

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

INSTITUTO DE BIOCÊNCIAS

CAMPUS DE BOTUCATU

NATÁLIA TAMAGUSKO MIURA

**Análise filogenética da família multigênica *Bet v 1* em
Eucalyptus grandis.**

MONOGRAFIA

BOTUCATU-SP

2012

Natália Tamagusko Miura

**Análise filogenética da família multigênica *Bet v 1* em
Eucalyptus grandis.**

Monografia apresentada ao Instituto de
Biotecnologia, *Campus* de Botucatu, UNESP,
para obtenção do título de Bacharel em
Ciências Biológicas.

Orientador: Prof. Dr. Celso Luis Marino

Co-orientadora: MSc. Maria Cecília Perantoni Fuchs

BOTUCATU-SP

2012

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: **ROSEMEIRE APARECIDA VICENTE**

Miura, Natália Tamagusko.

Análise filogenética da família multigênica *Bet v 1* em *Eucalyptus grandis* /
Natália Tamagusko Miura. – Botucatu : [s.n.], 2012

Trabalho de conclusão de curso (bacharelado - Ciências Biológicas) -
Universidade Estadual Paulista, Instituto de Biociências de Botucatu

Orientador: Celso Luis Marino

Coorientador: Maria Cecília Perantoni Fuchs

Capes: 20203004

1. *Eucalyptus grandis* – Resistência a doenças e pragas – Aspectos genéticos. 2. Eucalipto – Filogenia. 3. Plantas – Parasito. 4. Proteínas – Análise. 5. Genética vegetal.

Palavras-chave: Análises filogenéticas; *Eucalyptus grandis*; Família multigênica; Máxima Verossimilhança; Proteínas *pathogenesis-related*.

Dedicatória

Dedico este trabalho aos meus pais, Osvaldo e Neide, à minha irmã, Karen, à minha tia, Telma e à minha amiga, Noemia. Agradeço imensamente todo o apoio de vocês e por acreditar na minha força para seguir sempre em frente. Vocês ensinaram-me muito e, com certeza, a pessoa que sou é devido a todo o amor, carinho e compreensão de cada um. Agradeço, no fundo do meu coração, por tudo que fizeram e fazem por mim.

Também dedico esta monografia ao meu companheiro, maravilhoso amigo e namorado. Tudo isso em uma única pessoa: Massami. Muito obrigada pelo amor, carinho e compreensão. Você sempre me ajudou e continua ajudando muito. Sou grata por todos os momentos que passamos juntos e crescer ao seu lado foi, e continua sendo, maravilhoso.

Dedico este trabalho também aos meus amigos, por compartilharem os melhores momentos em Botucatu. Os verdadeiros amigos são os irmãos que a gente escolhe. Por isso, deixo registrado aqui os meus sinceros agradecimentos á Xei, Mim, Zi, Mah, Oxel, Cint e Esqueceu. Vocês sempre me deram incentivo e alegraram os meus dias. Pode ser que um dia nos separemos, mas a amizade é para a vida toda. Aqueles que não estiverem por perto permanecerão dentro de meu coração.

“O senhor... mire, veja: o mais importante e bonito, do mundo, é isto: que as pessoas não estão sempre iguais, ainda não foram terminadas - mas que elas vão sempre mudando. Afinam ou desafinam, verdade maior...” *Grande Sertão: Veredas* de Guimarães Rosa.

Agradecimentos

Á Deus, pela saúde, força e proteção.

Á Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Aos professores e funcionários do Departamento de Genética – Instituto de Biociências de Botucatu, principalmente ao Prof. Dr. Celso Luis Marino, pela oportunidade e ensinamentos.

Ao MSc. Bruno Melo e ao Prof. Dr. Claudio de Oliveira do Departamento de Morfologia - IBB, pelos ensinamentos relacionados à filogenia e pela imensa ajuda para concluir o trabalho.

Á todos do Laboratório CAGEN (Centro de Análises Genéticas e Moleculares), principalmente á MSc. Maria Cecília Perantoni Fuchs pela companhia, apoio e toda a orientação durante o bacharelado. E também pela ajuda, dicas e apoio que recebi de Cíntia, Karine, Lidia, Paola, Vanusa, Helena, Júlio e Leonardo.

“A mente que se abre a uma nova ideia
jamais volta ao seu tamanho original.”

Albert Einstein

Sumário

Resumo	6
Abstract	7
1. Introdução	8
2. Revisão da literatura	11
2.1. <i>Eucalyptus grandis</i>	11
2.2. Família multigênica.....	11
2.3. Família multigênica <i>Bet v I</i>	13
2.4. Análise filogenética de genes.....	13
3. Objetivo	16
4. Material e métodos	17
4.1. Busca de sequências gênicas.....	17
4.2. Construção de árvores filogenéticas de Máxima Verossimilhança.....	17
5. Resultados e discussão	19
6. Conclusão	28
7. Referências bibliográficas	29

Resumo

A caracterização molecular de uma mutação em *Eucalyptus grandis* resultou no desenvolvimento de um marcador SCAR (*Sequence Characterized Amplified Region*) que apresentou identidade com 19 regiões no genoma, que a jusante possuía sequências codificadoras de proteínas da família multigênica *Bet v 1*. Essas proteínas são essenciais na defesa das plantas, pois promovem resistências contra os patógenos através de sínteses de proteínas induzidas por *pathogenesis-related* (PR), o que caracteriza um mecanismo conhecido como hipersensibilidade. Dada a importância de tais proteínas no combate aos patógenos, assim como sua relação com o mutante de *E. grandis*, este trabalho teve como objetivo realizar análises filogenéticas dos genes dessa família multigênica em *E. grandis* e entre demais espécies vegetais. Em *E. grandis* foram encontrados 69 genes pertencentes à família multigênica *Bet v 1*, enquanto nas demais espécies o número de genes variou de 1 a 35. A maior quantidade de genes PRs em *E. grandis* pode conferir uma vantagem adaptativa com relação ao combate contra os patógenos, uma vez que a família multigênica *Bet v 1* está em contínuo processo de duplicação e mutação, garantindo uma coevolução patógeno-hospedeiro e auxiliando na capacidade de adaptação. Os genes da família multigênica *Bet v 1* encontrados, foram utilizados para inferir a história evolutiva dessa família através da construção de árvores de Máxima Verossimilhança entre espécies e em *E. grandis*. As análises de Máxima Verossimilhança demonstraram que a história evolutiva da família multigênica *Bet v 1* não seguiu o mesmo padrão evolutivo que as espécies vegetais, indicando a necessidade da inclusão de genes mais conservados na análise. A representação filogenética em *E. grandis* revela que a similaridade dos genes de uma mesma região é maior em relação aos genes localizados em regiões diferentes, provavelmente devido às sucessivas duplicações e mutações gênicas que ocorre através de processos de neofuncionalização e/ou subfuncionalização. Uma vez que os eventos mais antigos estão fracamente sustentados na representação filogenética de *E. grandis*, foram propostas duas hipóteses para a história evolutiva da família multigênica *Bet v 1* na espécie. Para esclarecer tal história evolutiva, torna-se necessário incluir outros genes, mais conservados, na análise filogenética.

Palavras-chave: *Eucalyptus grandis*, família multigênica, proteínas *pathogenesis-related*, análises filogenéticas, Máxima Verossimilhança.

Abstract

The molecular characterization of a mutation in *Eucalyptus grandis* resulted in the development of a SCAR marker (Sequence Characterized Amplified Region) that showed identity with 19 regions in the genome, which downstream coding sequences multigene family proteins *Bet v 1*. These proteins are essential in plant defenses, because they promote resistance against pathogens through proteins synthesis induced by pathogenesis-related (PR), which features a mechanism known as hypersensitivity. Due the importance of these proteins in fighting against pathogens, as well as its relationship with the mutant *E. grandis*, this study aimed to perform phylogenetic analyzis of the genes in this multigene family *E. grandis* and among other plant species. In *E. grandis* were found 69 genes belonging to multigene family *Bet v 1*, while in the other species the number of genes varied from 1 to 35. The largest amount of PRs genes in *E. grandis* may confer an adaptive advantage with respect to fight against pathogens, since the multigene family *Bet v 1* is in continuous process of duplication and mutation, ensuring a host-pathogen coevolution and supporting adaptability capacity. The gens of the multigene family found *Bet v 1* were used to infer evolutionary history of this family by constructing trees of Maximum Likelihood among species and *E. grandis*. The Maximum Likelihood analyzis demonstrated that the evolutionary history of the multigene family *Bet v 1* didn't follow the same evolutionary pattern as plant species, indicating the need for the inclusion of more conserved genes in the analysis. The phylogenetic representation in *E. grandis* reveals that the similarity of genes from the same region is higher compared to genes located in different regions, probably due to sucessive duplication and gene mutations that occurs through processes neofunctionalization and/or subfunctionalization. Since the oldest events are weakly supported phylogenetic representation of *E. grandis*, two hypotheses were proposed for the evolutionary history of the multigene family *Bet v 1* in the species. To clarify this evolutionary history, it becomes necessary to include other genes, more conserved, in phylogenetic analysis.

Keywords: *Eucalyptus grandis*, multigene family, pathogenesis- related proteins, phylogenetic analyzes, Maximum Likelihood.

1. Introdução

Estudos do grupo CAGEM (Centro de Análises Genéticas e Moleculares) para a caracterização molecular de uma mutação natural em *Eucalyptus grandis*, resultou no desenvolvimento de um marcador SCAR (*Sequence Characterized Amplified Region*) para a anomalia (Tambarussi, 2006; Lourenção, 2010). As plantas anômalas caracterizam-se por apresentar altura reduzida, superbrotamento caulinar, redução da área foliar, alteração do limbo e sobrevivência em poucos meses (Tambarussi, 2006; Lourenção, 2010). A sequência do marcador molecular SCAR desenvolvido apresentou identidade com 19 regiões no genoma, que logo a jusante possuía sequências codificadoras de proteínas com domínio *Bet v I* (proteína PR) com função de resposta a estímulo abiótico (Hooffmann-Sommergruber, 2000) e de defesa (Wen et al., 1997).

