura 2). A relação de mutações sinônimas/não-sinônimas foi de 1,48 (variando entre 0 e 5,00) (Tabela 3). Se comparado a outros estudos, resultados semelhantes foram encontrados em arroz (0,11-2,40, Kharabian-Masouleh et al., 2011), pinheiro do Canada (1.5, Pavy et al., 2006), *Arabidopsis thaliana* (2.0, Schmid et al., 2003), soja (1,28, Zhu et al., 2003), videira (1.0, Lijavetzky et al., 2007), mas resultados maiores do que os encontrados no trigo (0,50, Ravel et al., 2006). Já no gênero *Eucalyptus*, ocorreram resultados divergentes. Neste trabalho foram encontrados intervalos semelhantes ao de Hendre et al. (2012) na utilização de *E. camaldulensis*, (0 – 3,00) e divergindo dos resultados de Kulheim et al. (2009) que obtiveram como média (0,04-0,95). Novaes et al. (2008) utilizando a mesma espécie, *E. grandis*, através do sequenciamento de RNA encontraram uma média de mutações sinônimas/mutações não-sinônimas de 0,20, o que foi muito menor do que o relatado. No entanto, a estimativa variou significativamente de gene para gene.

Uma mutação ocorrida no gene Chiti-11 (Eucgr.J02519) foi classificada como STOP códon prematuro. Uma interrupção prematura da tradução, poderá ter, como

consequência, uma proteína resultante encurtada. Isso muitas vezes causa uma perda de função da proteína, uma vez que partes críticas da cadeia de aminoácidos não serão mais criadas. Devido a esta terminologia, códons de parada são também referidos como códons sem sentido (Gatfield et al., 2003).

5.1.4 Desenho e amplificação dos iniciadores específicos

Com a finalidade de se validar os SNPs e obter uma marca molecular, 10 genes foram selecionados para o desenho dos iniciadores específicos (tabela 4). O critério adotado para a seleção destes genes foi baseado na identificação *in silico* dos SNPs através dos dados gerados por RNAseq. As condições utilizadas para a confecção dos *primers* foram: (1) os *primers* somente seriam desenhados em regiões para amplificar genes que, através do alinhamento das sequências dos indivíduos contrastantes, possuíssem 100% de suas mutações de ponto em indivíduos anômalos, (2) e, somando-se a primeira condição, os indivíduos normais teriam o mesmo alelo do genoma utilizado como referência (*Eucalyptus grandis* Genome Project 2010, <http://www.phytozome.net/eucalyptus>).

Com a obtenção dos *primers*, além do objetivo de se encontrar algum marcador molecular que diferencie indivíduos normais de anômalos, também analisou-se a precisão da tecnologia de RNAseq na descoberta de SNPs.

A validação, em muitos casos, se faz necessária, principalmente quando se tratam de genes que possuem inúmeras cópias no genoma. Um falso SNP pode modificar observações feitas em contigs, e assim, modificar a classificação de um aminoácido como, por exemplo, a substituição por um STOP códon. Essa possibilidade pode ser consideravelmente reduzida com o desenho de iniciadores extremamente específicos para

uma perfeita condição de anelamento quando utilizada a técnica de PCR (Cánovas et al., 2010).

Desta forma, 10 pares de *primers* foram desenhados para amplificarem produtos de PCR dos genes alvo. Quatro pares de iniciadores foram desenvolvidos para genes da família Bet v1, cinco para a família quitinase classe I e um par de *primer* para a família da Taumatina (Tabela 4). O banco de dados do genoma de *Eucalyptus Grandis* (*Phytozome* v10.2) foi utilizado como base na confecção dos oligonucleotídeos. Para aumentar a precisão da especificidade no anelamento, cada iniciador forward e reverse, depois de desenhado, foi comparado (BLAST) com o genoma total de eucalipto com a finalidade de se obter, como retorno, apenas a cópia de interesse.

Os *primers* foram construídos para amplificar regiões que cobrem, no total 63 SNPs, segundo dados gerados por RNAseq.

Para a obtenção dos fragmentos de interesse de cada gene candidato, foram feitas reações de amplificação por PCR, utilizando o DNA total de folhas da mesma população F₁ segregante para a anomalia. Realizadas as reações de amplificação, os produtos gerados foram submetidos à eletroforese em gel de agarose, onde todos, apresentaram amplificação de banda única, variando a temperatura de anelamento entre 58 e 60 °C, e os fragmentos amplificados variando em tamanhos de 480 a 954 pb (Tabela 4).

Tabela 4 – *Primers* utilizados para validação dos SNPs identificados por RNAseq. São apresentados a identidade de cada iniciador; temperatura de anelamento, em graus Celsius (°C); tamanho do fragmento amplificado, em pares de base (pb); e as sequencias de cada *primer*.

Primer	Nome do locos (<i>Phytozome</i>)	Temperatura de anelamento	Amplicon	Forward / Reverse
BET-1	Eucgr.H04013	58 °C	760 pb	GATTGTTTCGTTTCTCCTTCAG/ AAAGCGTTGAGTTCATTGAG
BET-2	Eucgr.H04012	60 °C	480 pb	CCACAAGATACTCCTCAACAG/ CCAAACCCTTCCACTACC
BET-4	Eucgr.H04010	58 °C	480 pb	AGCATCGATCTTGTGTTTGAC/ GAGCATTGAGTTCATCGAGGG
BET-5	Eucgr.H04003	58 °C	480 pb	TTAGCAATCACGCCAAACC/ AAAGCTATCTGCATCGAGAG
CHITI-3	Eucgr.I01495	58 °C	954 pb	CTTGAACGAGACGACAGGG/ TGCTCTCCTAATCCGACCT
CHITI-4	Eucgr.K02166	60 °C	837 pb	TCCCAAATCAATACTGTCC/ GCCTTGAATAGAACCTCTTCC
CHITI-6	Eucgr.L00937	60 °C	804 pb	CAACACGAACCTCCAACCT/ CTGTGGTTTCATGCGATGTC
CHITI-9	Eucgr.H00329	60 °C	681 pb	CCATTATTCTTACCGTCGCC/ GCCCCGTAGTTGTAGTTCCA
CHITI-11	Eucgr.J02519	58 °C	396 pb	CTGTTCAACGAGATGCTG/ CATAGTATTTCTTGCCAGGAG
THAU-2	Eucgr.H03865	58 °C	678 pb	GTCAACCAGGGAGATGTC/ ACAGAGAAATCACACGCC

5.1.5 Sequenciamento pelo método de Sanger e validação dos SNPs

Pelo método de Sanger, a identificação de SNP através de genes candidatos, consiste em sequenciamento direto de indivíduos contrastantes para característica estudada, subsequente alinhamento utilizando ferramentas de bioinformática, para posterior comparação das sequências na procura de SNP. Esta metodologia também foi utilizada com sucesso em diversos tipos de estudos, como em humanos (Lai et al., 1998; Levy et al., 2009), bovinos (Calvo et al., 2006; Abatepaulo et al., 2008), plantas (Van et al., 2004; Novaes et al., 2008), além de outros organismos (Liao e Lee, 2010).

Para o sequenciamento, 20 indivíduos (10 normais e 10 anômalos) originados de um mesmo cruzamento F₁ segregante para uma anomalia, foram utilizados para identificação e validação dos SNPs.

As sequências resultantes apresentaram ótima qualidade, sendo posteriormente, alinhadas e editadas no programa ChromasPro versão 1.7.6. Após esse procedimento, realizou-se uma análise *in silico*, através da ferramenta BLAST (Altschul et al., 1997) no banco de dados do *Phytozome* v10.1, verificando-se que todos os fragmentos amplificados correspondem às enzimas em estudo.

A primeira constatação a ser notada, após o alinhamento das sequências, foi que grande parte das mutações ocorridas eram heterozigotas, corroborando com os dados de Savolainen e Pyhajarvi (2007) e Grattapaglia et al. (2011) que demonstraram que espécies de *Eucalyptus*, por possuírem pouca domesticação, encontram-se com alta incidência de alelos em heterozigose (Figura 8).

Na validação dos SNPs, pode-se verificar que das 63 mutações de ponto detectadas pela tecnologia de RNAseq, 7 (11,11%) não foram localizadas, utilizando a metodologia de Sanger. Os fragmentos amplificados pelo *primer* Bet-1 não acusaram as variações alélicas nas posições 286 (SNP1), 367 (SNP2) e 369 (SNP3). Já nos fragmentos amplificados pelo *primer* Bet-5, um único SNP não foi detectado na posição 108 (SNP22), os fragmentos amplificados pelo *primer* Chiti-11, identificado anteriormente, por possuir um Stop Códon na posição 333 (SNP73), não apresentaram tal polimorfismo e, por fim, 2 SNPs não foram detectados nos produtos de PCR do *primer* Chiti-3 nas posições 188 (SNP36) e 182 (SNP39).

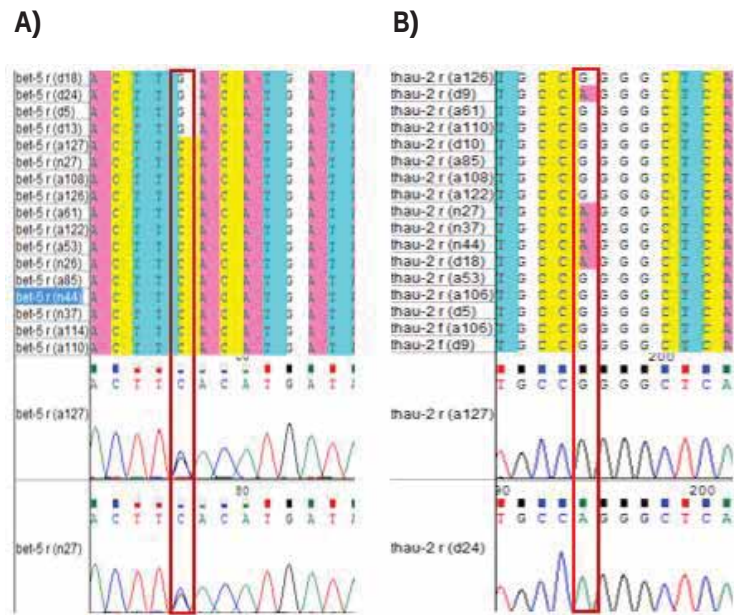


Figura 8 - Alinhamento realizado pelo programa ChromasPro versão 1.7.6, comparando genótipos de indivíduos normais e anômalos na procura de SNPs. A região destacada em vermelho mostra a localização dos SNPs nos indivíduos que possuem alelos em heterozigose (A), e indivíduos que possuem alelos diferentes em homozigose (B).

É importante notar que nas amostras utilizadas para RNAseq por Fuchs (2014), mesmo pertencentes aos mesmos parentais, não podemos afirmar que as progênes geradas pelo cruzamento eram as mesmas daquelas sequenciadas pelo método de Sanger, de modo que não se espera uma concordância de 100% dos resultados. No entanto, é notável que, apesar da diferença na composição da amostra na análise, apenas poucas mutações observadas em RNAseq, não foram detectadas em Sanger. Cánovas et al. (2010) fazendo esta mesma comparação entre tecnologias de sequenciamento, mas trabalhando com genes relacionados à fabricação de leite em bovinos, obtiveram resultados semelhantes, em que 11,6% de seus SNPs não foram detectados quando observadas nas análises de sequenciamento. Observando as regiões onde não foram identificadas as mutações através do sequenciamento de Sanger, verificou-se que outra hipótese pode ser levantada. Morozova e Marra (2008) através de ensaios, notaram que locais que possuem muitas regiões repetitivas, estão mais passíveis de erros de leitura. Outra hipótese, não menos provável, segundo Cirulli et al. (2010) apesar das novas tecnologias de

sequenciamento gerarem dados extremamente seguros, por se tratar de alinhamento de pequenos reads, 40 a 50 pb, em alguns casos, podem ocorrer erros na montagem do alinhamento, principalmente quando se tratam de regiões que possuem muitas cópias no genoma, que é o que ocorre neste trabalho.

Os resultados da genotipagem, utilizando as 20 amostras obtidas pelo cruzamento controlado, mostraram que não houve separação genotípica entre os indivíduos normais e anômalos quando amplificados pelos marcadores moleculares. Podemos levar em consideração, que Fuchs (2014) utilizou um número pequeno de material vegetal para análise transcricional (um pool com 5 indivíduos normais e 5 anômalos). Este número amostral reduzido pode interferir na fidedignidade dos resultados, ao se investigar uma marca molecular. Segundo Cirulli et al. (2010) outro fator limitante pode ser a cobertura, pois se tratando de transcriptomas, os genes são expressos em diferentes níveis numa amostragem contrastante, podendo haver um desequilíbrio. Este fator pode ter ocorrido na identificação dos SNPs por RNAseq, onde possíveis mutações que visualmente estariam ocorrendo somente nos indivíduos anômalos, poderiam estar também presentes, mas “escondidas” nos indivíduos normais, por não ter ocorrido a expressão em algumas amostras nestas regiões.

5.2 Identificação de miRNAs com potencial ligação aos mRNAs dos genes diferencialmente expressos

5.2.1 Processamento dos dados de sequenciamento para identificação dos prováveis miRNAs

Existe na literatura científica, relatos, demonstrando a atuação de miRNAs em plantas levando-as a uma ativação do sistema de defesa sem que ocorra ataque por patógenos, denominada autoimunidade. Li et al.(2012), identificaram miRNAs classificados como nta-miR6019 (22nt) e nta-miR6020 (21nt), os quais clivaram transcritos de receptores de imunidade NB-LRR do tabaco, responsáveis pela resistência ao vírus do mosaico do tabaco. Tais análises, apontaram que a expressão desregulada de genes NB-LRR poderiam estar desencadeando a autoimunidade, mesmo na ausência de infecção patogênica e, como consequência, a planta teve seu desenvolvimento inibido.

Tendo como possível hipótese o fenômeno ocorrido no trabalho de Li et al. (2012), e constatando a escassez de trabalhos relacionados a autoimunidade em plantas, principalmente quando o assunto é miRNA, a identificação e caracterização de miRNAs em *Eucalyptus* se tornou necessária, sendo realizada com o objetivo de dar suporte ao estudo na descoberta dos possíveis causadores da anomalia em desenvolvimento em *E. grandis*. O sequenciamento em larga escala de pequenos RNAs (smRNA-Seq), realizado por Levy et al. (2014), foi usado como ferramenta para a descoberta de miRNAs. A partir do grande número de dados gerados no sequenciamento, procedeu-se a identificação e caracterização dos miRNAs. Após o pré-processamento, as sequências únicas de pequenos RNAs foram usadas para uma avaliação de sequência de critérios antes de declarar uma sequência como miRNA. Esses critérios incluem o mapeamento no genoma

e a predição da estrutura secundária a partir da sequência genômica flanqueante. Após o pré-processamento, as sequências foram divididas em categorias de tamanho. Como esperado, a distribuição das sequências apresentaram um único pico demonstrando que a grande maioria das amostras tem em média 20 bases.

Após todo o processamento dos dados, tais como remoção dos adaptadores, filtragem das sequências, o número de sequências únicas de todas as amostras totalizaram 747.244. Desse total, 268 sequências foram identificadas como miRNAs resultantes da procura no banco de dados de miRNA, miRBase release 21 (<http://mirbase.org/>), distribuídas em 31 famílias.

Foram identificados, como esperado, vários miRNAs conservados como miR159, miR160, miR166, miR172 e miR396. Parece consenso na literatura que os miRNAs mais conservados estejam envolvidos em funções básicas como crescimento e desenvolvimento (Rajagopalan, Vaucheret et al., 2006; Cuperus, Fahlgren et al., 2011).

5.2.2 Predição de alvos para miRNA em *Eucalyptus*

A predição de alvos para os miRNA em *Eucalyptus*, utilizando o programa psRNATarget (<http://plantgrn.noble.org/psRNATarget/>) e o banco de dados de transcritos de *E. grandis* disponível no *Phytozome* (versão 10.2), resultaram em 6 transcritos preditos (Tabela 5). Dentre estes, estão transcritos sugeridos por Fuchs (2014) como possíveis causadores da anomalia em eucaliptos, tais como, os que originam proteínas com domínio Betv1 (Eucgr.H04017.1 e Eucgr.H04013.1) e proteínas similares à taumatina (Eucgr.E01384.1). Outros transcritos também foram identificados como possíveis alvos dos miRNAs, sendo eles a cupina (Eucgr.L03367.1), domínio RCD1-SRO-TAF4 (RST) (Eucgr.B00305.1) e a redutase férrica (Eucgr.I01144.1), todos, seguindo a nomenclatura gênica e suas respectivas palavras-chave, encontrada no banco *Phytozome* (Tabela 5).

Tabela 5: Transcritos preditos como alvos de miRNAs pelo programa psRNATarget. A seleção foi realizada entre os transcritos diferencialmente expressos relacionados a anomalia, desenvolvido por Fuchs (2014). São apresentados: Nome do transcrito segundo o banco *Phytozome*, código, nome e descrição da proteína resultante, segundo o banco Pfam.

