

**TESE DE DOUTORADO**

**DESCOBERTA E ANÁLISE DA VARIABILIDADE**

**INTERESPECÍFICA**

**DE PEQUENOS RNAs VIA SEQUENCIAMENTO DE NOVA**

**GERAÇÃO EM *EUCALYPTUS***

**Marília de Castro Rodrigues Pappas**

**Brasília - DF**

**Março, 2013**

**UNIVERSIDADE DE BRASÍLIA**

**Instituto de Ciências Biológicas**

**Departamento de Biologia Celular**

**DESCOBERTA E ANÁLISE DA VARIABILIDADE**

**INTERESPECÍFICA**

**DE PEQUENOS RNAs VIA SEQUENCIAMENTO DE NOVA**

**GERAÇÃO EM *EUCALYPTUS***

**Orientador: Dario Grattapaglia**

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

**BRASILIA – DF**  
**Março, 2013**

# TERMO DE APROVAÇÃO

Tese apresentada ao Departamento de Biologia Celular da Universidade de Brasília, como requisito parcial para obtenção de grau de Doutor em Ciências Biológicas, área de concentração Biologia Molecular. Tese defendida e aprovada em 08/03/2013 por:

---

Dr. Dario Grattapaglia - Orientador

Embrapa Recursos Genéticos e Biotecnologia

---

Dr. Rogério Margis

Universidade Federal do Rio Grande do Sul

---

Dra. Simone da Graça Ribeiro

Embrapa Recursos Genéticos e Biotecnologia

---

Dra. Ildinete Silva Pereira

Universidade de Brasília

---

Dr. Júlio Carlyle M. Rodrigues

Embrapa Recursos Genéticos e Biotecnologia

Dedico este trabalho aos meus grandes amores, Nicolas e Georgios.

Plagiando o meu orientador, colega de trabalho e amigo Dario Grattapaglia, Nicolas é a maior experiência genética já realizada por mim e pelo Georgios. Esse trabalho não é nada perto dele mas é, sem dúvida, por ele. Como ele diz, “o livro da mamãe” ficou pronto.

## **AGRADECIMENTOS**

Agradeço a minha mãe, que desde sempre me apoiou, deu força e incentivo não importando o que eu resolvesse fazer.

Ao meu pequeno príncipe Nicolas que teve paciência de aguentar a mãe trabalhando muito e sem tempo para ele, especialmente nas suas férias de verão de 2013.

À minha família que me ajudou a ter mais tranquilidade para trabalhar enquanto eles entretinham o Nicolas, especialmente, tio Sílvio e tia Simone e os primos André e Bernardo, vovó Suely, Iaiá Nancy e o papai sempre presente.

Ao meu “co-orientador” informal, conselheiro e, acima de tudo, parceiro de todas as horas, Georgios. Sem ele, essa tese não seria possível. Sem palavras para agradecer tanta dedicação.

Ao meu orientador, Dario Grattapaglia, pelo apoio e confiança incondicionais de sempre.

A Simone Ribeiro, cuja ajuda foi inestimável na implementação do *northern blot* e, entre uma hibridização e outra, acabou se tornando mais que uma colaboradora, uma amiga querida.

Ao Dr Scott Poethig, que gentilmente cedeu seus protocolos para o *northern blot*, compartilhou dados e foi sempre tão acessível.

A Alessandra Reis que foi a responsável pelo sequenciamento Illumina na Fasteris. E ao Laurent pela gentileza e disponibilidade.

Aos colegas do laboratório de Genética Vegetal da Embrapa Recursos Genéticos e Biotecnologia pela convivência, apoio e momentos de descontração. Especialmente, à Neide e ao Marcão, sempre disponíveis e dispostos a resolver qualquer problema e ajudar em tudo.

Aos colegas do LPPIII da Embrapa Recursos Genéticos e Biotecnologia pela convivência e apoio, em especial, à Lígia por me socorrer quando precisei.

|   |    |
|---|----|
| Introdução.....   | 5  |
| O gênero <i>Eucalyptus</i> .....                                    | 5  |
| Pequenos RNAs.....  | 8  |
| Motivação.....  | 9  |
| Objetivos e Metas.....  | 11 |
| Capítulo I.....   | 12 |
| Revisão Bibliográfica.....  | 13 |
| Pequenos RNAs não codificadores.....                                | 13 |
| Pequenos RNAs em plantas.....                                       | 15 |
| Descoberta de miRNAs em plantas.....                                | 19 |
| MiRNA em arbóreas.....  | 21 |
| Metodologia.....  | 24 |
| Material Biológico.....   | 24 |
| Extração de RNA.....  | 27 |
| Processamento dos dados de sequenciamento.....                      | 27 |
| Validação experimental.....   | 28 |
| Resultados e discussão.....   | 30 |
| Sequenciamento de pequenos RNAs.....                                | 30 |
| Análise geral de sequências.....                                    | 30 |
| Sequências mais abundantes.....                                     | 35 |
| A abundância da sequência não implica classificação como miRNA..... | 41 |
| Análise comparativa da abundância.....                              | 48 |
| Mapeamento contra o genoma de referência.....                       | 49 |
| Famílias conservadas de miRNAs.....                                 | 52 |
| Compatibilidade estrutural.....                                     | 55 |
| Predição de alvos de regulação por miRNAs.....                      | 58 |
| Validação experimental - <i>Northern blot</i> .....                 | 60 |
| Validação experimental – microarranjo.....                          | 65 |
| Dados consolidados de sequenciamento.....                           | 71 |

|  |     |
|--|-----|
| Conservação com <i>Eugenia uniflora</i> .....          | 74  |
| Conclusões.....  | 76  |
| Capítulo II.....                                       | 79  |
| Introdução.....  | 80  |
| Revisão bibliográfica.....                             | 81  |
| Origem dos RNAs em fase .....                          | 81  |
| Rede de regulação dos RNAs em fase .....               | 83  |
| Os RNAs em fase e genes de resistência a doenças ..... | 86  |
| Metodologia.....                                       | 89  |
| Sequenciamento .....                                   | 89  |
| Mapeamento e identificação de siRNAs em fase .....     | 89  |
| Predição de Alvos .....                                | 89  |
| Resultados.....  | 90  |
| Conclusões.....  | 101 |
| Perspectivas futuras .....                             | 103 |
| Anexo I.....   | 107 |
| Anexo II.....  | 116 |
| Bibliografia.....                                      | 121 |

## **RESUMO**

Há pouco mais de uma década foi descrita a função de mais uma classe de elementos regulatórios chamados de micro RNAs – miRNAs. Inicialmente descritos como reguladores pós transcricionais envolvidos no processo desenvolvimento, estes já tiveram papel validado em uma ampla gama de processos como germinação, determinação de morfologia e respostas a estresse biótico e abiótico. A descrição desta classe de reguladores em *Eucalyptus* contribui com o conhecimento científico deste importante e vasto gênero inserindo mais um relevante elemento na anotação do recente genoma de referência de *Eucalyptus grandis*.

Para a identificação em escala genômica de miRNAs em *Eucalyptus*, foi realizado um sequenciamento de pequenos RNAs em larga escala em amostras de folha e xilema em desenvolvimento nas duas principais espécies plantadas do gênero – *E. grandis* e *E. globulus*. Com o objetivo de otimizar o processo de descoberta de locos de pequenos RNAs, as amostras de folha e xilema de *E. grandis* usadas no sequenciamento foram da mesma árvore BRASUZ1, usada no sequenciamento do genoma de referência de *Eucalyptus*. O sequenciamento permitiu, além da descoberta de pequenos RNAs, uma análise da variabilidade interespecífica pelo mapeamento das sequências de ambas as espécies contra o genoma de *E. grandis*.

A caracterização *in silico* das sequências obtidas com o sequenciamento obedeceu a uma sequência de passos iniciada pela busca, por similaridade de sequência, visando a identificação de sequências de miRNAs de famílias conservadas em outras espécies e já descritas anteriormente. Em seguida, a identificação de potenciais novos miRNAs obedeceu a uma sequência de critérios já descritos. As sequências de pequenos RNAs foram mapeadas fisicamente no genoma identificando assim os potenciais locos dos genes MIR e permitindo a validação *in silico* pela análise da estrutura secundária compatível com os parâmetros de precursores, satisfazendo a primeira premissa para a validação de um novo miRNA.

Os números totais obtidos mostram claramente a cautela que deve existir na análise de dados de sequenciamento de pequenos RNAs. Dentre os grandes números obtidos, parte é representada por contaminantes como RNA ribossômico e de cloroplasto. Dentre os pequenos RNAs regulatórios, diferentes classes e subclasses estão representadas e os critérios para declarar uma sequência de pequeno RNA como um miRNA devem ser cuidadosamente observados.

A validação experimental das sequências de potenciais novos miRNAs foi, inicialmente, realizada via *northern blot* de algumas das sequências mais expressas em cada uma das amostras sequenciadas. Uma validação em larga escala foi realizada utilizando um microarranjo contendo milhares das sequências putativas de miRNAs descobertas no sequenciamento. Além da análise interespecífica, foram utilizadas amostras de tecidos distintos – folha, xilema, ovário e plântulas – visando uma análise da variabilidade de expressão entre tecidos.

A predição de alvos para os candidatos a novos miRNAs foi realizada *in silico* buscando por sequências que apresentem complementariedade às sequências dos miRNAs utilizando o banco de transcritos de *Eucalyptus* disponível publicamente no site que hospeda os genomas sequenciados pelo JGI ([phytozome.net](http://phytozome.net)).

A análise dos dados de sequenciamento em larga escala revelou aspectos interessantes do repertório de pequenos RNAs. Sequências de 21 nucleotídeos altamente expressas nas amostras sequenciadas foram caracterizadas como pequenos RNAs interferentes (siRNAs) secundários, ou *trans-acting* siRNAs (tasi-RNAs), sintetizados em fase a partir da clivagem de transcritos alvo por miRNAs. A identificação de alvos desses tasiRNAs revelou um mecanismo conservado de regulação da expressão de genes envolvidos na defesa contra patógenos, notadamente, os da família caracterizada pelos domínios protéicos NBS-LRR (*NB-ARC – leucine rich repeat*), e de genes da família PPR (*pentatricopeptide repeat*).

## **ABSTRACT**

A decade ago function of another class of regulatory elements called micro RNAs – miRNAs – was elucidated. Initially described as post-transcriptional regulators involved in development, these elements had been proved to regulate a wide range of processes such as germination, morphology and responses to biotic and abiotic stress. The description of miRNAs in *Eucalyptus* contributes to scientific knowledge of this important and vast genre inserting another important element in the annotation of *Eucalyptus grandis* reference genome.

For genome wide identification of *Eucalyptus* miRNAs, large scale sequencing was performed in leaves and developing xylem in the two most planted species for cellulose production - *E. grandis* and *E. globulus*. In order to optimize the discovery process of small RNAs loci, samples of *E. grandis* leaf and xylem used were from the same tree BRASUZ1, used for the reference genome. Interspecific analysis was conducted based on mapping of the sequences from both species against the genome of *E. grandis*.

*In silico* characterization started by searching for sequence similarity against miRBase conserved miRNAs. The identification of potential new miRNA sequence followed the criteria already described. Sequences of small RNAs derived from large-scale sequencing were mapped against the genome thereby identifying potential MIR loci and enabling validation by *in silico* analysis of secondary structure compatible with miRNA precursor parameters satisfying the first premise for the validation of a new miRNA.

Total number of sequences obtained clearly shows that caution must exist in small RNA sequencing data analysis. Great part of sequences is represented by contaminants such as ribosomal and chloroplast RNA. Among the small regulatory RNAs, different classes and subclasses are represented and the criteria for declaring a sequence of small RNA as a miRNA should be carefully observed.

Experimental validation of potentially new miRNA sequences was first performed via northern blot of some of the more expressed sequences in each of the samples. A large-scale validation was then performed using a microarray containing thousands of sequences

of putative miRNAs discovered in sequencing. Besides interspecific analysis, samples from different tissues - leaf, xylem, ovary and seedlings – were used to evaluate tissue variability.

Target prediction for putative new miRNAs was performed *in silico*. A search for complementary sequences to the miRNAs was performed using *Eucalyptus* transcript database publicly available on the website that hosts the genomes sequenced at JGI (phytozome.net).

Data analysis from large scale sequencing revealed interesting aspects of small RNAs repertoire. Highly expressed 21 nucleotide sequences were characterized as small interfering RNAs (siRNAs), or trans-acting siRNAs (tasi-RNAs), synthesized in phase from cleavage of transcripts targeted by miRNAs. Identification of targets for these tasiRNAs revealed a conserved mechanism regulating the expression of genes involved in defense against pathogens, especially those from the family characterized by NBS-LRR protein domains (NB-ARC - leucine rich repeat) and PPR gene family (pentatricopeptide repeat).

## INTRODUÇÃO

### O GÊNERO *EUCALYPTUS*

Reconhecida como a oitava maior família das angiospermas, a família Myrtaceae possui distribuição, predominantemente, no Hemisfério Sul (Grattapaglia, Vaillancour et al., 2012). Suas árvores ou arbustos apresentam grande diversidade na região tropical, especialmente na América do Sul, Austrália e Ásia tropical, com ocorrências na África e na Europa. A família é comumente encontrada em muitos dos *hot spots* de biodiversidade do mundo, como o sudoeste da Austrália, o Cerrado e a mata Atlântica no Brasil, onde até 90 espécies de Myrtaceae podem ser encontradas por hectare – muitas destas ainda não descritas (Govaerts, Sobral *et al.*, 2008). Um grande número de espécies da família tem importância ecológica e econômica. Os óleos essenciais são um produto economicamente importante, extraído das folhas de espécies de *Eucalyptus* por destilação a vapor. *Melaleuca alternifolia* tem uma elevada concentração de um agente ativo antimicrobiano e antiinflamatório. Além de vários óleos perfumados, como o de *Corymbia citriodora* e componentes mentolados do eucalipto. Algumas espécies de Myrtaceae são frutíferas como goiaba, *Psidium guajava*, que é provavelmente a fruta mais cultivada. Pitanga (*Eugenia uniflora*), jabuticaba (*Myrciaria cauliflora*) e jambo (*Syzygium malaccense*) são outros exemplos de frutas regionais importantes (Wilson, 2011). Muitas espécies são reconhecidas como uma fonte valiosa de madeira, especialmente aquelas do gênero *Eucalyptus*.

O gênero *Eucalyptus* é nativo da Austrália e ocorre desde regiões de alta pluviosidade a semi-áridas, desde o nível do mar a grandes altitudes e tem ampla distribuição geográfica com predominância no hemisfério sul (Ladiges, Udovicic *et al.*, 2003). São árvores perenes de vida longa que constituem o gênero florestal mais plantado no mundo (Ladiges, Udovicic *et al.*, 2003), excedendo 20 milhões de hectares na última avaliação oficial da FAO (Fao, 2007). O Brasil ocupa o primeiro lugar como produtor e exportador de polpa de celulose no mercado internacional. Em 2007, 50% de toda a polpa de celulose de fibra curta negociada no mundo era derivada de *Eucalyptus* e espera-se que,

até 2.020, atinja 60% desse total O eucalipto é, portanto, uma cultura florestal de grande impacto econômico e social. É um gênero chave na produção de polpa de celulose, energia na forma de carvão e fonte de madeira sólida para construção e movelaria com produtividade e qualidade crescente. É interessante destacar que, no Brasil, produtos de base florestal são, majoritariamente, provenientes de florestas plantadas. Papel e celulose tem 100% de sua produção derivada de florestas plantadas e carvão, 68% (Silvicultura, 2003), o que constitui importante papel de floresta de substituição preservando florestas de espécies nativas.

As espécies de *Eucalyptus* tem hábito predominantemente alógamo (Moran, J *et al.*, 1989; Gaiotto, Bramucci *et al.*, 1997) e apresentam, em geral, forte depressão por seleção contra endogamia (Potts e Savva, 1988). Hibridização interespecífica é observada em populações naturais (Pryor e Johnson, 1971; Griffin, Burgess *et al.*, 1988) e esta característica é explorada com sucesso no melhoramento genético em países tropicais onde se aproveita a variação genética natural para crescimento, adaptação e propriedades da madeira entre espécies para consolidar caracteres complementares em clones híbridos (Assis, 2000).

O gênero *Eucalyptus* possui mais de 700 espécies descritas, a maioria pertencente ao subgênero *Symphyomyrtus*. Grande parte das espécies plantadas comercialmente pertence a três seções deste subgênero como *Eucalyptus grandis* e *E. urophylla* (seção *Latoangulatae*), *E. globulus* (seção *Maidenaria*) e *E. camaldulensis* (seção *Exsertaria*) (Grattapaglia e Kirst, 2008). *E. globulus* é uma espécie de climas temperados e se distingue das demais espécies, mais comumente usadas em ambiente tropicais, por sua excelente qualidade da madeira em termos de densidade e qualidade de fibras representando uma importante fonte de variação genética atualmente explorada em híbridos interespecíficos. A espécie é reconhecida por possuir a melhor combinação de propriedades químicas e físicas para produção de papel e celulose dentre as espécies comercialmente plantadas de *Eucalyptus* (Grattapaglia, 2004). Contudo, sua origem em regiões de clima temperado faz com que tenha crescimento lento e pouca adaptabilidade em regiões tropicais. Nestes ambientes e em regiões subtropicais, *E. grandis* é a espécie mais adotada, o que fez desta a espécie florestal mais amplamente plantada em nível mundial. A possibilidade de

hibridização entre espécies dentro do gênero permite a introgressão de alelos para caracteres de interesse a partir de diferentes espécies para *E. grandis*. Notadamente, experimentos de hibridização com *E. globulus* em programas de melhoramento florestal em regiões tropicais tem gerado híbridos com excelente combinação de adaptabilidade e qualidade superior da madeira (Grattapaglia, 2004).

## PEQUENOS RNAs

A existência de RNAs não codificadores, ou que não codificam proteínas, além de RNA transportador e RNAs ribossômicos, é conhecida há, aproximadamente, duas décadas (Lee, Feinbaum *et al.*, 1993). A sua existência e importância não eram reconhecidas, provavelmente, por não se enquadrarem no fluxo do Dogma Central da Biologia. Entretanto, desde a sua descoberta, abriu-se uma nova e fascinante fronteira nas pesquisas biológicas pelo estabelecimento de sua participação fundamental nos processos de regulação dos componentes do Dogma Central. Regiões do genoma consideradas inativas ou sem função, o então chamado “DNA lixo”, na verdade, mostraram ser locais de atividade considerável de pequenos RNAs.

A regulação via pequenos RNAs (*small RNAs* – sRNAs) adiciona mais um componente ao vasto repertório celular de mecanismos de regulação da expressão gênica. À medida que novos sRNAs são caracterizados, avança o entendimento dos mecanismos que governam a expressão gênica e da atuação destes elementos regulatórios e seus alvos em diversos processos biológicos. Apesar de uma fração destes elementos apresentarem alta conservação desde plantas terrestres primitivas, a alta taxa evolutiva observada faz com que grande parte do repertório regulatório de pequenos RNAs seja espécie ou linhagem específica.