A infecção de variadas espécies de plantas por patógenos tais como vírus ou fungos, gera respostas no hospedeiro que vão de leves a graves (Van Loon, 1975). Em infecções sistêmicas, as partículas virais encontram-se espalhadas nas plantas e resulta no aparecimento de vários sintomas de doenças, como cortes necróticos nas folhas, lesões ao redor dos locais de entrada do patógeno (Van Loon, 1985) e uma resposta muito leve não resultam em mudanças significativas nos padrões de expressão dos genes da planta (Pieterse, 2005).

A reação de hipersensibilidade é uma defesa induzida que é comum em várias espécies vegetais tornando-as resistentes a ataques patogênicos subsequentes (Van Loon, 1985). Essa defesa é acompanhada por um grande aumento na expressão de diferentes proteínas, e entre elas, estão membros de um grupo de proteína PR (*pathogenesis-related*) (Van Loon, 1975). As proteínas PRs foram descobertas em cultivares de tabaco, que exibiram respostas de hipersensibilidade por infecção através do vírus do mosaico do tabaco (Gianinazzi et al., 1987; Van Loon et al., 1970). Estas proteínas PRs podem acumular-se sob várias condições ambientais alternadas e, em plantas de tabaco, elas têm sido encontradas no início da floração (Fraser, 1986).

Os genes das proteínas PRs foram inicialmente isolados a partir de plantas de tabaco e à medida que estas proteínas foram detectadas em outras espécies e correlacionadas com a reação de defesa da planta, uma possível função antiviral de proteína PR foi postulada (Pierpoint, W, 1983). Desde então, as proteínas com propriedades físicas semelhantes à de proteína PR têm sido descritas em muitas outras espécies de plantas após a infecção com

vírus, fungos ou bactérias (Van Loon, 1985). Os mutantes de *Arabidopsis thaliana*, com deficiência nos componentes de sinalização do ácido salicílico ou etileno, demonstraram uma maior susceptibilidade às doenças, quando infectados por patógenos específicos, indicando que proteínas PRs têm importante função na resistência às doenças em espécies vegetais (Kunkel e Brooks, 2002; Thomma et al., 2001). Além de tornar as plantas resistentes às doenças, as proteínas PRs possuem atividades enzimáticas, tais como quitinases, glucanases, endoproteínases e peroxidases (Hoffmann-Sommergruber, 2002) e também promovem alterações bioquímicas como aumento de atividades das enzimas envolvidas na via fenilpropanóide (Matthews, 1980).

A correlação do fenômeno de resistência adquirida com a síntese de proteínas induzidas iniciou estudos por diversos grupos de pesquisa para uma melhor compreensão do papel destas proteínas PRs na hipersensibilidade como, por exemplo, o estudo de Van Loon (1975) que se baseia na análise da resposta do hospedeiro e a função das proteínas PRs na resistência adquirida. A ocorrência de PRs em várias plantas, sua expressão em tecidos específicos e sua consistente localização no apoplasto, assim como em compartimento vacuolar, e sua indução por componentes de sinalização endógenos e exógenos, sugerem que as PRs podem ter funções importantes além de seu papel aparentemente limitado na defesa da planta. É possível, que elas sejam também proteínas de estresse direcionadas para aliviar efeitos nocivos dos produtos de degradação celular (Van Loon, 1999).

As proteínas PRs podem ser induzidas artificialmente em plantas sadias pela aplicação de produtos químicos, tais como o ácido salicílico, etileno (White e Ryals, 1997) e ácido abscísico que promovem efeito semelhante ao da infecção patogênica (Stintzi et al., 1993). Em ataques por insetos, as proteínas PRs são induzidas por componentes de sinalização do ácido salicílico ou etileno (Van Loon et al., 2006).

De acordo com sua estrutura primária, relações imunológicas e propriedades enzimáticas, as proteínas PRs são agrupadas em 17 famílias (PR-1 a PR-17), numeradas na ordem em que foram descobertas (Van Loon, 1999). Essas proteínas possuem várias características em comum: são solúveis em tampões ácidos, têm baixos pesos moleculares (de 10kD a 20 kD), ocorre nos espaços intercelulares das folhas e são resistentes a proteases (Van Loon, 1985) e evidências bioquímicas e genéticas sugerem que estas proteínas são codificadas por genes distintos (Antoniw et al., 1980; Van Loon, 1985; Gianinazzi et al., 1983).

Alguns membros da família PR-10 apresentam uma fraca atividade ribonuclease (Bufe et al., 1996). Os genes das PR-10 estão presentes abundantemente em espécies vegetais

(Wen et al., 1997; Liu e Ekramoddoullah, 2004) e correspondem às famílias multigênicas presentes, por exemplo, em *Malus domestica* (Gao et al., 2005). As proteínas PR-10 são também chamadas de intracelulares, uma vez que são expressas no citoplasma, ao contrário da maioria das PRs (Radauer et al., 2008).

O *Eucalyptus* é uma das principais fontes de madeira utilizada para a produção de celulose e é responsável por mais de 65% do volume produzido (Associação Brasileira de Celulose e Papel, BRACELPA). Em relação aos outros países, os programas de melhoramento genético das empresas florestais brasileiras encontram-se avançados e técnicas moleculares são utilizadas para avaliação da variabilidade genética de populações (Leite et al., 2002; Mezzena, 2003), elaboração de mapas genéticos (Verhagen et al., 1996; Brondani et al., 1998, 2002) e mapeamento genético de caracteres quantitativos de interesse na produção (Grattapaglia et al., 1996; Junghans et al., 2003). Tais técnicas moleculares podem ser utilizadas no aumento da resposta de hipersensibilidade a um agente patogênico invasor através da transformação de plantas com gene PR (Pfitzer et al., 1989). No entanto, é necessário que a sequência genética que codifica a proteína PR seja expressa e produzida como uma proteína ou polipeptídeo funcional na célula vegetal. Assim, a transformação compreende plantas contendo tanto proteínas homólogas de genes PR ou proteínas heterólogas de genes PR (Pfitzner et al., 1989). Dessa forma, essa técnica poderia conferir uma melhor resposta de hipersensibilidade ao *Eucalyptus*, uma vez que se trata de uma espécie de interesse econômico. Sabendo-se da importância das proteínas PRs com domínio *Bet v 1* na resistência contra patógenos e seu possível envolvimento na formação de plantas anormais em *Eucalyptus*, este trabalho teve como objetivo obter a caracterização molecular dessa família multigênica *Bet v 1* em *E. grandis*, assim como analisar seus similares em outras espécies vegetais.

2. Revisão da literatura

2.1. *Eucalyptus grandis*

O gênero *Eucalyptus* pertence à família Myrtaceae e foi descrito em 1788 por L' Héritier de Brutelle (Andrade, 1961). Esse gênero é originário da Austrália, com exceção de *E. urophylla* e *E. deglupta* que são naturais do Timor e Papua Nova Guiné, respectivamente (Pryor, 1985), e reúne mais de 900 espécies diferentes (Brooker e Kleinig, 2004).

O eucalipto foi introduzido no Brasil por Edmundo Navarro de Andrade em 1994 com finalidade comercial, inicialmente para produção de lenhas para as locomotivas das estradas de ferro paulista (Wilcken, 2000). No Brasil, o eucalipto encontrou ótimas condições de clima e solos favoráveis com ótimos resultados em ciclos silviculturais, com desenvolvimento mais acelerado em relações aos demais países (Bertola, 2000) e representa mais da metade de toda a área reflorestada no Brasil (Higa et al., 2000).

A madeira do eucalipto é utilizada como lenha ou carvão vegetal para produzir energia e como postes, telhados ou pisos para construção civil (Wilcken, 2000) e, há destaque na produção de celulose em comparação às demais espécies testadas, que não possibilitaram a produção de papel com qualidade satisfatória (Juvenal e Mattos, 2002). O setor florestal brasileiro é um dos mais desenvolvidos e competitivos do mundo e detém uma parcela significativa dos plantios globais, com 7,0 milhões de hectares, sendo 74,8% da área correspondente ao eucalipto de acordo com o Anuário Estatístico da ABRAF, 2012 e cerca de um terço dessa área (2,2 milhões de hectares) corresponde às florestas destinadas para produção de celulose e papel (BRACELPA, 2009). Atualmente, os investimentos em pesquisa estão direcionados para o melhoramento genético das espécies de eucalipto, visando o aumento da produtividade (Ferreira e Grattapaglia, 1998).

2.2. Família multigênica

Existe um grande número de famílias multigênicas localizadas em genomas de eucariotos (Wagner, 2002; Ohta, 2003) que correspondem a vários genes com fragmentos de alta similaridade que se apresentam através de repetições em *tandem*, com sequências espaçadoras que variam em tamanho conforme a espécie (Goes et al., 2011). Como exemplos de famílias multigênicas, há família multigênica do RNA ribossômico, das hemoglobinas,

imunoglobulinas, antígenos de histocompatibilidade, actinas, tubulinas, queratinas, histonas e genes de RNA de transferência (Pinhal, 2007).

Os membros da família multigênica têm origem através de duplicações gênicas ou de variações de algum gene ancestral (Graichen, 2011) e podem estar reunidos nos mesmos cromossomos ou dispersos em cromossomos diferentes (Oliveira et al., 2001). As duplicações e variações gênicas são consideradas as responsáveis pelo aumento no tamanho do genoma e por prover material para o surgimento de novas variedades protéicas, propiciando a diversificação de elementos regulatórios complexos e de padrões de desenvolvimento (Gogarten e Lorraine, 1999). Há vários tipos de duplicações gênicas que se classificam conforme a extensão da região genômica envolvida. Entre elas há a duplicação gênica completa que produz duas cópias idênticas, sendo que as seqüências podem evoluir de diferentes formas: uma das cópias pode originar um pseudogene por apresentar mutações deletérias ou ambas podem ser retidas (Thompson, 2006). A preservação na seqüência de aminoácidos das duas cópias originadas pela duplicação pode acarretar a subfuncionalização, onde cada cópia adota algumas das características do ancestral; ou a neofuncionalização, quando um gene mantém a função original, enquanto o outro adquire uma nova função (Gonzalez- Duarte e Albalat, 2005). A duplicação gênica, de modo geral, pode permitir o surgimento de novos genes e, conseqüentemente, dar origem a novidades fenotípicas (Thompson, 2006). Em *Zea mays*, por exemplo, os dois genes *Adh* diferem em seus padrões de expressão tecido-específicos, onde o gene *Adh 1* é expresso em tecidos da semente e pólen, enquanto *Adh 1* e *Adh 2* são expressos em tecidos de raiz (Thompson, 2006). Variação similar em padrões de expressão é encontrada em outras gramíneas, girassóis, eucaliptos e pinheiros (Gaut et al., 1999).