Transcrito alvo	ID PFAM	Domínio Proteico	Descrição PFAM
Eucgr.H04017.1	PF00407	Proteínas com domínio Bet v1	Resposta a defesa Resposta a estímulos
Eucgr.H04013.1	PF00407	Proteínas com domínio Bet v1	Resposta a defesa Resposta a estímulos
Eucgr.E01384.1	PF00314	Proteínas similares a taumatina	N/D
Eucgr.B00305.1	PF12174	Domínio de planta RCD1-SRO-TAF4 (RST)	Resposta a estresses e hormônios
Eucgr.L03367.1	PF00190	Cupina	N/D
Eucgr.I01144.1	PF01794	Redutase férrica (componente da membrana)	Atividade da redutase férrica

Os transcritos preditos, foram alvos de 5 diferentes classes de miRNAs. Podemos destacar que o euc-miR156 e euc-miR157 tiveram como alvo o mesmo transcrito o Eucgr.H04017.1. No entanto, duas isoformas do miRna167, nomeados de euc-miR167_1 e euc-miR167_2 tiveram como alvos diferentes mRNAs, respectivamente o Eucgr.E01384.1 e o Eucgr.B00305.1. Já o euc-miR166 e euc-mirR390, apresentaram como alvos, respectivamente, os mRNAs Eucgr.L03367.1 e Eucgr.H04013.1 (Tabela 6).

Vale ressaltar que, um dos transcritos selecionados como alvo em eucalipto, o Eucgr.I01144.1, quando analisado pelo preditor de miRNAs, psRNATarget, não obteve como retorno nenhum mRNA alvo. Podemos sugerir que a maior parte das famílias de genes MIR não seja conservada, constituindo assim, como únicas dentro de uma espécie ou entre espécies relacionadas (Cuperus, Fahlgren et al., 2011; Jones-Rhoades, 2012; Axtell, 2013).

MiRNAs linhagem ou espécie específicos já foram descritos, por exemplo, em *Arabidopsis*, arroz, *Populus*, *Physcomitrella patens*, *Pinus taeda*, *Medicago truncatula*,

trigo e sorgo (Arazi, TalmorNeiman et al., 2005; Lu, Tej et al., 2005; Lu, Sun et al., 2005; Sunkar, Girke et al., 2005; Rajagopalan, Vaucheret et al., 2006; Fahlgren, Howell et al., 2007; Jin, Li et al., 2008; Morin, Aksay et al., 2008; Wang, Chen et al., 2011; Wei, Yan et al., 2011; Zhang, Zheng et al., 2011). Sugere-se que a taxa de surgimento e desaparecimento de genes de miRNAs não conservados seja alta, facilitada pelo tamanho e simples arquitetura dos genes MIR, e que estes, estejam relacionados à regulação de processos sob alta pressão seletiva como relacionados à defesa contra patógenos e isolamento reprodutivo (Rajagopalan, Vaucheret et al., 2006; Tang, 2010; Jones-Rhoades, 2012). Não é raro que novos miRNAs não tenham alvos identificados ou talvez reconheçam o alvo por um mecanismo ainda não elucidado, o que não permite a compreensão de sua função biológica. Considera-se que os miRNAs mais recentes tenham distribuição filogenética restrita e que sejam evolutivamente mais sujeitos a alterações (Jones-Rhoades, 2012).

Tabela 6: MiRNAs de eucalipto preditos e seus respectivos mRNAs alvos. Resultado dos dados gerados pelo programa psRNATarget. São mostrados: Nome do miRNA, mRNA alvo, E value (expectation value), Tamanho do miRNA, posição inicial e final do sítio de ligação do miRNA no mRNA e o modo de inibição predita (clivagem ou bloqueio da tradução).

miRNA	Transcrito Alvo	E value	Tamanho fragmento	Posição do sítio de ligação do miRNA	Inibição
miR156 (euc-miR156)	Eucgr.H04017.1	4.0	20	565 - 584	Clivagem
miR157 (euc-miR157)	Eucgr.H04017.1	4.0	20	565 - 584	Clivagem
miR166 (euc-miR166)	Eucgr.L03367.1	4.5	20	43 - 62	Clivagem
miR167_1 (euc-miR167_1)	Eucgr.E01384.1	3.5	20	898 - 917	Tradução
miR167_2 (euc-miR167_2)	Eucgr.B00305.1	4.0	21	1012 - 1032	Tradução
miR390 (euc-miR390)	Eucgr.H04013.1	4.5	20	9 - 28	Clivagem

Euc-miR156 e euc-miR157 de *Eucalyptus* pertencem à mesma família de miRNA. Em uma análise comparativa, através de alinhamento de sequências realizados pelo programa ClustalX v2.1 (Thompson et al., 1997), Euc-miR156 apresentou alta identidade com os ortólogos de *Arabidopsis thaliana* e *Populus trichocarpa* (Figura 10a). Referindo-se ao miR156, alguns processos biológicos já tiveram a regulação desta família confirmada experimentalmente. A regulação da transição de fase vegetativa, por exemplo, é regulada por miR156 e seus alvos da família SQUAMOSA Promoter-Binding Protein-Like (SPLs), anteriormente mostrada em *Arabidopsis* (Wu e Poethig, 2006), mostrou-se ser um mecanismo regulatório extensamente conservado, desde plantas herbáceas anuais a perenes, após ser identificado em híbridos de *Populus* (Wang, Park et al., 2011). Outro exemplo de ação destes miRNAs foi demonstrado por Fu et al. (2012) em gramíneas (*Panicum virgatum*), onde, um precursor, miR156b, foi superexpresso. Os efeitos em genes SPL foram revelados por análises de RT-PCR e notou-se alterações morfológicas relacionadas a produção de biomassa, eficiência da sacarificação e digestibilidade da forragem nessas plantas transgênicas. MiR156 controla a dominância apical e a transição floral em gramíneas por suprimir seus genes SPL alvo. Quando houveram níveis relativamente baixos na expressão de miR156, ocorreu um aumento na produção de biomassa, produzindo plantas com tempo normal de floração. Em contrapartida, quando os níveis na expressão de miR156 foram elevados, resultou em crescimento severamente atrofiado da planta, característica morfológica semelhante à encontrada na anomalia em *Eucalyptus grandis*, descrita por Tambarussi (2006) e analisada morfológicamente por Revolti (2015) (Figura 9a) sugerindo, desta forma, a possível atuação do euc-miR156 na referida planta em estudo, prejudicando seu desenvolvimento.

Outro miRNA predito neste estudo foi o euc-miR166 (miR166), o qual já foi bem relatado por pesquisadores, atuando, em *Medicago truncatula*, *Populus* e *Arabidopsis thaliana* (Boualem et al., 2008; Ko et al., 2005; Jung e Park, 2007). Basicamente, a função principal desses miRNAs nessas plantas é atuar em genes alvos que regulam diversos aspectos do desenvolvimento da planta, incluindo a formação de meristema apical e lateral, polaridade da folha e desenvolvimento vascular. Ko et al. (2005) demonstraram, através da clonagem do fator de transcrição “*leucine-zipper protein*” (HD-Zip classe III) e da família de miR166 utilizando a técnica de análise de expressão por *Northern blot*, que a expressão de HD-ZIP estava intimamente associada à formação de madeira, pois em estações climáticas onde o desenvolvimento é alto, o miR166 estava sendo pouco expresso, porém, em épocas de inverno, onde o desenvolvimento da madeira é baixo, miR166 estava sendo altamente expresso. Este fenômeno, levaram os pesquisadores a conclusão de que a expressão de miR166 é uma “ferramenta” utilizada pela planta para controlar seu desenvolvimento. A expressão deste miRNA nas plantas anômalas pode estar relacionada ao nanismo encontrado na mesma, como relatado por Tambarussi (2006), na Figura 9a, e por Bomblies et al. (2007), este último autor, estudando fenótipos de plantas similares à anomalia. Euc-miR166 de *Eucalyptus* apresentou alta identidade com os ortólogos de *Arabidopsis thaliana*, *Populus trichocarpa* e *Medicago truncatula* (Figura 10b).

Outra característica morfológica fundamental encontrada nas plantas anômalas de eucalipto é o volume da raiz pouco desenvolvida. Estudos de caracterização morfológica realizados por Revolti (2015), comprovam que estes indivíduos anômalos possuem suas ramificações laterais das raízes pouco desenvolvidas quando comparados com indivíduos normais (Figura 9b e 9c). Este fator pode estar relacionado à atuação do miRNA euc-miR167. Segundo a literatura, miR167 está intimamente ligado a tolerância da planta

contra a seca. Um trabalho realizado pelos pesquisadores Kinoshita et al. (2012), utilizando pequenos RNAs e a técnica de *microarray*, demonstraram que miR167a de *Arabidopsis thaliana* atua clivando o mRNA *IAR3*, conhecido pela sua função no desenvolvimento de raízes. A alta taxa de expressão de miR167 atua inibindo o crescimento de ramificações laterais das raízes, e como consequência, prejudicando a planta na absorção de água e nutrientes. Em contraste com o tipo selvagem, mutantes transgênicas de *IAR3* conseguiram inibir a clivagem de seus mRNAs, conduzindo assim, a uma planta com ramificações de raízes mais desenvolvidas e com melhor aproveitamento na absorção de água e nutrientes. Euc-miR167_1 e Euc-miR167_2 de *Eucalyptus* apresentaram alta identidade com o ortólogo de *Arabidopsis thaliana* (Figura 10c).



Figura 9: (A) Imagem retirada do trabalho de Tambarussi (2006) comparando o desenvolvimento do eucalipto normal (esquerda) e anômalo (direita). Nota-se pouco desenvolvimento do indivíduo anômalo. (B) e (C) Imagem cedida por Revolti (2015) comparando o desenvolvimento de raízes entre um eucalipto normal (B) e anormal (C). Em (C) ocorreu pouco desenvolvimento de ramificações laterais, quando comparado a (B).

O florescimento também pode ser influenciado por miRNAs. Trabalhos realizados com a planta *Physcomitrella patens* mostraram que a ação em conjunto de dois mirRNAs, miR156 e miR390, em um mesmo alvo de um transcrito primário “trans-acting small interfering RNAs” (tasiRNAs) regula o tempo de florescimento e influencia no tempo de

desenvolvimento da planta. Segundo Cho et al. (2012), a superexpressão de miR390 causa uma formação mais lenta do gametóforo, e consequentemente, levando a planta a um prejuízo em sua reprodução. Já o miR156 promove uma mudança no desenvolvimento dos protonemas filamentosos para gametóforos frondosos na planta *Physcomitrella patens*, inibindo assim, o desenvolvimento da floração, causando a esterilidade nesses vegetais. No caso das plantas anômalas, não podemos constatar se o florescimento seria prejudicado ou não, pois como relatado anteriormente por Tambarussi (2006), elas sobrevivem por somente 4 meses, não atingindo a fase de reprodução. Euc-miR390 de *Eucalyptus* apresentou alta identidade com os ortólogos de *Physcomitrella patens* (Figura 10d).

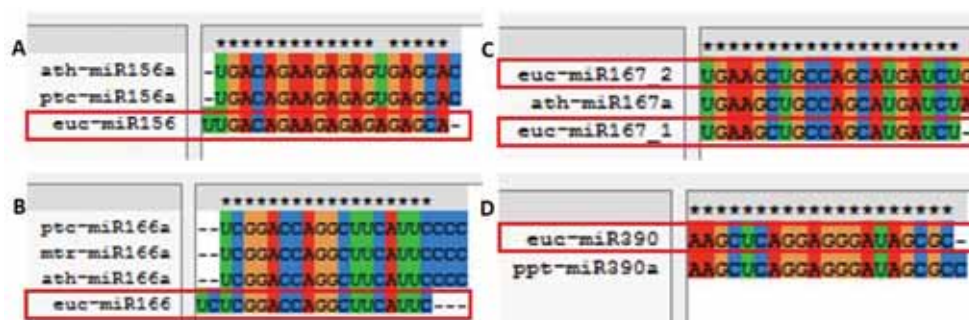


Figura 10: Alinhamento para comparação dos miRNAs e seus respectivos ortólogos. Destaque em vermelho para os miRNAs de *Eucalyptus*. A) Alinhamento do miRNA de eucalipto euc-miR156 com miR156a de *A. thaliana* (ath), *P. trichocarpa* (ptc). B) Alinhamento de Euc-mi166 com miR166a de *A. thaliana* (ath), *M. truncatula* (mtr) e *P. trichocarpa*. C) Alinhamento de Euc-miR167_1 e euc-miR167_2 com miR167a de *A. thaliana* (ath). D) Alinhamento de Euc-miR390 com euc-miR390a de *P. patens* (ppt).

5.3 Verificação da possível influência dos produtos dos genes diferencialmente expressos sobre o interatoma.

5.3.1 Considerações iniciais

Anteriormente, usando a análise de enriquecimento funcional de termos do GO, Fuchs (2014) constatou que os termos significativamente frequentes no grupo de genes diferencialmente expressos nas plantas anormais em relação a todo o genoma do *E. grandis* eram aqueles que, conjuntamente, sugerem que essas plantas estão deflagrando uma resposta de defesa mesmo na ausência de patógenos. Como essas plantas não sobrevivem até a idade reprodutiva, essa suposta resposta de defesa é tão intensa que torna a planta inviável ainda na fase de plântula.

Esse enriquecimento funcional realizado anteriormente, por esta mesma autora, foi feito somente com os genes diferencialmente expressos. Levando em consideração que os genes e proteínas não atuam isolados nas células, isto é, eles interagem entre si e formam uma rede complexa de interações moleculares, seria interessante saber como a expressão diferencial desses genes/proteínas estaria influenciando outros genes/proteínas nessa rede de interação, ou seja, seria interessante fazer essa análise de enriquecimento funcional levando em consideração outros genes/proteínas sob influência dos diferencialmente expressos. Trabalhos anteriores levando em consideração a análise de redes de interações entre proteínas, por exemplo, mostram que os efeitos da expressão diferencial podem ser críticos sobre genes com mesmo nível de expressão nos casos controle e alterado.

Nesta seção, mostramos justamente os resultados da análise de enriquecimento funcional levando em consideração a rede de interações entre proteínas (rede PPI) na qual as proteínas dos genes diferencialmente expressos estão inseridos. Na realidade, essa análise foi feita em *Arabidopsis thaliana* devido à escassez de dados de interações em *E.*

grandis. A *A. thaliana*, por outro lado, é a espécie do grupo das plantas com maior abundância de dados de interações. Por exemplo, até o momento, no banco de dados de interação molecular IntAct versão 4.1.7 (<http://www.ebi.ac.uk/intact>), enquanto a *A. thaliana* tem cerca de 19 mil interações experimentalmente verificadas, a próxima planta com maior número de interações é *Oryza sativa* com 676 interações. Portanto, os resultados encontrados na rede da *A. thaliana* devem ser extrapolados para o *E. grandis* como se a *A. thaliana* estivesse sob a mesma condição dos plantas anormais de *E. grandis*.

Para a realização dessa análise de enriquecimento funcional indireta, primeiro foi feita a determinação das proteínas em *A. thaliana* homologas às codificadas pelos genes diferencialmente expressos do *E. grandis* (Eu2Ara), depois construiu-se o interatoma da *A. thaliana* (Ara700) e mapearam-se as Eu2Ara nessa Ara700. Através da determinação dos caminhos geodésicos entre as Eu2Ara na Ara700, foi construído o subgrafo dos diferencialmente expressos (Gdif) contendo as proteínas mais influenciadas pela Eu2Ara e, finalmente, análises de enriquecimento funcionais foram realizados para a descoberta dos processos sob maior influência das Eu2Ara na Gdif. Nas seções seguintes encontram-se os resultados e a discussão detalhada de uma dessas partes.

5.3.2 Determinação das proteínas em *A. thaliana* homologas às codificadas pelos genes diferencialmente expressos do *E. grandis* (Eu2Ara)

Como não há dados de interação disponíveis para *E. grandis*, como já mencionada anteriormente, optou-se em encontrar na *A. thaliana*, planta que tem dados abundantes de interação, as proteínas homólogas às codificadas pelos genes diferencialmente expressos entre as plantas normais e anormais de *E. grandis* (Eu2Ara). Utilizando blastp e considerando como homóloga as proteínas em *A. thaliana* com os menores E-values para cada proteína de *E. grandis*, foram encontradas 48 proteínas homólogas às codificadas pelos genes diferencialmente expressos (Eu2Ara), como mostrado na Tabela 7.

Tabela 7: Proteínas homologas identificadas em *A. thaliana* (Eu2Ara) utilizando, como referência, os genes diferencialmente expressos obtidos no trabalho de Fuchs (2014).