Estudos iniciais em *A. thaliana*, utilizando tecnologias de sequenciamento em massa, revelaram que o componente de pequenos RNAs do genoma e seu papel regulatório é maior e mais complexo do que previamente relatado (Wang, Reyes *et al.*, 2004). Portanto, a identificação e caracterização de novos pequenos RNAs, incluindo micro RNAs (miRNAs) e pequenos RNAs interferentes (siRNAs), para diferentes espécies é o primeiro passo para a compreensão destes mecanismos de regulação da expressão gênica.

## MOTIVAÇÃO

Após o desenvolvimento de vários projetos de investigação da porção funcional de genomas e da redução de custo do sequenciamento, milhares de sequências gênicas das mais diversas espécies já foram descritas e estão disponíveis em bancos públicos. O que se observa é que grande parte das sequências gênicas de espécies filogeneticamente próximas é conservada, especialmente, aquelas relacionadas a funções fisiológicas básicas, como desenvolvimento.

No gênero *Eucalyptus*, particularmente, a experiência gerada no âmbito do Projeto Genolyptus (Grattapaglia, 2004) com o sequenciamento de ESTs (*Expressed Sequence Tags*) e fragmentos de genes de cerca de seis espécies distintas mostra altíssima conservação entre sequências codificadoras. A alta similaridade observada nas sequências gênicas das enzimas envolvidas na via de lignificação, por exemplo, no banco de dados de sequências expressas do projeto Genolyptus, bem como em bases de ESTs de diferentes espécies florestais e mesmo de *Arabidopsis*, corrobora esta observação. Acredita-se, portanto, que boa parte da ampla variação fenotípica intra e interespecífica observada no gênero seja decorrente, principalmente, de regulação da expressão gênica mais do que polimorfismos em sequências codificadoras e, portanto, em sequências protéicas, reforçando a relevância do estudo destas moléculas.

A descrição de miRNAs do gênero *Eucalyptus* é uma contribuição para o conhecimento científico na área desses componentes regulatórios do gênero e de grande relevância para análises comparativas com outras espécies florestais e plantas modelo como *Arabidopsis thaliana* e *Populus trichocarpa*. Além disso, espera-se que novas sequências, potencialmente específicas do gênero *Eucalyptus*, sejam descritas. Os conhecimentos assim gerados poderão, futuramente, servir às mais diversas áreas onde a composição química e qualidade da madeira se fazem relevantes: indústrias de papel e celulose, produção de carvão vegetal, como fonte de energia renovável, na construção civil, indústria moveleira e na potencial utilização de lignina e celulose como biomassa para produção de etanol (Lynd, Laser *et al.*, 2008).

Além disso, com o recente sequenciamento do genoma de referência de *Eucalyptus grandis*, a identificação e mapeamento de pequenos RNAs, especialmente miRNAs, contribui com a anotação do genoma caracterizando esses pequenos elementos regulatórios de papel fundamental para a homeostase genômica, desenvolvimento, adaptação a variações ambientais e defesa contra patógenos.

O foco central desta tese envolve, portanto, uma análise ampla do genoma contribuindo para o conhecimento científico de novos elementos regulatórios identificando, em larga escala, pequenos RNAs, especialmente micro RNAs, em espécies plantadas do gênero *Eucalyptus*.

## **OBJETIVOS E METAS**

O objetivo geral deste trabalho foi realizar uma análise genômica ampla para a identificação e caracterização de pequenos RNAs nas duas principais espécies plantadas do gênero *Eucalyptus* e realizar a anotação de miRNAs no genoma de *Eucalyptus grandis*. Para tanto, foram executadas diferentes etapas experimentais, listadas a seguir:

- ✓ Anotação de genes de miRNAs no genoma de *E. grandis*
- ✓ Descoberta experimental de pequenos RNAs de *E. grandis* e *E. globulus* via sequenciamento em larga escala
- ✓ Análise da variabilidade interespecífica da expressão de miRNAs entre as duas principais espécies comercialmente plantadas de *Eucalyptus*
- ✓ Descoberta computacional de potenciais novos miRNAs de *Eucalyptus* por análise da estrutura secundária dos precursores a partir da sequência flanqueante no genoma
- ✓ Desenvolvimento de um microarranjo para validação experimental em larga escala de miRNAs em *Eucalyptus*
- ✓ Identificação *in silico* de mRNAs alvo de regulação pelos miRNAs identificados

## **CAPÍTULO I**

*Descoberta e análise da variabilidade inter-específica de pequenos RNAs por sequenciamento em larga escala em Eucalyptus sp.*

## **REVISÃO BIBLIOGRÁFICA**

### **PEQUENOS RNAs NÃO CODIFICADORES**

Em 1990, Jorgensen e colaboradores (Napoli, Lemieux *et al.*, 1990) procederam a introdução de um gene homólogo de Chalcona sintetase (CHS) em petúnia, visando aumentar a expressão desta enzima limitante na produção de antocianinas, compostos flavonóides responsáveis em grande parte pela coloração de flores. Com a superexpressão do gene, a expectativa era que a coloração das flores se tornasse mais intensa. No entanto, o efeito foi justamente o oposto, com uma dramática redução (de até 50 vezes) na expressão do mRNA de CHS, em alguns casos resultando em flores brancas, completamente sem pigmentação. Apesar de não poder precisar o mecanismo molecular, os autores sugeriram o envolvimento da metilação de DNA como um fator no silenciamento da expressão gênica. O mesmo comportamento foi observado no fungo *Neurospora crassa* quando foi realizada a transformação por plasmídeos com construções para dois genes endógenos envolvidos na síntese de carotenóides. Os fenótipos albinos resultantes indicavam a supressão da expressão gênica por sequências homólogas, sendo este fenômeno denominado “*quelling*” (Romano e Macino, 1992).

Nesses casos, ocorreu o silenciamento gênico induzido por transgenes, fenômeno que na época não tinha as suas bases moleculares compreendidas. O trabalho pioneiro de Fire e Mello, que lhes rendeu o prêmio Nobel em 2006, mostrou que a injeção de RNA fita dupla em indivíduos de *C. elegans* induzia uma potente inibição nos níveis do mRNA endógeno correspondente, o que não acontecia quando fitas simples, tanto a senso quanto a anti-senso, eram injetadas (Fire, Xu *et al.*, 1998). Este fenômeno foi denominado de interferência por RNA (RNAi) e impulsionou as pesquisas para caracterização funcional de pequenos RNAs silenciadores endógenos.

Ambros e colaboradores descreveram o primeiro micro RNA em *Caenorhabditis elegans* (Lee, Feinbaum *et al.*, 1993) que, contudo, só foi reconhecido como uma classe de regulador de expressão gênica uma década mais tarde (Zhang, Pan *et al.*, 2005). Demonstrou-se que o loco que controla o tempo de desenvolvimento larval em *C. elegans*,

lin-4, não codificava uma proteína, mas sim produzia um par de pequenos RNAs. Um deles com cerca de 22 nucleotídeos e outro maior com cerca de 61 nucleotídeos que, acreditava-se, seria o precursor do menor. Notaram, ainda, que esses pequenos RNAs apresentavam complementariedade a múltiplas posições na sequência 3'UTR (*untranslated region* ou região não-traduzida) do gene lin-14, região anteriormente proposta como mediadora da repressão de lin-14 pelo produto do gene lin-4. A investigação deste mecanismo de regulação mostrou que a redução nos níveis da proteína lin-14 produzidos não era acompanhada por uma redução no nível do RNA mensageiro (mRNA) transcrito. Todas essas informações corroboram o modelo onde o RNA lin-4 se liga à região 3'UTR do mRNA de lin-14 reprimindo a sua tradução, como parte dos mecanismos que iniciam a transição do primeiro para o segundo estágio larval (Bartel, 2004). Desde então, o RNA lin-4 é reconhecido como o primeiro membro da classe de pequenos RNA regulatórios denominados micro RNAs – miRNAs.

Pequenos RNAs dirigem silenciamento gênico transcricional e pós-transcricional e, assim, modulam os transcritomas eucarióticos e, conseqüentemente, as proteínas traduzidas a partir destes (Jones-Rhoades, Bartel *et al.*, 2006; Mallory e Vaucheret, 2006; Rajagopalan, Vaucheret *et al.*, 2006). A edição da revista *Nature Genetics* de junho de 2006 foi inteiramente dedicada aos miRNAs, tratando o tema como uma “Micro revolução na Biologia”. A repercussão acerca do tema chegou à imprensa popular, sendo tema de capa da revista *The Economist*, que anunciou pequenos RNAs como o “Big Bang da Biologia” (Matranga e Zamore, 2007).

Desde então, esta classe de moléculas vem sendo estudada com o objetivo de se compreender sua participação na regulação da expressão gênica em diversas espécies animais e vegetais. Estas moléculas atuam regulando negativamente a expressão de seus RNAs mensageiros alvo por diferentes mecanismos, seja em nível transcricional, por metilação de DNA e, conseqüente, alteração da cromatina, ou pós-transcricional, pela degradação do mRNA alvo ou repressão de sua tradução (Voinnet, 2009). Além da contribuição para diversos aspectos da Biologia básica, pequenos RNAs silenciadores vêm sendo utilizados como novas ferramentas genéticas para a biotecnologia e melhoramento genético de espécies vegetais (Frizzi e Huang, 2010; Osakabe, Kajita *et al.*, 2011).

Em humanos, a literatura relacionada a miRNAs é vastíssima e destaca-se, especialmente, o papel de miRNAs na pluripotência de células tronco (Anokye-Danso, Snitow *et al.*; Pfaff, Moritz *et al.*), no desenvolvimento de diversos tipos de câncer (Kunej, Godnic *et al.*, 2012; Nikitina, Urazova *et al.*, 2012), de doenças auto-imunes como lúpus eritromatoso (Amarilyo e La Cava, 2012) e cardiovasculares (D'alessandra, Pompilio *et al.*, 2012; Papageorgiou, Tousoulis *et al.*, 2012). Após a observação que miRNAs são secretados, podem ser translocados e apresentam alta estabilidade conferida por microvesículas (Zhu, Xia *et al.*, 2012), miRNAs passaram a ser sugeridos como marcadores circulantes para diagnóstico de doenças como câncer e doenças cardiovasculares (Corsini, Bronte *et al.*, 2012; Wan, Wang *et al.*, 2012; Zen e Zhang, 2012).

## PEQUENOS RNAs EM PLANTAS

Em plantas, os pequenos RNAs silenciadores compreendem os micro RNAs (Bartel, 2004) e várias classes de pequenos RNAs interferentes (siRNAs) (Rajagopalan, Vaucheret *et al.*, 2006). Ambos são pequenos RNAs não codificadores, de 20 a 24 nucleotídeos, responsáveis por silenciamento de alvos, sejam RNA ou DNA, por complementariedade de sequência. A distinção entre estas duas classes é fundamental para o melhor entendimento de sua função biológica e identificação de novos elementos em espécies ainda inexploradas neste sentido. Essencialmente, estas diferem na sua biogênese e nos componentes efetores recrutados para sua atividade (Axtell, Westholm *et al.*, 2011; Axtell, 2013).

MiRNAs são componentes regulatórios endógenos transcritos por RNA polimerase II na forma de um precursor primário (pri-miRNA) fita simples poliadenilada e com uma estrutura secundária em alça (*foldback*) característica com pareamento parcial. O processamento do pri-miRNA dá origem a um precursor intermediário (pre-miRNA) em *stem loop*, que é clivado por um endonuclease *Dicer like 1* (DCL1) dando origem ao duplex miRNA-miRNA\* (Voinnet, 2009; Axtell, 2013). O miRNA maduro, fita simples, é

integrado a um complexo protéico de silenciamento induzido por RNA (RISC – *RNA induced silencing complex*), ligando-se a uma proteína da família argonauta (usualmente AGO1, mas também AGO 2 ou 7) e será responsável pelo direcionamento deste ao ácido RNA mensageiro alvo por complementariedade de sequência (Voinnet, 2009; Axtell, 2013).

SiRNAs podem ter origem endógena, transcritos a partir de regiões repetitivas (*inverted repeats*), elementos transponíveis e transcrição bidirecional, ou exógena, a partir de transgenes ou RNA viral (Tang, 2010; Axtell, 2013). Os siRNAs são processados a partir de precursores de RNA fita dupla, totalmente complementares (Carthew e Sontheimer, 2009) e clivados por uma endonuclease *Dicer like* (DCL) DCL2 ou 3 (Axtell, 2013). O direcionamento do siRNA maduro ao ácido nucléico alvo se dá pela ligação a uma proteína AGO4 (Axtell, 2013).

MiRNAs de animais e plantas apresentam alguns pontos distintos na sua biogênese e nas enzimas envolvidas no processamento de seus precursores (Ambros, Lee *et al.*, 2003; Bartel e Chen, 2004; Cuperus, Fahlgren *et al.*, 2011). Destaca-se o fato de, em plantas, existir complementariedade quase perfeita (com poucos erros de pareamento – 1 a 3, normalmente) entre o miRNA e seu mRNA alvo que, usualmente, apresenta um único sítio de ligação (Baumberger e Baulcombe, 2005; Cuperus, Carbonell *et al.*, 2010). Em animais, os mRNAs alvo, em geral, possuem múltiplos sítios de ligação ao miRNA com complementariedade parcial (Carthew e Sontheimer, 2009). A implicação funcional dessas características é que, em plantas, o principal, embora não único, mecanismo de regulação por miRNAs é a degradação dos alvos (Addo-Quaye, Eshoo *et al.*, 2008; Mallory, Elmayan *et al.*, 2008) e, em animais, é a repressão da tradução e desestabilização do transcrito (Carthew e Sontheimer, 2009; Voinnet, 2009). A complementariedade parcial entre o miRNA e o mRNA alvo observada em metazoários, supostamente, colabora para o grande número de alvos para um miRNA observados nestas espécies, ao contrário do que ocorre em plantas.

A regulação por miRNAs pode, também, ocorrer em nível transcricional, silenciando um loco alvo no genoma por metilação. Esse mecanismo foi, inicialmente,

descrito no musgo *Physcomitrella patens* e, até então, não se sabia se era um mecanismo conservado em plantas (Voinnet, 2009; Khraiwesh, Arif *et al.*, 2010). Contudo, a descrição, em arroz, de uma subclasse de miRNAs de 24 nucleotídeos mediando a metilação de DNA nos seus locos alvo em *trans* sugere que este seja uma via de regulação gênica conservada e de origem antiga (Wu, Zhou *et al.*, 2010). Assim como siRNAs, miRNAs de 24 nucleotídeos são processados por DCL3 e associados a AGO4. Contudo, além de origens distintas, siRNAs de 24 nucleotídeos são conhecidos por silenciar alvos por metilação de alvos em *cis* ou locos muito similares, ao contrário, de miRNAs (Wu, Zhou *et al.*, 2010). O mecanismo de metilação de DNA dirigida por RNA (*RNA directed DNA methylation – RdDm*) foi, inicialmente, associado exclusivamente a siRNAs com função de silenciamento de transposons e manutenção de estabilidade do genoma. RdDm é uma marca epigenética dinâmica responsável pelo silenciamento transcricional de locos alvo com papel relevante também em *imprinting* e regulação gênica durante desenvolvimento e resposta a estresse abiótico (Chinnusamy e Zhu, 2009).

A plasticidade da regulação da expressão via miRNAs vai além da simples redução da expressão dos níveis de tradução de uma proteína. Esses pequenos RNAs podem atenuar flutuações do mRNA alvo e influenciar variações temporais (Voinnet, 2009). Mecanismos de regulação por *feedback* parecem ser comuns na manutenção de miRNAs (Baulcombe, 2004). Isto pode permitir um mecanismo fino de regulação da expressão dos genes alvo e, conseqüentemente, também das proteínas codificadas pelos genes alvo (Schwab, Palatnik *et al.*, 2005).

A clivagem de mRNAs alvo por miRNAs parece ser um mecanismo típico de processos que necessitem regulação permanente como na embriogênese e no desenvolvimento de meristemas. Por outro lado, a repressão da tradução de um transcrito alvo seria um mecanismo de maior flexibilidade que permite aumentar ou diminuir a expressão de um determinado gene, por exemplo, em resposta a estresses, voltando à condição padrão após o estímulo. Além disso, estes mecanismos não são, necessariamente, mutuamente exclusivos e podem dar maior versatilidade aos circuitos regulados por miRNAs (Voinnet, 2009).

Vários trabalhos mostram a importância de miRNAs em processos tão fundamentais como crescimento, desenvolvimento, florescimento e resposta a estresse biótico e abiótico (Bartel, 2004; Lu, Sun *et al.*, 2005; Sunkar, Girke *et al.*, 2005; Ko, J.H., Prassinis, C. *et al.*, 2006; Chen, 2009). Em *Arabidopsis*, grande parte dos genes alvo de miRNAs conhecidos codificam proteínas com papel descrito, ou sugerido, no controle do desenvolvimento. O papel crucial da regulação por miRNAs no desenvolvimento já foi claramente comprovado em animais e plantas (Kidner e Martienssen, 2005; Chen, 2009; 2012). A observação, em mutantes com síntese de miRNA alterada, mostra anormalidades no padrão e morfologia de folhas e flores, no desenvolvimento do sistema vascular, na separação de órgãos, redução da fertilidade e perda do meristema apical (Axtell e Bartel, 2005; Zhang, Zou *et al.*, 2010; Arif, Fattash *et al.*, 2012; Varkonyi-Gasic, Lough *et al.*, 2012).

O impacto da regulação gênica por miRNAs no desenvolvimento pode ser exemplificada pela participação do miR156 na determinação da transição de fase vegetativa e da floração. MiR156 silencia fatores de transcrição SPL (*squamosa promoter binding protein like*) que são responsáveis pela manutenção das características fenotípicas juvenis em plantas. MiR156 é abundante em plântulas e sua expressão diminui conforme a planta atinge a fase adulta, enquanto miR172 exibe um padrão exatamente oposto (Wu e Poethig, 2006; Wu, Park *et al.*, 2009; Zhang, Zou *et al.*, 2010; Wang, Park *et al.*, 2011). SPL9 e, possivelmente SPL10, ativam miR172 responsável pela determinação de características adultas ao reprimir seus alvos (Wu, Park *et al.*, 2009). A super-expressão de miR156 causa o prolongamento da expressão de caracteres juvenis e floração extremamente tardia (Schwab, Palatnik *et al.*, 2005; Wu e Poethig, 2006; Wang, Schwab *et al.*, 2008).