As pesquisas sobre famílias multigênicas começaram na década de 70, buscando compreender a variabilidade genética gerada, ou existente, em indivíduos de uma população e está relacionada à redundância gênica. (Wagner, 2002; Ohta, 2003). A redundância do material genético é uma das mais peculiares características de todos os genomas, sendo que esta vai aumentando com o tamanho e a complexidade do genoma (Santamaria et al., 2004 apud Thompson, 2006).

2.3. Família multigênica *Bet v 1*

A família multigênica *Bet v 1*, estudada neste trabalho, está relacionada à um marcador SCAR produzido para uma anomalia de desenvolvimento em *Eucalyptus grandis* (Lourenção, 2010). Este marcador possui similaridade com 19 regiões no genoma, que logo a jusante apresenta sequências codificadoras de proteínas com domínio *Bet v1*.

Proteínas da família multigênica *Bet v 1* são alergênicos, causando reações de hipersensibilidade do sistema imunológico (Gajhed et al., 1996). Sua estrutura compreende 6 estruturas secundárias com forma de folhas-beta anti-paralelas e 3 estruturas alfa-hélices (Rosch et al., 1996). Essas proteínas atuam como o principal alérgeno do pólen em *Betula verrucosa* (Breitener, 1992; Valenta et al., 1991), *Alnus glutinosa* (Breitener, 1992; Valenta et al., 1991), *Apium graveolens* (Breitener et al., 1995), *Corylus avelana*, *Carpinus betulus* (Valenta et al., 1991) e *Malus domestica* (Gao et al., 2005). As alergias causadas por esses alérgenos são mediadas por imunoglobulina E (IgE) que afetam uma em cada cinco pessoas na Europa e América do Norte e sintomas incluem febre, dermatite, asma e, em casos mais graves, choque anafilático (Kungl et al., 1996).

Os genes pertencentes à família *Bet v 1* possuem “ORFs” (*Open Reading Frames*) de 465 a 480pb (Wen et al., 1997) e codificam proteínas relacionadas à defesa da planta (Wen et al., 1997). Swoboda et al. (1995) sugeriram um papel importante das proteínas *Bet v 1* durante a germinação do pólen no estigma, protegendo os tecidos reprodutivos femininos de infecção por patógenos. Devido à similaridade das proteínas *Bet v 1* à família 10 das proteínas “*pathogenesis-related*”(PR), estas e suas homólogas foram incorporadas à classe das proteínas PR-10, que são induzidas por estresse, ataque de patógenos e estímulo abiótico (Hoffmann-Sommergruber, 2000).

2.4. Análise filogenética de genes

A filogenia é uma hipótese das relações existentes, as quais são expressas por meio de representação em forma de árvores (Stearns e Hoekstra, 2003). Devido ao desenvolvimento de ferramentas de bioinformática, os estudos de evolução molecular avançaram consideravelmente (Setubal, 2004), sendo a análise filogenética de sequências de DNA, uma importante ferramenta para estes estudos (Nei e Kumar, 1987). As árvores filogenéticas podem ser construídas tanto a partir de dados qualitativos, tais como a presença

ou ausência de uma estrutura morfológica, ou a presença ou ausência de um dos quatro nucleotídeos em uma localidade particular do DNA; ou traços quantitativos, como razões entre os comprimentos das porções corpóreas ou a porcentagem de similaridade entre sequências de DNA (Stearns e Hoekstra, 2003). Sequências repetidas de DNA, como DNAs satélites, minisatélites, microsatélites e genes repetidos em tandem, como as famílias multigênicas, têm sido descritas para muitas espécies e podem ser utilizadas como marcadores cromossômicos úteis em estudos de evolução e de organização genômica (Martins, 2006).

No curso da história evolutiva, os genes e suas regiões regulatórias sofrem variadas modificações, incluindo substituições nucleotídicas, duplicações, recombinações e eventos de reparo (Liberles, 2001) e processos de duplicação gênica, podem dar origem às famílias multigênicas, assim como novidades genéticas ao produzir novos genes (Gonzalez-Duarte e Albalat, 2005; Thompson, 2006).

Os genes ortólogos realizam a mesma função molecular ou celular e podem ser identificados em genomas completos através das identidades entre todos os genes dos organismos analisados (Dominguez, 1999). Os genes parálogos podem realizar a mesma função, conduzindo a redundância funcional (Dominguez, 1999) e apresentam uma extensa semelhança entre suas sequências primárias, o que se pressupõe que há um evento de duplicação em série de um gene ancestral comum (Doyle e Gaut, 2000), seguido por divergência durante o processo evolutivo (Reis e Teixeira, 2006; Fani et al., 2000). Os genes parálogos podem ser duplicados em regiões codificadoras e podem se agrupar em famílias (Thompson, 2006). A redundância gênica dos genes parálogos é importante na evolução por promover inovações genéticas, dessa forma, o entendimento das funções específicas de cada membro de uma família gênica, assim como a comparação das mesmas entre as várias espécies constituem problemas relevantes (Santamaria et al., 2004).

Diferentes modelos de evolução tentam explicar a evolução de famílias multigênicas. Dentre eles, o modelo da evolução divergente implica na divergência gradual de proteínas filogeneticamente relacionadas que, posteriormente, podem adquirir novas funções gênicas (Thompson, 2006). Um segundo modelo, evolução em concerto, abrange todos os membros de uma família multigênica que evoluem conjuntamente, de forma que uma mutação espalha-se por todos os membros da família através de permutação ou conversão gênica. Esse modelo foi utilizado para explicar a evolução de várias famílias, porém, sua aplicabilidade começou a ser questionada (Thompson, 2006). Conseqüentemente, um novo modelo, chamado evolução por nascimento e morte, foi proposto e este modelo, propõem que

novos genes são criados por eventos de duplicação gênica, sendo que alguns são mantidos no genoma durante um longo período, enquanto outros são deletados ou tornam-se não funcionais (Nei e Hughes, 1992).

Muitos genes nucleares são membros de famílias multigênicas (Thompson, 2006). A formação dessas famílias, devido a recorrentes duplicações e a divergências funcionais, têm sido fundamental na evolução de plantas (Thompson, 2006). Vários estudos têm sido conduzidos para avaliar a dinâmica evolutiva das famílias multigênicas em plantas, como por exemplo, as proteínas regulatórias R, as proteínas de ligação à clorofila a e b, entre outras (Fernandes, 2009). A maioria dessas famílias possui diversos *locis* e apresenta uma grande variação no número de cópias entre as espécies (Thompson, 2006). Análises filogenéticas indicam que muito dessa variação pode ser atribuída a duplicações recentes e dessa forma, a dinâmica evolutiva dessas famílias apresenta flutuações no número de cópias através de eventos múltiplos de duplicação e deleção (Morton et al., 1996). No entanto, os processos evolutivos que controlam a estrutura dinâmica de tais famílias são relativamente pouco compreendidos (Small e Wendel, 2000).

3. Objetivo

Realizar análises prévias dos genes da família multigênica *Bet v1* encontradas em *Eucalyptus grandis* e analisar seus similares presentes em outras espécies vegetais, utilizando técnicas da bioinformática molecular. Além dessas análises, interpretar a história evolutiva dos genes envolvidos, utilizando ferramentas da filogenia para entender o processo de diversificação dessa família multigênica.

4. Material e métodos

4.1. Busca de sequências gênicas

Foram realizadas análises da sequência do marcador SCAR, com a ferramenta BLAST (Altschul et al., 1997), no genoma de *Eucalyptus grandis* presente no banco de dados Phytozome (www.phytozome.net). As sequências CDS (*Coding DNA Sequence*) dos genes da família multigênica *Bet v 1* (Pfam:00407) a jusante às regiões com similaridade ao SCAR, foram alinhadas através do *software* MUSCLE (Edgar, 2004) e a partir do alinhamento das sequências foi obtida uma sequência consenso utilizando o *software* Bioedit Sequence Aligment (Hall, 1999).

A sequência consenso obtida foi utilizada para a busca de outros genes da família multigênica *Bet v 1* em *Eucalyptus grandis*, assim como em outras espécies vegetais através da ferramenta BLAST (Altschul et al., 1997) no banco de dados Phytozome (www.phytozome.net). As espécies que demonstraram mais de uma sequência CDS similar ao consenso de *E. grandis* tiveram suas sequências alinhadas pelo *software* MUSCLE (Edgar, 2004) e as sequências consensos de cada espécie foram obtidas utilizando o *software* Bioedit Sequence Aligment (Hall, 1999).

4.2. Construção de árvores filogenéticas de Máxima Verossimilhança

Os consensos e as cópias únicas dos CDS dos genes similares à família multigênica *Bet v 1* das espécies vegetais analisadas foram alinhados pelo *software* MUSCLE (Edgar, 2004), assim como as sequências CDS dos genes similares à família multigênica *Bet v 1* encontradas em *Eucalyptus grandis*, juntamente com um grupo externo (espécie mais próxima ao *E. grandis*). Esses alinhamentos foram utilizados para a construção de árvores de Máxima Verossimilhança para o estudo dos genes desta família multigênica entre as espécies e dentro de *E. grandis*.