Número de Acesso	Nome do gene/proteína	Função do gene/proteína
AT1G02335	GL22/ GERMIN-LIKE PROTEIN SUBFAMILY 2 MEMBER 2 PRECURSOR	Encodes a plasodesmata-located protein involved in regulating primary root growth
AT1G09560	GLP5/ GERMIN-LIKE PROTEIN 5	Encodes a plasodesmata-located protein involved in regulating primary root growth
AT1G14950	POLYKETIDE CYCLASE/DEHYDRASE AND LIPID TRANSPORT SUPERFAMILY PROTEIN	Involved in: response to biotic stimulus, defense response
AT1G17860	KUNITZ FAMILY TRYPSIN AND PROTEASE INHIBITOR PROTEIN	endopeptidase inhibitor activity
AT1G20030	PATHOGENESIS-RELATED THAUMATIN SUPERFAMILY PROTEIN	
AT1G21270	SERINE/THREONINE PROTEIN KINASE	Involved in: response osmotic stress
AT1G24020	MLP423/ MLP-LIKE PROTEIN 423	Involved in: response to biotic stimulus, defense response
AT1G43800	FTM1/ FLORAL TRANSITION AT THE MERISTEM1	oxidoreductase activity
AT1G59960	NAD(P)-LINKED OXIDOREDUCTASE SUPERFAMILY PROTEIN	oxidoreductase activity
AT1G64160	DIR5/ DIRIGENT PROTEIN 5	Encodes a dirigent protein involved in the synthesis of pinorexinol
AT1G65680	EXPB2/ EXPANSIN B2	member of beta-expansins
AT1G78860	CURCULIN-LIKE (MANNOSE-BINDING) LECTIN FAMILY PROTEIN	
AT2G15620	NIR1/ NITRITE REDUCTASE	Involved in the second step of nitrate assimilation
AT2G16060	HB1/ HEMOGLOBIN 1	Encodes a class 1 nonsymbiotic hemoglobin induced by low oxygen levels

Tabela 7: Proteínas homologas identificadas em *A. thaliana* (Eu2Ara) utilizando, como referência, os genes diferencialmente expressos obtidos no trabalho de Fuchs (2014).

(continua)

Número de Acesso	Nome do gene/proteína	Função do gene/proteína
AT2G40610	EXPA8/ EXPANSIN A8	member of Alpha-Expansin Gene Family
AT2G44480	BGLU17/ BETA GLUCOSIDASE 17	cation binding, hydrolase activity
AT2G45180	BIFUNCTIONAL INHIBITOR/LIPID- TRANSFER PROTEIN/SEED STORAGE 2S ALBUMIN SUPERFAMILY PROTEIN	lipid binding
AT3G01680	SEOR1/ SIEVE-ELEMENT- OCCLUSION-RELATED 1	Encodes a protein localized to phloem filaments
AT3G05950	RMLC-LIKE CUPINS SUPERFAMILY PROTEIN	manganese ion binding, nutrient reservoir activity
AT3G10920	MSD1/ MANGANESE SUPEROXIDE DISMUTASE	
AT3G12500	HCHIB/ BASIC CHITINASE	encodes a basic chitinase involved in ethylene/jasmonic acid mediated
AT3G18000	XPL1/ XIPOTL 1	Reduce transmission through pollen
AT3G20380	TRAF-LIKE FAMILY PROTEIN	
AT3G23430	PHO1/ PHOSPHATE 1	Encodes a protein with the mainly hydrophilic N-terminal and the C-terminal containing 6 potential membrane-spanning domains
AT3G25180	CYP82G1/ CYTOCHROME P450,20152015 FAMILY 82, SUBFAMILY G, POLYPEPTIDE 1	Encodes a cytochrome P450 monooxygenase (CYP82G1)
AT3G54420	EP3/ HOMOLOG OF CARROT EP3-3 CHITINASE	encodes an EP3 chitinase
AT4G01470	TIP1;3/ TONOPLAST INTRINSIC PROTEIN 1;3	Encodes AtTIP1;3, functions as water and urea channels in pollen
AT4G11650	OSM34/ OSMOTIN 34	
AT4G12520	BIFUNCTIONAL INHIBITOR/LIPID- TRANSFER PROTEIN/	lipid binding
AT4G25000	AMY1/ ALPHA-AMYLASE- LIKE	Predicted to be secreted protein based on signalP prediction

Tabela 7: Proteínas homologas identificadas em *A. thaliana* (Eu2Ara) utilizando, como referência, os genes diferencialmente expressos obtidos no trabalho de Fuchs (2014).

(continua)

Número de Acesso	Nome do gene/proteína	Função do gene/proteína
AT4G38970	FBA2/ FRUCTOSE-BISPHOSPHATE ALDOLASE 2	modulated in response to ABA in <i>Arabidopsis thaliana</i> seeds
AT5G14780	FDH/ FORMATE DEHYDROGENASE	Encodes a NAD-dependent formate dehydrogenase
AT5G17220	GSTF12/ GLUTATHIONE S-TRANSFERASE PHI 12	Encodes glutathione transferase belonging to the phi class of GSTs
AT5G24090	CHIA/ CHITINASE A	Chitinase A (class III) expressed exclusively under environmental stress conditions
AT5G25940	EARLY NODULIN-RELATED	molecular_function unknown
AT5G26340	MSS1	Encodes a protein with high affinity, hexose-specific/H ⁺ symporter activity
AT5G28010	POLYKETIDE CYCLASE/DEHYDRASE AND LIPID TRANSPORT SUPERFAMILY PROTEIN	Involved in: response to biotic stimulus, defense response
AT5G35525	PLAC8 FAMILY PROTEIN	
AT5G36110	CYP716A1/ "CYTOCHROME P450, FAMILY 716, SUBFAMILY A, POLYPEPTIDE 1"	
AT5G38410	RBCS3B/ RUBISCO SMALL SUBUNIT 3B	Encodes a member of the Rubisco small subunit (RBCS) multigene Family
AT5G39160	RMLC-LIKE CUPINS SUPERFAMILY PROTEIN	manganese ion binding, nutrient reservoir activity
AT5G40780	LHT1/LHT1/ LYSINE HISTIDINE TRANSPORTER	Encodes LHT1 (lysine histidine transporter), a high-affinity transporter
AT5G41040	RWP1/ REDUCED LEVELS OF WALL-BOUND PHENOLICS 1	Encodes a feruloyl-CoA transferase required for suberin synthesis
AT5G44640	BGLU13/ BETA GLUCOSIDASE 13	cation binding, hydrolase activity
AT5G48430	EUKARYOTIC ASPARTYL PROTEASE FAMILY PROTEIN	aspartic-type endopeptidase activity

Tabela 7: Proteínas homologas identificadas em *A. thaliana* (Eu2Ara) utilizando, como referência, os genes diferencialmente expressos obtidos no trabalho de Fuchs (2014).

(continua)

Número de Acesso	Nome do gene/proteína	Função do gene/proteína
AT5G62520	SRO5/ SIMILAR TO RCD ONE 5	encodes a protein with similarity to RCD1 but without the WWE domain
AT5G64120	2015PRX71/ PEROXIDASE 71	Encodes a cell wall bound peroxidase that is induced by hypo-osmolarity

5.3.3 Construção do interatoma da *Arabidopsis thaliana* (Ara700)

Para que a influência das Eu2Ara sobre outras proteínas possa ser analisada, é necessário descobrir com quais outras proteínas elas interagem e como é a estrutura de rede de interações entre as Eu2Ara e essas proteínas. Para isso, é necessário primeiramente construir o interatoma, isto é, a rede de interações entre proteínas do organismo de interesse, no caso, a *A. thaliana*. Como mencionado no início dessa seção, portanto, a primeira etapa para isso foi a construção da Ara700. A Ara700 é uma rede de interações com 17.749 proteínas e 420.032 interações. Essas interações incluem aquelas experimentalmente verificadas e computacionalmente inferidas.

A distribuição de conectividades (Figura 11) do Ara700 indica que sua estrutura é semelhante à estrutura de outras redes biológicas, ou seja, poucas proteínas interagem com muitas proteínas e muitas proteínas interagem com poucas proteínas. Considerando como *hubs*, ou seja, proteínas altamente conectadas, aquelas proteínas entre as 20% mais conectadas (Zotenko et al., 2008), então temos na Ara700 3.538 proteínas consideradas *hubs*. Essas proteínas interagem no mínimo com 68 e no máximo com 849 interações. Dentre as 48 Eu2Ara, três (6%) são *hubs*.

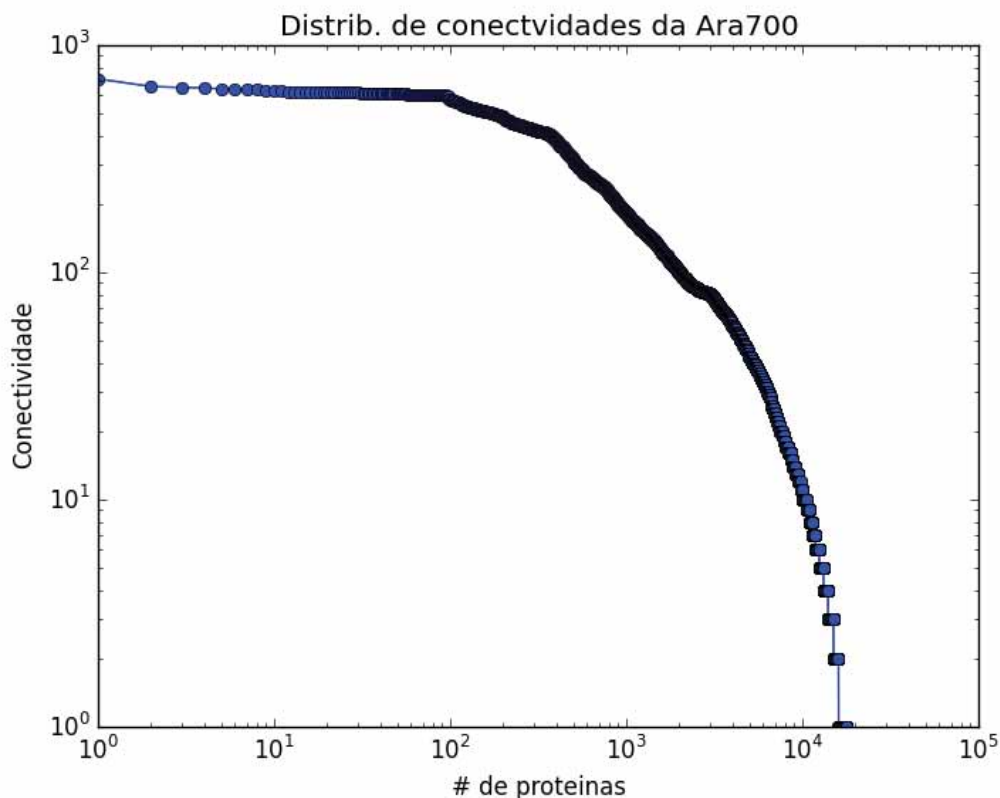


Figura 11: Distribuição de conectividades das proteínas da Ara700 indicando que sua rede é semelhante à estrutura de outras redes biológicas, ou seja, poucas proteínas interagem com muitas proteínas e muitas proteínas interagem com poucas proteínas. O eixo x representa o número de proteínas e o eixo y, o grau de conectividade, isto é, a quantidade de interações. Ambos os eixos estão em escala logarítmica.

5.3.4 Construção e análise da estrutura do subgrafo dos diferencialmente expressos (Gdif)

Com a Ara700 construída e como se busca nessa parte do trabalho descobrir a influência das Eu2Ara sobre outras proteínas, as 48 Eu2Aras encontradas foram mapeadas na Ara700 e, a partir de então, a Gdif foi construída como descrito na metodologia. As proteínas localizadas nos caminhos geodésicos entre as Eu2Aras e que formam a Gdif são as proteínas que, embora seus genes no *E. grandis* não se mostraram diferencialmente expressas entre plantas normais e anormais, podem sofrer algum tipo de influência pelas Eu2Aras, assim como, consequentemente, os processos biológicos aos quais elas participam.

A Gdif contém 3.811 proteínas, o que corresponde a 21,5% da quantidade de proteínas presentes na Ara700, e 35.542 interações, o que representa cerca de 9% da quantidade de interações da Ara700. A distribuição de conectividades do Gdif (Figura 12) indica que sua estrutura é semelhante à estrutura de outras redes biológicas, ou seja, poucas proteínas interagem com muitas proteínas e muitas proteínas interagem com poucas proteínas (Zotenko et al., 2008), assim como a Ara700.

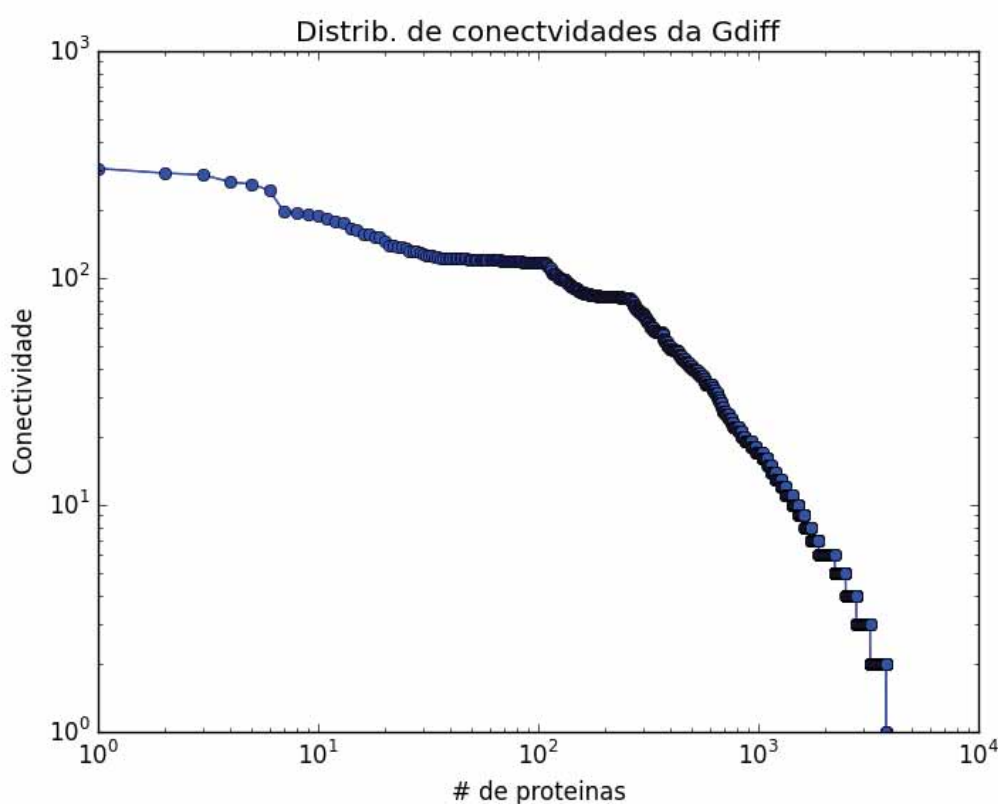


Figura 12: Distribuição de conectividades das proteínas do GDIF indicando que sua rede é semelhante à estrutura de outras redes biológicas, ou seja, poucas proteínas interagem com muitas proteínas e muitas proteínas interagem com poucas proteínas. O eixo x representa o número de proteínas e o eixo y, o grau de conectividade, isto é, a quantidade de interações. Ambos os eixos estão em escala logarítmica.

A Gdif tem 772 proteínas consideradas *hubs*. Essas proteínas interagem no mínimo com 23 e no máximo com 339 interações. Aproximadamente 18,75% (9) das 48 Eu2Ara, são *hubs* (Tabela 8). Esses nove *hubs* identificados possuem número de interações bem distintas, variando a partir do mais conectado, com 123 interações

(AT4G38970, FRUCTOSE-BISPHOSPHATE ALDOLASE) até o menos conectado (N-methyltransferase - AT3G18000) possuindo 33 interações. Os Eu2Ara *hubs* da Gdif participam de diversos processos biológicos relacionados a resposta a estímulos abióticos, bióticos e hormonais, lignificação da parede celular e estágios de desenvolvimento das plântulas.

Tabela 8: Proteínas mais conectadas na rede Gdif (*hubs*) com seu número de interações, nome do gene/proteína e processos biológicos do *Gene Ontology* associados.

Hubs	Interações	Proteína	Processos Biológicos
AT1G21270	82	SERINE/THREONINE PROTEIN KINASE	Defense response to bacterium Response to salt stress
AT2G15620	62	NIR1/ NITRITE REDUCTASE	seedling development stage
AT3G10920	41	MSD1/ MANGANESE SUPEROXIDE DISMUTASE 1	defense response to bacterium, response to salt stress
AT3G18000	33	N-METHYLTRANSFERASE	Phosphatidylcholine biosynthetic process pollen development unidimensional cell growth choline biosynthetic process post-embryonic root development pollen tube growth pollen tube guidance
AT4G25000	34	AMY1/ ALPHA-AMYLASE- LIKE	response to abscisic acid response to gibberellin
AT4G38970	123	FRUCTOSE-BISPHOSPHATE ALDOLASE	response to cadmium ion response to abscisic acid
AT5G14780	45	FDH/ FORMATE DEHYDROGENASE	response to cadmium ion response to wounding
AT5G17220	40	GSTF12/ GLUTATHIONE S- TRANSFERASE PHI 12	anthocyanin-containing compound metabolic process
AT5G64120	39	PRX71/ PEROXIDASE 71	involved in the lignification of cell walls

5.3.5 Análise de enriquecimento funcional

Para obter mais informações sobre os processos biológicos relacionados à patogênese da anomalia genética sob investigação, buscou-se pesquisar profundamente, através da análise de enriquecimento funcional, que termos específicos do GO relacionados a processos biológicos estão associados aos Eu2Ara e genes/proteínas que estejam sob sua influência (ou seja, a Gdif). Esta análise pode revelar quais vias específicas para a resposta podem estar associadas às características morfológicas da anomalia.