Apesar de a expressão de miRNAs ser tempo e tecido dependente, estes podem ser translocados e atuar em células diferentes da sua origem (Voinnet, 2005b; Dunoyer, Schott *et al.*, 2010; Molnar, Melnyk *et al.*, 2010). Este mecanismo foi também observado para siRNAs que podem ser translocados a maiores distâncias que miRNAs e formar um gradiente de silenciamento importante, por exemplo, para determinação de polaridade de folhas (Chitwood, Nogueira *et al.*, 2009; Schwab, Maizel *et al.*, 2009). Apesar de,

inicialmente, se acreditar que miRNAs não pudessem ser translocados entre células, estudos recentes mostram que estes podem atuar como sinais intercelulares em plantas (Gursansky, Searle *et al.*, 2011).

#### DESCOBERTA DE miRNAs EM PLANTAS

Apesar de locos de genes MIR estarem presentes nos genomas de plantas, animais e vírus, as suas sequências não são conservadas entre estes, evocando a possibilidade de que miRNAs surgiram em dois eventos evolutivamente distintos, provavelmente, a partir de uma via ancestral de RNA interferente (Matranga e Zamore, 2007; Tarver, Donoghue *et al.*, 2010). Sugere-se que miRNAs tenham evoluído independentemente em eucariotos como uma exaptação da maquinaria de biossíntese comum (Tarver, Donoghue *et al.*, 2010). Contudo, dentro dos reinos, vários miRNAs parecem ter uma origem comum como se pode inferir pela conservação entre miRNAs de *Arabidopsis*, arroz e até mesmo de algumas briófitas (Kidner e Martienssen, 2005). Dentre as famílias de miRNAs já descritas em plantas, 21 parecem ser universais em angiospermas (Axtell e Bowman, 2008). Sete destas são altamente conservadas sendo encontradas até em plantas terrestres primitivas como *Selaginella* e *Physcomitrella* (Axtell, Snyder *et al.*, 2007). A descoberta de miRNAs na alga unicelular *Chlamydomonas reinhardtii* sugere ainda que a origem de miRNAs em plantas remonte à origem dos eucariotos fotossintéticos (Matranga e Zamore, 2007).

A extensão da conservação de genes MIR e seus transcritos alvo parece ser mantida em parte, contudo, não exclui algumas variantes entre espécies. Há exemplos de famílias de miRNAs conservados como miR159 e miR396 que regulam fatores de transcrição MYB e GRF (*growth-regulating factor*), respectivamente, mas tiveram alvos não relacionados validados com grande impacto no desenvolvimento em grupos taxonômicos restritos (Axtell, 2013).

Por outro lado, sugere-se que a maior parte das famílias de genes MIR não seja conservada, sejam 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, Talmor-Neiman *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). O seu processamento, a partir do precursor é heterogêneo, são expressos em menor quantidade e codificados por locos únicos. Sugere-se que possam ser (Jones-Rhoades, 2012) transientes e não funcionais (Fahlgren, Montgomery *et al.*, 2006; Fahlgren, Howell *et al.*, 2007; Fahlgren, Jogdeo *et al.*, 2010; Cuperus, Fahlgren *et al.*, 2011).

A utilização de sequenciamento em larga escala, especialmente na plataforma Illumina, permitiu uma amostragem mais ampla de miRNAs em várias espécies (Morin, Aksay *et al.*, 2008; Ding, Zhang *et al.*, 2009; Martinez, Forment *et al.*, 2011; Schreiber, Shi *et al.*, 2011; Song, Liu *et al.*, 2011; Wang, Chen *et al.*, 2011; Wei, Yan *et al.*, 2011; Lertpanyasampatha, Gao *et al.*, 2012; Wang, Huang *et al.*, 2012), levando à descoberta de novos genes, até então desconhecidos (Breakfield, Corcoran *et al.*, 2011). Inicialmente, a descoberta de miRNAs por clonagem e sequenciamento Sanger levou à descoberta dos miRNAs mais abundantemente expressos que, em sua maioria, compõem as famílias de miRNAs conservados. A obtenção de sequências com tecnologias de larga escala permitiu a detecção de sequências raras, expressas em menor quantidade e ainda aquelas específicas de um tecido, estágio ou condição, predominantemente, de origem mais recente (Voinnet, 2009). Nos últimos anos, a aplicação crescente de tecnologias de sequenciamento em larga escala tem se provado extremamente útil na descoberta e caracterização de pequenos

RNAs nas mais diversas condições como, por exemplo, estresse biótico e abiótico, desenvolvimento de sementes e de flor (Kulcheski, De Oliveira *et al.*, 2011; Song, Liu *et al.*, 2011; Wang, Chen *et al.*, 2011; Wang, Huang *et al.*, 2012). E em diversas espécies como soja, pepino, cevada, trigo, sorgo, cedro chinês, maçã e pêsego (Kulcheski, De Oliveira *et al.*, 2011; Martinez, Forment *et al.*, 2011; Schreiber, Shi *et al.*, 2011; Yao e Sun, 2011; Zhang, Zheng *et al.*, 2011; Wan, Wang *et al.*, 2012; Xia, Zhu *et al.*, 2012; Zhou, Song *et al.*, 2012; Zhu, Xia *et al.*, 2012).

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, Wanchana *et al.*, 2011; Xuan, Guo *et al.*, 2011; Tempel e Tahi, 2012).

#### MIRNA EM 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 (Khraiwesh, Zhu *et al.*, 2011; Kruszka, Pieczynski *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, Kajita *et al.*, 2011; Sun, Shi *et al.*, 2011).

Atualmente, dentre as espécies de plantas representadas no banco de dados público de micro RNAs, miRBase (versão 19, agosto de 2012) (Kozomara e Griffiths-Jones, 2011), são poucos os exemplos de espécies arbóreas. Além de maçã (Xia, Zhu *et al.*, 2012) e *Citrus* (Song, Fang *et al.*, 2009; Song, Wang *et al.*, 2010; Song, Yu *et al.*, 2012), representados neste banco, foram realizados experimentos de descoberta *in silico* e por sequenciamento em pêssigo, potencial modelo para a família Rosaceae (Eldem, Celikkol Akcay *et al.*, 2012; Zhang, Bai *et al.*, 2012).

Entre os representantes de espécies florestais no miRBase estão *Hevea brasiliensis* (28 miRNAs maduros) (Gebelin, Argout *et al.*, 2012), *Acacia mangium* e *A. auriculiformis* (3 e 7 miRNAs, respectivamente) (Wong, Cannon *et al.*, 2011; Ong e Wickneswari, 2012), as coníferas *Pinus taeda* (38 miRNAs) (Lu, Sun *et al.*, 2007), *Pinus densata* (31 miRNAs) (Wan, Zhang *et al.*, 2012) e *Picea abies* (41 miRNAs) (Yakovlev, Fossdal *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, Difazio *et al.*, 2006), com cerca de 370 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, Almerao *et al.*, 2012).

Alguns processos biológicos já tiveram a regulação por miRNAs confirmada experimentalmente. A regulação da transição de fase vegetativa, por exemplo, é regulada por miR156 e seus alvos da família SPL, 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).

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 – madeira – e reguladas ao longo do desenvolvimento da planta e das estações do ano em um híbrido de *Populus tremula* x *Populus Alba* (Ko, J.H., Prassinis, C. *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, Miura *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, Du *et al.*, 2011).

Com o objetivo de buscar expressão diferencial em xilema secundário (madeira) de *Acacia mangium*, foi realizado sequenciamento em larga escala de pequenos RNAs de árvores com conteúdo de lignina contrastantes onde 12 famílias de miRNAs conservados e 82 candidatos foram identificados (Ong e Wickneswari, 2012).

A identificação de pequenos RNAs em coníferas sugere que angiospermas e gimnospermas possuam mecanismos distintos de biossíntese. Ao contrário de angiospermas, a classe predominante de pequenos RNAs em gimnospermas é a de 21 bases. Gimnospermas não possuem a enzima DCL3, envolvida no processamento de siRNAs de 24 nt. Em contrapartida, foi identificada uma nova isoforma dessa enzima. A relação entre siRNAs de 24 bases e silenciamento transcricional por metilação pode estar relacionada com silenciamento de sequências redundantes, de grandes famílias gênicas, de regiões duplicadas do genoma em angiospermas e a baixa expressão desta classe em gimnospermas sugere um potencial mecanismo para a evolução dos seus grandes genomas (Ahuja e Neale, 2005).

Pela relevância, já demonstrada, da participação de miRNAs na regulação da expressão gênica e com o sequenciamento do genoma de referência de *Eucalyptus grandis*, este trabalho se propõe a identificar miRNAs novos e conservados de *Eucalyptus*, recurso ainda inexistente para o gênero, colaborando com a anotação do referido genoma e para o conhecimento científico deste vasto gênero.

## **METODOLOGIA**

### **MATERIAL BIOLÓGICO**

As amostras utilizadas no experimento de sequenciamento em larga escala de pequenos RNAs foram de árvores adultas. Xilema em desenvolvimento e folhas de *E. grandis* (árvore BRASUZ 1, cujo genoma foi sequenciado para geração da referência) e xilema em desenvolvimento de *E. globulus* indivíduos A2 e C3. As amostras utilizadas na validação por *northern blot* incluem folhas e xilema em desenvolvimento de árvores adultas coletadas na coleção de espécies de *Eucalyptus* do horto da Universidade Católica de Brasília (*E. grandis*, *E. dunnii*, *E. urophylla*) (Figura 1) e mudas de oito meses de idade das demais espécies do gênero *Eucalyptus*. Para estudos de conservação entre diferentes taxa da família Myrtaceae foram utilizadas amostras de dezoito espécies sendo oito delas, espécies do gênero *Eucalyptus*:

1. *Eucalyptus grandis*
2. *Eucalyptus botryoides*
3. *Eucalyptus brassiana*
4. *Eucalyptus dunnii*
5. *Eucalyptus globulus*
6. *Eucalyptus pellita*
7. *Eucalyptus resinifera*
8. *Eucalyptus urophylla*
9. *E. urophylla* X *E. grandis* (híbrido)

Espécies de outros gêneros incluíram:

1. *Corymbia citriodora*
2. *Melaleuca lateritia*
3. *Campomanesia pubescens* (gabirola)
4. *Syzygium malaccense* (jambo roxo)

5. *Syzygium cumini* (jamelão)
6. *Eugenia uniflora* (pitanga)
7. *Eugenia calycina* (pitanga do cerrado)
8. *Psidium guajava* (goiaba)
9. *Psidium* sp. (araçá)
10. *Eugenia dysenterica* (cagaita)

Amostras de *Glycine max* (soja), *Oryza sativa* (arroz), *Pinus taeda* e *Solanum lycopersicum* (tomate) foram usadas como grupos externos. Folhas completamente desenvolvidas foram utilizadas para minimizar a variação de expressão de pequenos RNAs.

As amostras utilizadas na validação via microarranjo foram:

- *E. grandis* BRASUZ1  
folha e xilema em desenvolvimento de árvore adulta
- *E. globulus*  
xilema em desenvolvimento e ovário de árvore adulta
- *E. grandis* e *E. globulus*  
*pool* de tecidos de 10 plântulas de 30 dias de idade

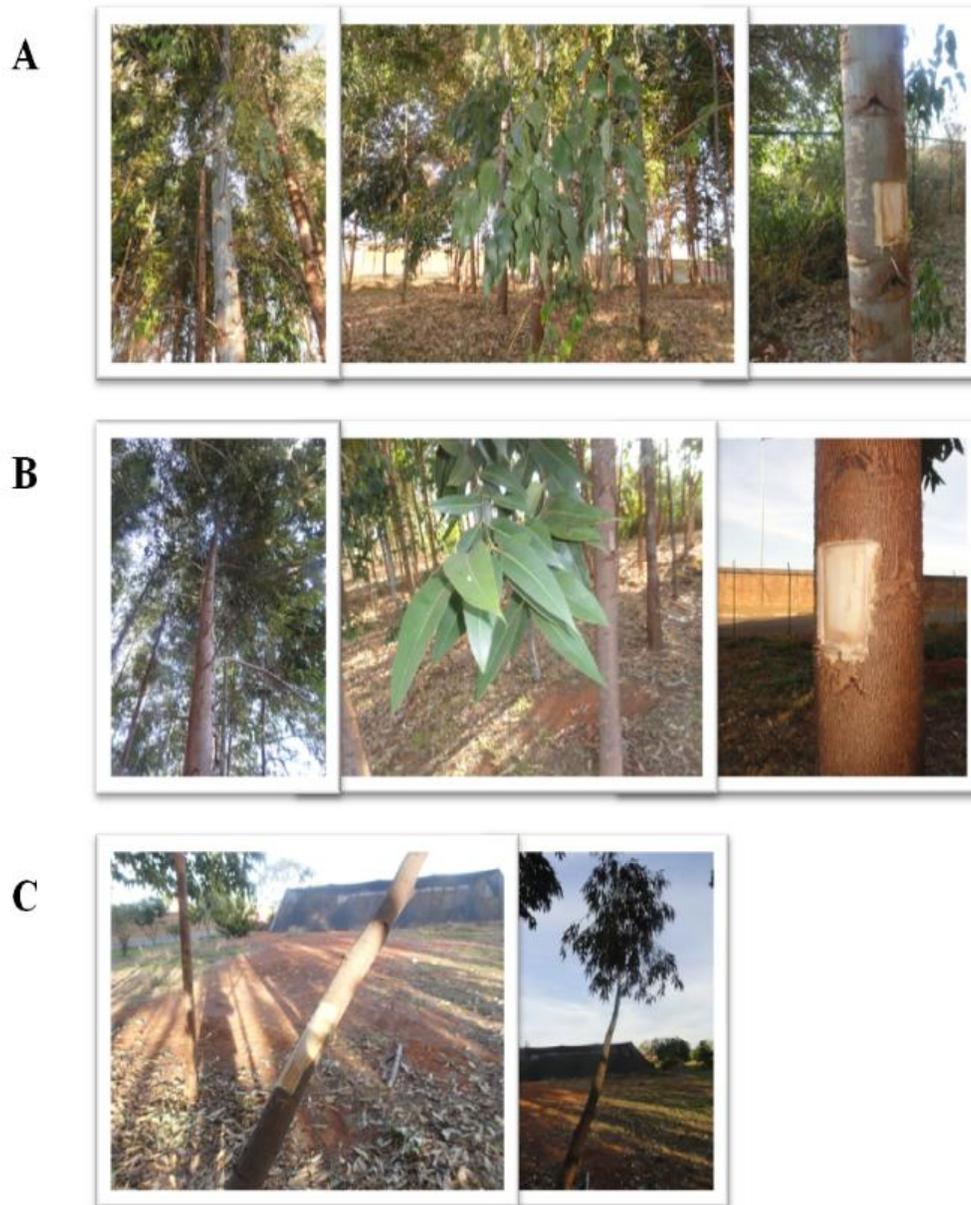


Figura 1: Árvores de *Eucalyptus* que forneceram amostras de tecido de folha e xilema em desenvolvimento para a execução da validação de pequenos RNAs via *northern blot*: (A) *E. dunnii*; (B) *E. urophylla* e (C) *E. grandis* – BRASUZI

## EXTRAÇÃO DE RNA

A metodologia utilizada para extração de RNA seguiu o protocolo para extração de RNA de *Pinus* baseado no método de CTAB (*Cetyl Trimethyl Ammonium Bromide*) (Chang, Puryear *et al.*, 1993) com adaptações. Como a precipitação com cloreto de lítio, incluída no protocolo, favorece a precipitação de RNAs de maior peso molecular, especialmente RNAs mensageiros, e grande parte dos pequenos RNAs são perdidos nesse processo e, por isso, essa etapa foi excluída da metodologia.

Esse material foi tratado de formas distintas de acordo com a metodologia a ser seguida. No caso do sequenciamento em larga escala, a seleção da fração de pequenos RNAs foi realizada pela empresa Fasteris (Suíça), que realizou o sequenciamento em plataforma Illumina GA II. Basicamente, a fração de pequenos RNAs é isolada a partir do RNA total por eluição, a partir de gel de acrilamida desnaturante. Estas moléculas são ligadas a adaptadores 3' e 5' para a síntese de cDNA.

No caso de *northern blot*, o material extraído, sem tratamento com DNase, foi aplicado em gel desnaturante de acrilamida 15%, sem separação prévia de frações.

## PROCESSAMENTO DOS DADOS DE SEQUENCIAMENTO

Um protocolo computacional foi desenvolvido especificamente para o processamento dos dados de sequenciamento em larga escala. A etapa de pré-processamento filtra as sequências por análise da qualidade, e executa a remoção da sequência de adaptadores e checagem de contaminantes. As sequências que passam por esta etapa são, então, agrupadas de acordo com o número de bases, quantificadas e usadas para criar um grupo de sequências não redundantes. O programa BWA (Li e Durbin, 2009) foi utilizado para mapear estas sequências contra o genoma de *E. grandis* (Phytozome v8.0, disponível em [phytozome.net](http://phytozome.net)).

A identificação de sequências de miRNAs conservados foi realizada usando PatMan (Prüfer, Stenzel *et al.*, 2008) para busca por similaridade contra o miRBase, versão

19 (Griffiths-Jones, Saini *et al.*, 2008), permitindo até três bases não coincidentes (*mismatches*) para a validação de uma sequência como nova isoforma.

A busca por mRNAs alvo dos miRNAs foi realizada com o programa psRNATarget (Dai e Zhao, 2011) que utilizou o banco de transcritos de *E. grandis* (phytozome 8.0). A anotação funcional prévia desses transcritos permitiu a inferência de quais as classes de genes e quais vias metabólicas estão sendo reguladas pelos potenciais miRNAs identificados. A análise da anotação destes genes alvo, assim como os domínios apresentados por estes, foram obtidos utilizando a ferramenta BioMart do Phytozome.

## VALIDAÇÃO EXPERIMENTAL

A validação experimental foi, inicialmente, realizada por *northern blot* como descrito anteriormente (Wu e Poethig, 2006). Para cada amostra, 30 microgramas de ácido nucleico total (DNA mais RNA), quantificados por espectrofotometria em Nanodrop (ThermoScientific), foram usados para eletroforese em gel desnaturante de acrilamida 15% (uréia 8 M) e transferidos para membrana Hybond N+. As sondas utilizadas são oligonucleotídeos de DNA com a sequência do complemento reverso das sequências completas dos miRNAs – equivalente ao miRNA\*. As sondas foram marcadas com gama 32<sup>P</sup> ATP e a hibridização realizada a 40°C em tampão de hibridização ULTRAhyb-oligo (Ambion). As membranas foram lavadas duas vezes em SSC 2X e SDS 0.5% por 30 minutos a 40°C e expostas por, pelo menos, 24 horas.