As construções das árvores de Máxima Verossimilhança foram realizadas pelo *software* MEGA 5.05 (Tamura et al., 2011). Para o estudo filogenético da família multigênica *Bet v 1* entre as diferentes espécies e dentro de *E. grandis*, foram construídas árvores filogenéticas de Máxima Verossimilhança baseadas no modelo de Kimura-2-Parâmetros (Kimura, 1980) com distribuição *gamma*. Os padrões de divergência foram representados por

um dendograma de Neighbor- Joining (NJ) com 1000 replicações no teste *bootstrap*, realizado no software MEGA 5.05 (Tamura et al., 2011). O modelo de Kimura-2-Parâmetros calcula a média presente nas sequências de nucleotídeos (Ratnasingham e Hebert, 2007), analisa a distância entre as espécies e também os caracteres diagnósticos e determina as distâncias genéticas intraespecíficas e interespecíficas (Ratnasingham e Hebert, 2007), além de levar-se em consideração a taxa de transversão e transição (Ribeiro, 2012). Análise *bootstrap* estima erros estatísticos em situações onde distribuição de uma determinada amostra é desconhecida (Tamura et al., 2011) e o melhor modelo de substituição nucleotídica foi estimado pelo software MEGA 5.05 (Tamura et al., 2011).

O organograma a seguir representa, de forma simplificada, as etapas realizadas durante o trabalho (FIGURA 1).

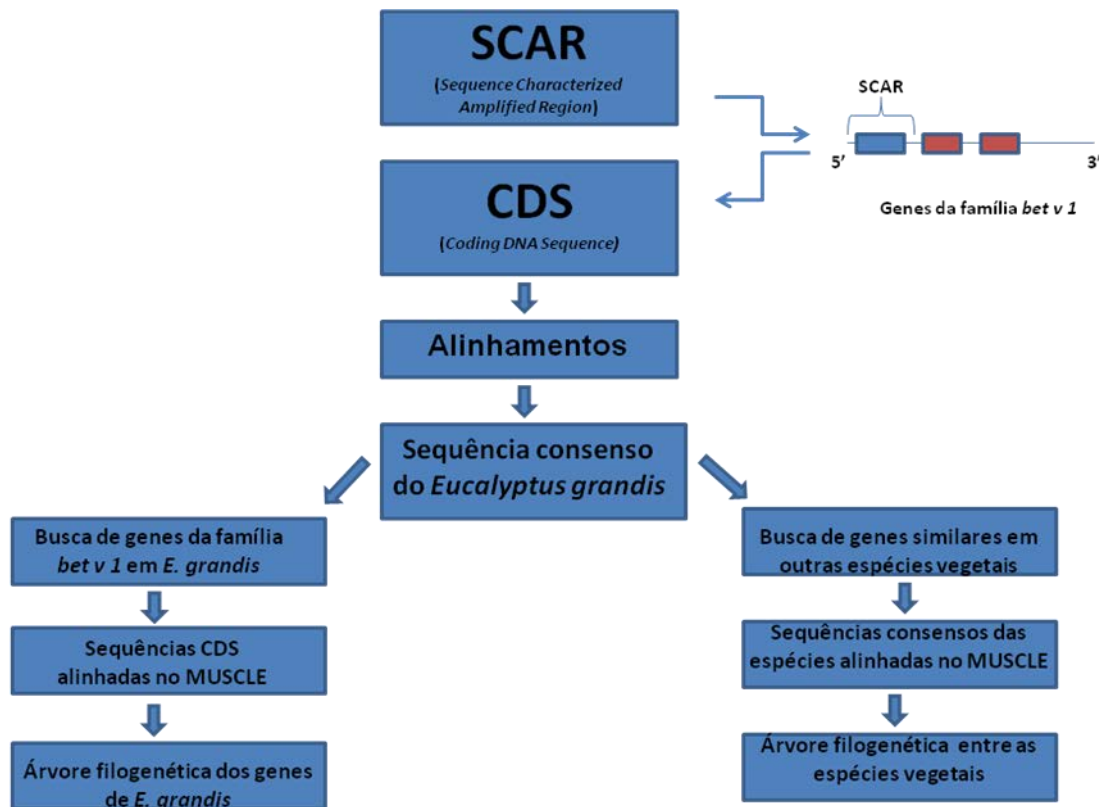


Figura 1: Organograma representando as etapas realizadas durante a execução do trabalho.

5. Resultados e discussão

A caracterização molecular de uma mutação em *Eucalyptus grandis* resultou no desenvolvimento de um marcador SCAR (*Sequence Characterized Amplified Region*) (Lourenção, 2010). Ao analisar a sequência do marcador SCAR no genoma de *Eucalyptus grandis*, foram encontradas 19 regiões com similaridade ao SCAR (FIGURA 2). Dessas 19 regiões, 14 apresentaram sequências relacionadas à proteína patogênica da família *Bet v 1* (Pfam: 00407) a jusante.

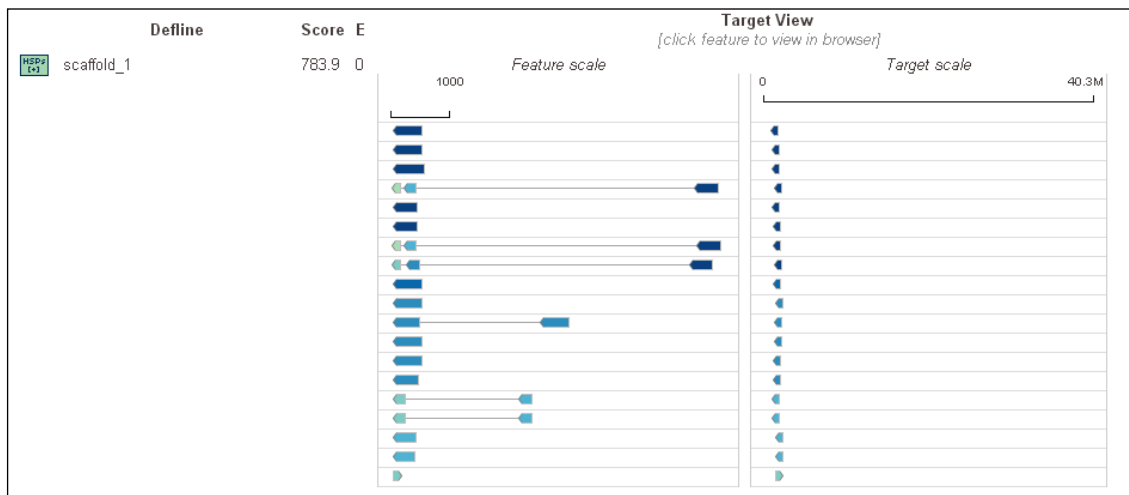


Figura 2: Regiões no genoma de *E. grandis* com similaridades à sequência do marcador SCAR relacionado à uma anomalia em *E. grandis*.

A partir do consenso obtido do alinhamento dos CDSs (*Coding DNA Sequence*) dos genes a jusante das regiões similares à sequência SCAR para anomalia de desenvolvimento em *Eucalyptus grandis* (Lourenção, 2010), foram encontrados 69 genes similares pertencentes à família multigênica *Bet v 1*. Os genes da família multigênica *Bet v 1* codificam proteínas com função de resposta à estímulos abióticos (Hooffmann-Sommergruber, 2000) e de defesa da planta (Wen et al., 1997). De acordo com o banco de dados Phytozome (www.phytozome.net), estes genes em eucaliptos, estão localizados em três *scaffolds* diferentes e se encontram organizados em *tandem*, o que corrobora com o fato destes genes pertencerem a uma família multigênica (Goes et al., 2011).

Esta mesma sequência consenso foi utilizada para a busca de genes similares em outras espécies vegetais pelo Phytozome (www.phytozome.net). Foram encontrados genes similares em 14 espécies com o número de genes variando de 1 a 35 (TABELA 1). Em

eucariotos, as famílias multigênicas podem variar no tamanho e no número de membros (Wagner, 2002; Ohta, 2003). As espécies que demonstraram mais de um gene com similaridade ao consenso de eucalipto, tiveram as sequências CDSs alinhadas (FIGURA 3) e restritas a um consenso para cada espécie. Os consensos e as sequências únicas das espécies foram utilizados para as análises filogenéticas de todos os representantes, assim como as sequências dos genes encontrados em *Eucalyptus grandis*.

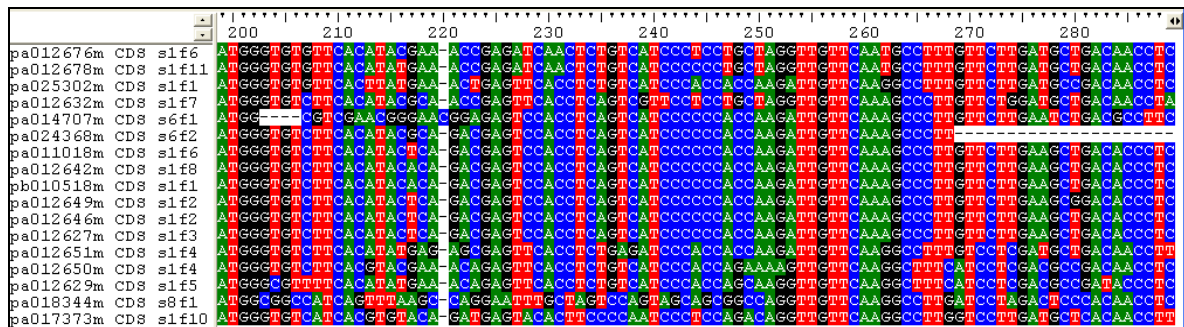


Figura 3: Sequências CDS de genes da família multigênica *Bet v 1* de *Prunus persica* alinhadas para a obtenção de uma sequência consenso única.

Tabela 1: Espécies com genes de sequências similares ao consenso de eucalipto, número de genes encontrados por espécie e porcentagem de similaridade com o consenso de *E. grandis*.