Usando o teste hipergeométrico com as correções de Benjamini e Hochberg através do Genetrail (Backes et al., 2007), foram comparadas as frequências de termos do GO relacionados a processos biológicos para os genes/proteínas participantes da GDIF (lembrando que os diferencialmente expressos também fazem parte da GDIF) com os de todo o conjunto de genes/proteínas de *A. thaliana*. Determinou-se quais termos do GO foram enriquecidos significativamente ($p < 0,05$ corrigido). Levando em consideração a planta anormal, destacaremos os processos biológicos alterados que possam ser mais importantes em relação a esta condição de interesse. Nas seções 5.11.1 e 5.11.2 fazemos uma breve descrição desses processos biológicos e na seção 5.11.3 discutiremos a ligação entre os termos e a morfofisiologia da planta anômala.

5.3.6 Análise de enriquecimento funcional da Gdif contra o proteoma de *A. thaliana*

Com base no enriquecimento funcional entre a GDIF e o proteoma de *A. thaliana*, foram obtidos, no total, 163 termos relacionados à categoria de processos biológicos enriquecidos significativamente. Esses termos foram subdivididos em grupos, de acordo

com sua categoria funcional. Destes, 34% participam de processos relacionados ao metabolismo energético (fotossíntese, glicólise e afins), 9% em processos ligados a replicação de DNA, 7% em processos catabólicos de proteínas, 12% em processos relacionados a RNAs não codificadores, 5% a processos relacionados à miRNAs, 28% com resposta à estímulos (resposta de defesa da planta) e 5% à sinalização celular. Vale ressaltar que, os termos do GO são únicos e cada termo é e corresponde exclusivamente a uma categoria funcional. Portanto, é correto afirmar que um gene/proteína pode estar associado a diversos termos do GO, pois um gene/proteína pode participar de múltiplos processos (Figura 13).

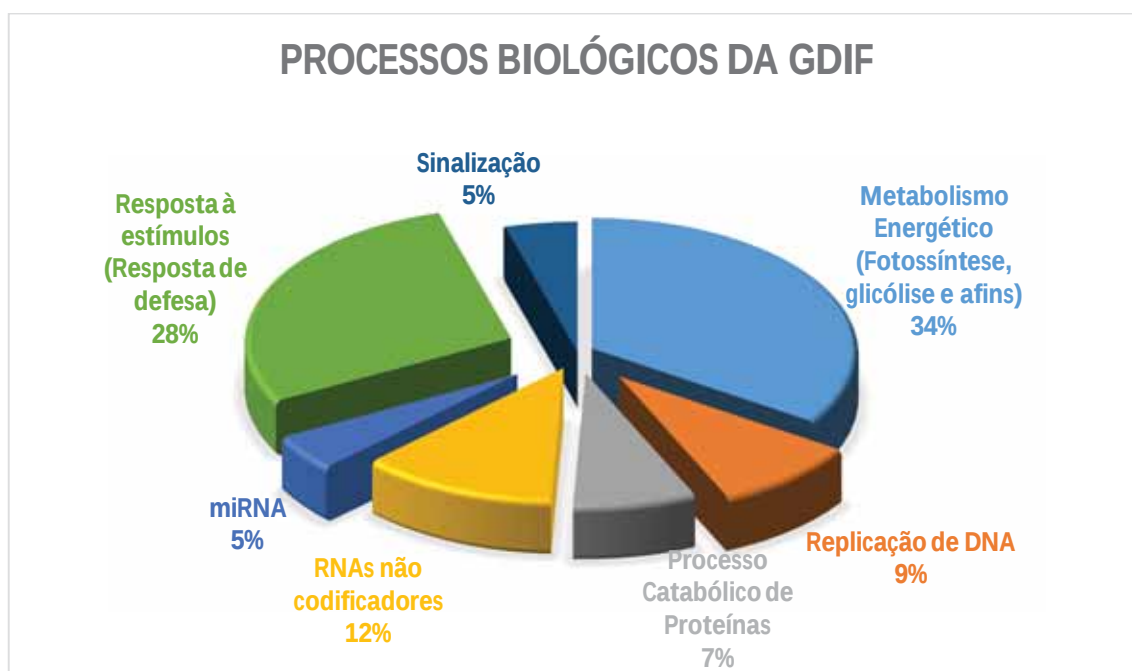


Figura 13: Frequência de grupos de processos biológicos do *Gene Ontology* encontrados na análise de enriquecimento funcional na rede de interação GDIF.

Dentre os processos biológicos enriquecidos significativamente na rede de interação molecular GDIF (Tabela 9), podemos relacionar os processos e fases da fotossíntese que contabilizam 14 GO possivelmente alterados (*photosynthesis, oxidative phosphorylation, NADPH regeneration, pentose phosphate shunt, hexose biosynthetic process, oxidoreduction coenzyme metabolic process, NADP metabolic process, nicotinamide nucleotide metabolic process, ATP synthesis coupled electron transport, mitochondrial ATP synthesis coupled electron transport, respiratory electron transport chain, energy derivation by oxidation of organic compounds e cellular respiration*).

Esses termos GO destacados refletem processos que são responsáveis pela geração de energia na planta. Seu evento principal é a fotofosforilação, que é a adição de fosfato ao difosfato de adenosina (ADP). A energia luminosa é captada por meio de pigmentos fotossintetizantes, capazes de conduzi-la até o centro de reação (P_{700}). Os elétrons excitados da P_{700} saem da molécula e são transferidos para uma primeira substância aceptora de elétrons, a ferroxidina. Esta logo os transfere para outra substância, e assim por diante, formando uma cadeia de transporte de elétrons. Nessa transferência, os elétrons vão liberando energia gradativamente e esta é aproveitada para transportar hidrogênio iônico de fora para dentro do tilacóide, reduzindo o pH do interior deste. A redução do pH ativa o complexo proteico “ATP sintetase”. O fluxo de hidrogênios iônicos através do complexo a fosforilação de moléculas de adenosina difosfato dando origem à adenosina trifosfato (ATP). Porém, existe outra forma de fosforilação, onde os elétrons das moléculas (P_{700}), excitados pela luz, são captados pela ferredoxina, mas ao em vez de passarem pela cadeia transportadora são captados pelo NADP (nicotinamida adenina dinucleotídeo fosfato) e não retornam para o P_{700} . Este fica temporariamente deficiente de elétrons. Esses elétrons são repostos por outras vias. Este processo de transporte também promove a síntese do (ATP) (Blankenship, 2008).

A etapa química, ciclo de Calvin ou ciclo das pentoses, que ocorre no estroma do cloroplasto, o tilacóide fornece ATP e NADPH₂ ao estroma do cloroplasto, onde se encontra a pentose (ribulose fosfato), essa pentose ativada por um fosfato, fixa o carbono que provém do dióxido de carbono do ar sob a ação catalisadora da "Rubisco" (ribulose bifosfato carboxilase-oxidase) e em seguida é hidrogenada pelo NADPH₂ formando o aldeído que dará origem à glicose (Bidlack e Jansky, 2003).

Termos envolvidos com o metabolismo da glicose, um dos principais produtos sintetizados através do ciclo fotossintético, e também, necessária para a atividade da planta, também estão enriquecidos, o que sugere possíveis alterações no metabolismo dessa molécula quando plantas normais e anormais são comparadas. Termos relacionados a sua formação (*gluconeogenesis*), utilização (*glucose metabolic process*, *glycolytic process*) e degradação (*glucose catabolic process*) estão enriquecidos. Outros termos relacionados a síntese de energia foram significativamente enriquecidos, tais como a “*glyceraldehyde 3 phosphate metabolic process*”, necessário para a síntese de carboidratos complexos, além do “*pyruvate metabolic process*” para a síntese de ATP.

Alguns termos de processo biológico foram significativamente enriquecidos no conjunto de proteínas/genes relacionado aos carboidratos (*carbohydrate derivative biosynthetic process*, *single organism carbohydrate catabolic process*) e, conforme o tamanho, podem ser considerados monossacarídeos (*monosaccharide metabolic process*).

Carboidratos, gorduras e óleos são importantes formas de armazenamento de carbono em plantas. O produto final na quebra de lipídios (ou ácidos graxos) é acetil-CoA, um intermediário básico do metabolismo central enviado para o ciclo de ácido cítrico. O NADH e FADH₂ gerado no ciclo do ácido cítrico fornece a energia para a

síntese de ATP, na síntese de fosforilação oxidativa na cadeia transportadora de elétrons (Taiz e Zeiger, 2013).

Termos relacionados a processos que modulam a frequência, a velocidade ou a extensão da replicação do DNA também foram significativamente enriquecidos entre os genes/proteínas da GDIF, o que sugere que esses processos podem estar alterados nas plantas anormais em decorrência das Eu2Aras. Podemos destacar (*cell cycle DNA replication*, *DNA endoreduplication*, *nucleotide biosynthetic process*, *nucleoside phosphate biosynthetic process*, *ribose phosphate metabolic process* e o *ribonucleotide metabolic process*). Estes processos são essenciais no ciclo de vida da planta como qualquer outro organismo vivo. A replicação do material genético se torna indispensável pela necessidade da duplicação celular que tem como finalidade, na mitose, repor células mortas, mantendo a homeostase do organismo, além da participação direta no crescimento e desenvolvimento das plantas, regeneração de lesões e renovação dos tecidos, e na meiose, para formação dos gametas (Taiz e Zeiger, 2013).

Referindo-se a macromoléculas, principalmente proteínas, notamos a ocorrência de termos do GO enriquecidos relacionados a processos catabólicos (*protein catabolic process*, *proteolysis involved in cellular protein catabolic process*, *cellular macromolecule catabolic process*, *macromolecule catabolic process*, *cellular modified amino acid metabolic process*). Catabolismo de proteínas é a degradação de proteínas em aminoácidos e outros compostos derivados, podendo ser utilizado para diversos fins, como por exemplo, para a polimerização em novas proteínas através do RNA e a “maquinaria” ribossômica. Catabolismo de proteínas, que é a quebra de macromoléculas, é essencialmente um processo de digestão, efetuado, principalmente pelas proteases (Alberts, 2008).

Termos relacionados a RNAs não codificadores de proteínas também mostram-se significativos na análise de enriquecimento funcional da GDIF. Podemos mencionar, por exemplo, o termo do GO “*ncRNA processing*” que atua na conversão do transcrito primário de ncRNA em moléculas maduras de ncRNA. Outros processos significativos na GDIF estão envolvidos na conversão de um transcrito de RNA ribossomal primário (rRNA) em uma ou mais moléculas de rRNA maduras (*rRNA metabolic process* e *rRNA processing*). O funcionamento correto do rRNA é de muita importância para a síntese de proteínas em um organismo. Seu mal funcionamento prejudica a interação correta entre o mRNA e o anticódon do RNA transportador (tRNA) e isso perturba a formação de proteínas essenciais. Termos do GO relacionados a fatores que modulem a atividade no silenciamento de genes em processos transcricionais ou pós-transcricionais resultando na inativação do gene a longo prazo foram enriquecidos (*gene silencing*, *production of siRNA involved in RNA interference*, *posttranscriptional gene silencing*, *RNA interference*, *posttranscriptional gene silencing by RNA*).

Em destaque, foram enriquecidos significativamente termos relacionados a produção e ação de miRNA (*gene silencing by miRNA*, *production of miRNAs involved in gene silencing by miRNA* e *posttranscriptional regulation of gene expression*). Os miRNAs pertencem a categoria dos RNAs não codificadores de proteínas, mas nesse caso, merece ser discutido separadamente. A ação de um miRNA, tanto pela clivagem quanto por mecanismos de repressão da tradução, requer uma complementaridade de pares de bases entre o miRNA e o seu RNA mensageiro (mRNA) alvo. Em plantas, os miRNAs e seus alvos apresentam alto grau de complementariedade e a inibição ocorre, de um modo geral, via clivagem do mRNA (Voinnet, 2009).

De acordo com a Tabela 9, podemos destacar os genes/proteínas da GDIF representativos na categoria funcional 'Resposta à estímulos'. A maioria dos termos desta categoria estão relacionados a defesa da planta contra patógenos. Os termos “*immune effector process*” e “*systemic acquired resistance*”, não fazem parte da categoria funcional resposta à estímulos, mas estão diretamente relacionados a todos os tipos de sistemas imunitários que possam contribuir na defesa da planta contra estresses bióticos e, atuam em conjunto, com o termo “*systemic acquired resistance salicylic acid mediated signaling pathway*”. O ácido salicílico (AS) interage com enzimas catalase e peroxidase, resultando no aumento dos níveis de H₂O₂ pela inibição da atividade dessas enzimas (Chen et al., 1993; Hammond-Kosack e Jones, 1996; Durner et al., 1997; Durrant e Dong, 2004). Com isso, o AS intensifica a produção de espécies reativas de oxigênio (ROS) e, consequentemente, a resposta de hipersensibilidade (Levine et al., 1994; Durrant e Dong, 2004). A participação do AS na sinalização e indução da resistência sistêmica adquirida (SAR) é confirmado pela sua presença no floema após a infecção pelo patógeno (Ward et al., 1991; Durner et al., 1997; Durrant e Dong, 2004). Outros termos GO enriquecidos significativamente relacionados a resposta de defesa podem ser mencionados, como o “*defense response incompatible interaction*”, que por se tratar de um termo genérico, engloba todos os tipos de defesa relacionados a patogenicidade, tais como vírus, bactérias, fungos ou outros agentes patogênicos. Quando se trata especificamente de resposta de defesa contra fungos, os termos “*defense response to fungus*”, “*response to chitin*” e “*response to fungus*” aparecem enriquecidos. 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). Considerando a defesa contra

vírus, os termos “*response to virus*”, “*defense response to virus*”, “*dsRNA fragmentation*”, “*response to dsRNA*”, “*production of small RNA involved in gene silencing by RNA*” e “*cellular response to dsRNA*” também apareceram significativamente enriquecidos. Os vírus dsRNAs se replicam no citosol. Quando exposto ao ambiente citoplasmático, o genoma de RNA dupla fita destes vírus podem induzir uma série de mecanismos de defesa celular, como a apoptose, a produção de interferons e RNAs de interferência (Carter e Saunders, 2007).

Alguns outros termos GO (*respiratory burst involved in defense response e regulation of response to stress*) significativamente frequentes estão envolvidos em processos de defesa abióticos. O primeiro tem a função de atuar em uma fase de atividade metabólica elevada, durante o qual aumenta o consumo de oxigênio, criados como parte de uma resposta de defesa; isto conduz à produção de um sistema dependente de NADH, de peróxido de hidrogênio (H_2O_2), ânion superóxido e radicais hidroxila (Kawano, 2003), e o segundo, modular a taxa da resposta ao estresse osmótico. A Glicina betaína é um membro representativo deste grupo osmólito que é sintetizado no cloroplasto a partir de colina (Wang et al., 2003). A colina e glicina betaína são aumentados quando as plantas são submetidas a um stress abiótico, fazendo um balanceamento hídrico na planta (Storey e Jones, 1975; Russel et al, 1998).

Termos GO enriquecidos significativamente relacionados a sinalização celular podem ser mencionados, como o (*intracellular signal transduction, ethylene activated signaling pathway, phosphorelay signal transduction system*). Atuam basicamente transmitindo um sinal para elementos receptores no interior da célula, que se tornam ativados para propagar-se adiante e, finalmente, o sinal desencadear uma alteração na função ou estado da célula. Podemos citar a recepção à etileno (C_2H_4), onde uma das finalidades é modular um processo celular, por exemplo a transcrição.

Tabela 9: Termos GO enriquecidos dos genes/proteínas da rede de interação molecular GDIF.

Termos GO	Número de hits	p-value (corrigido)
carbohydrate catabolic process(5)	52	7,45E-013
single organism carbohydrate catabolic process(5)	51	7,45E-013
single organism carbohydrate catabolic process(5)	51	1,02E-011
protein catabolic process(5)	51	1,02E-011
proteolysis involved in cellular protein catabolic process(6)	51	1,02E-011
cellular macromolecule catabolic process(5)	58	2,07E-011
macromolecule catabolic process(5)	59	2,07E-011
generation of precursor metabolites and energy(4)	48	7,04E-011
pyruvate metabolic process(8)	47	7,66E-011
glucose metabolic process(7)	45	1,02E-010
glycolytic process(4)	28	1,27E-010
nucleotide metabolic process(6)	65	8,69E-010
nucleoside phosphate metabolic process(5)	65	1,07E-009
response to cadmium ion(6)	39	2,27E-009
coenzyme metabolic process(5)	39	3,36E-009
hexose metabolic process(6)	45	6,01E-009
nucleobase containing small molecule metabolic process(4)	65	6,01E-009
ubiquitin dependent protein catabolic process(8)	39	6,83E-009
modification dependent macromolecule catabolic process(6)		
modification dependent protein catabolic process(7)	39	9,96E-009
cofactor metabolic process(4)	46	1,69E-008
monosaccharide metabolic process(5)	46	3,06E-008
proteasomal protein catabolic process(6)	23	3,06E-008
oxidation reduction process(4)	42	4,60E-007
response to metal ion(5)	39	1,02E-006
response to topologically incorrect protein(4)	36	1,13E-006
response to misfolded protein(5)		
proteasome assembly(7)	23	3,74E-006
copper ion transport(8)	10	4,66E-006
gluconeogenesis(8)	19	4,66E-006
hexose biosynthetic process(7)	19	4,66E-006
monosaccharide biosynthetic process(6)	19	8,39E-006
NADPH regeneration(5)	22	1,59E-005
pentose phosphate shunt(6)	22	1,59E-005
oxidoreduction coenzyme metabolic process(6)	25	2,00E-005
NADP metabolic process(8)	22	2,27E-005
nicotinamide nucleotide metabolic process(7)	22	2,27E-005
pyridine nucleotide metabolic process(6)	22	2,27E-005
cellular protein complex assembly(6)	37	2,95E-005
glucose catabolic process(8)	26	2,95E-005
monosaccharide catabolic process(6)		
hexose catabolic process(7)	26	2,95E-005

Tabela 9: Termos GO enriquecidos dos genes/proteínas da rede de interação molecular GDIF.