Uma validação em larga escala de miRNAs foi feita por meio de um experimento baseado no desenvolvimento e construção de um microarranjo customizado contratado junto à empresa LC Sciences (USA). Foi utilizado um chip de 30 mil sondas incluindo controles internos. Todas as sequências de miRNAs maduros de *Arabidopsis thaliana* do banco miRBase (versão 18 – novembro de 2011) foram inseridas como sondas no arranjo para análise de conservação, além de 9.344 sequências obtidas a partir do sequenciamento de *Eucalyptus* e de candidatos preditos *de novo in silico*, provenientes de outro trabalho (dados não publicados). Sete amostras foram usadas no experimento: xilema e duas amostras de folha (com morfologia juvenil e adulta) de *E. grandis* BRASUZ1 – todas as

amostras coletadas de uma mesma árvore simultaneamente; xilema e ovário de *E. globulus*; e um *pool* de tecidos de 10 plântulas de 30 dias de idade de *E. grandis* e um outro *pool* de igual constituição porém de *E. globulus*.

A normalização dos dados de cada ponto no arranjo foi realizada incluindo controles internos inseridos pela empresa, as triplicatas técnicas de cada sequência e pela subtração do *background* da fluorescência.

## **RESULTADOS E DISCUSSÃO**

### SEQUENCIAMENTO DE PEQUENOS RNAs

#### ANÁLISE GERAL DE SEQUÊNCIAS

A identificação e caracterização de miRNAs em *Eucalyptus* foi realizada com o objetivo de dar suporte à anotação do genoma de *E. grandis*. O sequenciamento em larga escala de pequenos RNAs (smRNA-Seq) 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 utilizando o programa miRDeep (Friedlander, Mackowiak et al., 2011). A predição de transcritos alvo, utilizando o programa psRNAtarget (Dai e Zhao, 2011). E, finalmente, a validação experimental via northern blot e, em larga escala, via microarranjo.

O smRNA-Seq foi realizado utilizando a plataforma Illumina GA II. As quatro bibliotecas utilizadas são provenientes de *E. grandis* (folha e xilema) e *E. globulus* (xilema) conforme descrito na metodologia. Cada uma das quatro amostras seguiu o protocolo padrão de construção de bibliotecas indicado pelo fabricante, com adaptadores únicos ligados a cada uma como códigos de barra. A corrida foi realizada com todas as bibliotecas em uma canaleta (*lane*) de uma célula de sequenciamento (*flowcell*) utilizando os *kits* Illumina para sequenciamento *single-end* com 36 bases por leitura. No total, foram obtidas 6.104.498 leituras (*reads*) que constituíram o ponto de partida para uma análise computacional com diversas etapas de processamento.

Os dados brutos de sequenciamento foram pré-processados de forma a excluir sequências com baixa qualidade, sequências sem adaptadores e sequências de RNA ribossômico (rRNA), transportador (tRNA) e de cloroplasto – e realizar a contagem de sequências únicas, descartando a redundância (Tabela 1). A identificação de sequências de

rRNA, tRNA e RNA de cloroplasto mostrou que estas correspondem, em média, a cerca de 1,8% das sequências únicas obtidas para cada amostra.

Tabela 1: Contagens e porcentagens de sequências geradas e analisadas no experimento de sequenciamento em larga escala com a plataforma Illumina GAI. São apresentados os números totais de sequências geradas, o número de sequências descartadas por não apresentarem sequências de adaptadores, o número total de sequências submetidas à análise, o número de sequências únicas, o número de sequências únicas após a remoção de sequências de rRNA, tRNA e RNA de cloroplasto (contaminantes) e o número de sequências únicas mapeadas no genoma.

| <i>Amostra</i>  | <i>E. grandis<br/>folha</i> | <i>E. grandis<br/>xilema</i> | <i>E. globulus A2</i> | <i>E. globulus gloC3</i> | <i>Total</i>         |
|---|-----------------------------|------------------------------|-----------------------|--------------------------|----------------------|
| <i>Total de sequências geradas</i>                        | 1.484.867<br>(100%)         | 1.766.355<br>(100%)          | 1.737.872<br>(100%)   | 1.115.404<br>(100%)      | 6.104.498<br>(100%)  |
| <i>Sequências sem adaptadores</i>                         | 34.652<br>(2,3%)            | 70.813<br>(4,0%)             | 34.066<br>(2,0%)      | 21.774<br>(1,9%)         | 161.305<br>(2,6%)    |
| <i>Número total de sequências submetidas à análise</i>    | 1.450.215<br>(97,7%)        | 1.695.542<br>(96%)           | 1.703.806<br>(98,0%)  | 1.093.630<br>(98,1%)     | 5.943.193<br>(97,4%) |
| <i>Número de sequências únicas</i>                        | 283.375<br>(19,1%)          | 402.158<br>(22,8%)           | 789.762<br>(45,4%)    | 505.663<br>(45,3%)       | 1.980.958<br>(32,5%) |
| <i>Número de sequências após remoção de contaminantes</i> | 265.579<br>(17,9%)          | 377.958<br>(21,4%)           | 754.898<br>(43,4%)    | 476.080<br>(42,7%)       | 1.874.515<br>(30,7%) |

Após o pré-processamento, as sequências foram divididas em categorias de tamanho. Como esperado, a distribuição das sequências apresenta dois picos em 21 e 24 bases em todas as amostras. Destaca-se o fato de que, na análise comparativa, a fração de

24 bases apresenta-se maior em todas as quatro amostras (Figura 2), chegando a ser 3,75 vezes maior que a categoria de 21 bases na amostra de xilema da árvore BRASUZ1 de *E. grandis*. Pequenos RNAs de 24 bases são, predominantemente, siRNAs envolvidos em metilação de DNA dirigida por RNA em locos complementares do genoma, resultando em silenciamento gênico e de transposons (Voinnet, 2009). Este é um padrão comum em angiospermas, que apresentam maior abundância de pequenos RNAs de 24 nucleotídeos. Curiosamente, gimnospermas, ao contrário, não apresentam essa classe de pequenos RNAs. A ausência da enzima DCL3 (*Dicer like 3*), responsável pelo processamento de pequenos RNAs de 24 nucleotídeos, explicaria a ausência desses elementos no grupo (Dolgoshina, Morin *et al.*, 2008; Morin, Aksay *et al.*, 2008).

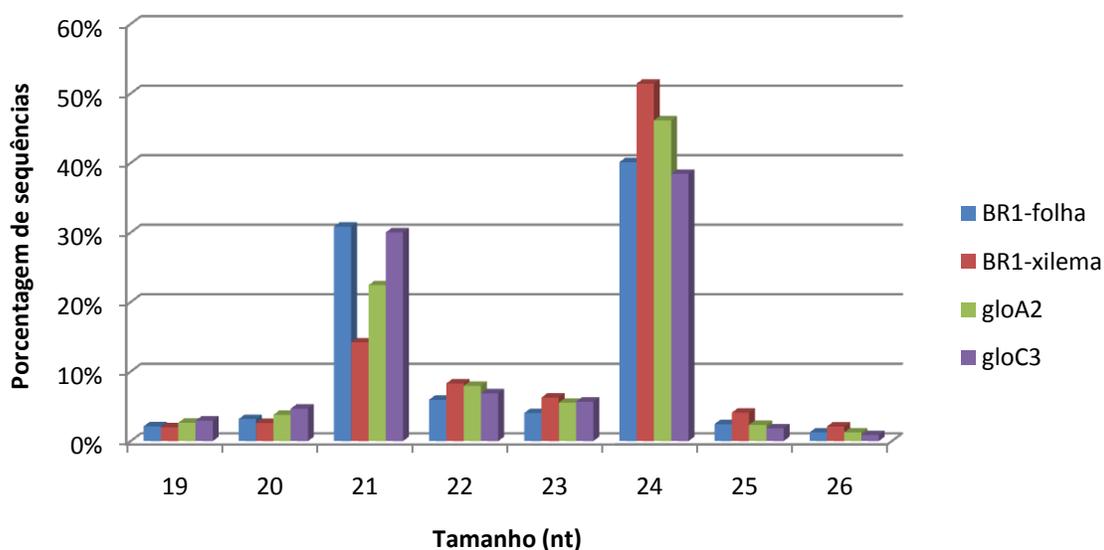


Figura 2: Distribuição da porcentagem de sequências, em diferentes classes de tamanho em nucleotídeos (nt), nas quatro amostras sequenciadas: xilema de *E. grandis* BRASUZ1 (BR\_xilema), folhas de *E. grandis* BRASUZ1 (BR\_folhas), xilema de *E. globulus* A2 (glo\_A2) e xilema de *E. globulus* C3 (glo\_C3).

Em termos de contagem do número total de sequências entre as classes de tamanho, verifica-se que a fração de 21 bases contem representantes com alta redundância, o que pode ser um indicativo de maior nível de expressão. É possível observar que, dentre as sequências de 21 bases, várias apresentam menos de dez leituras (*reads*) repetidas no sequenciamento enquanto algumas poucas sequências se destacam muito em relação à quantidade, passando de 50.000 leituras repetidas. Já as sequências de 24 bases, embora apresentem, praticamente, todas as sequências com menos de 100 repetições, mostram um marcante destaque na diversidade de grupos com sequências distintas.

Este padrão pode ser melhor observado no gráfico de perfil de agrupamento de sequências (*cluster profile*), onde os grupos de sequências únicas são ordenados pela sua redundância (Figura 3). Nesta figura, observa-se que, para todas as amostras, a classe de 24 nt é a que apresenta a maior diversidade de sequência, o que é uma característica comum em plantas, onde o repertório de pequenos RNAs é dominado por uma vasta diversidade destes siRNAs (Voinnet, 2009). Em termos de número de leituras por sequência, consistentemente observa-se que algumas sequências de 21 nt possuem níveis de redundância muito superiores a vasta maioria das outras classes, um indicativo de que está sendo detectada a intensidade de expressão inerente ao seu papel biológico. Nas amostras de xilema, também se observa que algumas classes de sRNAs, principalmente de 22 nt, apresentam alguns grupos com mais de 1.000 leituras, e sua identidade e implicação funcional serão descritos adiante.

O número de sequências únicas de todas as quatro amostras totalizou 1.857.986. Os números por classe de tamanho podem ser observados na Tabela 2 onde se destaca, mais uma vez, a diversidade de sequências de 24 nucleotídeos.

Tabela 2: Número total de sequências únicas por classe de tamanho consolidando os dados de todas as amostras sequenciadas.

| <b><i>Tamanho<br/>(bases)</i></b>      | <b>20</b> | <b>21</b> | <b>22</b> | <b>23</b> | <b>24</b> |
|--|-----------|-----------|-----------|-----------|-----------|
| <b><i>Número de<br/>sequências</i></b> | 45.411    | 169.661   | 94.027    | 125.499   | 1.230.102 |

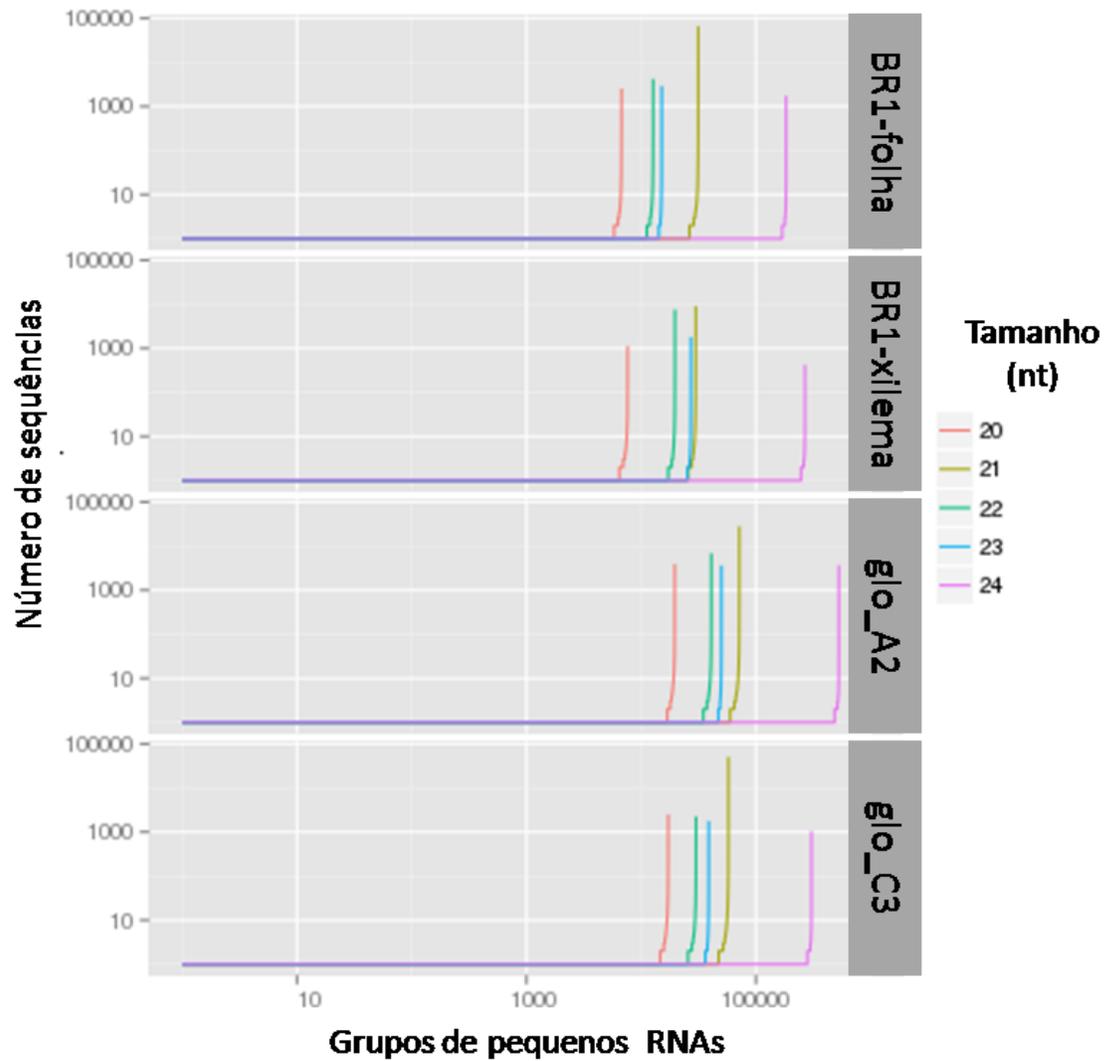


Figura 3: Gráfico do perfil de agrupamento (*cluster profile*) de seqüências com classificação dos grupos pelo número de seqüências. O número de grupos de seqüências únicas (eixo x) e o número de seqüências no grupo (eixo y) são plotados separadamente por classe por tamanho (nt). Distribuição por amostras: *E. grandis* BRASUZ1-folhas (BR1-folha), *E. grandis* BRASUZ1-xilema (BR1- xilema), xilema de *E. globulus* A2 (gloA2) e xilema de *E. globulus* C3 (gloC3)

## SEQUÊNCIAS MAIS ABUNDANTES

A análise das sequências de 21 bases mais abundantes em busca de potenciais novos miRNAs revelou aspectos interessantes. Foram identificados, como esperado, vários miRNAs conservados como miR159, miR160, miR166, miR172 e miR396. Parece consenso na literatura que os miRNAs mais expressos são os conservados, envolvidos em funções básicas como crescimento e desenvolvimento (Rajagopalan, Vaucheret *et al.*, 2006; Cuperus, Fahlgren *et al.*, 2011).

A análise comparativa das sequências de 21 nucleotídeos mostrou que o miR159 é a sequência de sRNA mais expressa. Na amostra de folha de *E. grandis* BRASUZ1, o euc\_miR159a representa 40% do total de sequências únicas (Figura 4A). Em xilema de *E. grandis* BRASUZ1 e de *E. globulus* C3 (Figura 4B e D), miR159a também foi a sequência mais expressa com 10,5% e 18,2% do total, respectivamente. Em *E. globulus* A2, embora o miRNA mais expresso seja miR172, miR159 é o segundo mais expresso com cerca de 7% (Figura 4C).

A sequência do euc\_miR159a de *Eucalyptus* apresenta 100% de identidade com os ortólogos de *Arabidopsis thaliana*, *Glycine max* e *Populus trichocarpa* (Figura 5A). Em destaque no alinhamento, a primeira base, na extremidade 5', de todos os miR159, sendo uma uracila, condição tida como preferencial para a ligação do miRNA à proteína argonauta 1 (AGO1) do complexo de silenciamento (Mi, Cai *et al.*, 2008; Montgomery, Howell *et al.*, 2008).

Em *A. thaliana*, foi observada alta expressão de miR159 em folhas e a expressão não foi detectada em caule (Park, Li *et al.*, 2002). Ao contrário, em árvores de um ano de idade de *Populus trichocarpa*, miR159 foi altamente expresso em xilema em desenvolvimento e pouco em folha (Lu, Sun *et al.*, 2005). Estudos com mutantes de miR159, em *A. thaliana*, sugerem que a falta regulação de genes de fatores de transcrição MYB por miR159 diminua a proliferação celular (Alonso-Peral, Li *et al.*, 2010).

A predição de alvos para este miRNA em *Eucalyptus*, utilizando o programa psRNAtarget e o banco de dados de transcritos de *E. grandis* disponível no Phytozome (versão 9.0), resultou em 10 transcritos preditos (Tabela 3). Dentre estes, estão dois

transcritos com domínio MYB – Eucgr.G03183.1 e Eucgr.E01581.1, seguindo a nomenclatura gênica encontrada no banco Phytozome. Uma busca por transcritos similares usando o BLAST (Altschul, Madden et al., 1997) contra o banco não redundante de nucleotídeos do NCBI, mostrou similaridade com um transcrito predito como MYB de *Vitis vinifera* (XM\_002275810.2). O alinhamento par a par desses transcritos mostra a conservação do sítio de ligação do miR159 que, por sua vez, é fora do domínio MYB (Figura 5B). Proteínas MYB são conhecidas como uma das principais classes de fatores de transcrição envolvidas nas vias relacionadas com a formação da madeira (Zhong, Richardson et al., 2007). A família R2R3 de fatores MYB tem papel central no controle do metabolismo, especialmente, a síntese de classes específicas de fenilpropanóides como flavonóides (Stracke, Ishihara et al., 2007), antocianinas (Deluc, Barrieu et al., 2006) e ligninas (Patzlaff, Mcinnis et al., 2003; Bedon, Grima-Pettenati et al., 2007). Os resultados de sequenciamento em *Eucalyptus* apontam para um sistema altamente regulado entre miR159 e fatores de transcrição MYB. Apesar de não ser possível precisar exatamente se os membros regulados são os mesmos descritos nos estudos supracitados, abre-se uma perspectiva para a inserção dos miRNAs no controle da via biossintética de lignina. A confirmação experimental desses alvos requer experimentos que confirmem a clivagem do mRNA como por PCR-RACE ou sequenciamento de degradoma.