Espécie	Família	Nº de genes	Porcentagens de similaridade
<i>Aquilegia coerulea</i>	Ranunculaceae	1	42,3% a 85,3%
<i>Mimulus guttatus</i>	Scrophulariaceae	4	25,9% a 51,9%
<i>Populus trichocarpa</i>	Salicaceae	11	35% a 97,5%
<i>Carica papaya</i>	Caricaceae	1	34,4% a 68%
<i>Phaseolus vulgaris</i>	Fabaceae	1	57,4% a 82,4%
<i>Citrus sinensis</i>	Rutaceae	5	29,4% a 80,6%
<i>Citrus clementina</i>	Rutaceae	4	35,6% a 73,1%
<i>Prunus persica</i>	Rosaceae	17	39,8% a 84,1%
<i>Malus domestica</i>	Rosaceae	35	18,9% a 100%
<i>Vitis vinifera</i>	Vitaceae	13	34% a 91,2%
<i>Manihot esculenta</i>	Euphorbiaceae	7	29,9% a 77,1%
<i>Ricinus communis</i>	Euphorbiaceae	11	29,1% a 75,3%
<i>Setaria italica</i>	Poaceae	1	39,4% a 94,4%
<i>Selaginella moellendorffii</i>	Selaginellaceae	1	39,4% a 63,5%

Pode-se observar que o *E. grandis* possui um elevado número de genes da família multigênica *Bet v 1* com relação às demais espécies. Assim como em *Populus trichocarpa*, *Malus domestica* e *Vitis vinifera* (Chagne et al., 2012; Myburg et al., 2011; Tuskan et al.,

2006), estudos sugerem que o eucalipto tenha sofrido eventos de duplicação genômica ancestral (Myburg et al., 2011). Tais eventos são frequentes em angiospermas e compreendem de ciclos recorrentes de poliploidização/diploidização (Paterson et al., 2010).

As análises filogenéticas são indicações das relações de ancestralidade (Miyaki et al., 2001; Stearns e Hoekstra, 2003). Na árvore de Máxima Verossimilhança entre as diferentes espécies (FIGURA 4), a *Selaginella moellendorffii* foi utilizada para enraizar a árvore, pois, ela pertence a um grupo mais primitivo de plantas vasculares sem sementes do Filo Lycophyta (Smith et al., 2006). As outras espécies vegetais as quais encontramos genes similares pertencem ao grupo das angiospermas (FIGURA 5).

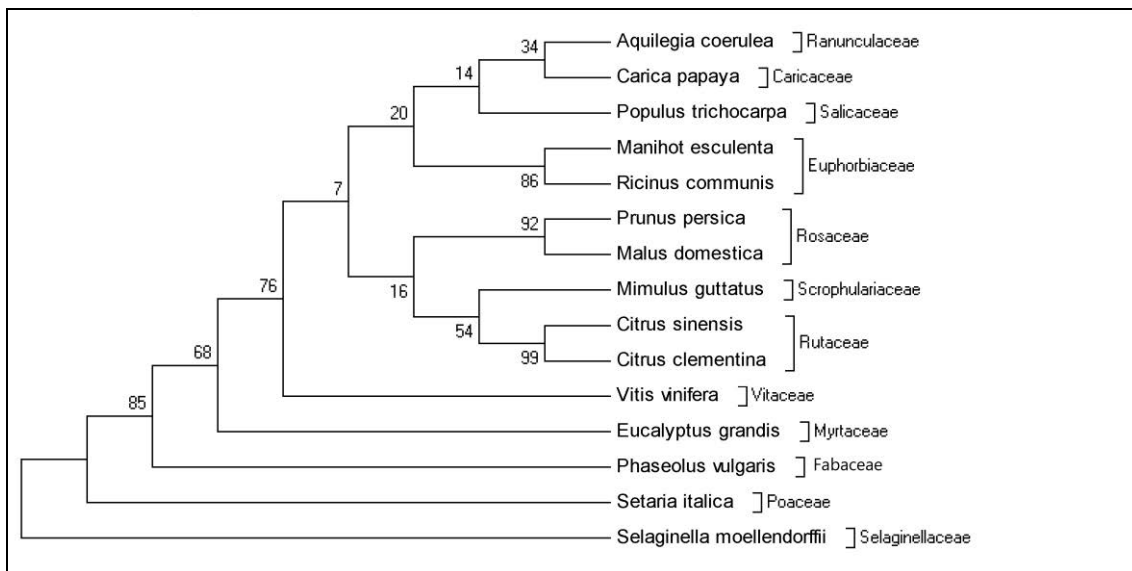


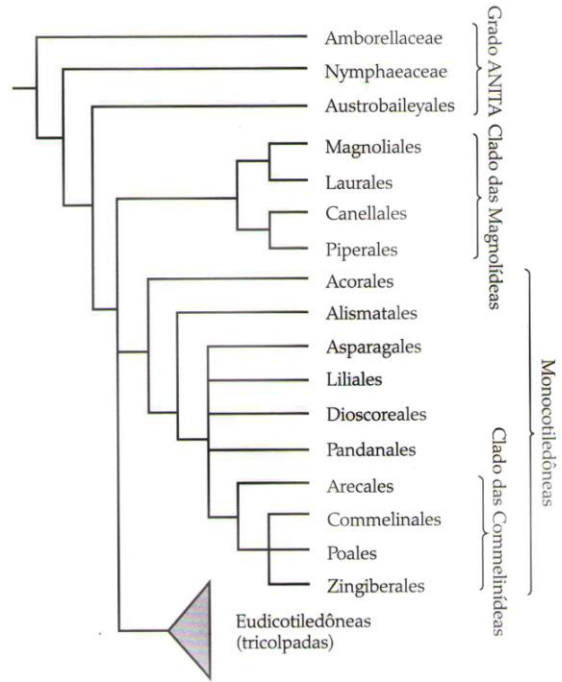
Figura 4: Reconstrução filogenética entre famílias de plantas da divisão Traqueófitas (Judd et al., 2009) utilizando o método de Máxima Verossimilhança.

Conforme pode-se verificar, a história evolutiva da família multigênica *Bet v 1* não seguiu o mesmo padrão de história evolutiva das espécies vegetais (FIGURA 4 e 5). Essa diferença pode ser explicada pelo fato de que as famílias multigênicas são formadas por genes mais variáveis (Hood et al., 1975), ou seja, ao longo do tempo acumulam mais mutações que em genes conservados (Dover e Tautz, 1986). Os genes menos conservados não fornecem o melhor tipo de dado para inferências filogenéticas entre táxons mais divergentes (Figueiredo e Leal, 2008) e, segundo Page e Holmes (1998), a filogenia de organismos pode ser inferida somente se houver um conjunto de genes e não apenas de um único gene, corroborando com a divergência entre as histórias evolutivas da família multigênica *Bet v 1* e das espécies.

Em *Eucalyptus grandis*, os 69 genes encontrados pertencentes à família multigênica *Bet v 1* foram utilizados para inferir a história evolutiva dessa família através da construção de uma árvore de Máxima Verossimilhança (FIGURA 6). Estes genes estão localizados em três regiões (*scaffolds* 1, 8 e 9) diferentes no genoma de *E. grandis*. Juntamente à análise dos genes de *E. grandis*, foi introduzido as sequências consenso das espécies de *Citrus* (grupo Malvídeas, clado Sapindales) (Judd et al, 2009) para que estas sejam utilizadas para enraizar a árvore. O gênero *Citrus* foi escolhido para enraizar a árvore, pois seu grupo taxonômico é considerado como grupo-irmão de Myrtales (clado de *E. grandis*) (eurosídeas II, Jansen et al., 2006 apud Judd et al., 2009) (FIGURA 5).

A análise da reconstrução filogenética dos genes de *E. grandis* (FIGURA 6), demonstrou clados fortemente sustentados com relação aos eventos mais recentes. Já os nós de eventos mais antigos estão fracamente sustentados, indicando que os genes desta família multigênica estão em processo contínuo de mutação. Uma vez que os genes estudados estão relacionados à patogenicidade, tal processo pode ser em decorrência a uma melhora adaptativa, conhecida por coevolução patógeno-hospedeiro (Camargo, 2002). Esse processo de coevolução ocorre quando as espécies influenciam na evolução uma da outra e há uma pressão seletiva mútua e contínua entre elas de tal forma que o ambiente de cada uma das espécies se modifica conforme as competições pela sobrevivência. Isso acaba por criar novos mecanismos de adaptação (Ridley, 2006) que pode se transformar em uma “corrida armamentista” (Camargo, 2002). Pressupõe-se que o hospedeiro e o patógeno coevoluem através de substituições gênicas sucessivas e que mutações ocorrem na mesma taxa em ambos (Stearns e Hoekstra, 2003). Segundo Vermeij (1987, 1999) e Camargo (2002), a relação patógeno-hospedeiro apresenta um padrão evolutivo por escalada, sugerindo que o patógeno cria novos mecanismos de ataque enquanto o hospedeiro responde com criação de novos mecanismos de defesa. A coevolução patógeno-hospedeiro é um processo antagônico e pode levar a um equilíbrio dinâmico conhecido como “Rainha Vermelha” (Ridley, 2006). Essa hipótese da “Rainha Vermelha” se refere à obra literária de Lewis Carrol em *Alice no país das maravilhas*: “é preciso correr tanto quanto se consegue para ficar no mesmo lugar” e a “corrida armamentista” é uma analogia às mudanças coevolutivas (Ridley, 2006).