(continua)

Termos GO	Número de hits	p-value (corrigido)
pyridine containing compound metabolic process(5)	22	3,21E-005
protein complex assembly(5)	39	3,31E-005
protein complex biogenesis(4)	39	3,31E-005
photosynthesis(4)	25	3,39E-005
proteasome mediated ubiquitin dependent protein catabolic process(7)	11	3,62E-005
cellular macromolecular complex assembly(6)	38	5,34E-005
defense response incompatible interaction(5)	37	1,15E-004
immune effector process(3)	28	1,46E-004
ncRNA metabolic process(6)	34	1,65E-004
macromolecular complex assembly(5)	39	2,34E-004
response to mechanical stimulus(4)	9	2,43E-004
proteasome core complex assembly(8)	17	2,68E-004
plastid organization(4)	26	3,19E-004
response to wounding(4)	21	3,43E-004
cofactor biosynthetic process(5)	20	4,38E-004
methylation(3)		
macromolecule methylation(4)	49	6,04E-004
response to chitin(5)	38	6,23E-004
response to organonitrogen compound(4)	38	6,23E-004
ribose phosphate metabolic process(5)		
ribonucleotide metabolic process(6)	29	7,82E-004
single organism membrane organization(3)	36	8,86E-004
cytoskeleton organization(4)	29	9,42E-004
isoprenoid biosynthetic process(5)	32	9,42E-004
isoprenoid metabolic process(5)	32	9,42E-004
nucleotide biosynthetic process(6)	35	9,42E-004
nucleoside phosphate biosynthetic process(5)	35	1,10E-003
systemic acquired resistance(6)	32	1,14E-003
purine nucleotide metabolic process(6)	15	1,21E-003
sulfur compound metabolic process(4)	33	1,54E-003
RNA methylation(5)	32	1,66E-003
purine containing compound metabolic process(5)	15	1,77E-003
organic hydroxy compound biosynthetic process(5)	37	1,78E-003
response to bacterium(4)	24	2,23E-003
rRNA metabolic process(7)	22	2,33E-003
rRNA processing(6)	22	2,33E-003
defense response to bacterium(5)	13	2,59E-003
chloroplast organization(5)	16	2,73E-003
sulfur compound biosynthetic process(5)	28	3,21E-003
electron transport chain(4)	9	4,48E-003
DNA endoreduplication(6)	6	5,82E-003
cell cycle DNA replication(5)	6	5,82E-003

Tabela 9: Termos GO enriquecidos dos genes/proteínas da rede de interação molecular GDIF.

(continua)

Termos GO	Número de hits	p-value (corrigido)
pteridine containing compound metabolic process(5)		
folic acid containing compound metabolic process(6)	6	5,82E-003
carbohydrate derivative biosynthetic process(5)	48	5,83E-003
RNA modification(6)	42	6,96E-003
defense response to fungus(5)	19	8,66E-003
plastid membrane organization(4)		
thylakoid membrane organization(5)	15	9,00E-003
cellular response to organic cyclic compound(6)	27	1,00E-002
chromatin organization(4)	32	1,00E-002
positive regulation of cellular protein metabolic process(5)	4	1,00E-002
positive regulation of protein metabolic process(5)	4	1,00E-002
positive regulation of translation(6)	4	1,00E-002
purine ribonucleotide metabolic process(7)	10	1,00E-002
alcohol biosynthetic process(6)	24	1,09E-002
chlorophyll biosynthetic process(4)	9	1,18E-002
histone lysine methylation(6)	20	1,18E-002
microtubule based process(4)	20	1,18E-002
oxidative phosphorylation(5)		
ATP synthesis coupled electron transport(6)		
mitochondrial ATP synthesis coupled electron transport(7)	7	1,38E-002
respiratory electron transport chain(5)	7	1,38E-002
respiratory burst involved in defense response(4)		
respiratory burst(4)	15	1,42E-002
response to virus(4)	14	1,42E-002
ncRNA processing(7)	22	1,48E-002
glyceraldehyde 3 phosphate metabolic process(5)		
isopentenyl diphosphate biosynthetic process methylethanol 4 phosphate pathway(6)	19	1,49E-002
isopentenyl diphosphate biosynthetic process(6)		
isopentenyl diphosphate metabolic process(6)	19	1,49E-002
coenzyme biosynthetic process(6)	9	1,74E-002
intracellular signal transduction(5)	24	1,74E-002
defense response to virus(4)	13	1,80E-002
cellular modified amino acid metabolic process(5)	10	2,05E-002
ribose phosphate biosynthetic process(6)		
ribonucleotide biosynthetic process(7)	22	2,08E-002
glycosyl compound metabolic process(4)	18	2,41E-002
sulfur amino acid biosynthetic process(6)	20	2,49E-002
production of ta siRNAs involved in RNA interference(7)	10	2,91E-002
cellular aldehyde metabolic process(4)	25	2,96E-002
sulfur amino acid metabolic process(5)	20	2,96E-002
RNA splicing via endonucleolytic cleavage and ligation(8)	14	2,97E-002
regulation of hydrogen peroxide metabolic process(6)	11	2,97E-002
regulation of reactive oxygen species metabolic process(5)	11	2,97E-002
regulation of gene expression epigenetic(6)	26	2,98E-002
		3,15E-002
alcohol metabolic process(5)	24	

Tabela 9: Termos GO enriquecidos dos genes/proteínas da rede de interação molecular GDIF.

(continua)

Número de hits	p-value (corrigido)	
gene silencing(4)	24	3,15E-002
ATP metabolic process(8)	7	3,26E-002
glycerolipid biosynthetic process(5)	1	3,26E-002
glycerolipid metabolic process(5)	1	3,26E-002
glycerophospholipid biosynthetic process(6)	1	3,26E-002
glycerophospholipid metabolic process(6)	1	3,26E-002
histone methylation(5)	21	3,26E-002
nucleoside triphosphate metabolic process(6)	7	3,26E-002
phosphatidylinositol biosynthetic process(7)	0	3,26E-002
phosphatidylinositol metabolic process(7)	0	3,26E-002
protein methylation(5)		
protein alkylation(7)	21	3,26E-002
purine nucleoside triphosphate metabolic process(7)	7	3,26E-002
ribonucleoside triphosphate metabolic process(7)		
purine ribonucleoside triphosphate metabolic process(8)	7	3,26E-002
sphingolipid metabolic process(5)	7	3,26E-002
steroid biosynthetic process(6)	15	3,26E-002
steroid metabolic process(5)	15	3,26E-002
systemic acquired resistance salicylic acid mediated signaling pathway(6)	14	3,26E-002
pigment biosynthetic process(5)	17	3,27E-002
dsRNA fragmentation(4)		
response to dsRNA(5)		
production of small RNA involved in gene silencing by RNA(5)		
cellular response to dsRNA(6)	10	3,27E-002
gene silencing by miRNA(7)	10	3,27E-002
posttranscriptional gene silencing(5)	13	3,27E-002
production of miRNAs involved in gene silencing by miRNA(6)	10	3,27E-002
production of siRNA involved in RNA interference(6)	10	3,27E-002
transition metal ion transport(7)	18	3,51E-002
response to fungus(4)	29	3,58E-002
alpha amino acid metabolic process(5)	30	3,60E-002
ethylene activated signaling pathway(6)	8	3,77E-002
nucleotide sugar metabolic process(5)	3	3,77E-002
phosphorelay signal transduction system(6)	8	3,77E-002
photosynthesis light reaction(5)	14	3,77E-002
protein secretion(5)	3	3,77E-002
secretion by cell(4)	3	3,77E-002
secretion(4)	3	3,77E-002
RNA interference(7)	12	4,10E-002
posttranscriptional gene silencing by RNA(6)	12	4,10E-002
energy derivation by oxidation of organic compounds(4)		
cellular respiration(5)	11	4,13E-002

Tabela 9: Termos GO enriquecidos dos genes/proteínas da rede de interação molecular GDIF.

(continua)

Número de hits	p-value (corrigido)	
porphyrin containing compound biosynthetic process(6)	11	4,13E-002
tetrapyrrole biosynthetic process(6)	11	4,13E-002
pyrimidine ribonucleotide metabolic process(7)		
pyrimidine ribonucleotide biosynthetic process(8)	19	4,21E-002
photorespiration(6)	17	4,21E-002
posttranscriptional regulation of gene expression(6)	17	4,21E-002
regulation of response to stress(4)	28	4,21E-002
secondary metabolic process(4)	26	4,63E-002
pyrimidine nucleotide metabolic process(6)		
pyrimidine nucleotide biosynthetic process(7)	19	4,89E-002
histone modification(4)	23	4,95E-002
beta glucan biosynthetic process(7)	8	4,95E-002
cellular metabolic compound salvage(5)	17	4,95E-002
cellulose biosynthetic process(8)	8	4,95E-002

5.3.7 Análise de enriquecimento funcional dos *hubs* da Gdif contra o proteoma de *A. thaliana*

Baseado no enriquecimento funcional entre os 772 *hubs* da GDIF em relação ao proteoma de *A. thaliana*, foram obtidos, no total, 24 termos relacionados à categoria de processos biológicos, enriquecidos significativamente. Estes termos foram subdivididos em grupos, de acordo com sua categoria funcional. Destes, 19% participam do metabolismo energético, 23% participam na resposta à estímulos (resposta de defesa da planta), 35% estão relacionados ao catabolismo de proteínas, 11% participam da sinalização celular e 12% relacionados a secreção (Figura 14). Da mesma forma como o ocorrido em processos biológicos da GDIF, os termos do GO são únicos e cada termo é e corresponde exclusivamente a uma categoria funcional. Portanto, é correto afirmar que um gene/proteína pode estar associado a diversos termos do GO, pois um gene/proteína pode participar de múltiplos processos.

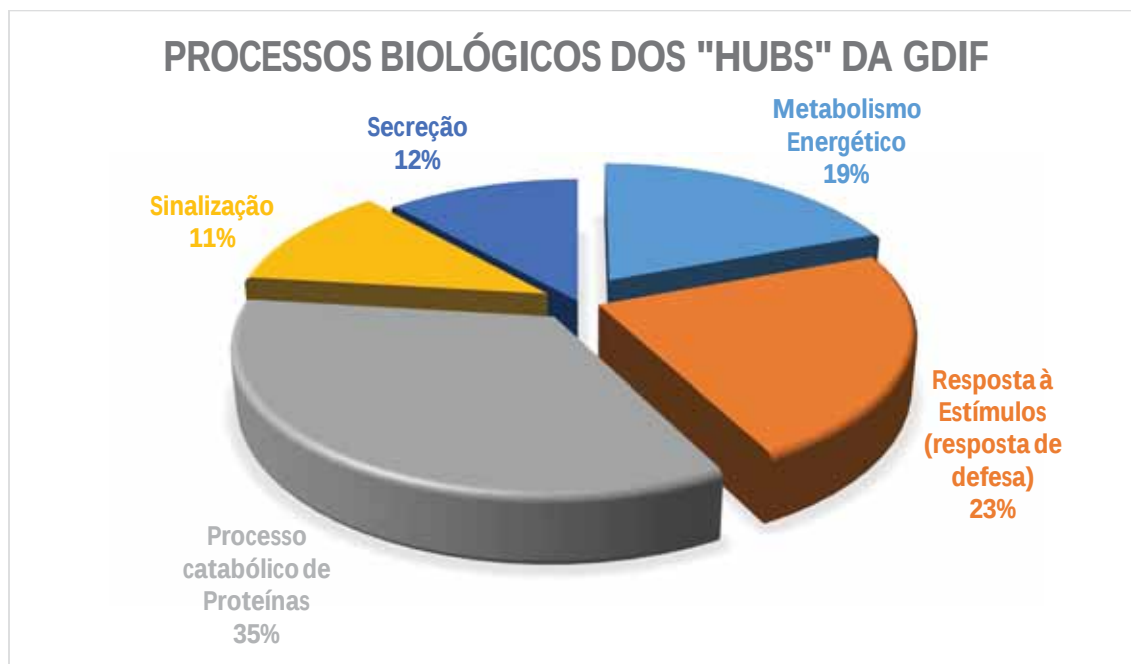


Figura 14: Frequência de grupos de processos biológicos do *Gene Ontology* encontrados na análise de enriquecimento funcional dos *hubs* na rede de interação GDIF.

Dentre os processos biológicos enriquecidos significativamente nos *hubs* da GDIF (Tabela 10), podemos destacar processos relacionados a obtenção de energia, nos termos do GO (*glycolytic process, generation of precursor metabolites and energy, pyruvate metabolic process, single organism carbohydrate catabolic process, carbohydrate catabolic process*). Na glicólise, a glicose é oxidada produzindo duas moléculas de piruvato, duas moléculas de ATP e dois equivalentes reduzidos de NADH^+ , que são introduzidos na cadeia respiratória. A glicólise é uma das principais rotas para geração de ATP nas células e está presente em todos os tipos de tecidos (Nelson e Cox, 2000). Já o processo de catabolismo de carboidratos, é a quebra de carboidratos para produzir ATP, sendo também, uma alternativa na produção de energia.

Dois grupos predominam na análise de enriquecimento funcional destes *hubs*: resposta de defesa e catabolismos de proteínas. No primeiro grupo, 6 termos de GO (*regulation of defense response, regulation of response to stress, response to bacterium,*

response to molecule of bacterial origin, *detection of stimulus* e *detection of biotic stimulus*) pertencem a respostas à estresses bióticos causados por fungos, bactéria ou vírus. Basicamente, esses genes/proteínas reagem à presença de corpos estranhos ou na ocorrência de lesões, com o objetivo de restringir ou prevenir os danos causados as plantas, contribuindo na recuperação da infecção causada pelo ataque.

O outro grupo predominante na análise de enriquecimento funcional destes *hubs*, correspondem aos processos de catabolismo de proteínas, com 9 termos do GO enriquecidos significativamente (*cellular macromolecule catabolic process*, *macromolecule catabolic process*, *modification dependent macromolecule catabolic process*, *ubiquitin dependent protein catabolic process*, *modification dependent protein catabolic process*, *proteasomal protein catabolic process*, *cellular protein catabolic process*, *protein catabolic process*, *proteolysis involved in cellular protein catabolic process*). Catabolismo de proteínas é a degradação de proteínas em aminoácidos e outros compostos derivados, podendo ser utilizados para diversos fins, como por exemplo, para a polimerização em novas proteínas através do RNA e a “maquinaria” ribossômica. Catabolismo de proteínas, que é a quebra de macromoléculas, é essencialmente um processo de digestão, efetuado, principalmente pelas proteases (Alberts, 2008).

Quando dependente de ubiquitina, as quebras de uma proteína ocorrem através de hidrólise das suas ligações peptídicas, iniciados pela ligação covalente de um grupo de ubiquitina ou vários grupos de ubiquitina, para a proteína. Os proteossomas também mediam atividades catabólicas em proteínas, através de hidrolises de suas ligações peptídicas (Lodish et al., 2004).

Três termos relacionados a secreção foram identificados na análise de enriquecimento funcional (*protein secretion*, *secretion by cell*, *secretion*). A secreção é o processo de elaboração, liberando produtos químicos, ou uma substância química

secretada a partir de uma célula ou glândula. Em contraste com a excreção, a substância pode ter uma determinada função, em vez de ser um produto de resíduos. (Lee et al., 2012).

Outros termos enriquecidos significativamente, com relativa importância são os que correspondem atuando na sinalização celular (*enzyme linked receptor protein signaling pathway, transmembrane receptor protein tyrosine kinase signaling pathway, cell surface receptor signaling pathway*). Um receptor ligado à enzima, também conhecido como um receptor catalítico, é um receptor transmembranar, em que a ligação do ligante extracelular causa a atividade enzimática intracelular. Assim, um receptor catalítico é uma proteína integral de membrana possuindo tanto atividade enzimática catalítica, quanto funções de receptor (Dudek, 2006). Os receptores tirosina quinase (RTK) são receptores de elevada afinidade na superfície da célula aos muitos fatores de crescimento polipeptídicos, citocinas e hormônios. Os receptores tirosina quinase iniciam uma cascata de sinalização que culmina na ativação ou inibição de fatores de transcrição que, por sua vez, ativam e/ou inibem a expressão de vários genes envolvidos em vários processos biológicos diferentes, com destaque a proliferação celular. Os receptores tirosina quinase fazem parte da grande família de proteínas tirosina quinase, abrangendo os receptores, que contêm um domínio transmembranar, bem como as tirosina quinase não receptores que não possuem domínios transmembranares (Hubbard e Till, 2000).