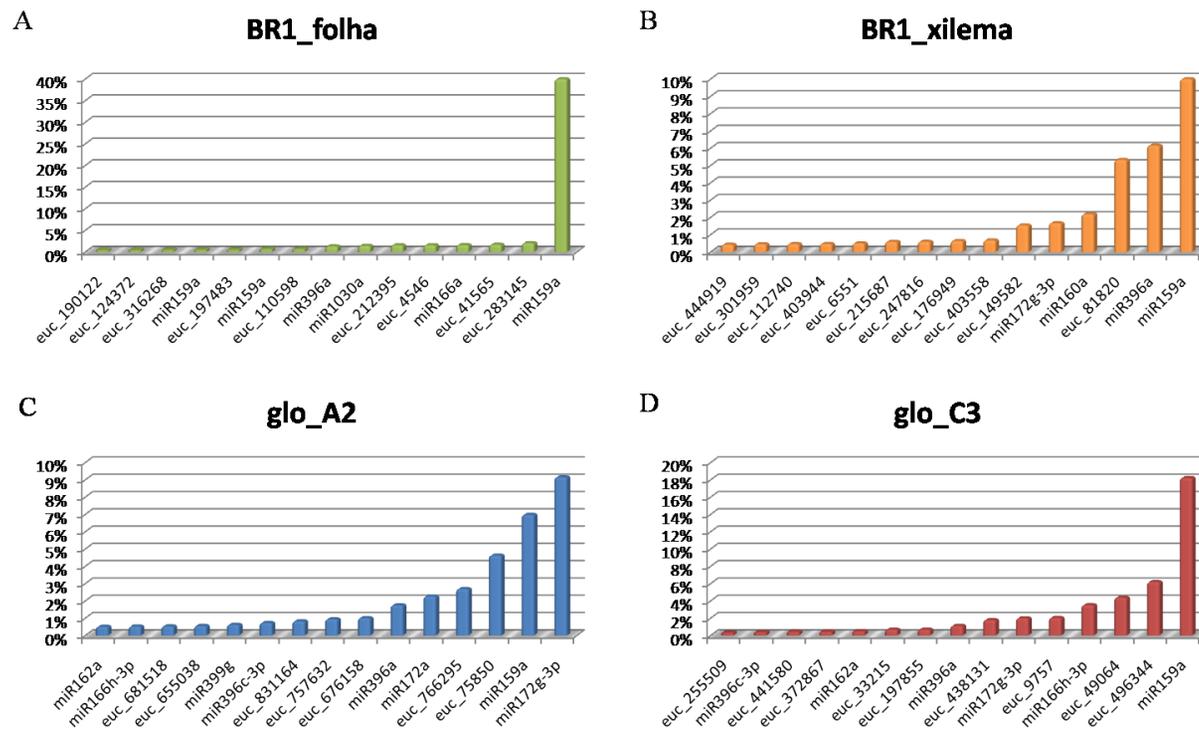


Figura 4: Distribuição de frequência das 15 sequências de pequenos RNAs de 21 nucleotídeos mais abundantes nos dados do sequenciamento, por amostra: (A) folhas de BRASUZ1 (BR1\_folha); (B) xilema de BRASUZ1 (BR1xilema); (C) xilema de *E. globulus* A2 (glo\_A2) e (D) xilema de *E. globulus* (glo\_C3). MiRNAs conservados são indicados e as demais sequências identificadas pelo código temporário *euc\_*, não tem ortólogos identificados no miRBase.

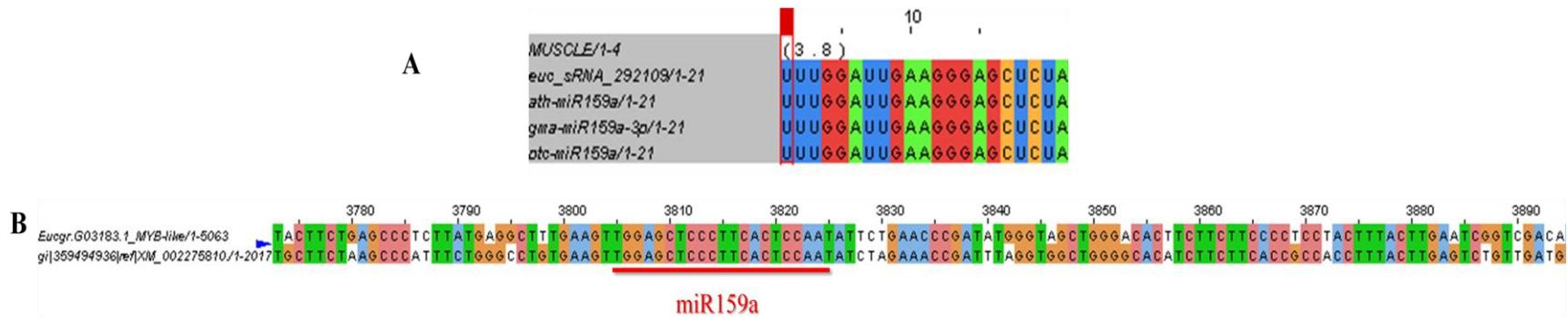


Figura 5: Conservação das sequências de miR159. (A) Alinhamento do *euc\_miR159a* (*euc\_sRNA\_292109*) com os ortólogos de *Arabidopsis thaliana* (*ath*), *Glycine max* (*gma*) e *Populus trichocarpa* (*ptc*). Em destaque, a primeira base na extremidade 5' (uracila). (B) Conservação do sítio de ligação de miR159 em um transcrito de *Eucalyptus grandis* com domínio MYB-like, predito como alvo pelo programa psRNAtarget, e em um transcrito de fator de transcrição GAMYB-like *Vitis vinifera*.

Tabela 3: Transcritos preditos como alvos de euc\_miR159a pelo programa psRNATarget. Nome do transcrito segundo o banco Phytozome, *E value* (*expectation value*), modo de inibição predita (clivagem ou bloqueio da tradução), código, nome e descrição do domínio protéico segundo o banco Pfam.

| <i>Nome do transcrito</i> | <i>E value</i> | <i>Inibição</i> | <i>ID PFAM</i>                | <i>Domínio Protéico</i>  | <i>Descrição PFAM</i>                                  |
|---------------------------|----------------|-----------------|-------------------------------|--|--|
| Eucgr.B03766.1            | 3.0            | Clivagem        | -                             | -  | -  |
| Eucgr.C02627.1            | 2.5            | Tradução        | PF00931<br>PF00560<br>PF01582 | Domínio NB-ARC<br>LRR( <i>leucine-rich repeat</i> )<br>Domínio TIR | apoptose<br>ligação a proteína<br>resposta imune inata |
| Eucgr.E01581.1            | 2.5            | Clivagem        | PF00249                       | Domínio de ligação a DNA Myb-like                                  | Domínio de ligação a DNA Myb-like                      |
| Eucgr.F01065.1            | 2.5            | Clivagem        | -                             | -  | -  |
| Eucgr.F01202.1            | 3.0            | Clivagem        | -                             | -  | -  |
| Eucgr.F03131.1            | 3.0            | Tradução        | PF02990                       | Proteína 70 endomembrana   | integral de membrana                                   |
| Eucgr.G03183.1            | 3.0            | Clivagem        | PF00249                       | Domínio de ligação a DNA Myb-like                                  | Domínio de ligação a DNA Myb-like                      |

| <i>Nome do transcrito</i> | <i>E value</i> | <i>Inibição</i> | <i>ID PFAM</i> | <i>Domínio Protéico</i>       | <i>Descrição PFAM</i>        |
|---------------------------|----------------|-----------------|----------------|-------------------------------|------------------------------|
| Eucgr.H04493.1            | 2.5            | Clivagem        | PF00501        | Enzima ligante de AMP         | atividade catalítica         |
| Eucgr.J00789.1            | 2.5            | Clivagem        | PF02535        | transportador <i>ZIP Zinc</i> | transporte de íons metálicos |

O único caso em que o miR159a não foi a espécie mais abundante ocorreu na amostra de xilema de *E. globulus* A2 onde miR172g foi o pequeno RNA de 21 bases com maior contagem (9%) (Figura 4C). O miR172 está envolvido na transição de fase vegetativa e na determinação da identidade de órgãos florais, funções conservadas em mono e dicotiledôneas (Zhu e Helliwell, 2011). Foi mostrado em *A. thaliana*, *Zea mays* e, posteriormente, em árvores – *Populus*, *Acacia* e *Eucalyptus globulus* – que a atividade sequencial de miR156 e miR172 controlam a transição de fase vegetativa (Wang, Park *et al.*, 2011). MiR172 também regula o gene AP2 (*apetala2*) que determina a formação de sépalas e pétalas em flores. A superexpressão de miR172 em *A. thaliana* causa um fenótipo acentuado de florescimento precoce e flores anormais incluindo a ausência de pétalas, mimetizando o que acontece em mutantes nulos para AP2 (Aukerman e Sakai, 2003).

Dentre as sequências mais expressas algumas são comuns a mais de uma amostra, como miR166 que tem papel reconhecido na formação de madeira desenvolvimento de meristema lateral e apical, assim como no desenvolvimento vascular e florescimento (Ko, J H, Prassinis, C *et al.*, 2006; Jung e Park, 2007; Du, Miura *et al.*, 2011; Robischon, Du *et al.*, 2011; Zhu, Hu *et al.*, 2011); e miR396, relacionado com divisão celular e desenvolvimento foliar (Wang, Gu *et al.*, 2011; Mecchia, Debernardi *et al.*, 2012).

#### A ABUNDÂNCIA DA SEQUÊNCIA NÃO IMPLICA CLASSIFICAÇÃO COMO miRNA

Como pode ser observado na Figura 4, diversos novos pequenos RNAs identificados entre as sequências mais abundantes não possuíam ortólogos depositados no banco miRBase. Alguns destes foram testados experimentalmente, através de *northern blot*, para confirmar sua expressão e tamanho de 21 bases. Confirmou-se a expressão e o tamanho dos sRNAs euc\_sRNA\_149582, euc\_sRNA\_75850, euc\_sRNA\_438131, euc\_sRNA\_33215 e euc\_sRNA\_372867 (Figura 6). Para dar suporte a hipótese de que estes sejam novos miRNAs utilizou-se o genoma de referência de *E. grandis* para extrair uma região de aproximadamente 100 bases flanqueando a região genômica onde foram mapeadas estas sequências. Contudo, a predição da estrutura secundária destas não

resultou em uma conformação compatível com precursores de miRNAs (Figura 7), com a sequência do miRNA maduro presente no braço de um *stem loop* (Meyers, Axtell *et al.*, 2008).

Um exemplo desses pequenos RNAs com alta expressão é o pequeno RNA euc\_sRNA\_438131, não conservado em outras espécies, que tem como alvos preditos um fator de transcrição da família GRAS e transcritos envolvidos com resistência a doenças contendo os domínios protéicos TIR-NBS-LRR (Tabela 4). Euc\_sRNA\_9757 tem como alvos preditos transcritos com domínio KH de ligação a RNA e euc\_sRNA\_372867 tem um único alvo predito, com o mesmo domínio KH.

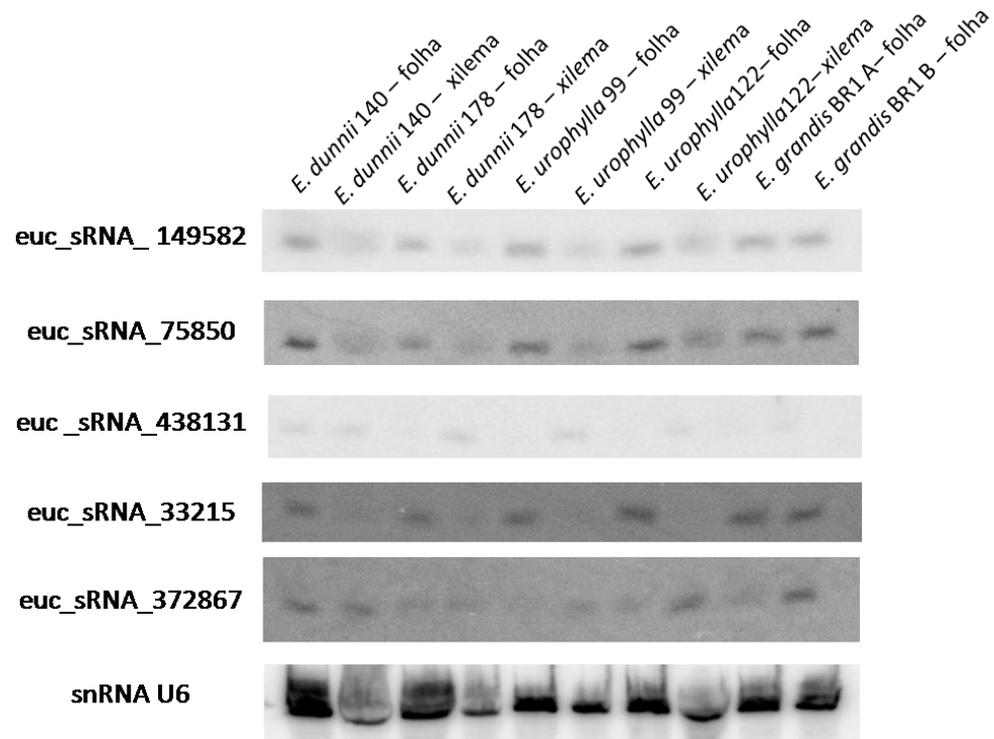


Figura 6: Northern blot de pequenos RNAs de *Eucalyptus* destacados pela alta contagem nos dados de sequenciamento. para *E. dunnii* 140 – folha, *E. dunnii* 140 – xilema, *E. dunnii* 178 – folha, *E. dunnii* 178 – xilema, *E. urophylla* 99 – folha, *E. urophylla* 99 – xilema, *E. urophylla* 122 – folha, *E. urophylla* 122 – xilema, *E. grandis* BRASUZ1 a – folha e *E. grandis* BRASUZ1 b – folha.

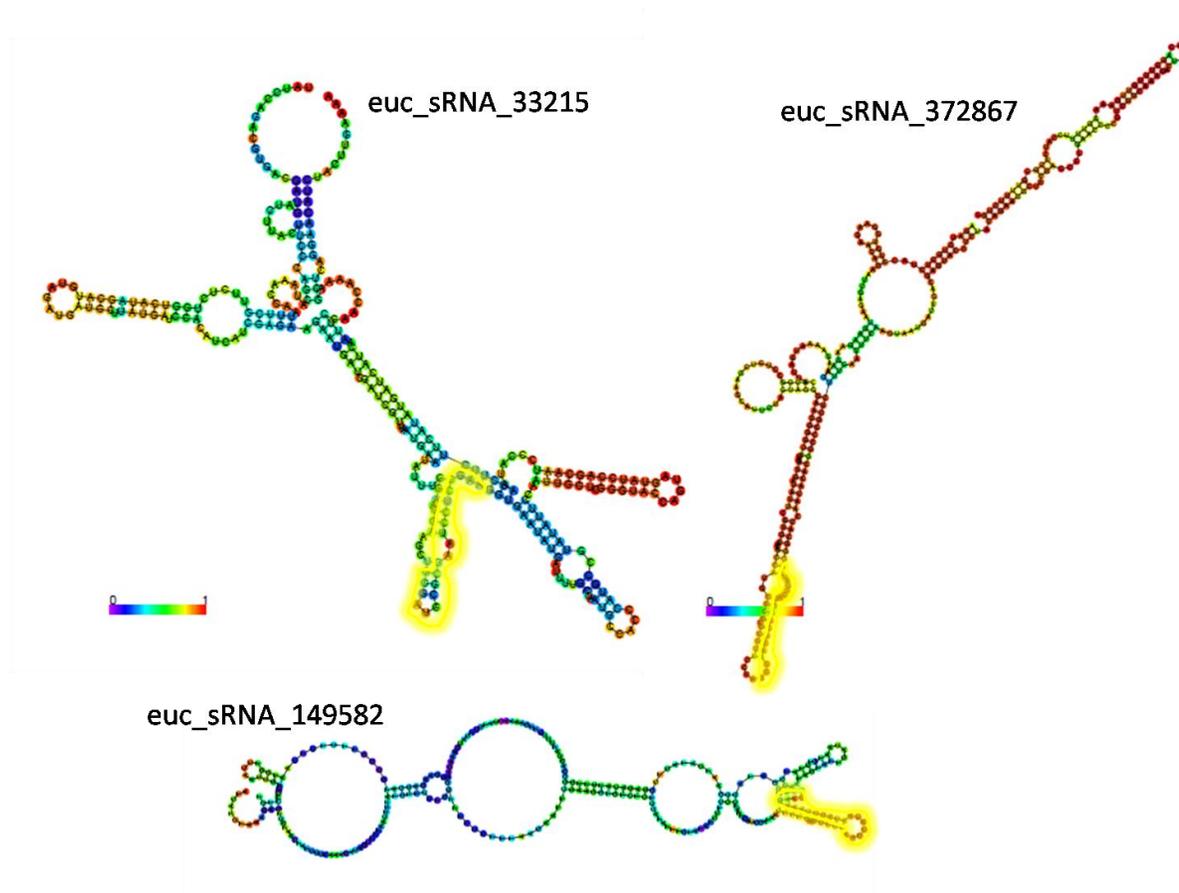


Figura 7: Exemplos de estruturas secundárias previstas para pequenos RNAs de 21 bases com alta contagem no sequenciamento, mapeadas no genoma mas sem estrutura compatível com precursores de miRNAs. As seqüências dos pequenos RNAs euc\_sRNA\_33215, euc\_sRNA\_372867 e euc\_sRNA\_149582 estão destacadas em amarelo.