ANGIOSPERMAS



ANGIOSPERMAS EUDICOTILEDÓNEAS (TRICOLPADAS)

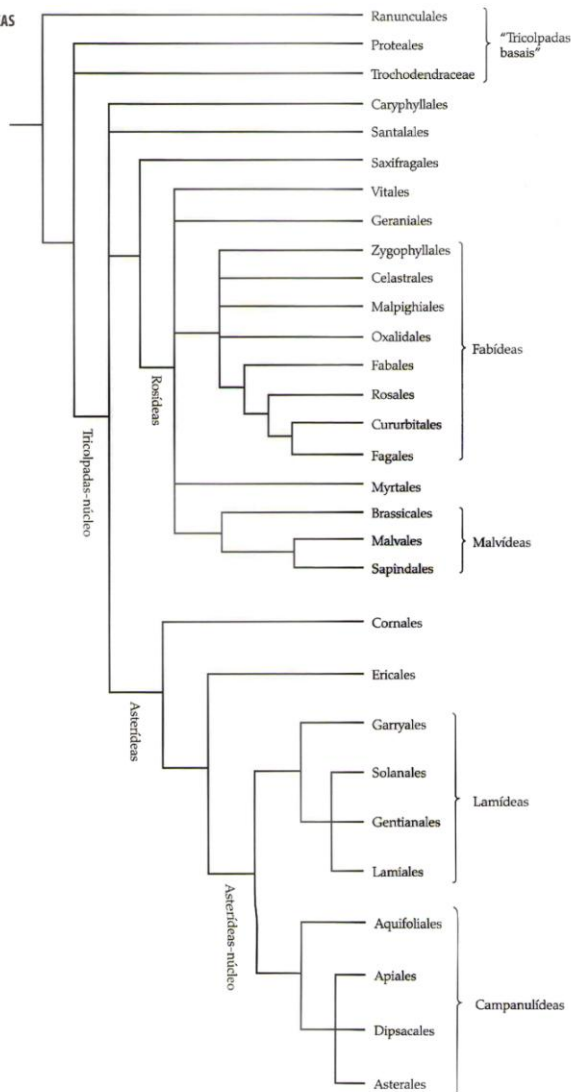


Figura 5: Cladogramas mostrando as relações simplificadas entre os principais grupos de angiospermas (Judd et al., 2009).

As duplicações genômicas ancestrais, juntamente com os recorrentes processos evolutivos, possuem um papel importante na adaptação dos organismos (Paterson et al., 2010). Uma vez que os diferentes ambientes possuem uma ampla variedade de patógenos, a grande diversidade dos genes da família multigênica *Bet v 1*, a qual pertence ao grupo de proteínas PRs (*pathogenesis-related*), pode conferir vantagem na defesa e adaptação. A alta capacidade de adaptação do gênero *Eucalyptus* em variados ambientes (Bertola, 2000) pode estar correlacionada com o grande número de genes da família multigênica *Bet v 1*, auxiliando-os na defesa contra o estresse biótico e, conseqüentemente, na adaptação.

De acordo com os grupos fortemente sustentados, podemos inferir que os genes do *scaffold 8* são ancestrais em relação aos genes do *scaffold 9*, pois no clado sustentado com 94% verificou-se a presença de genes do *scaffold 9* e apenas um gene do *scaffold 8* (FIGURA 6). Os genes do *scaffold 8* também estão presentes em outros clados com genes do *scaffold 1* ou apenas genes do *scaffold 8* (FIGURA 6).

Com relação à similaridade gênica, a análise de Máxima Verossimilhança sugere que os genes de uma mesma região (*scaffold*) são mais próximos que os genes das outras regiões. A maior similaridade de genes localizados em uma mesma região genômica pode ser explicada por sucessivas duplicações gênicas com possíveis processos de neofuncionalização e/ou subfuncionalização. As duplicações gênicas produzem cópias idênticas de um mesmo gene e estas podem progredir de diferentes formas: originar um pseudogene através de mutações deletérias; adquirir a subfuncionalização, onde as duas cópias preservam as sequências de aminoácidos, mas cada cópia adota algumas das características do ancestral; ou a neofuncionalização, no qual um gene mantém a função original e o outro adquire nova função (Gonzalez- Duarte e Albalat, 2005).

Para esclarecer as questões mais antigas da história evolutiva, seria necessário incluir outros genes mais conservados na análise. A utilização de genes com menores taxas de mutação, associados aos genes PRs pode ser mais eficiente para um melhor entendimento sobre o surgimento de vários genes similares em *E. grandis*. Dessa forma, de acordo com a análise da reconstrução filogenética foi possível propor duas hipóteses sobre a evolução desses genes em *E. grandis*.

A primeira hipótese (FIGURA 7) sugere que o gene ancestral está localizado no *scaffold 8*. Este gene teria sofrido duplicações e subseqüentes mutações em suas cópias. Uma ou mais cópias podem ter sido transferidas para o *scaffold 1*, e neste sofrido, novamente, duplicações e mutações, nas quais parte dos genes do *scaffold 1* manteve-se similar ao

original (*scaffold 8*). Esse processo pode ser representado pelos cladogramas sustentados em 68% e 49% que possuem, respectivamente, genes do *scaffold 1* e genes dos *scaffolds 1 e 8* (FIGURA 6).

O surgimento dos genes no *scaffold 9* seria um evento mais recente, sendo explicado pelo reduzido número de genes que são muito similares aos do *scaffold 8* (94% de sustentabilidade). Da mesma forma que ocorreu no *scaffold 1*, um ou mais genes do *scaffold 8* foram transferidos para o *scaffold 9* e neste sofrido recentes duplicações e mutações.

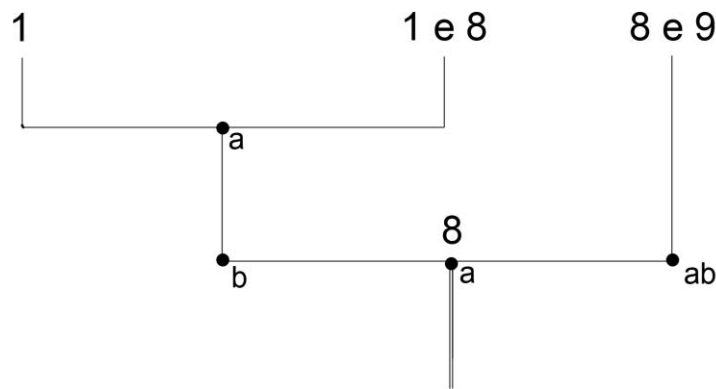


Figura 7: Esquema representativo da história evolutiva dos genes da família multigênica *Bet v 1* em *E. grandis*, de acordo com a hipótese na qual o gene ancestral estaria no *scaffold 8*. (1) genes do *scaffold 1*; (8) genes do *scaffold 8*; (9) genes do *scaffold 9*; (a) duplicações e subsequentes mutações; (b) transferência de um ou mais genes do *scaffold 8*.

A segunda hipótese (FIGURA 8) sugere que o gene ancestral esteja localizado no *scaffold 1*, o qual poderia ter sofrido duplicações e subsequentes mutações em suas cópias. Uma ou mais cópias deste *scaffold* podem ter sido transferidas para o *scaffold 8*. As duplicações e mutações ocorreram concomitantemente em ambos *scaffolds*, assim apenas parte dos genes dos *scaffolds 8 e 1* são similares. Na árvore filogenética (FIGURA 6), esse processo é representado pelos cladogramas sustentados com 68%, 99% e 49% que possuem, respectivamente, genes do *scaffold 1*, genes do *scaffold 8* e genes dos *scaffolds 1* agrupados com os genes do *scaffold 8*. Assim como explicado na hipótese anterior, os genes do *scaffold 9* teriam surgido a partir de um ou mais genes do *scaffold 8*.

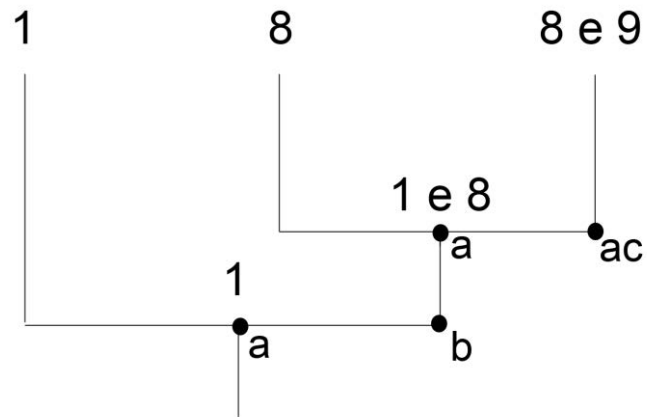


Figura 8: Esquema representativo da história evolutiva, de acordo com a hipótese na qual o gene ancestral da família multigênica *Bet v 1* em *E. grandis* estaria no *scaffold 1*. (1) genes do *scaffold 1*; (8) genes do *scaffold 8*; (9) genes do *scaffold 9*; (a) duplicações e subsequentes mutações; (b) transferência de um ou mais genes do *scaffold 1*; (c) transferência de um ou mais genes do *scaffold 8*.

6. Conclusão

De acordo com as representações filogenéticas, pode-se inferir que a história evolutiva da família multigênica *Bet v 1* não segue o mesmo padrão evolutivo que as espécies vegetais. O mesmo padrão poderia ser obtido acrescentando-se um representante de cada grupo vegetal na análise filogenética, o que não foi possível devido ao reduzido número de genomas vegetais sequenciados. Além disso, a inclusão de mais genes pode contribuir para uma melhor resolução filogenética.

Na representação filogenética dos genes de *Eucalyptus grandis*, pode-se observar que apenas os clados de eventos mais recentes estão fortemente sustentados. Isto sugere que os genes da família multigênica *Bet v 1* estão em contínuo processo de evolutivo, ou seja, possuem alta taxa de mutação, conferindo grande variabilidade desses genes em *E. grandis*. O grande número de genes de *E. grandis* (69) com relação às demais espécies (1 a 35) pode estar associado à alta capacidade de adaptação do eucalipto em diferentes ambientes, conferindo vantagem adaptativa ao estresse biótico. Além disso, o *E. grandis* possui alta taxa de heterozigose devido à sua domesticação recente, o que garante alelos diferentes para um mesmo gene.