Tabela 10: Termos GO enriquecidos no grupo de genes/proteínas *hubs* da GDIF

Termos GO	Número de hits	p-value (corrigido)
protein phosphorylation(7)	22	2,99E-007
cellular protein catabolic process(6)	19	3,28E-007
protein catabolic process(5)	19	3,28E-007
proteolysis involved in cellular protein catabolic process(6)	19	3,28E-007
enzyme linked receptor protein signaling pathway(6)		
transmembrane receptor protein tyrosine kinase signaling pathway(7)	14	7,63E-007
cell surface receptor signaling pathway(5)	14	8,02E-007
cellular macromolecule catabolic process(5)	20	1,41E-006
macromolecule catabolic process(5)	20	2,05E-006
glycolytic process(4)	11	1,03E-005
generation of precursor metabolites and energy(4)	16	1,79E-005
ubiquitin dependent protein catabolic process(8)	14	4,39E-005
modification dependent macromolecule catabolic process(6)		
modification dependent protein catabolic process(7)	14	4,66E-005
single organism carbohydrate catabolic process(5)	15	5,53E-005
carbohydrate catabolic process(5)	15	8,54E-005
response to bacterium(4)	12	1,27E-004
pyruvate metabolic process(8)	14	2,53E-004
proteasomal protein catabolic process(6)	8	1,41E-003
protein secretion(5)	3	2,09E-003
secretion by cell(4)	3	2,09E-003
secretion(4)	3	2,09E-003
response to molecule of bacterial origin(5)	7	6,68E-003
regulation of defense response(5)	11	3,00E-002
regulation of response to stress(4)	11	3,11E-002
detection of biotic stimulus(4)	4	4,01E-002
detection of stimulus(3)	4	4,01E-002

5.3.8 Relação entre os processos biológicos possivelmente alterados e a morfofisiologia da planta anômala

Na análise de enriquecimento funcional da GDIF, notou-se a presença de um grupo de genes/proteínas relacionados a síntese e regulação pós-transcricional por miRNAs. Como esse grupo de genes/proteínas interage com genes/proteínas Eu2Ara,

supõe-se que, as alterações causadas no desenvolvimento da planta possam estar sofrendo influência destes miRNAs. Estudos realizados por Li et al. (2012), demonstraram a atuação irregular de miRNAs em plantas, podem leva-la a uma ativação descontrolada do sistema de defesa sem que ocorra ataque por patógenos, denominada autoimunidade. Estes miRNAs foram identificados como nta-miR6019 (22 nt) e nta-miR6020 (21 nt), os quais clivaram transcritos de receptores de imunidade NB-LRR do tabaco, responsáveis pela resistência ao vírus do mosaico do tabaco. Tais análises apontaram que a expressão desregulada de genes NB-LRR poderiam estar desencadeando a autoimunidade, mesmo na ausência de infecção patogênica e, como consequência, a planta obter a inibição de seu desenvolvimento. Curiosamente, uma das metodologias aplicadas para investigar a causa da anomalia, neste trabalho, foi, exatamente, a realização de predições miRNA em RNA mensageiros alvos, na população anômala do eucalipto em estudo. Estes mRNAs, foram alvos de 5 diferentes classes de miRNAs. Tendo conhecimento da função de cada um desses miRNAs, baseado na literatura, sugeriu-se, que estes estariam prejudicando o desenvolvimento da planta anormal, causando, desta forma, atrofiamento, inibição no crescimento de ramificações laterais das raízes, e como consequência, prejudicando a planta na absorção de água e nutrientes essenciais a sua sobrevivência.

Outra informação relevante obtida da análise de enriquecimento funcional da GDIF e nos *hubs* da GDIF é que a anomalia pode estar associada a alterações na resposta de defesa contra estresses bióticos e abióticos, fato que confirma as análises obtidas por Fuchs (2014). Como já mencionado por esta mesma autora, as progênies anômalas investigadas não mostraram nenhuma evidência de ataques por patógenos e estavam sob um ambiente controlado. A ativação da resposta de defesa de plantas anormais pode ser devido a algum processo autoimune causada por incompatibilidade genética. Esta incompatibilidade, associada à ativação inapropriada do sistema imunitário das plantas

(sistemas auto-imune) é denominado como necrose híbrida (Bomblies & Weigel, 2007; Bomblies et al, 2007; Mizuno et al., 2010). Nas análises de enriquecimento funcional, notou-se um número estatisticamente elevado de termos GO relacionados a fotossíntese. O que nos chamou a atenção é que estes termos biológicos destacados participam da maioria dos processos da fotossíntese até a obtenção do produto final. Associado ao enriquecimento funcional em gene/proteínas relacionado a defesa da planta, mudanças na atividade fotossintética são descritos em plantas sob estresses biótico e abiótico. É relatado, que a atividade fotossintética pode ser reduzida ou aumentada (Alam, 1999; Scharte et al, 2005; Chaves et al, 2009; Bilgin et al, 2010). Segundo Ayres (1991) e Ney et al. (2013), o aumento da atividade fotossintética pode ocorrer como um mecanismo de compensação onde a taxa fotossintética reduzida no local da infecção induz uma atividade fotossintética em outros tecidos não infectados. As plantas sob estresse não só ajustam a sua atividade fotossintética, mas também aumentam a sua capacidade em absorver nutrientes (Alam, 1999), o que poderia explicar o enriquecimento significativo do termo GO em atividades relacionadas a síntese de glicose e carboidratos.

O enriquecimento dos termos relacionados à sinalização sugere que as vias responsáveis pela captação dos sinais externos e a transmissão desses sinais via cascatas de sinalização até um provável fator de transcrição ou grupo de fatores de transcrição também possam estar alterados nas plantas anormais em decorrência das Eu2Aras. As proteínas tirosina quinase (PTKs) são responsáveis pela fosforilação de proteínas e, como tal, exercem papel central fundamental na transdução de sinais, pois elas atuam como ponto de apoio na rede de moléculas sinalizadoras independentes, cuja função é a regulação da expressão gênica. As PTKs estão relacionadas a diversos processos fundamentais, como proliferação, diferenciação, sobrevivência ou morte celular (Hubbard et al., 1998).

Outros exemplos de processos bem estudados de vias de sinalização mediadas por PTKs são desencadeados pela percepção de hormônio brassinosteróides (grupo dos esteróides), pelas plantas, que tem como função promover o crescimento da planta (Zhu et al., 2013), ou de padrões moleculares associados a patógeneses (PAMPs), que conduzem a respostas imunes que constituem a primeira camada ativa de imunidade da planta (Boller e Felix, 2009). A sinalização alterada em plantas anormais pode estar ligada ao processo de auto-imunidade, o qual parece estar ocorrendo nas plantas em estudo.

6 CONCLUSÃO

- Uma porcentagem pequena de SNPs detectados pela técnica de RNAseq não foram detectadas pelo sequenciamento de Sanger, mostrando a eficiência das duas técnicas.

- Não houveram separação genotípica entre os indivíduos normais e anômalos quando amplificados pelos marcadores moleculares.

- Por se tratar de genes que, segundo a literatura, apresentam um grande impacto à resposta da planta na defesa, os marcadores tipo SNP desenvolvidos e validados, neste trabalho, poderão ser avaliados em progênies obtidas de cruzamentos controlados, segregantes para característica relacionadas a estresses bióticos e abióticos.

- Cinco diferentes classes de miRNAs apresentaram transcritos preditos relacionados aos genes diferencialmente expressos.

- Os microRNAs euc-miR156, euc-miR166 e euc-miR167 foram considerados candidatos, atuando na regulação gênica causando prejuízo no desenvolvimento das ramificações laterais da raiz e no tamanho da planta anômala.

- A análise de enriquecimento funcional do interatoma relacionado as eu2ara (proteínas de *A. thaliana* homologas às codificadas pelos genes diferencialmente expressos do *E. grandis*) indica que muitos processos biológicos que não estão envolvidos com os diferencialmente expressos, estão alterados em decorrência deles. Levando em consideração esses processos, concluímos que a morfofisiologia das plantas anormais parecem ser uma propriedade emergente (são propriedades que surgem somente quando ha interação entre os componentes do sistema) das alterações desses múltiplos processos

biológicos. Deve se levar em consideração que essas análises foram feitas de forma puramente computacional e que experimentos de bancada devem ser realizados para sua confirmação.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Abatepaulo AR, Caetano AR, Mendes JR. Detection of SNPs in bovine immune-response genes that may mediate resistance to the cattle tick *Rhipicephalus (Boophilus) microplus*. *Animal Genetics*. 2008; 39: 328-329.

Abeles FB, Bosshart RT, Forrense LE, Habig WH. Preparation and purification of glucanase and chitinase from bean leaves. *Plant Physiol*. 1970; 47: 129-134.

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.

Agrios, G N. *Plant pathology*. San Diego: Academic Press. 1997; 4: 635p.

Alam SM. Nutrient uptake by plants under stress conditions. In: Pessarakli M, *Handbook of Plant and Crop Stress*. New York, US: Marcel Dekker. 1999; 285–314.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. **Biologia Molecular da Célula**. Editora Artmed. 2010; 5ed.

Alberts B. *Molecular Biology of the Cell*. New York: Garland Science. 2008; 5ed.

Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nature Genetics*. 2004; 36: 1282-1290.

Allmer J, Yousef M. Computational methods for ab initio detection of microRNAs. *Front Genet*. 2012. 3(209).

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.

An C, MOU Z. Host defense response in a novel (*Arabidopsis Xanthomonas citri*) subsp. *Citri* pathosystem. *Plos one*. 2012; 7: 31130.

Anguelova MV, Westhuizen VD, Pretorius ZA. Beta-1,3-glucanase and chitinase activities and the resistance response of wheat to leaf rust. *J. Phytopathol*. 2001; 149: 381-384.

Anthonisse J. The rush in a directed graph. [S.l.], 1971. Technical Report BN 9/71.

Antikainen M, Griffith M, Zhang J, Hon WC, Yang DSC, Pihakashi-Maunsbach K. Immuno localization of antifreeze proteins in winter rye leaves, crowns, and roots by tissue printing. *Plant Physiol*. 1996; 110: 845–857.

Arazi T, Talmor-Neiman M, et al. Cloning and characterization of micro-RNAs from moss. *Plant Journal*. 2005, 43(6) p.837-848.

Assis TF, Resende MDV. Genetic improvement of forest tree species. *Crop Breeding and Applied Biotechnology*. 2011; 1: 44-49.

Assis TF. Production and use of Eucalyptus hybrids for industrial purposes. In: DUNGEY HS, DIETERS MJ, NIKLES DG. *Hibrid Breeding and Genetics of Forest trees: Proceedings of QFRI/CRC-SPF Symposium*. 2000; 63-74.

Aukerman MJ, Sakai H. Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2* -like target genes. *Plant Cell*. 2003; 15: 2730-2741.

Axtell MJ. Classification and Comparison of Small RNAs from Plants. *Annu Rev Plant Biol*. 2013; 16.

Ayres PG. Growth responses induced by pathogens and other stresses. In: Mooney HA, Winner WE, Pell EJ, eds. *Response of Plants to Multiple Stresses*. New York, US: Academic Press. 1991; 227-246.

Backes C, et al. GeneTrail—advanced gene set enrichment analysis. *Nucleic Acids Res.* 2007; 35: 86–92.

Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004; 5(2): 101–113.

Barber MS, Bertram RE, Ride JP. Chitin oligosaccharides elicit lignifications in wounded wheat leaves. *Physiol. Mol. Plant Pathol.* 1989; 34: 3-12.

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.

Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell, Cambridge.* 2009; 136: 215-233.

Bartnicki-Garcia S. Cellwall chemistry, morphogenesis, and taxonomy of fungi. *Ann. Rev. Microbiol.* 1968; 22: 87-108.

Batalia MA, Monzingo AF, Ernst S, Roberts W, Robertus JD. The crystal structure of the antifungal protein zeamatin, a member of the thaumatin-like, PR-5 protein family. *Nat Struct Biol.* 1996; 3: 19–23.

Bennett, S. Solexa Ltd. *Pharmacogenomics.* 2004; 5: 433-8.

Bidlack JE; Stern KR, Jansky S. *Introductory plant biology.* New York: McGraw-Hill, 2003.

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.

Blankenship RE. *Molecular Mechanisms of Photosynthesis*. John Wiley & Sons Inc. 2008; 2 ed.

Boller T, Felix, G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern recognition receptors. *Annu. Rev. Plant Biol.* 2009; 60: 379–406.

Bomblies K, Lempe J, Eppele 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.

Bormann C, Baier D, Horr I, Raps C, Berger J, Jung G, Schwarz H. Characterization of a novel, antifungal, chitin binding protein from (*Streptomyces tendae*) Tu901 that interferes with growth polarity. *J Bacteriol.* 1999; 181: 7421–7429.

Boualem A, Laporte P, Jovanovic M, Laffont C, Plet J, Combier J, Niebel A, Crespi M, Frugier F. MicroRNA166 controls root and nodule development in *Medicago truncatula*. *The Plant Journal*. 2008; 54: 876–887.

Broekaert WF, Van Parijs J, Allen AK, Peumans WJ. Comparison of some molecular, enzymatic and antifungal properties of chitinases from thorn-apple, tobacco and wheat. *Physiol. Mol. Plant Pathol.* 1988; 33: 319-331.

Brooker MIH. A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Australian Systematic Botany*. 2000; 13: 79 – 148.

Brookes AJ. The essence of SNPs. *Gene*. 1999; 234: 177-186.

Bundock PC, Elliott, FG, Ablett G, Benson AD, Casu RE, Aitken KS, Henry RJ. (2009) Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnol.* 2009; 7: 347–354.

Calvo JH, Martínez-Royo A, Silveri L. Isolation, mapping and identification of SNPs for four genes (ACP6, CGN, ANXA9, SLC27A3) from a bovine QTL region on BTA3. *Cytogenetics and Genome Research.* 2006; 114(1): 39-43.

Cànovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, Medrano J. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm Genome.* 2010; 21:592–598.

Carter J, Saunders, V. *Virology: Principles and Applications.* Chichester: Wiley. 2007; 382p).

Cato, AS, Gardner RC, Kent J, Richardson TE. A rapid PCR-based method for genetically mapping ESTs. *Theoretical and Applied Genetics.* New York. 2001; 102: 296-306.

Chandrasekar A, Riju A, Sithara K, Anoop S, Eapen SJ. Identification of single nucleotide polymorphism in ginger using expressed sequence tags. *Bioinformation.* 2009; 4: 119-122.

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 K, Baxter T, Muir WM. Genetic Resources, Genome Mapping and evolutionary Genomics of the Pig (*Sus scrofa*). *International Journal of Biological Sciences.* 2007; (3).

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.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ. (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet.* 2002; 3: 19p.

Cho SH, Coruh C, Axtell MJ. miR156 and miR390 Regulate tasiRNA Accumulation and Developmental Timing in *Physcomitrella patens*. *The Plant Cell*. 2012; 24: 4837–4849.

Cirulli ET, Singh A, Shianna KV, Ge D, Smith JP et al. Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol*. 2010; 11: 57p.

Cohen-Kupiec R, Chet I. The molecular biology of chitin digestion. *Curr. Opinion Biotechnol*. 1998; 9: 270-277.

Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic acids research*. 2008; 36(19): 122p.

Cui H, Xiang T, Zhou JM. Plant immunity: a lesson from pathogenic bacterial effector proteins. *Cell Microbiol*. 2009; 11(10): 1453-1461.

Dagan T, Talmor Y, Grau D. Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. *Mol. Biol*. 2002; 19: 1022–1102.

Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res*. 2011; 39.

Dangl JL, Jones JDG. Plant pathogens and integrated defense responses to infection. *Nature*. 2001; 411: 826-833.

Dantec LL, Chagne D, Pot D, Cantin O, Garnier-Gere, P, Frank Bedon F, Frigerio JM, Chaumeil P, Leger P, Garcia V, Laigret F, de Daruvar A, Plomion C. Automated SNP detection in expressed sequence tags: statistical considerations and application to maritime pine sequences. *Plant Mol. Biol*. 2004; 54: 461–470.

Daurelio LD, Romeroe M, Petrocelli S, Merelo P, Cortadi A, Talón M, Tadeo F, Orelano E. Characterization of *Citrus sinensis* transcription factors closely associated with the non-host response to (*Xanthomonas campestris*) pv.vesicatoria. J Plant Physiol. 2013.

De Souza AA, Takita MA, Coletta Filho HD, Campos MA, Teixeira JEC, Targon MLPN, Carlos EF, Ravasi JR, Fischer CN, Machado MA. Comparative analysis of differentially expressed sequence tags of sweet orange and mandarin infected with *Xylella fastidiosa*. Genetic and Molecular Biology. 2009; 30(3): 965-971.

Derckel J, Audran J, Haye B, Lambert B, Legendre L. Characterization, induction by wounding and salicylic acid, and activity against *Botrytis cinerea* of chitinases and β -1,3-glucanases of ripening grape berries. Physiol. Plant.1998; 104: 56-64.

Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. Cell. 2006; 127: 1309–1321.

Domingues DS. Identificação de marcador RAPD e SCAR relacionado ao caráter florescimento precoce em *Eucalyptus grandis*. Botucatu, 2003, 24p. Trabalho científico (exigência parcial para obtenção do título de Bacharel em Ciências Biológicas) – Instituto de Biociências, Universidade Estadual Paulista.

Domingues MN. Caracterização de proteínas de *Citrus sinensis* que interagem com a proteína efetora PthA, indutora do cancro cítrico. 2011. UNICAMP.

Doughty RW. The eucalyptus: a natural and commercial history of the gum tree: Johns Hopkins University Press. 2000.

Du J, Miura E, et al. The Populus Class III HD ZIP transcription factor POPCORONA affects cell differentiation during secondary growth of woody stems. PLoS One. 2011, 6(2), 17458.

Dudek RW. High-yield cell and molecular biology. Lippincott Williams & Wilkins. 2006; 19p.

Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. *Nature*. 1980; 287: 560–561.

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.

Ebrahim S, Usha K, Singh B. Pathogenesis related (PR) proteins in plant defense mechanism science against microbial pathogens: communicating current research and technological advances. Formatex Research Center, Badajoz. 2011.

Edens L, Hesling L, Klock R, Ledeboer AM, Maat J, Toonen MY, Visser C, Verrips CT. Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its expression in (*Escherichia coli*). *Gene*. 1982; 18: 1–2.

Eldridge K. Eucalypt domestication and breeding. Oxford, New York: Clarendon Press; Oxford University Press, 1993.

Eldridge KJ, Davidson, et al. Eucalypt Domestication and Breeding: Oxford University Press, USA. 1994.

Emara MG, Kim H. Genetic Markers and their Application in Poultry Breeding. *Poultry Science*. 2003; 82: 952–957.

Emrich SJ, Barbazuk WB, Li L, Schnable PS. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome research*. 2007; 17(1): 69-73.

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.

Ewing B, Hillier L, Wendl MC et al. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*. 2008; 8(3): 175-85.

Fahlgren N, Howell MD, et al. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE*. 2007; 2(2): 219p.

FAO. *Eucalypts for Planting: Forestry Ser.ed.*: Fao, 1981.

Febres VJ, Khalaf A, Gmitter JRFG, Moore GA. Production of disease resistance in Citrus by understanding natural defense pathways and pathogen interactions. *Tree and Forestry Science and Biotechnology*. 2009; 3:30-39.

Feltus FA, Wan J, Schulz SR, Estill JC, Jiang N, Paterson AH. An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res*. 2004; 14: 1812–1819.

Ferreira ME, Grattapaglia D. Introdução ao uso de marcadores moleculares em análises genéticas. Brasília. Embrapa-Cenargen. 1998; 3: 220p.

Ferreira, M. Melhoramento e a Silvicultura intensiva clonal. IPEF. 1992; 45: 8p.

Flach J, Pilet PE, Jollès P. What's new in chitinase research? *Experientia*. 1992; 48: 701-716.

Fladung M, Buschbom J. Identification of single nucleotide polymorphisms in different *Populus* species. *Trees*. 2009; 23(6): 1199-1212.

Freeman S, Herron JC. *Análise Evolutiva*. Porto Alegre: Artmed; 2009.

Fu C, Sunkar R, Zhou C, Shen H, Zhang J, Matts J, Wolf J, Mann D, Stewart Jr N, Tang Y, Wang Z. Overexpression of miR156 in switchgrass (*Panicum virgatum* L.) results in various morphological alterations and leads to improved biomass production *Plant Biotechnology Journal*. 2012; 10: 443–452.

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.

Futuyma DJ. *Evolutionary Biology*. 3rd ed. Sunderland, MA: Sinauer; 1997.

Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, Van Neerven RJJ, Schou C, Lowenstein H, Spangfort MD. X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nature Structural Biology*. 1996; 3(12): 1040-1045.

Ganal MW, Altmann T, Roder MS. SNP identification in crop plants. *Plant Biol*. 2009; 12(2):211p.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucl Acids Res*. 2003; 31: 3784-3788.

Gatfield D, Unterholzner L, Ciccarelli FD, Bork P, Izaurralde E. Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *The EMBO Journal*. 2003; 22(15): 3960-3970.

Gebelin V, Argout X, et al. Identification of novel microRNAs in *Hevea brasiliensis* and computational prediction of their targets. *BMC Plant Biol*. 2012; v.12, p.18.

Giri A P, Harsulkar A M, Patankar A G, Gupta VS, Sainani M N, Deshpande V V, Ranjekar PK. Association of induction of protease and chitinase in chickpea roots with resistance to (*Fusarium oxysporum* f. sp.ciceri). *Plant Pathol*. 1998; 47: 693-699.

Golle DP, Reiniger LRS, et al. "Melhoramento florestal: ênfase na aplicação da biotecnologia." *Ciência Rural*. 2009; 39(5): 1607-1614.

Gonçalves FMA, Rezendo GD, et al. "Progresso Genético por Meio da Seleção de Clones de Eucalipto em Plantios Comerciais." *Revista Árvore*. 2001; 25(3): 289-293.

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.

Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB. Association genetics in (*Pinus taeda*L.) I. wood property traits. *Genetics*. 2007; 175: 399–409.

Gordon H. FASTX-Toolkit: FASTQ/A Short-Reads Pre-processing Tools. 2009.

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

Grattapaglia D, Silva-Junior OB, Kirst M, de Lima BM, Faria DA, Pappas-Jr GJ. High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Plant Biol*. 2011; 11: 65p.

Grewal RK, Gupta S, Das S. (*Xanthomonas oryzae* pv. *Oryzae* triggers) immediate transcriptomic modulations in rice. *BMC Genomics*. 2012; 13: 49-61.

Griffiths-Jones S, Saini HK, et al. miRBase: tools for microRNA genomics. *Nucleic Acids Res*, v.36, n.Database issue. 2008; p.154-8.

Gupta, PK, Roy JK, Prasad M. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science*, Washington, D.C. 2001; 80: 524-535.

Guzman FM Almerao P, et al. Identification of microRNAs from *Eugenia uniflora* by high-throughput sequencing and bioinformatics analysis. *PLoS One*. 2012; 7(11): 49811.

Ha M, Pang MX, Agarwal V, Chen Z.J. Interspecies regulation of microRNAs and their targets. *Biochimica et Biophysica Acta -Gene Regulatory Mechanisms*, Austin. 2008; 1779:735- 742.

Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Pasadena, CA USA: [s.n.], 2008. p. 11–15.

Han MH, Goud S, Song L, Fedoroff N. The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, Washington. 2004; 101:1093-1098.

Harfouche A, Meilan R, Kirst M, Morgante M, Boerjan W, Sabatti M, Mugnozza GS. Accelerating the domestication of forest trees in a changing world. *Trend in Plant Science*. 2012; 17(2): 64-72.

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol*. 2009; 10: 32p.
Hejgaard J, Jacobsen S, Svendsen I. Two antifungal thaumatin-like proteins from barley grain. *FEBS Lett*. 1991; 291: 127–131.

Hendre PS, Kamalakannan R, Varghese M. High-throughput and parallel SNP discovery in selected candidate genes in *Eucalyptus camaldulensis* using Illumina NGS platform *Plant Biotechnology Journal*. 2012; 10: 646–656.

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.

Hubbard SR, Till JH. Protein tyrosine kinase structure and function. *Annu. Rev. Biochem*. 2000; 69: 373–98.

Hubbard SR., Mohammadi M., Schlessinger J. Autoregulatory mechanisms in protein-tyrosine kinases, *J Biol Chem.* 1998; 15(20):11987-90.

Hyten DL, Choi IY, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang, EY. A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. *Crop Science.* 2010; 50(3): 960p.

IBÁ. Anuário estatístico IBÁ – ano base 2014. Brasília.

Illumina, Inc. 2011. Quality scores for next-generation sequencing. Technical Note: Sequencing (http://res.illumina.com/documents/products/technotes/technote_qscores.pdf).

Ispolatov I, Doebeli M. Speciation due to hybrid necrosis in plant-pathogen models. *Evolution.* 2009; 63:3076–3084.

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8 - A Global View on Proteins and their Functional Interactions in 630 Organisms. PubMed, 2009.

Jiang Z, Wu XL, Zhang M, Michal JJ, Wright-Jr RW. The complementary neighborhood patterns and methylation-to-mutation likelihood Structures of 15,110 single-nucleotide polymorphisms in the bovine genome. *Genetics.* 2008; 180: 639–647.

Jin W, Li N, et al. Identification and verification of microRNA in wheat (*Triticum aestivum*). *Journal of Plant Research.* 2008; 121: 351-355.

Jones JD, Dangl JL. The plant immune system. *Nature.* 2006; 444: 323–329.

Jones-Rhoades M, Bartel D, Bartel B. MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology*, Palo Alto, Cambridge. 2006; 57:19-53.

Jung J, Park C. MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and Xoral development in Arabidopsis. *Planta*. 2007; 225: 1327–1338.

Jwa NS, Kumar Agrawal G, Rakwal R, Park CH, Agrawal VP. Molecular cloning and characterization of a novel jasmonate inducible pathogenesis-related class 10 protein gene, JIOsPR10, from rice (*Oryza sativa* L.) seedling leaves. *Biochem Biophys Res Commun*. 2001; 286:973–983.

Kanazin V, Marek LF, Shoemaker RC. Resistance gene analogs are conserved and clustered in soybean. *Proc Natl Acad Sci USA*. 1996; 93: 11746–11750.

Kasprzewska A. Plant chitinases - Regulation and function. *Cellular & Molecular Biology Letters*. 2003; 8: 809 – 824.

Kawano T. Roles of the reactive oxygen species-generating peroxidase reactions in plant defense and growth induction. *Plant Cell Rep*. 2003; 21(9): 829-37.

Keane PJ, Kile GA, et al. Diseases and pathogens of eucalypts: CSIRO. 2000

Keller I, Bensasson D, Nichols RA. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genetics*. 2007; 3: 22p.

Kharabian-Masouleh A, Waters DLE, Reinke RF, Henry RJ. Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. *Plant Biotechnol*. 2011; 9: 1074–1085.

Khraiwesh BJ, Zhu k, et al. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim Biophys Acta*. 2011; n2 p.137-48.

Kinoshita N, Wang H, Kasahara H, Liu J, MacPherson C, Machida Y, Kamiya Y, Matthew A, Chua N. IAA-Ala Resistant3, an Evolutionarily Conserved Target of miR167, Mediates Arabidopsis Root Architecture Changes during High Osmotic Stress. *The Plant Cell*. 2012; 24: 3590–3602.

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

Ko J, Prassinis C, Han K. Developmental and seasonal expression of PtaHB1, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). *New Phytologist.* 2006; 169: 469–478.

Ko JH, Prassinis C, et al. Developmental and seasonal expression of PtaHB1, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). *New Phytol.* 2006; 169(3): 469-78.

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.

Kruglyak L. The use of a genetic map of biallelic markers in linkage studies. *Nature genetics.* 1997; 17(1): 21-4.

Kruszka K, Pieczynski M, et al. Role of microRNAs and other sRNAs of plants in their changing environments. *J Plant Physiol.* 2012; v.169, 16: 1664-72.

Kulheim C, Yeoh SH, Maintz J, Foley WJ, Moran GF. Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics.* 2009; 10: 452p.

Kuwabara C, Takezawa D, Shimada T, Hamada T, Fujikawa S, Arakawa K. Bscisic acid- and cold-induced thaumatin-like protein in winter wheat has an antifungal activity against snow mould, *Microdochium nivale*. *Physiol Plant.* 2002; 115: 101–110.

Lai E, Riley J, Purvis I. et al. A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. *Genomics*. 1998; 54(1): 31-38.

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: 25p.

Lawrence CB, Singh NP, Qui J, Gardner RG, Tuzun S. Constitutive hydrolytic enzymes are associated with polygenic resistance of tomato to (*Alternaria solani*) and may function as an elicitor release mechanism. *Physiol. Mol. Plant Pathol*. 2000; 57: 211-220.

Lebel S, Schellenbaum P, Walter B, Malliot P. Characterisation of *Vitis vinifera* PR10 mutigene family. *BMC Plant Biol*. 2010; 10:184p.

Lee Y, Kim M, Han JJ, Yeom KH, Lee S, Baek SH, Kim VN. MicroRNA genes are transcribed by RNA polymerase II. *Embo Journal*, Seoul. 2004; 23: 4051-4060.

Lee, Jin-Sook; Jeremic, Aleksandar; Shin, Leah; Cho, Won Jin; Chen, Xuequn; Jena, Bhanu P. (2012). "Neuronal porosome proteome: Molecular dynamics and architecture". *Journal of Proteomics*. 2012; 75(13): 3952–62.

Lepoittevin C, Frigerio JM, Garnier-Gere P, Salin F, Cervera MT, Vornam B, Harvengt L, Plomion C. In vitro vs in silico detected SNPs for the development of a genotyping array: what can we learn from a non-model species? *PLoS One*. 2010; 5,1103.

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 A, Szwerdszarf D, Abu-Abied M, Mordehaev I, Yaniv Y, Riov J, Arazi T, Sadot E. Profiling microRNAs in *Eucalyptus grandis* reveals no mutual relationship between alterations in miR156 and miR172 expression and adventitious root induction during development. *BMC Genomics*. 2014; 15:524.

Levy D, et al. Genome-wide association study of blood pressure and hypertension. *Nature Genetics*. 2009; 41: 677-687.

- Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Baker B. MicroRNA regulation of plant innate immune receptors. PNAS. 2012; 109(5): 1790–1795.
- Li H, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–2079.
- Li WH, Graur D. Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA. 1991; 284p.
- Li Y, Grit Haseneyer G, Chris-Carolin Scho CC, Ankerst D, Korzun V, Wilde P, Baue E. High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale*L.) genes involved in frost response. BMC Plant Biol. 2011; 11: 6p.
- Liao P, Lee KH. From SNPs to functional polymorphism: The insight into biotechnology applications. Biochemical Engineering Journal. 2010; 149-158.
- Lijavetzky D, Cabezas JA, Rodríguez V, Martínez-Zapater JM. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics. 2007; 8: 424p.
- Lin KC, Bushnell WR, Szabo LJ, Smith AG. Isolation and expression of a host response gene family encoding thaumatin-like proteins in incompatible oat–stem rust fungus interactions. Mol Plant–Microbe Interact. 1996; 9: 511–522.
- Linthorst, H.J.M. Pathogenesis related proteins of plants. Critical Reviews in Plant Science. 1991; 10: 123-150.
- Liu JJ, Ekramoddoullah AKM. The family 10 of plant pathogenesis -related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. 2006; 68: 3-13.

Lobbes D, Rallapalli G, Schmidt DD, Martin C, Clarke J. Serrate: a new player on the plant microRNA scene. *Embo Reports*, Colney. 2006; 7:1052- 1058.

Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J. *Molecular cell biology*. New York: W.H. Freeman and CO. 2004; 5: 66–72.

Lorito M, Woo SL, Fernandez IG, Colucci G, Harman GE, Pintor-Toro JA, Filippone E, Muccifora S, Lawrence CB, Zoina A, Tuzun S, Scala F. Genes from mycoparasitic fungi as a source for improving plant resistance to fungal pathogens. *Proc. Natl. Acad. Sci.* 1998; 95: 7860-7865.

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.

Lu C, Tej SS, et al. Elucidation of the small RNA component of the transcriptome. *Science*. 2005; 309: 1567-1569.

Lu S, Sun YH, et al. MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. *Plant J.* 2007; 51(6), 1077p.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembien LA, Berka J, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437: 376-380.

Marx GA. Developmental mutants in some annual seed plants. *Ann. Rev. Plant Physiol.* 1983; 34: 389-417.

Mauch F, Mauch-Mani B, Boller T. Antifungal hydrolases in pea tissue. II. Inhibition of fungal growth by combinations of chitinase and β -1,3-glucanase. *Plant Physiol.* 1988; 88: 936-942.

Megraw M, Baev V, Rusinov V, Jensen ST, Kalantidis K, Hatzigeorgiou AG. MicroRNA promoter element discovery in *Arabidopsis*. *Rna*. a Publication of the Rna Society, Pennsylvania. 2006; 12: 1612-1619.

Michelmore RW, Paran I, Kesseli RV. 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*. 1991; 88: 9828-9832.

Mitsuhara I, Iwai T, Seo S, Yanagawa Y, Kawahigasi H, Hirose S, Ohkawa Y, Ohashi Y. Characteristic expression of twelve rice PR1 family genes in response to pathogen infection, wounding, and defense-related signal compounds. *Mol Gen Genom*. 2008; 279(4): 415–427.

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 A, Garcia CH. A cultura do eucalipto no Brasil. São Paulo: Sociedade Brasileira de Silvicultura, 2000 .112 p.

Morin RD, Aksay G, et al. Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Res*. 2008; 18(4), 571p.

Moriyama EN, and Powell JR. Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution*. 1996; 13: 261–277.

Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics*. 2008; 92: 255–264.