Tabela 4: Transcritos preditos como alvos do pequeno RNA de *Eucalyptus euc\_sRNA\_438131*. Nome do transcrito, número do cromossomo onde se encontra o loco correspondente, *E value* (*expectation value*), modo de inibição predita (clivagem ou bloqueio da tradução), código do domínio protéico no PFAM e sua descrição.

| <i>Nome do gene</i> | <i>Cromossomo</i> | <i>E value</i> | <i>Inibição</i> | <i>ID PFAM</i> | <i>Descrição PFAM</i>              |
|---------------------|-------------------|----------------|-----------------|----------------|------------------------------------|
| Eucgr.B00380        | scaffold_2        | 3.0            | Tradução        | PF00069        | Domínio proteína kinase            |
| Eucgr.B00685        | scaffold_2        | 3.0            | Clivagem        | PF00560        | LRR ( <i>Leucine Rich Repeat</i> ) |
| Eucgr.D02129        | scaffold_4        | 2.0            | Clivagem        | PF00320        | GATA <i>zinc finger</i>            |

| <i>Nome do gene</i> | <i>Cromossomo</i> | <i>E value</i> | <i>Inibição</i> | <i>ID PFAM</i>                | <i>Descrição PFAM</i>   |
|---------------------|-------------------|----------------|-----------------|-------------------------------|---|
| Eucgr.E01049        | scaffold_5        | 3.5            | Clivagem        | PF00010                       | Domínio de ligação a DNA hélice-loop-hélice                         |
| Eucgr.E03125        | scaffold_5        | 3.5            | Clivagem        | PF00560                       | <i>Leucine Rich Repeat</i>  |
| Eucgr.E03126        | scaffold_5        | 3.5            | Clivagem        | PF00931<br>PF00560            | Domínio NB-ARC<br>LRR ( <i>Leucine Rich Repeat</i> )                |
| Eucgr.E03335        | scaffold_5        | 3.5            | Clivagem        | PF01582<br>PF00931<br>PF00560 | Domínio TIR<br>Domínio NB-ARC<br>LRR ( <i>Leucine Rich Repeat</i> ) |

| <i>Nome do gene</i> | <i>Cromossomo</i> | <i>E value</i> | <i>Inibição</i> | <i>ID PFAM</i>     | <i>Descrição PFAM</i>                                   |
|---------------------|-------------------|----------------|-----------------|--------------------|---|
| Eucgr.E03840        | scaffold_5        | 3.0            | Tradução        | PF00069            | Domínio proteína kinase                                 |
| Eucgr.G02940        | scaffold_7        | 2.5            | Clivagem        | PF03514            | Fator de transcrição – família GRAS                     |
| Eucgr.K00971        | scaffold_11       | 2.0            | Clivagem        | PF00023            | Repetição <i>Ankyrin</i>                                |
| Eucgr.K01839        | scaffold_11       | 3.0            | Clivagem        | PF01453<br>PF00069 | Lectina ligante de D-mannose<br>Domínio proteína kinase |

## ANÁLISE COMPARATIVA DA ABUNDÂNCIA

Diferenças quantitativas na abundância de miRNAs podem ser observadas pela comparação par a par do número de sequências contrastando intra especificamente (xilema de *E. globulus* A2 e C3), inter especificamente (xilema de *E. grandis* e *E. globulus*) e entre tecidos do mesmo indivíduo (xilema e folhas de *E. grandis*). Os resultados revelaram que os repertórios de pequenos RNAs mais similares são, como esperado, das amostras intra-específicas derivadas do mesmo tecido – xilema de duas árvores de *E. globulus* – e os mais diversos, entre tecidos distintos – folha e xilema – do mesmo indivíduo (Figura 8). A análise da expressão diferencial entre tecidos mostra que cerca de 36% dos miRNAs conservados observados em cada tecido são mutuamente exclusivos e aqueles presentes em ambas as amostras variam em quantidade em até duas vezes ( $p\text{-value} \leq 0.05$ ). Estes resultados corroboram os dados da literatura que citam a ocorrência de expressão diferencial de miRNAs entre espécies, tecido e tempo de desenvolvimento (Morin, Aksay *et al.*, 2008; Wang, Liu *et al.*, 2011; Wei, Yan *et al.*, 2011; Chen, 2012).

Apesar da indicação de expressão diferencial, o presente experimento não foi especificamente desenhado para tal propósito, pela falta de réplicas amostrais adequadas. Para análise comparativa dos dados seria necessário haver réplicas técnicas e as amostras deveriam ser de mesma idade, coletadas em um mesmo ambiente de modo a minimizar as variáveis influenciando a expressão de pequenos RNAs. Embora não possamos usar os dados de sequenciamento quantitativamente, os resultados da análise par a par é consistente com o padrão de expressão esperado. Abre-se assim a possibilidade para um estudo futuro dos sRNA diferencialmente expressos.

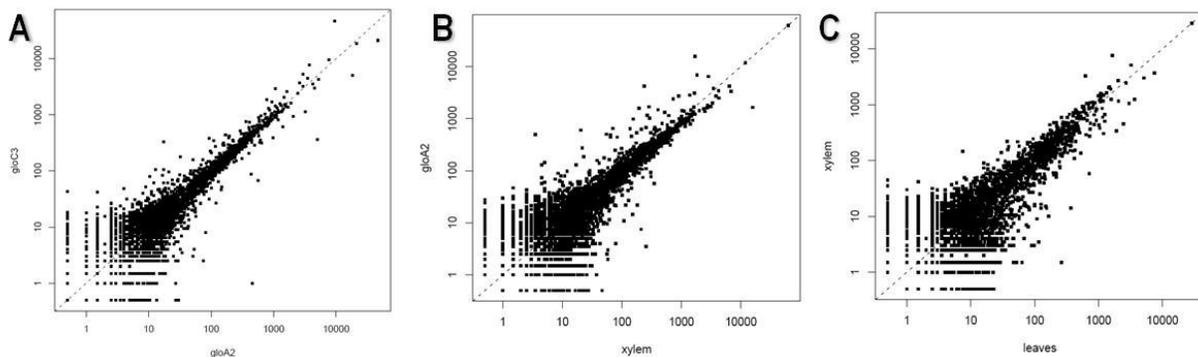


Figura 8: Comparação par a par do número de seqüências (escala log) entre amostras de: (A) xilema de duas árvores de *E. globulus* (gloA2 e gloC3); (B) entre xilema de *E. globulus* (gloA2) e de *E. grandis* BRASUZ1; (C) entre tecidos (folha e xilema) de *E. grandis* BRASUZ1.

## MAPEAMENTO CONTRA O GENOMA DE REFERÊNCIA

As 1.857.986 seqüências únicas de pequenos RNAs foram mapeadas no genoma de referência (Phytozome versão 9.0) utilizando o programa BWA (Li e Durbin, 2009). Os resultados do mapeamento revelaram uma variação no número de seqüências mapeadas por amostra entre 67,2% a 87,3% (Tabela 5).

Tabela 5: Resultados do mapeamento de seqüências únicas de pequenos RNAs no genoma de *E. grandis* usando o programa BWA por amostra analisada.

| <i>Amostra</i>             | <i>BRI- folha</i>  | <i>BRI- xilema</i> | <i>E. globulus A2</i> | <i>E. globulus C3</i> |
|----------------------------|--------------------|--------------------|-----------------------|-----------------------|
| <b>Total de seq.</b>       | 265.579            | 377.958            | 754.898               | 476.080               |
| <b>Seqüências mapeadas</b> | 231.893<br>(87,3%) | 327.580<br>(86,7%) | 507.430<br>(67,2%)    | 321.403<br>(67,5%)    |

As amostras de *E. grandis* BRASUZ apresentam maior número (em média 20% maior) de sequências mapeando contra o próprio genoma em comparação a ambas as amostras de *E. globulus*. Este resultado indica a existência de variabilidade interespecífica na fração de pequenos RNAs do genoma, consistente com a variabilidade de conteúdo genômico total entre as duas espécies. Ainda assim, a proporção de sequências mapeadas de *E. globulus* contra o genoma de *E. grandis* mostra a forte conservação de pequenos RNAs no gênero.

Os dados consolidados de mapeamento das sequências de todas as classes de tamanhos (19 a 26 bases) revelam a tendência de mapeamento de pequenos RNAs em regiões repetitivas (Tabela 6). Pela natureza dos pequenos RNAs não codificadores de 24 nucleotídeos, que silenciam, predominantemente, transposons e regiões repetitivas (Voinnet, 2009), espera-se que estes mapeiem, mais frequentemente, em regiões repetitivas que as demais classes. Um exemplo do mapeamento em regiões repetitivas por classe de tamanho pode ser observado na Figura 9 que ilustra a amostra de xilema *E. grandis* BRASUZ1. Pode ser observado, contudo, que o mapeamento em regiões repetitivas é consistentemente maior para todas as classes de tamanho. O mesmo comportamento foi observado para as outras amostras.

Tabela 6: Dados de mapeamento de sequências de pequenos RNAs, por amostra, em regiões não repetitivas e repetitivas.

|  | <i>E. grandis</i> BR1 -<br>folha | <i>E. grandis</i> BR1 -<br>xilema | <i>E. globulus</i><br>A2 | <i>E. globulus</i><br>C3 |
|--|----------------------------------|-----------------------------------|--------------------------|--------------------------|
| <i>Sequências mapeadas</i>                     | 231.893<br>(100%)                | 327.580<br>(100%)                 | 507.430<br>(100%)        | 321.403<br>(100%)        |
| <i>Mapeadas em regiões<br/>não repetitivas</i> | 69.376<br>(29,9%)                | 133.262<br>(40,7%)                | 181.046<br>(35,7%)       | 112.914<br>(35,1%)       |
| <i>Mapeadas em<br/>regiões repetitivas</i>     | 162.517<br>(70,1%)               | 194.318<br>(59,3%)                | 326.384<br>(64,3%)       | 208.489<br>(64,9%)       |

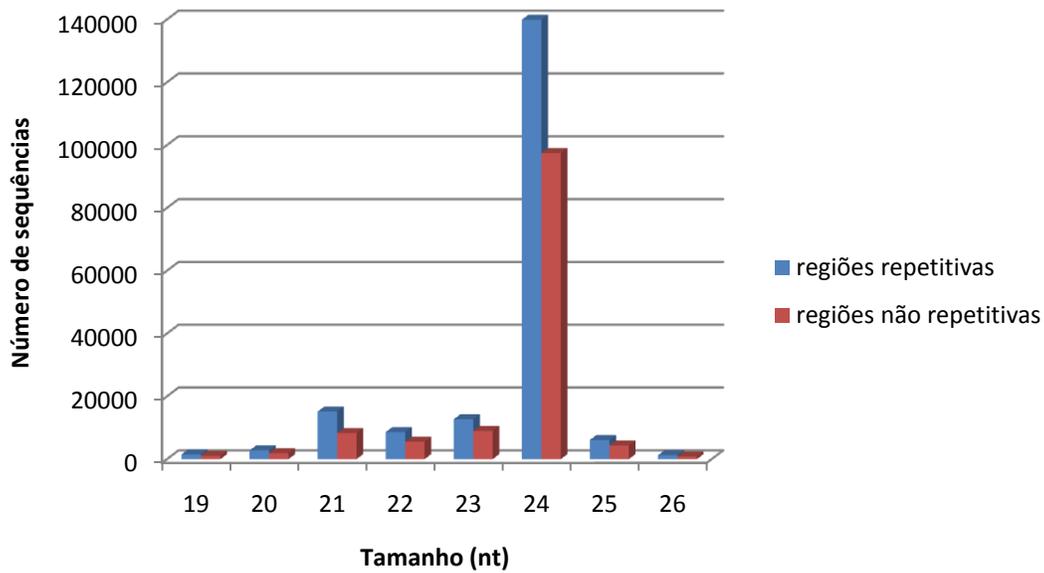


Figura 9: Número de seqüências mapeadas, por classe de tamanho em nucleotídeos (nt), em regiões repetitivas e não repetitivas da amostra de xilema de *E. grandis* BRASUZ1.

Analisando os dados de mapeamento, por amostra, das seqüências de 21 e 24 nucleotídeos, isoladamente, nota-se que o grande número de seqüências de 24 nucleotídeos representadas nos dados de sequenciamento se confirma no mapeamento, mostrando que, de fato, essa é a classe mais diversa de pequenos RNAs (Tabela 7).

Tabela 7: Resultados do mapeamento de seqüências únicas de pequenos RNAs de 21 e 24 nucleotídeos, por amostra, contra o genoma de *E. grandis* usando o programa BWA.

| <i>Amostra</i> | <i>BR1- folha</i>  | <i>BR1- xilema</i> | <i>E. globulus A2</i> | <i>E. globulus C3</i> |
|----------------|--------------------|--------------------|-----------------------|-----------------------|
| <b>21 nt</b>   | 24.942<br>(9,3%)   | 23.419<br>(6,2%)   | 47.805<br>(6,3%)      | 37.391<br>(7,8%)      |
| <b>24 nt</b>   | 163.798<br>(61,4%) | 237.389<br>(62,8%) | 346.968<br>(45,9%)    | 200.457<br>(42,1%)    |

## FAMÍLIAS CONSERVADAS DE MIRNAS

Uma das primeiras tarefas em relação à anotação funcional dos genes de miRNAs é a busca por ortólogos de miRNAs conservados em outras espécies. Além de uma validação quanto à anotação do gene, também podem ser inferidas as suas funções, potencialmente, conservadas.

Foi realizada uma busca por miRNAs conservados, por similaridade, com todas as sequências únicas de pequenos RNAs de 20 a 22 bases de todas as amostras do sequenciamento, totalizando 379.085, usando o programa PatMan (Prüfer, Stenzel *et al.*, 2008). A busca foi realizada contra todas as sequências de miRNAs maduros do banco miRBase (versão 19 – agosto de 2012). Um total de 470 miRNAs de 21 nucleotídeos (nt) e 69 miRNAs de 22 nt de *Eucalyptus* apresentaram similaridade com uma sequência ortóloga no miRBase, permitindo até três diferenças de bases no alinhamento (Tabela 8

Tabela 8). Os miRNAs de 21 nt conservados com total identidade em relação ao miRBase totalizam 95 isoformas pertencentes a 25 famílias de miRNAs (Tabela 9) (dados completos no Anexo I). Relaxando a estringência do alinhamento, foram observadas 35 famílias de genes MIR com até uma base de diferença e, com três diferenças, 74 famílias. Ressaltando que as sequências que não apresentam 100% de identidade, ou seja, com 1 a 3 diferenças de bases, representam variação espécie-específica nas famílias conservadas. Dentre os miRNAs de 22 nt, os 16 com 100% de identidade correspondem a 11 famílias de miRNAs e os 23 com uma base distinta, a 15 famílias. Sequências de miRNAs de 22 nucleotídeos não são comuns ou abundantes como as de 21 bases mas são reconhecidas como uma importante subclasse de miRNAs.

Tabela 8: Número de sequências de pequenos RNAs de 21 e 22 nucleotídeos (nt) de *Eucalyptus* com identidade e similaridade com isoformas de miRNAs do banco miRBase (versão 19) com 0 a 3 bases de diferença.

| <i>Diferenças(nt)</i> | <i>Número de sequências</i> |              |
|-----------------------|-----------------------------|--------------|
|                       | <i>21 nt</i>                | <i>22 nt</i> |
| <b>0</b>              | 95                          | 16           |
| <b>1</b>              | 115                         | 23           |
| <b>2</b>              | 111                         | 30           |
| <b>3</b>              | 149                         | -            |
| <b>Total</b>          | 470                         | 69           |

Dentre as famílias de miRNAs conservadas que tiveram isoformas com 100% de identidade detectadas nos dados de sequenciamento de *Eucalyptus*, diversas tem um número relevante de relatos indicando a sua função, como esperado para miRNAs conservados em processos básicos como desenvolvimento, morfologia, reprodução e imunidade. Por exemplo, o miR164 está envolvido na determinação da morfologia de folhas (Nikovics, Blein *et al.*, 2006), o miR319 é crítico para o crescimento de pétalas e desenvolvimento (Nag, King *et al.*, 2009), e os miR160 e miR398 estão relacionados com imunidade (Zhang, Zou *et al.*, 2010; Zhu, Ding *et al.*, 2011).

Tabela 9: Famílias de miRNAs conservados com isoformas com identidade de 100% confirmada nos dados de sequenciamento em larga escala de *Eucalyptus* com base em busca por similaridade contra o miRBase.

| <i>Famílias de miRNAs de 21 nt identificadas no smRNA-Seq de Eucalyptus</i> |        |         |        |        |
|---|--------|---------|--------|--------|
| miR156  | miR164 | miR171  | miR393 | miR399 |
| miR157  | miR166 | miR172  | miR395 | miR477 |
| miR159  | miR167 | miR2111 | miR396 | miR479 |
| miR160  | miR168 | miR319  | miR397 | miR535 |
| miR162  | miR169 | miR390  | miR398 | miR858 |

## COMPATIBILIDADE ESTRUTURAL

Dados da literatura até hoje indicam a existência de algumas centenas de genes de miRNAs em genomas de plantas. Isso implica que a grande maioria das sequências obtidas em experimentos de smRNA-Seq não correspondem a miRNAs, mas sim outras classes de pequenos RNAs. Para a correta caracterização destes elementos, foram observados diversos critérios estabelecidos (Meyers, Axtell *et al.*, 2008). O primeiro critério é a presença da sequência do miRNA maduro no braço de uma estrutura em *stem loop* do precursor, de onde deve ser excisado pela enzima DCL1. Neste aspecto, a disponibilidade do genoma completo de *E. grandis* BRASUZ1 ofereceu as condições ideais para que se realizasse uma busca computacional ampla para verificar a aderência de cada uma das sequências aos critérios de estrutura secundária do precursor. Em espécies onde esse recurso não está disponível, a predição da estrutura secundária de precursores depende de sequências de RNAs expressos (como por exemplo, ESTs – *expressed sequence tags*) que, nem sempre, são adequadas e disponíveis para a tarefa.

A predição de miRNAs *de novo*, a partir de dados de sequenciamento, foi realizada pelo pacote de programas miRDeep (Friedlander, Mackowiak *et al.*, 2011). Em termos gerais, o programa usa um modelo probabilístico da biogênese de miRNAs para anotar a compatibilidade da posição e frequência do pequeno RNA sequenciado com a estrutura secundária do precursor. A partir do mapeamento das sequências do smRNA-Seq no genoma, a sequência é estendida para incluir cerca de 100 bases flanqueando o início e o final da região mapeada e utilizada na predição da estrutura secundária do potencial precursor utilizando o programa RNAFold (Hofacker, 2003). A compatibilidade da estrutura com um precursor de miRNA e a estabilidade energética são avaliadas.

A partir de um total de 1.405.134 sequências únicas de 21 nucleotídeos mapeadas no genoma de *E. grandis*, 178 tiveram estrutura secundária compatível com precursores primários de miRNAs. Interessante destacar que, diante do grande número de sequências mapeadas no genoma, apenas uma minoria apresenta a estrutura secundária compatível (1,3%). O número de estruturas compatíveis preditas por amostra de sequenciamento se mostrou bastante consistente entre as amostras de xilema de *E. globulus* – 74 e 73 para *E. globulus* A2 e C3, respectivamente. Na amostra de folha de *E. grandis* foram identificadas 82 sequências e na de xilema, 55 sequências com estrutura compatível. Destas, apenas 11 mapeiam em pseudo-moléculas (*scaffolds*) distintos dos 11 *scaffolds* equivalentes aos cromossomos de *Eucalyptus*. Alguns exemplos de sequências que atenderam às premissas necessárias para um precursor de miRNA e mapeadas no *scaffold* 3 são mostrados na Figura 10.

Usualmente, quando se realiza um experimento de smRNA-Seq para descoberta de miRNAs, existe a tendência de se buscar a identificação dentre as sequências mais abundantes no sequenciamento. Foi realizada uma inspeção visual das estruturas secundárias preditas para as 140 sequências com maior contagem no sequenciamento da amostra de folha de *E. grandis* e, apenas, 3,5 % destas são compatíveis com precursores de miRNAs. Destaca-se assim o fato que miRNAs representaram apenas uma pequena parte dos pequenos RNAs abundantes de 21 nt identificados no experimento e, portanto, reforça a necessidade da criteriosa caracterização de novos miRNAs.