7. Referências bibliográficas

- ABRAF, anuário estatístico 2012. Disponível em: <<http://www.abraflor.org.br>>. Acesso em: 29 de set.
- ALTSCHUL, S.F.; MADDEN, T.L.; SCHÄFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, v.25, n.17, p.3389-3402, 1997.
- ANDRADE, E. D. O eucalipto. 2.ed. S. Paulo: Jundiaí, 1961, p.50-64.
- ANTONIW, J. F; RITTER, C. E.; PIERPOINT, W. S; VAN LOON, L. C. Comparasion of three pathogenesis-related proteins from plants of two cultivars of tobacco infected with TMV. *J. Gen. Virol*, n.47, p.79-87.
- BERTOLA, A. Eucalipto – 100 anos de Brasil. Setor de Inventário Florestal: V&M Florestal Ltda., 2000, 89p.
- BRACELPA, relatório anual 2009. Disponível em: <<http://www.bracelpa.org.br>>. Acesso em: 29 de set.
- BREITENEDER, H.; FERREIRA, F.; REIKERSTORFER, A.; DUCHEN, M.; VALENTA R.; HOFFMANN-SOMMERGRUBER , K.; EBNER, C.; BREITENBACH, M.; KRAFT, D.; SCHEINER, O. Complementary DNA cloning and expression in *Escherichia coli* of *Aln g 1*, the major allergen in pollen of alder (*Alnus glutinosa*). *Journal of Allergy and Clinical Immunology*, v.90, p.909-917, 1992.
- BREITENEDER, H; HOFFMANN-SOMMERGRUBER, K.; O'RIORDAIN, G.; SUSANI, M.; AHORN, H.; EBNER, C.; KRAFT, D.; SCHEINER, O. Molecular characterization of *Api g 1*, the major allergen of celery (*Apium graveolens*) and its immunological and structural relationships to a group of 17kDa tree pollen allergens. *European Journal of Biochemistry*, v.233, p.484-489, 1995.
- BRONDANI, R. P. V.; BRONDANI, C.; GRATTAPAGLIA, D. Toward a genus-wide reference linkage map for *Eucalyptus* basead exclusively on highly informative microsatellite markers. *Molecular Genetics and Genomics*, v.267, p.338-347, 2002.
- BRONDANI, R. P. V.; BRONDANI, C; TARCHINI, C., GRATTAPAGLIA, D. Development characterization and mapping of microsatellite markers in *eucalyptus grandis*. *Theoretical and Applied Genetics*, v.97, p.816-827, 1998.
- BROOKER, M. I. H.; KLEINING, D. A. Field guide to *Eucalyptus*, 2.ed., vol.3, Melbourne, Australia: *Bloomings Books*, 2004.

- BUFE, A.; SPANGFORT, M.D.; KAHLERT, H.; SCHLAAK, M.; BECKER, W.M. The major birch pollen allergen, *Bet v 1*, shows ribonuclease activity. *Planta*, v.199, p.413-415, 1996.
- CAMARGO, L. E. A. Manual de Fitopatologia, Universidade Estadual do Sudoeste da Bahia, v.1, 2002.
- CHAGNE, D.; LIN-WANG, K.; ESPLEY, R V.; MARRA, M.; SANDBERG, G.; ROKHSAR, D. An ancient duplication of apple MYB factors is responsible for novel red fruit-flesh phenotypes. *American Society of Plant Biologists*, p. 112.206771, 2012.
- DOMINGUEZ, A.; Anotação de genomas e bases de dados de fungos. Universidade de Salamanca, 1999. Disponível em: <<http://www.esa.ipb.pt>> Acessado em: 15 de set.
- DOVER, G.A.; TAUTZ, D. Conservation and divergence in multigene families: alternatives to selection and drift. *Biological Sciences*, v.312, n.1154, p.275-289, 1986.
- DOYLE J. J., GAUT B. S. Evolution of genes and taxa: a primer, *Plant Mol Biol*, v.42, p. 1-23, 2000.
- EDGAR, R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity *BCM Bioinformatics*, v.5, p.113, 2004.
- FERNANDES, D. Morfogênese *in vitro* em tomateiro e berinjela e silenciamento gênico da sintase do *mio*-inositol-fosfato por RNAi em tomateiro. Dissertação de Mestrado-Universidade Federal de Viçosa, 2009.
- FERREIRA, M.E.; GRATTAPAGLIA, D. Introdução ao uso de marcadores moleculares em análises genéticas. 3 ed. Brasília: Embrapa-Cenargen, 1998, 220p.
- FIGUEIREDO, H. C. P.; LEAL, C. A. G. Tecnologias aplicadas em sanidade de peixes. *Revista Brasileira de Zootecnia*, Viçosa, v.37, 2008.
- FRASER, R.S.S. Plant physiol pathol in tobacco mosaic virus induced in abscisic acid concentration in tobacco leaves, v. 168, n.4, p. 592-598, 1986.
- GAJHEDE, M.; OSMARK, P.; POULSEN, F. M.; IPSEN, H; LARSEN, J. N; JOOST, R. J.; SCHOU, C.; LOWENSTEIN, H. X-ray and NMR structure of *Bet v 1*, the origin of birch pollen allergy. *Nature structural biology*, v.3, p. 1040-1045, 1996.
- GAO, Z.S.; VAN de WEG, W.E.; SCHAART, J.G.; SCHOUTEN, H.J.; TRAN, D.H.; KODDE,L.P.; VAN der MEER, I.M.; VAN der GEEST, A.H.; KODDE, J.; BREITENEDER, H.; HOFFMANN-SOMMERGRUBER , K.; BOSCH, D.; GILISSEN, L.J. Genomic cloning and linkage mapping of the *Mal d 1* (PR-10) gene family in apple (*Malus domestica*). *Theoretical and Applied Genetics*, v.111, n.1, p.171-183, 2005.

- GAUT B. S., PEEK, A. S., CLEGG, M.T., Patterns of genetic diversification within the *Adh* gene family in the grasses (Poaceae), *Mol. Biol. Evol*, v. 16, p.1086-1097, 1999.
- GIANINAZZI, S.; VAN LOON, L.C.; VAN KAMMEN; MARTIN, C.; VALLEE, J. C. Pathogenesis-related proteins are chitinases. *Virology*, v.84, p. 6750-6754, 1987.
- GIANINAZZI, V. P; The physiology of vesicular-arbuscular mycorrhizal roots. *Plant Soil*, v.71, p. 197-209, 1983.
- GOES, P. R. N.; MONTEIRO, M. A.; OLIVEIRA, A. V.; GUEDES, A. A. S. Padronização das condições de amplificação da técnica rDNA 5S, para obtenção de marcadores moleculares espécie-específicos em populações de *Hypostomus* do Rio Ivaí. *Iniciação Científica Cesumar*, v.13, n. 1, p.31-35, 2011.
- GOGARTEN J. P., LORRAINE, O. Orthologs, paralogs and genome comparisons, *Curr Op Genet & Developm*, v. 9, p.630-636, 1999.
- GONZALES- DUARTE, R., ALBALAT, R. Merging protein, gene and genomic data: the evolution of the MDR-ADH family, *Heredity*, v.95, p. 184-197, 2005.
- GRAICHEN, D. A. S. Antígenos B de *Echinococcus*: instabilidade genômica, variação no verme adulto e anticorpos policlonais. 2011. 159f. Dissertação de Doutorado- Universidade Federal do Rio Grande do Sul, 2011.
- GRATTAPAGLIA, D.; BERTOLUCCI, F.L.G.; PENCHEL, R.; SEDEROFF, R.R. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics*, v. 144, p. 1205 -1214, 1996.
- HALL, T. H. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Oxford University Press: *Nucleic Acids Symposium Series*, n.41, p.95-98, 1999
- HIGA, R. C. V.; MORA, A. L.; HIGA, A. R. Plantio de eucalipto na propriedade rural. Embrapa Florestas: Curitiba, 1. ed., 2000, 31p.
- HOFFMANN-SOMMERGRUBER, K. Pathogenesis-related (PR)-proteins identified as allergens. *Biochemical Society Transactions*, v.30, p.930-935, 2002.
- HOFFMANN-SOMMERGRUBER, K. Plant allergens and pathogenesis-related proteins. What do they have in common? *International Archives of Allergy and Immunology*, v.122, p.155-166, 2000.
- HOOD, L.; CAMPBELL, J. H., ELGIN, S. C. R. The organization, expression, and evolution of antibody genes and other multigene families. *Annual review of Genetics*, v. 9, p.305-353, 1975.

- JUDD, W. S.; CAMPBELL, C.; KELLOGG, E.; STEVENS, P F.; DONOGHUE, M. J. *Sistemática vegetal: um enfoque filogenético*. 3.ed., Porto Alegre: Artmed, 2009.
- JUNGHANS, D.T.; ALFENAS, A.C.; BROMMONSCHENKEL, S.H.; ODA, S.; MELLO, E.J.; GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. *Theoretical and Applied Genetics*, v.108, n.1, p.175-180, 2003.
- JUVENAL, T. L.; MATTOS R. L. G. O setor de celulose e papel: *Banco nacional do desenvolvimento* (BNDES), 2002.
- KIMURA, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Molecular Evol.* v. 16, p. 111-120, 1980.
- KUNGL, A. J., KRAFT, D., LIDEMANN, A., SUSANI, M., BREITENBACH, M., SCHEINER, O., AUER, M., MACHIUS, M., VISSER, A.J. Evidence for an alpha helical T cell epitope in the C-terminus of the mais birch pollen allergen *Bet v 1*. *Biochem. Biophys. Res. Commun.* v.223, p. 187-192, 1996.
- KUNKEL, B.N.; BROOKS, D.M. Cross talk between signaling pathways in pathogen defense. *Current Opinion in Plant Biology*, v.5, p.325-331, 2002.
- LEITE, S. M. M.; BONINE, C. A.; MORI, E. S.; VALLE, C. F.; MARINO. Genetic variability in a breeding population of *Eucalyptus urophylla*, *Silvae Genetica*, v.51, n.5-6, p.253-256, 2002.
- LIBERLES D. A., Evaluation of methods for determination of a reconstructed history of gene sequence evolution, *Mol Bio Evol*, v.18, p.2040-2047, 2001.
- LIU, J.J.; EKRAMODDOULLAH, A.K.M. Characterization, expression and evolution of two novel subfamilies of *Pinus monticola* cDNAs encoding pathogenesis-related (PR)-10 proteins. *Tree Physiology*, v.24, n.12, p.1377-1385, 2004.
- LOURENÇÃO, J.C. Caracterização de uma região genômica relacionada a uma anomalia de viveiro em *Eucalyptus*. 2010. 96f. Dissertação de Mestrado- Instituto de Biociências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, 2010.
- MARTINS, C. UNIVERSIA in Notícias: pesquisadores da UNESP- Botucatu tentam explicar mecanismos evolutivos através de genoma de peixes, 2006.
- MATTHEWS, R. E. F. Host plant responses to virus infection. In: *Comprehensive Virology: virus-host interactions*. New York: Plenum Press, 1980, p. 297-359.