Moxon S, Schwach F, Dalmay T, Maclean D, Studholme DJ, Moulton V. A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics*. 2008; 24: 2252–2253.

Mur LA, Kenton P, Lloyd AJ, Ougham H, Prats E. The hypersensitive response; the centenary is upon us but how much do we know? *J Exp Bot.* 2008; 59: 501–520.

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.

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L e Minobe Y. Search for and Analysis of Single Nucleotide Polymorphisms (SNPs) in Rice (*Oryza sativa*, *Oryza rufipogon*) and Establishment of SNP Markers. *DNA Research.* 2002; 9: 163–171.

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.

Nelson DL; Cox MM. *Lehninger Principles of Biochemistry*. USA: Worth Publishers, 2000).

Neuhaus JM, Fritig B, Linhorst HJM, Meins F Jr, Mikkelsen JD, Ryals J. A revised nomenclature for chitinase genes. *Plant Mol Biol Rep.* 1996; 14: 102–104.

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.

Nielsen KK, Mikkelsen JD, Dragh KM, Bojsen K. An acidic class III chitinase in sugar beet: induction by *Cercospora beticola*, characterization, and expression in transgenic tobacco plants. *Mol. Plant Microbe Interact.* 1993; 6: 495-506.

Nimchuk Z, Eulgem T, Holt BF, Dangl JL. Recognition and response in the plant immune system. *Annual Review of Genetics.* 2003; 37: 579-609.

Nogueira FTS, Madi S, Chitwood D, Juarez M, Timmermans M. Two small regulatory RNAs establish opposing fates of a developmental axis. *Genes & Development*. 2007; 21: 750-755.

Novaes E, Drost DR, Farmerie WG, Pappas-Jr, GJ, Grattapaglia D, Sederoff RR, Kirst M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics*. 2008; 9: 312p.

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

Oh Y, Donofrio N, Pan H, Coughlan S, Brown DE, et al. Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus (*Magnaporthe oryzae*). *Genome*. 2008; 9:85p.

Oliveira ACB, Caixeta ET, Zambolim EM, Zambolim L, Sakiyama NS. Aplicação técnica de marcadores moleculares no melhoramento de plantas:. Campinas: Instituto Agrônômico, 17p. (Documentos IAC, 81, ISSN: 1809 – 7693), 2007.

Ong SS, Wickneswari R. Expression profile of small RNAs in *Acacia mangium* secondary xylem tissue with contrasting lignin content - potential regulatory sequences in monolignol biosynthetic pathway. *BMC Genomics*, 2012; 3:13.

Oraguzie NC. Association mapping in plants. NEW YORK; USA: Springer; 2007, 277p.

Ori N, Cohen AR, Etzioni A, Brand A, Yanai O, Shleizer S, Menda N, Amsellem Z, Efroni I, Pekker I, Alvarez JP, Blum E, Zamir D, Eshed Y. Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. *Nature Genetics*. 2007; 39:787-791.

Osakabe Y, Kajita S, et al. Genetic engineering of woody plants: current and future targets in a stressful environment. *Physiol Plant*. 2011; 142, 2: 105-17.

Park WLJ, Song R, Messing J, Chen X. Nuclear processing and export of microRNAs in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, Madison. 2005; 102:3691-3696.

Paszkiewicz K, Studholme DJ. High-throughput sequencing data analysis software: current state and future developments. In *Bioinformatics for high throughput sequencing* (Rodriguez-Ezpeleta, N., Hackenberg, M. and Aransay, A.M., eds), New York: Springer Science. 2012; 231-248.

Paula RC, Pires IE, et al. "Predição de ganhos genéticos em melhoramento florestal." *Pesquisa Agropecuária Brasileira*. 2002; 37(2): 159-165.

Pavy N, Parsons LS, Paule C, MacKay J, Bousquet J. Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. *BMC Genomics*. 2006; 7: 174p.

Pegg GF. Chitinase from *Verticillium alba-atrum*. *Methods in Enzymology*. 1988; 161: 474-479.

Pieterse CMJ, Van Wees SCM, Van Pelt JA, Knoester M, Laan R, Gerrits H, Weisbeek PJ, Van Loon LC. A novel signaling pathway controlling induced systemic resistance in Arabidopsis. *Plant Cell*. 1998; 10: 1571–1580.

Pinto G, Loureiro J, Santos C. Analises of the genetic stability of *Eucalyptus globulus* Labill. Somatic embryos by flow cytometry. *Theoretical and Applied Genetics*. 2004; 109: 580-587.

Poke FS, Vaillancourt RE, Potts BM, Reid JB. Genomic research in Eucalipto. *Genetica*. 2005; 125: 79-101.

Potts BM, Pederick LA. Morphology, phylogeny, origin, distribution and genetic diversity of the eucalypts. In: B. N. Brown (Ed.). *Diseases and Pathogens of Eucalypts*. Collingwood: CSIRO Publishing. Morphology, phylogeny, origin, distribution and genetic diversity of the eucalypts. 2000; 11-34.

Potts BM. Genetic improvement of eucalypts. Encyclopedia of forest sciences. Oxford UK: Elsevier. 2004; 1480-1490.

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.

Quintin, K, 2001 Queensland Term Wildlife Field Guide: *Eucalyptus grandis* (Flooded gum). Disponível em: <<http://people.hws.edu/fieldguide/delete.asp?ID=120&dlete=no>>. Acesso em: 24 de maio de 2004.

Rajagopalan R, Vaucheret H, et al. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes and Development. 2006; 20: 3407-3425.

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Beckert M, Charmet G. Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome. 2006; 49: 1131–1139.

Reimann C, Dudler R. cDNA cloning and sequence analysis of a pathogen-induced thaumatin-like protein from rice (*Oryza sativa*). Plant Physiol. 1993; 101: 1113–1114.

Reiss E, Horstmann C. Drechslera teres-infected barley (*Hordeum vulgare* L.) leaves accumulated eight isoforms of thaumatin-like proteins. Physiol Mol Plant Pathol. 2001; 58: 183–188.

Resende MDV. Melhoramento de espécies perenes. Recursos genéticos e melhoramento de plantas. L. L. NASS. Rondonópolis, MT, Fundação MT. 2001; 357-421.

Revolti PM. Caracterização morfológica de um mutante em *Eucalyptus* [Tese]. Botucatu, Instituto de Ciências Biológicas, Universidade Estadual Paulista "Júlio de Mesquita Filho", 2015.

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel BE, Bartel DP. Prediction of plant microRNAs targets. *Cell*. 2002; 110: 513-520.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nature biotechnology*. 2011; 29(1): 24-6.

Robischon M, Du J, et al. The *Populus* class III HD ZIP, popREVOLUTA, influences cambium initiation and patterning of woody stems. *Plant Physiol*. 2011; 155(3) 1214-25.

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.

Russell BL, Rathinasabapathi B, Hanson AD. Osmotic stress induces expression of choline monooxygenase in sugar beet and amaranth. *Plant Physiol*. 1998; 116: 859–865.

Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner H, Hunt MD. Systemic Acquired Resistance. *The Plant Cell*. 1996; 8: 1809-1819.

Ryan CA. The systemin signaling pathway: differential activation of plant defensive genes. *Biochimica Biophysica Acta*. 2000; 1477: 112-121.

Savolainen O, Pyhajarvi T. Genomic diversity in forest trees. *Current opinion in plant biology*, [S.l.]2007; 10: 162-167.

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.

Schenk MF, Cordewener JHG, America AHP, Westende WPC, Smulders MJM, Gilissen LJWJ Characterization of PR-10 genes from eight *Betula* species and detection of Bet v 1 isoforms in birch pollen. *BMC Plant Biol.* 2009; 9:24.

Schlumberg A, Mauch F, Vogeli U, Boller T. Plant chitinases are potent inhibitors of fungal growth. *Nature.* 1986; 324: 365-367.

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weissshaar B. Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res.* 2003; 13: 1250–1257.

Schneeberger K, Weigel D. Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci.* 2011; 16: 282–288.

Sela-Buurlage MB, Ponstein AS, Bres-Vloemans SA, Melchers LS, Van Den Elzen PJ M, Cornelissen BJC. Only specific tobacco (*Nicotiana tabacum*) chitinases and β -1,3-glucanases exhibit antifungal activity. *Plant Physiol.* 1993; 101: 857-863.

Selitrennikoff CP. Antifungal proteins. *Appl Environ Microbiol.* 2001; 67: 2883–2894.

Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol Biochem.* 2008; 46: 941–950.

Silva JC. Cresce a presença do eucalipto no Brasil. *Revista da Madeira*, v. 92, p.61-66, 2005.

Simko I, Haynes KG, Jones RW. Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics.* 2006; 173: 2237–2245.

Simon A. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010.

Singh NK, Bracker CA, Hasegawa PM, Handa AK, Buckel S, Hermodson MA, Pfankoch E, Regnier FE, Bressan RA. Characterization of osmotin. A thaumatin-like protein associated with osmotic adaptations in plant cells. *Plant Physiol.* 1987; 85: 529–536.

Souza, A.P. Biologia molecular aplicada ao melhoramento. In: Nass, LL, Valois, ACC, Mello, IS, Valadares-Inglis, MC. (Eds.). Recursos genéticos e melhoramento - plantas. Rondonópolis: Fundação MT, 2001. p. 939-965.

Stearns SC, Hoekstra RF. Evolução: uma introdução. São Paulo: editora Atheneu; 2003.
Storey R, Jones RGW. Betaine and choline levels in plants and their relationship to NaCl stress. *Plant Science Letters*. 1975; 4: 161-168.

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

Suarez V, Staehelin C, Arango R, Holtorf H, Hofsteenge J, Meins JF. Substrate specificity and antifungal activity of recombinant tobacco class I chitinases. *Plant Mol. Biol*. 2001; 45: 609-618.

Sun YH, Shi R, et al. MicroRNAs in trees. *Plant Mol Biol*. 2011; 80, 1: 37-53.

Sunkar R, Girke T, et al. Cloning and Characterization of MicroRNAs from Rice. *Plant Cell*. 2005; 17(5): 1397-1411.

Swoboda I, Dang TCH, Heberle-Bors E, Vicente O. Expression of Bet v 1, the major birch pollen allergen, during anther development. An in situ hybridization study. *Protoplasma*. 1995; 187: 103-110.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, Doerks T, Stark M, Müller J, Bork P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011; 39: 561–568.

Taheri P, Tarighi S. Cytomolecular aspects of rice sheath blight caused by *Rhizoctonia solani*. *Eur J Plant Pathol*. 2011; 129:511–528.

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.

Tang G. Plant microRNAs: an insight into their gene structures and evolution. *Semin Cell Dev Biol.* 2010; 21(8): 782p.

Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, London. 1989; 17: 6463-6471.

Tempel S, Tahi F. A fast ab-initio method for predicting miRNA precursors in genomes. *Nucleic Acids Res.* 2012 40(11), 80p.

Thakur V Wanchana S, et al. Characterization of statistical features for plant microRNA prediction. *BMC Genomics.* 2011; 12, 108.

Thumma B, Pegg G, Warburton P, Brawner J, Macdonell P , Yang X, Southerton S. Molecular Tagging of Rust Resistance Genes in Eucalypts. PlantHealth Australia Ltd. 2012a.

Thumma BR, Sharma N, Southerton SG. Transcriptome sequencing of *Eucalyptus camaldulensis* seedlings subjected to water stress reveals functional single nucleotide polymorphisms and genes under selection. *BMC Genomics.* 2012b; 13:364.

Tor M, Gordon P, Cuzick A, Eulgem T, Sinapidou E, Mert-Turk F, Can C, Dangl JL, Holub EB. Arabidopsis SGT1b is required for defense signaling conferred by several downy mildew resistance genes. *Plant Cell.* 2004; 14: 993–1003.

Tuskan GA, Difazio S, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science.* 2006; 313: 1596-1604.

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

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, et al. A high-resolution, nucleosome position map of (*C. elegans*) reveals a lack of universal sequence-dictated positioning. *Genome research*. 2008; 18(7): 1051-63.

Van K, et al. Discovery of single nucleotide polymorphisms in soybean using primers designed from ESTs. *Euphytica*. 2004; 139: 147-155.

Van Loon LC, Van Strein EA. The families of pathogenesis-related protein, their activities, and comparative analysis of PR-1 type proteins. *Physiol Mol Plant Pathol*. 1999; 55:85–97.

Vaucheret H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Development*, Cold Spring Harbor. 2006; 20:759-771.

Velazhahan R, Muthukrishnan S. Transgenic tobacco plants constitutively overexpressing a rice thaumatinlike protein (PR5) show enhanced resistance to *Alternaria alternata*. *Biol Plant*. 2003; 47: 347–354.

Velazhahan R, Samiyappan R, Vidhyasekaran P. Purification of an elicitor-inducible antifungal chitinase from suspensioncultured rice cells. *Phytoparasitica*. 2000; 28: 131-139.

Vera P, Conejero V. Pathogenesis-related proteins of tomato P-69 as an alkaline endoproteinase. *Plant Physiol*. 1988; 87:58–63.

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.

Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics*. 2002; 34(3): 275-305.

Voinnet O. Origin, biogenesis, and activity of plant microRNAs. *Cell*. Cambridge. 2009; 113: 669-687.

Wan LC, Zhang H, et al. Transcriptome-wide identification and characterization of miRNAs from *Pinus densata*. *BMC Genomics*. 2012; (13): 132.

Wang JW, Park MY, et al. miRNA control of vegetative phase change in trees. *PLoS Genet*. 2011; 7(2): 100p.

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.

Wei LQ, Yan LF, et al. Deep sequencing on genome-wide scale reveals the unique composition and expression patterns of microRNAs in developing pollen of *Oryza sativa*. *Genome Biol*. 2011; 12(6): 53p.

Wicker T, Narechania A, Sabot F, Stein J, Vu GTH, Graner A, Ware, D, et al. Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC genomics*. 2008; 9(1), 518.

Willadino, LG, Cavalcante, UMT, Estresses ambientais: danos e benefícios em plantas, p.95-105, 2005.

Willians JGK, Kubelik AR, Livak KJ, Rafalski JA, Scott V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, London. 1990; 18: 6531-6535.

Wong MM, Cannon CH, et al. Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via de novo transcriptome sequencing. *BMC Genomics*. 2011; v.12, p.342.

Wu G, Park My, Conway SR, Wang J, Weigel D, Poethig RS. The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. *Cell*. 2009; 138: 750-759.

Wu G, Poethig RS. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development*. 2006; 133 (18) 3539p.

Xu H, Reddy ASN. Cloning and expression of a PR5-like protein from *Arabidopsis*: inhibition of fungal growth by bacterially expressed protein. *Plant Mol Biol*. 1997; 34: 949–959.

Xuan P, Guo M, et al. Plant MiRNAPred: efficient classification of real and pseudo plant pre-miRNAs. *Bioinformatics*. 2011; 27(10), p.1368-76.

Yakovlev IA, Fossdal CG, et al. MicroRNAs, the epigenetic memory and climatic adaptation in Norway spruce. *New Phytol*. 2010; v.187, 1154-69.

Yamamoto T, Iketani H, Leki H, Nishizawa Y, Notsuka K, Hibi T, Hayashi T, Matsuta N. Transgenic grapevine plants expressing a rice chitinase with enhanced resistance to fungal pathogens. *Plant Cell Reports*. 2000; 19: 639-646.

Yang T, Xue L, An L. Functional diversity of miRNA in plants. *Plant Science, Limerick*. 2007; 172: 423-432.

Yeh S, Moffatt BA, Griffith M, Xiong F, Yang D S C, Wiseman S B, Sarhan F, Danyluk J, Xue Y Q, Hew C L, Doherty-Kirby A, Lajoie G. Chitinase genes responsive to cold encode antifreeze proteins in winter cereals. *Plant Physiol*. 2000; 124:1251-1263.

Yu B, Yang Zy, Li JJ, Minakhina S, Yang MC, Padgett RW, Steward R, Chen XM. Methylation as a crucial step in plant microRNA biogenesis. *Science, Washington*. 2005; 307:932-935.

Yun D, D'Urzo M P, Abad L, Takeda S, Salzman R, Chen Z, Lee H, Hasegawa P M, Bressan RA. Novel osmotically induced antifungal chitinases and bacterial expression of an active recombinant isoform. *Plant Physiol.*1996; 111: 1219-1225.

Zhang J, Zhou JM. Plant Immunity Triggered by Microbial Molecular Signatures. *Mol Plant.* 2010; 3(5): 783-793.

Zhang L, Zheng Y, et al. Identification and temporal expression analysis of conserved and novel microRNAs in Sorghum. *Genomics.* 2011; 98(6): 460p.

Zhu B, Chen THH, Li PH. Analysis of late blight disease resistance and freezing tolerance in transgenic potato plants expressing sense and antisense genes for an osmotin-like protein. *Planta.* 1986; 198: 70–77.

Zhu T. Expression profile matrix of arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell.* 2002; 14: 559–574.

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB. Single-nucleotide polymorphisms in soybean. *Genetics.* 2003; 163: 1123– 1134.

Zhu JY. et al. Brassinosteroid signalling. *Development* 140. 2013; 1615–1620.

Zotenko E, Mestre J, O'Leary DP, Przytycka TM. Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. *PLoS Comput Biol.* 2008; 4(8): 100-140.