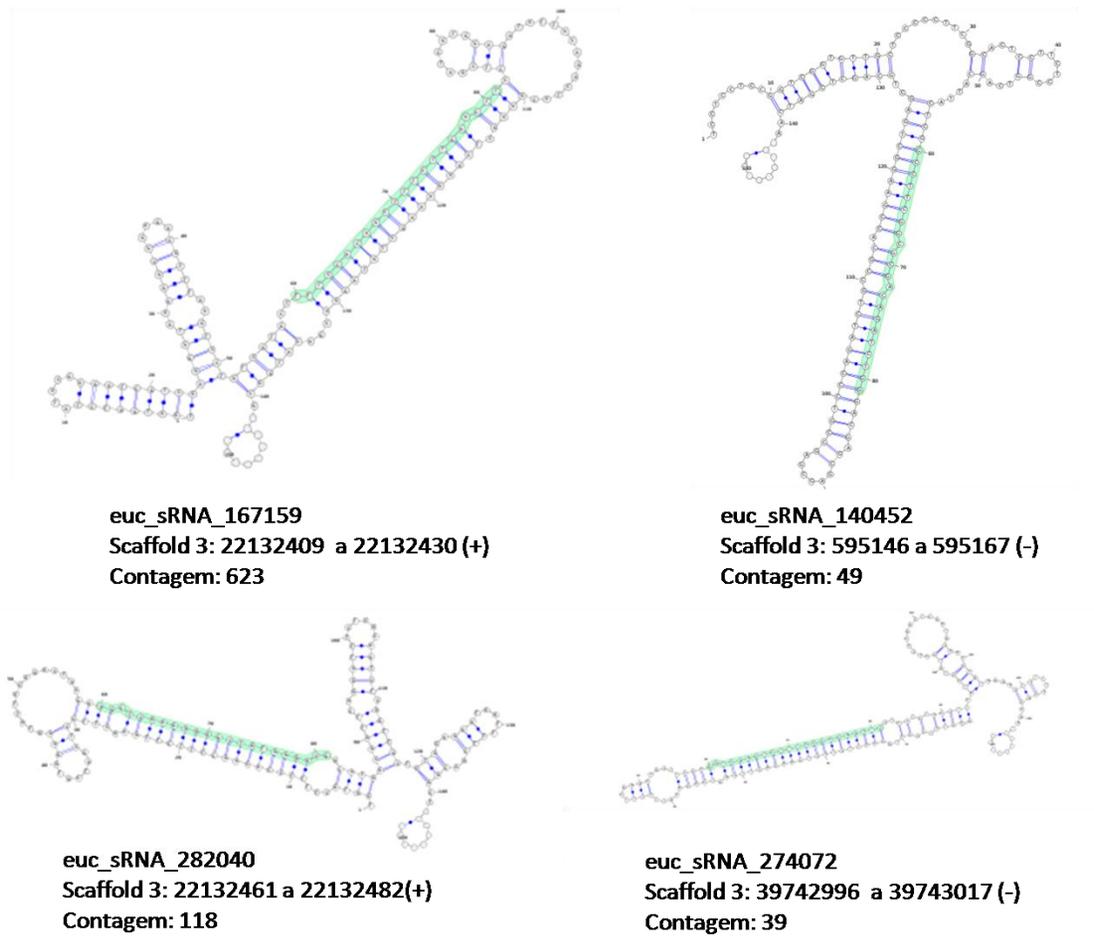


Figura 10: Exemplos da estrutura secundária de precursores de miRNAs preditos pelo programa RNAfold a partir de sequências mapeadas no *scaffold 3* do genoma de *Eucalyptus grandis* mais 130 bases da região flanqueante. Destacado em verde, a sequência do miRNA maduro.

## PREDIÇÃO DE ALVOS DE REGULAÇÃO POR MIRNAS

Uma característica distintiva de miRNAs é a ligação a mRNAs alvo para exercer a sua função regulatória. Foi realizada uma busca em larga escala de alvos para todas as leituras de 21 e 22 nt oriundas do experimento de smRNA-Seq utilizando o programa psRNATarget (Dai e Zhao, 2011).

Por se tratar de uma predição computacional, optou-se por focalizar na análise qualitativa das classes funcionais dos alvos. A Tabela 10 mostra que uma vasta gama de processos biológicos está representada pelos transcritos preditos como alvo. Dentre as famílias mais abundantemente representadas estão, entre outros, transportadores de íons, fatores de transcrição de SBP (*squamosa binding protein*), proteínas da família com repetição pentatricopetídeo (PPR), sequências com função de genes de resistência à patógenos da superfamília NBS-LRR (NB-ARC, *leucine rich repeat*) incluindo a família TIR-NBS-LRR (*Toll Interleukin-like domain*, NB-ARC e *Leucine Rich Repeats*). Os transcritos preditos como alvo para miRNAs conservados também mostram conservação entre espécies, como já havia sido relatado para *Populus trichocarpa* (Tuskan, Difazio *et al.*, 2006).

Além dos alvos preditos comumente encontrados, confirmando a conservação também dos alvos de miRNAs, alguns novos incluem transcritos de interesse para a formação da madeira como celulose sintetases e citocromo P450 envolvidos no metabolismo de uma série de substratos que podem ser divididos, de maneira geral, em vias biossintéticas e de detoxificação. Citocromos P450 são catalisadores importantes nas vias de biossíntese de fenilpropanóides, como antocianinas, hormônios como a giberelina, ácidos graxos, compostos de defesa como isoflavonóides e terpenos e polímeros estruturais como ligninas (Schuler e Werck-Reichhart, 2003).

**Tabela 10:** Exemplos de domínios protéicos presentes em transcritos alvo preditos, pelo programa psRNATarget, para miRNAs de *Eucalyptus*.

| <i>Cromossomo</i> | <i>Início (base)</i> | <i>Fim (base)</i> | <i>Nome do gene</i> | <i>ID Pfam</i> | <i>Descrição PFAM</i>  |
|-------------------|----------------------|-------------------|---------------------|----------------|--|
| scaffold_2        | 5066982              | 5068746           | Eucgr.B00373        | PF00249        | Domínio de ligação a DNA MYB-like                                  |
| scaffold_3        | 38728066             | 38731008          | Eucgr.C02141        | PF01582        | Domínio TIR  |
| scaffold_3        | 54096878             | 54099701          | Eucgr.C02857        | PF07649        | Domínio C1-like  |
| scaffold_3        | 54096878             | 54099701          | Eucgr.C02857        | PF00130        | Domínio de ligação a ésteres de Forbol/diacilglicerol (Domínio C1) |
| scaffold_3        | 54096878             | 54099701          | Eucgr.C02857        | PF00628        | PHD-finger   |
| scaffold_3        | 77900131             | 77904149          | Eucgr.C04268        | PF01582        | Domínio TIR  |
| scaffold_3        | 77900131             | 77904149          | Eucgr.C04268        | PF00931        | Domínio NB-ARC   |
| scaffold_3        | 77900131             | 77904149          | Eucgr.C04268        | PF00560        | <i>Leucine Rich Repeat</i>   |
| scaffold_8        | 72355247             | 72364511          | Eucgr.H05056        | PF03470        | Domínio XS zinc finger   |

## VALIDAÇÃO EXPERIMENTAL - *NORTHERN BLOT*

Uma confirmação da expressão observada nos dados de sequenciamento foi realizada por outra metodologia. Essa validação experimental, inicialmente, em pequena escala para poucos pequenos RNAs por meio de *northern blot* que, além de permitir a validação de sua expressão, possibilita uma avaliação do nível de expressão em diferentes amostras, embora sem um dado quantitativo absoluto ter sido atribuído a estes.

A validação foi realizada, a princípio, em amostras de três espécies de *Eucalyptus*: *E. dunnii*, *E. urophylla* e *E. grandis* – todas espécies do subgênero *Symphyomyrtus*, sendo a primeira, da seção *Maidenaria* e as outras duas da seção *Latoangulatae*. Foram incluídas amostras de RNA total de xilema e folhas completamente desenvolvidas. Foram utilizadas duas árvores de *E. dunnii* (140 e 178), duas árvores de *E. urophylla* (99 e 122) e dois clones da árvore *E. grandis* BRASUZ1, todos coletados no horto da Universidade Católica de Brasília (Figura 1).

Com exceção de BRASUZ1, com duas amostras de RNA total de folha, os demais indivíduos tiveram amostras de RNA total de folha e xilema incluídas no experimento (Figura 11). Essa amostragem permitiu a observação de expressão diferencial, entre tecidos, de alguns miRNAs como, por exemplo, o miR156, onde a expressão somente foi observada em amostras de folha. No caso dos novos sRNAs de *Eucalyptus*, foi possível observar maior expressão em amostras de folha para *euc\_sRNA\_149582*, *euc\_sRNA\_75850* e *euc\_sRNA\_33215*. Ainda foi possível observar que, apesar da baixa expressão de *euc\_sRNA\_438131* em todas as amostras, observou-se sua expressão diferencial entre tecidos com tendência de maior expressão em folhas.

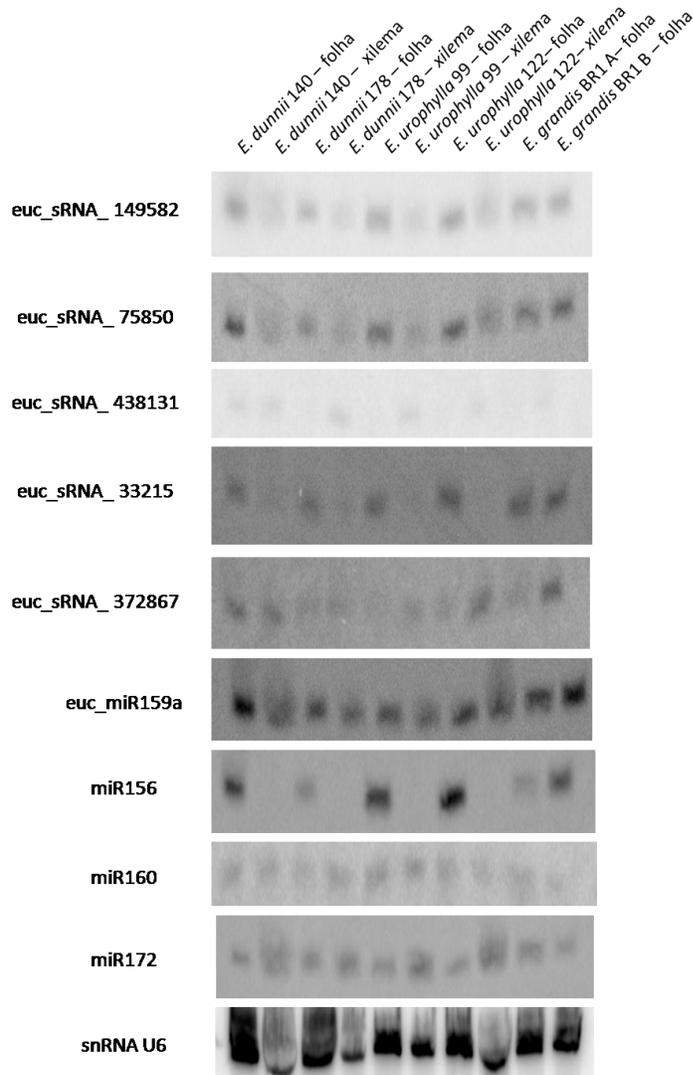


Figura 11: Resultados da validação experimental dos sRNAs e miRNAs descobertos em *Eucalyptus* por meio de *Northern blot* nas amostras de RNA de: *E. dunnii* árvore 140 – folha, *E. dunnii* árvore 140 – xilema, *E. dunnii* árvore 178 – folha, *E. dunnii* árvore 178 – xilema, *E. urophylla* árvore 99 – folha, *E. urophylla* árvore 99 – xilema, *E. urophylla* árvore 122 – folha, *E. urophylla* árvore 122 – xilema, *E. grandis* BRASUZ1 réplica A – folha e *E. grandis* BRASUZ1 réplica B – folha. Os miRNAs conservados miR156, miR160 e miR172 foram usados como controles positivos. O pequeno RNA nuclear snRNA U6 foi usado como controle de carregamento.

Com o objetivo de testar a extensão da conservação dos sRNAs que apresentaram resultado positivo na validação por *northern blot*, foi usada uma amostra de RNA total de

folha de *E. grandis* – BRASUZ1, como controle positivo, e amostras de mais duas espécies da família Myrtaceae. *Corymbia citriodora*, uma espécie florestal, anteriormente classificada como sendo do gênero *Eucalyptus* e *Eugenia uniflora* (pitanga), árvore frutífera nativa da América do sul. Na qualidade de grupo externo, membros de famílias filogeneticamente não relacionadas foram também utilizados no experimento: *Glycine max* (soja), *Oryza sativa* (arroz), *Pinus taeda* e *Solanum lycopersicum* (tomate) (Figura 12A). Destaca-se que a expressão dos sRNAs euc\_sRNA\_149582 e euc\_sRNA\_75850 parece ser específica da família Myrtaceae, e ausente em todos os membros não relacionados cobrindo uma ampla diversidade filogenética, incluindo mono e dicotiledôneas além de uma gimnosperma.

Interessante observar que a expressão dos candidatos euc\_sRNA\_438131, euc\_sRNA\_332215 e euc\_sRNA\_372867 ocorreu, exclusivamente, nas amostras de *Eucalyptus*, o que poderia ser um indicativo de siRNAs exclusivos do gênero, já que sua expressão já havia sido confirmada também em *E. dunnii* e *E. urophylla*. Vale destacar que estes três candidatos estavam entre os 15 mais expressos em todas as quatro amostras sequenciadas.

A exclusividade de expressão tanto na família quanto, principalmente, no gênero sugere a presença de sRNAs de origem evolutiva recente, posterior a separação das diferentes unidades taxonômicas testadas. Para testar tal hipótese, o experimento de validação incluiu mais seis amostras de diferentes espécies de *Eucalyptus* – *E. botryoides*, *E. brassiana*, *E. globulus*, *E. pellita*, *E. resinifera* – e um híbrido *E. urophylla* X *E. grandis*, além de outras oito espécies da família Myrtaceae: *Psidium* sp. (araçá), *Psidium guajava* (goiaba), *Syzygium cumini* (jamelão), *Melaleuca lateritia* (melaleuca), *Eugenia calycina* (pitanga do cerrado), *Eugenia dysenterica* (cagaita), *Campomanesia pubescens* (gabirola) e *Syzygium malaccense* (jambo roxo) (Figura 12B). Em cada uma das membranas, foram hibridizados miRNAs conservados como controles positivos além do pequeno RNA nuclear U6 como controle de carregamento do gel. Os sRNAs que tiveram a expressão confirmada nas três espécies de Myrtaceae citadas anteriormente foram testados e tiveram a expressão confirmada em todas as espécies indicando a descoberta de sequências, aparentemente, específicas da família. Destaca-se ainda a maior expressão do

euc\_sRNA\_75850 nas espécies de *Eucalyptus* quando comparada com as demais espécies da família.

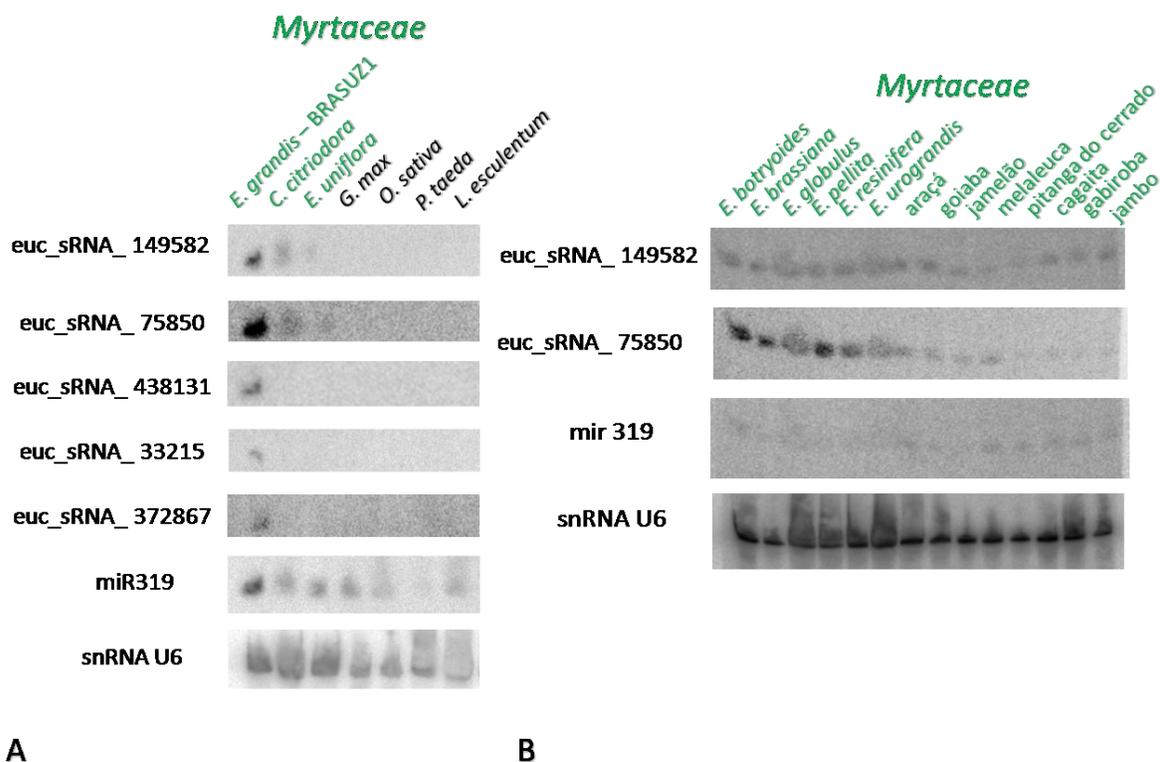


Figura 12: Resultados da validação experimental dos sRNAs descobertos em *Eucalyptus* por meio de Northern blot em amostras de RNA de folha de: (A) *E. grandis* – BRASUZ1, *Corymbia citriodora*, *Eugenia uniflora*, *Glycine max*, *Oryza sativa*, *Pinus taeda* e *Solanum lycopersicum*; (B) *Eucalyptus botryoides*, *E. brassiana*, *E. globulus*, *E. pellita*, *E. resinifera*, *E. urograndis* (*E. urophylla* x *E. grandis*), *Psidium* sp. (araçá), *Psidium guajava* (goiaba), *Syzygium cumini* (jamelão), *Melaleuca lateritia* (melaleuca), *Eugenia calycina* (pitanga do cerrado), *Eugenia dysenterica* (cagaita), *Campomanesia pubescens* (gabirola) e *Syzygium malaccense* (jambo roxo). O miRNA conservado miR319 foi usado como controle positivo e o pequeno RNA nuclear snRNA U6 foi usado como controle de carregamento.