- MEZZENA, L. Monitoramento da variabilidade genética de populações de melhoramento de *Eucalyptus grandis* por marcadores RAPD e SSR. 2003. 68f. Dissertação de Mestrado-Instituto de Biociências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, 2003.
- MYBURG, A.; GRATTAPAGLIA, D.; TUSKAN, G.; JENKINS, J.; SCHUMUTZ, J.; MIZRACHI, E.; HEFER, C.; STERCK, L.; HAYES, R.; ROKHSAR, D. The *Eucalyptus grandis* genome project: genome and transcriptome resources for comparative analysis of woody plant biology. IUFRO Tree Biotechnology Conference, Bahia-Brasil, 2011.
- MORTON, B. R., GAUT, B. S., CLEGG, M. T. Evolution of alcohol dehydrogenase genes in the Palm and Grass families, *Proc Natl Acad Sci USA*, v. 93, p. 11735- 11739, 1996.
- NEI, M., HUGHS, A. L. Balanced polymorphism and evolution by the birth-and-death process en the MHC loci. In: TSUJI, K., AIZAWA, M., SASAZUKI, T. *Histocompatibility workshop and conference*. Oxford Univ. Press, Oxford, UK, pp. 27-38, 1992.
- NEI, M., KUMAR, S. Molecular Evaluation and Phylogenetics, 2000. Disponível em: <<http://mep.asu.edu/text/overview.shtml>> Acessado em: 23 de agost.
- OHTA, T. Gene families: multigene families and superfamilies. *Encyclopedia of the human genome*, 2003.
- OLIVEIRA, N.; BATISTA, A. N.; RIGDEN, D. J.; SILVA, R. O. FRAGOSO, R. R.; GOMES, E. A.; MONNERAT, R. G.; GROSSI, M. F. Clonagem de uma família multigênica de proteinase serina em bicudo do algodoeiro (*Anthonomus grandis*) In: *Encontro do talento estudantil da Embrapa recursos genéticos e biotecnologia*, Brasília, 2001, 31p.
- PAGES, R. D. M; HOLMES, E. C. Molecular evolution: a phylogenetic approach. *Blackwell Science*, Oxford, U. K., 1998.
- PATTERSON, A. H.; FREELING, M.; TANG, H.; WANG, X. Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology*, v.61, p.349-372, 2010.
- PIERPOINT, W. S., Trends in Biochem: pathogenesis related protein in plants, 1983.
- PIETERSE, Indução de resistência sistêmica por rizobactérias e comunicação na rota de sinalização, *RAPD*, v.13, p. 277-295, 2005.
- PFITZNER, U. M.; PFITZNER, A.P.; GOODMAN, H. M. Pathogenesis related proteins in plants, 1989.

- PINHAL, D. DNA ribossomal 5S: sequência nucleotídica, organização genômica e potencial como biomarcador para análises moleculares em espécies de tubarões. 2007. 71f. Dissertação de Mestrado- Instituto de Biociências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, 2007.
- PRYOR, L. *Eucalyptus*. In: *Halevy, A. CRC Handbook of Flowering*. 1.ed. Boca Raton: Franklin Book Company, v.2, 1985, p. 476-482.
- RADAUER, C.; LACKNER, P.; BREITENEDER, H. The *Bet v 1* fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evolutionary Biology*, v.8, p.286, 2008.
- RATNASINGHAM, S.; HEBERT, P. D. N.; BOLD: the barcode of life data system. *Mol. Ecol.*, v.7, p.355-367, 2007.
- REIS, V M.; TEIXEIRA, K. R. S Fixação biológica de nitrogênio- estado da arte. Embrapa, 2006.
- RIBEIRO, A. O. Identificação molecular de peixes marinhos das regiões Sudeste e Sul do Brasil com ênfase no estado de São Paulo. 2012. 78f. Dissertação de Mestrado- Instituto de Biociências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, 2012.
- RIDLEY, M. Evolução. São Paulo: editora Artmed, 3ª edição, p.633-644, 2006.
- ROSCH P., KRAFT, D., FABER, C., LIDEMANN, A., STICHT, H., EJCHART, A., KUNGL, A. J., SUSANI, M., FRANK, R. M., BREITENBACH, M. Secondary structure and tertiary fold of the birch pollen allergen *Bet v 1* in solution. *J. Biol. Chem.* v.32, p. 19243-19250, 1996.
- SANTAMARIA, M., LANAVE, C., SACCONI, C. The evolution of the adenine nucleotide translocase family, *Gene*, v.333, p.51-59, 2004.
- SETUBAL, J. C. Conselho de Informações sobre Biotecnologia: genômica. São Paulo: Editora Atheneu, p. 105, 2004.
- SMALL, R. L., WENDEL J. F. Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). *Genetics*, v.155, p.1913-1926, 2000.
- SMITH, A. R.; PYERS, K. M.; SCHUETTPELZ, E.; KORALL, P.; SCHNEIDER, H.; WOLF, P. G. A classification of extant ferns. *Taxon*, v.55, p.705-731, 2006.
- STEARNS, S. C., HOEKSTRA, R. F. Evolução: uma introdução. São Paulo: editora Atheneu, p.240-241, 2003.

STINTZI, A.; HEITZ, V.; PRASAD, V.; WIEDEMANN-MERDINOGLU, S.; KAUFFMANN, S.; GEOFROY, P.; LEGRAND, M.; FRITIG, B. Plant 'pathogenesis-related' proteins and their role in defense against pathogens. *Biochimie*, v.75, 687-706, 1993.

SWOBODA, I.; DANG, T.C.H.; HEBERLE-BORS, E.; VICENTE, O. Expression of *Bet v 1*, the major birch pollen allergen, during anther development. An in situ hybridization study. *Protoplasma*, v.187, p.103-110, 1995.

TAMBARUSSI, E. V. Associação de marcador RAPD e desenvolvimento de marcador SCAR para um tipo de anomalia de viveiro em *Eucalyptus grandis*. 2006. 31f. Monografia (Bacharelado em Engenharia Florestal)- Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista "Júlio de Mesquita Filho", Botucatu, 2006.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M. KUMAR, S. MEGA 5: molecular evolutionary genetics analysis using Maximum Likelihood, evolutionary distance, and Maximum Parsimony Methods. *Mol Biol Evol*, 2011.

THOMMA, B.P.H.J; PENNINCKX, I.A.M.A.; BROEKAERT, W.F.; CAMMUE, B.P.A. The complexity of disease signaling in Arabidopsis. *Current Opinion in Immunology*. v.13, p.63-68, 2001.

THOMPSON, C. E., Divergência funcional da família gênica da álcool desidrogenase em plantas. 2006. 122f. Dissertação de mestrado- Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

TUSKAN, G. A; DIFAZIO, S.; JANSSON, S.; The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, v.313, n.5793, p.1596-1604, 2006.

VALENTA, R; BREITENEDER, H; PETTERNBURGER, K; BREITENBACH, M; RUMPOLD, H; KRAFT, D; SCHEINER, O. Homology of the major birch-pollen allergen, *Bet v 1*, with the major pollen allergens of alder, hazel, and hornbeam at the nucleic acid level as determined by cross-hybridization. *Journal of Allergy and Clinical Immunology*, v.87, n.3, p.677-682, 1991.

VAN LOON, L. C. *Occurrence and properties of plant pathogenesis-related proteins*. In: *Pathogenesis-Related Proteins in Plants*. Boca Raton: CRC Press, 1999, p.1-19.

VAN LOON, L. C. Pathogenesis-related proteins. *Plant Molecular Biology*, v. 4, p.111-116, 1985.

VAN LOON, L. C. Similarity of qualitative changes of specific proteins after infection with different viruses and their relationship to acquired resistance. *Virology*, v. 67, p.566-575, 1975.

VAN LOON, L.C.; REP, M.; PIETERSE, C.M. Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology*, v.44, p.135-162, 2006.

VAN LOON, L. C.; VAN KAMMEN, A. Polyacrylamide disc electrophoresis of the soluble leaf proteins from *Nicotiana tabacum* var. "Samsun" and "Samsun NN". Changes in protein constitution after infection with tobacco mosaic virus. *Virology*, v.40, p.190-211, 1970.

VERHAEGEN, D.; PLOMION, C.; GION, J.M.; POITEL, M.; COSTA, P.; KREMER, A. Quantitative trait dissection analysis in Eucalyptus using RAPD markers, 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression cross different ages. *Theoretical and Applied Genetics*, v. 95, p. 597- 608, 1997.

VERMEIJ, G. J. *Evolution and Escalation*. Princeton University Press, 1987.

VERMEIJ, G. J. Inequality and the directionality of history. *American Naturalist*, 1999.

WAGNER, A. Selection and gene duplication: a view from the genome, *Genome Biology*, v.3, p. 1012.1-1012.3, 2002.

WEN, J.; VANEK-KREBITZ, M.; HOFFMANN-SOMMERGRUBER, K.; SCHEINER, O.; BREITENEDER, H. The Potential of *Bet v 1* Homologues, a Nuclear Multigene Family, as Phylogenetic Markers in Flowering Plants. *Molecular Phylogenetics and Evolution*, v.8, n.3, p.317-333, 1997.

WHITE, R. F.; RYALS, J. A. Inducible herbicide resistance. *Virology*, 1997.

WILCKEN, C. F.; LIMA, A. C V.; DIAS, T. K. R.; MASSON, M. V; FILHO, P. J F.; POGETTO, M. H. F. A. D. Guia prático de manejo de plantações de eucaliptos. FEPAF, 2008, 19p