Vale destacar que o miR159a (Figura 11), é uma nova isoforma da família miR159 observada em *Eucalyptus*, com uma base distinta das isoformas miR159a, b e c de *Populus trichocarpa* e miR159a de *Glycine max* e *Arabidopsis thaliana* (Figura 13). As isoformas de *P. trichocarpa* tem a sequência do miRNA maduro idênticas, como observado no alinhamento. Há alguns exemplos de famílias de miRNA quando as diferenças entre isoformas estão apenas na sequência do precursor e a sequência do miRNA maduro é conservada. Embora com menor intensidade, foi observado sinal de hibridização em espécies filogeneticamente distantes pertencentes a outras famílias (Figura 14), possivelmente resultante de hibridização cruzada, uma vez que só uma base é distinta das demais isoformas descritas de miR159. Os resultados indicaram ainda haver uma tendência de maior expressão do miR159 nas amostras de folha de árvores de *Eucalyptus* adultas em relação ao xilema (Figura 11).

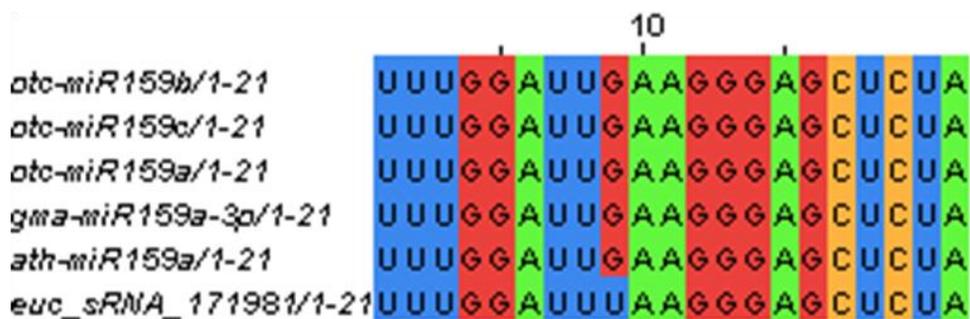


Figura 13: Alinhamento múltiplo das sequências de miR159a de *G. max* (gma), *P. trichocarpa* (ptc) e *A. thaliana* (ath) com a nova isoforma identificada em *Eucalyptus* (euc\_sRNA\_171981).

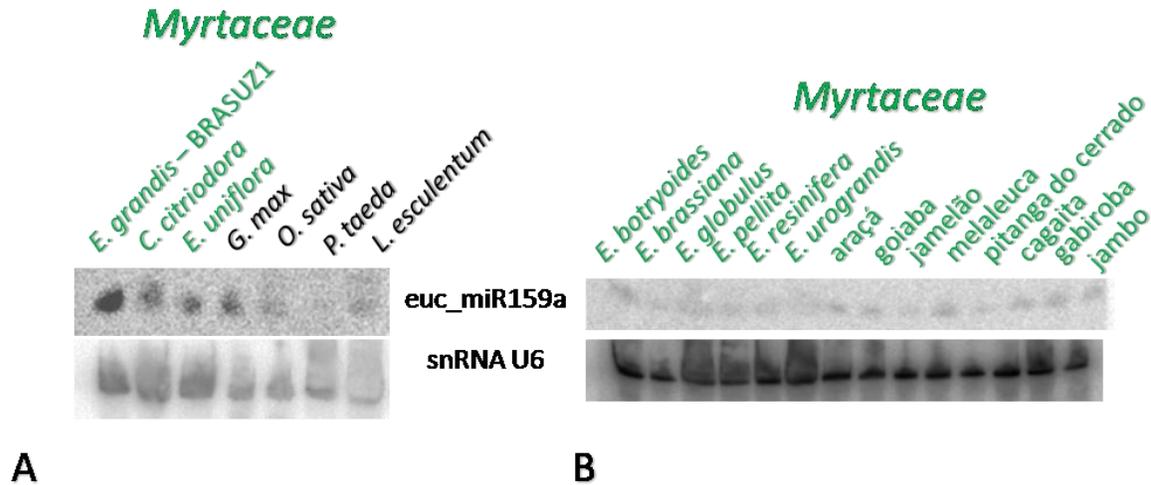


Figura 14: Resultados da validação experimental dos miR159a de *Eucalyptus* por meio de Northern blot em amostras de RNA de folha de: (A) *E. grandis* – BRASUZ1, *Corymbia citriodora*, *Eugenia uniflora*, *Glycine max*, *Oryza sativa*, *Pinus taeda* e *Solanum lycopersicum*; (B) *Eucalyptus botryoides*, *E. brassiana*, *E. globulus*, *E. pellita*, *E. resinifera*, *E. urograndis* (*urophylla* x *grandis*), araçá, goiaba, jamelão, melaleuca, pitanga do cerrado, cagaita, gabirola e jambo roxo.

## VALIDAÇÃO EXPERIMENTAL – MICROARRANJO

Visando ampliar a validação de expressão de pequenos RNAs, optou-se por adotar a tecnologia de microarranjo, que pode avaliar, simultaneamente, milhares de candidatos. Em particular, utilizou-se a tecnologia  $\mu$ Paraflo<sup>®</sup> de biochips microfluídicos que permite uma identificação direta e específica de miRNAs, com grande sensibilidade (LC Sciences<sup>™</sup>) (Zhou, Zhu *et al.*, 2012).

O microarranjo foi construído utilizando diversos candidatos oriundos do experimento de smRNA-Seq, principalmente sequências de 21 nt que apareceram com mais de uma contagem em diversas amostras. Além disso, foi incluída uma amostragem de sequências de 22 e 24 nt com alta contagem de leituras e sequências preditas com compatibilidade estrutural preditas pelo programa miRDeep (pág. 51). No total, 9.443 sondas distintas foram incluídas no microarranjo cada uma em triplicata, para fornecer

robustez estatística na análise de sinal. Além disso, foram incluídas 269 sondas com todos miRNA descritos para *A. thaliana* na versão 18 do miRBase.

O principal objetivo deste experimento é a validação da expressão de pequenos RNAs em uma escala significativamente maior que a possível pela técnica de *northern blot*. Isto permite uma melhor mineração entre o grande número de candidatos do smRNA-Seq como um dos pré-requisitos para a anotação de genes de miRNAs (Meyers, Axtell *et al.*, 2008).

Para ampliar o potencial de detecção, foram incluídas amostras de outros tecidos, com programa de expressão significativamente distinto das utilizadas para o smRNA-Seq (folha e xilema). Três amostras da árvore *E. grandis* BRASUZ1 foram utilizadas. Duas amostras de folha em estágios de desenvolvimento distintos, coletadas de um mesmo galho de árvore adulta foram usadas com o objetivo de observar se haveria variação no repertório de pequenos RNAs expressos ou se estas funcionariam como réplicas biológicas. Amostra de xilema foi coletada simultaneamente da mesma árvore. Amostras de xilema e ovário de árvore adulta de *E. globulus* foram também utilizados. Ovário foi utilizado por ser um órgão com expressão distinta daqueles tecidos usados no experimento de descoberta com o objetivo de observar a extensão dos pequenos RNAs expressos em tecidos diversos. Pool de 10 plântulas de 30 dias de idade de *E. grandis* e *E. globulus* foram também incluídos no experimento com o intuito de maximizar os tecidos amostrados em uma fase especialmente ativa de crescimento e desenvolvimento. O desenho do experimento não permite uma análise de expressão diferencial entre as amostras por falta de réplicas biológicas. Sumarizando, o foco do experimento de microarranjo é qualitativo e a expectativa é acumular mais indícios sobre pequenos RNAs que estejam sendo expressos nas condições testadas e ter uma estimativa mais completa do universo de miRNAs no genoma de *E. grandis*.

Com os resultados brutos dos microarranjos foram realizadas análises estatísticas para verificar se a intensidade de hibridização era acima do *background* para todas as três repetições da sonda em pelo menos uma das sete amostras analisadas nos sete microarranjos. São gerados p-valores para cada sonda e só foram consideradas aquelas

com  $p < 0.01$ . A normalização dos dados de intensidade foi realizada usando método LOWESS (*locally weighted scatterplot smoothing*) (Cleveland e Devlin, 1988; Bolstad, Irizarry *et al.*, 2003) nos dados subtraídos do *background*. Os números dos resultados positivos do experimento por amostra, de um total de 9.443 sondas para *Eucalyptus* e 269 para *Arabidopsis thaliana*, são mostrados na Tabela 11. A amostra de folha de *E. grandis* foi a que mostrou maior número de sondas com sinal significativo de expressão.

O menor número de sondas positivas observadas foi de 440 para a amostra de ovário de *E. globulus* e é consistente com o esperado uma vez que a descoberta não foi realizada neste tecido (mas em folhas e xilema) e, ainda, possivelmente por se tratar de um tecido com padrão muito específico de expressão.

Dentre todas as sondas, uma média de 6,5% dos novos candidatos a pequenos RNAs foram validados. Deve ser ressaltado que esta porcentagem refere-se ao número total de sondas utilizadas o que representa um universo de descoberta e validação de diversas classes de pequenos RNAs e não especificamente da classe dos miRNAs. Uma média de 12% das sondas de miRNAs de *A. thaliana* foram consideradas como positivas nas amostras de *Eucalyptus* e reforça a tese de que, embora haja uma extensa conservação de miRNAs entre plantas mesmo pouco relacionadas filogeneticamente, parte dos miRNAs expressos é linhagem ou espécie específica (Cuperus, Fahlgren *et al.*, 2011; Jones-Rhoades, 2012; Axtell, 2013).

Os microarranjos foram também utilizados para validar a descoberta computacional de novos genes MIR no genoma de *E. grandis*. Estes candidatos possuem estrutura secundária compatível com precursores de miRNA e exibem leituras no experimento de smRNA-Seq. Para um total de 178 sondas derivadas desta predição pelo programa miRDeep, 14 tiveram sinal significativo identificado no microarranjo em, pelo menos, uma das amostras ( $p \leq 0,01$ ). A busca por mRNAs alvo destes pelo programa psRNATarget, revelou que todas possuem alvos preditos, que incluem exemplos de fatores de transcrição das famílias MYB e GRAS, hidrolases, metiltransferases e oxidoredutases. Somando-se ao fato de não possuírem homólogos identificados no miRBase, ter indicação de expressão

(smRNA-Seq e microarranjos) e estrutura secundária, estes se destacam como fortes candidatos a novos genes MIR em *Eucalyptus*.

Tabela 11: Números absolutos de sondas de *Eucalyptus* e *Arabidopsis* com expressão validada no experimento de microarranjo para cada uma das sete amostras: ovário de *E. globulus* (GL\_OV); plântula de *E. globulus* (GL\_PL); xilema de *E. globulus* (GL\_XY); folha juvenil de *E. grandis* – BRASUZ1 (GR\_L1); folha adulta de *E. grandis* – BRASUZ1 (GR\_L2); plântula de *E. grandis* (GR\_PL); e xilema de *E. grandis* (GR\_XY). O número total de sonda utilizadas no experimento foi de 9.443 para *Eucalyptus* e 269 para *Arabidopsis thaliana*.

| <b>Amostra</b> | <b>Número de sondas com expressão positiva</b> |                    |
|----------------|--|--------------------|
|                | <i>Eucalyptus</i>                              | <i>Arabidopsis</i> |
| <b>GL_OV</b>   | 440  | 19                 |
| <b>GL_PL</b>   | 516  | 27                 |
| <b>GL_XY</b>   | 555  | 25                 |
| <b>GR_L1</b>   | 786  | 43                 |
| <b>GR_L2</b>   | 797  | 45                 |
| <b>GR_PL</b>   | 546  | 33                 |
| <b>GR_XY</b>   | 641  | 33                 |

Procurou-se verificar como as diferentes amostras compartilhavam as sequências expressas de acordo com os resultados do microarranjo. Diagramas de Venn foram construídos para explorar a expressão das sequências testadas para diversas combinações de amostras. Primeiramente, a análise de sequências comuns às quatro amostras de *E. grandis*, duas amostras de folha (em estágios de desenvolvimento distintos, denominadas juvenil e adulta), xilema e plântula (Figura 15A). Foram identificadas 61 e 46 sequências exclusivamente nas amostras de folha, juvenil e adulta, respectivamente, e 97 em comum entre estas. As quatro amostras apresentaram 449 sequências em comum, 48 exclusivas de

xilema e 10 de plântula. A comparação entre as amostras de folha e xilema de *E. grandis* mostram 554 sequências comuns (Figura 15B).

A comparação entre os três tecidos de *E. globulus* mostra 387 sequências comuns, 104 exclusivas de xilema, 50 de plântula e 15 de ovário (Figura 15C). A comparação entre as amostras de plântulas revelou que as duas espécies tem 447 sequências expressas em comum, 92 exclusivamente expressas em *E. grandis* e 59 em *E. globulus* (Figura 15D).

Esta análise reforça também a tese de que grande parte do repertório de miRNAs é comum a diferentes tecidos e espécies por estarem relacionados a regulação de vias comuns associadas a processos como desenvolvimento e manutenção da homeostase do genoma. A menor parte deste repertório é espécie, tecido e estágio específica. As alterações específicas de cada tecido, tipo celular e as adaptações a condições específicas como uma fase de desenvolvimento ou resposta a estresse seriam controladas por essa fração de pequenos RNAs. A observação das sequências mutuamente exclusivas nas duas amostras de folhas coletadas simultaneamente em um mesmo galho de uma mesma árvore de BRASUZ1 mas em estágios de desenvolvimento distintos demonstra um exemplo de expressão estágio específica.

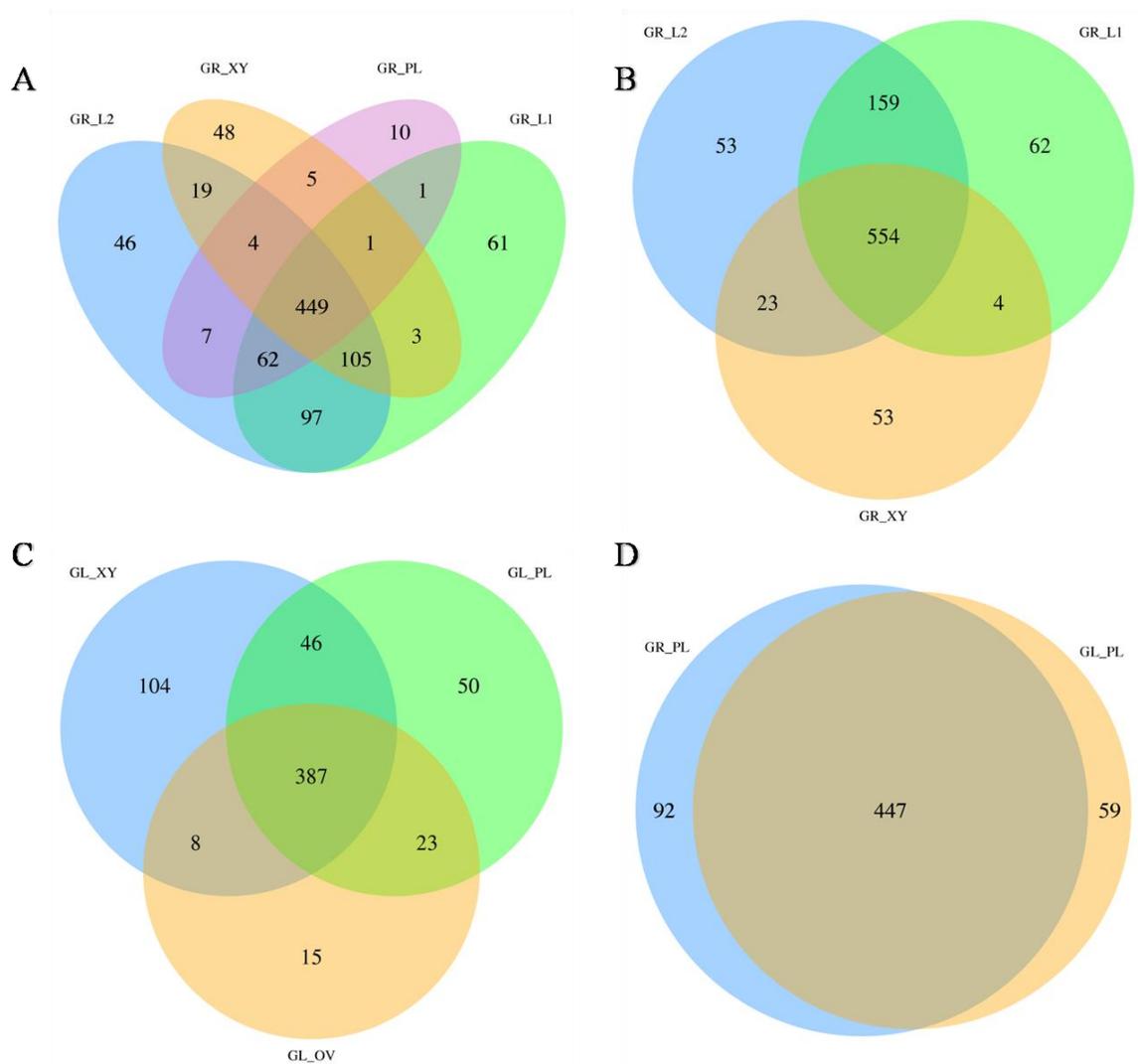


Figura 15: Compartilhamento da ocorrência de seqüências validadas de pequenos RNAs no experimento de microarranjo entre: (A) as quatro amostras de *E. grandis* – duas réplicas independentes de folha (GR\_L1 e GR\_L2), xilema (GR\_XY) e plântula (GR\_PL); (B) amostras de folha (GR\_L1 e GR\_L2) e xilema (GR\_XY) de *E. grandis*; (C) as três amostras de *E. globulus* – xilema (GL\_XY), plântula (GL\_PL) e ovário (GL\_OV); (D) duas amostras de plântula – *E. grandis* (GR\_PL) e *E. globulus* (GL\_PL).

## DADOS CONSOLIDADOS DE SEQUENCIAMENTO

Os resultados consolidados de todas as etapas experimentais executadas, destacando os números de sequências geradas no experimento de sequenciamento, e os quantitativos resultantes das diversas etapas subsequentes de mapeamento, análise de conservação, predição de estrutura secundária do precursor pela sequência genômica flanqueante, predição de transcritos alvo e, finalmente, validação via microarranjo foram sumarizados em um fluxograma (Figura 16). Esta análise consolidada demonstra claramente a dimensão de dados que podem ser obtidos via sequenciamento em larga escala e, principalmente, a crucial importância de se executar análises subsequentes rigorosas destes dados para se aproximar de resultados numéricos finais provavelmente mais próximos da realidade biológica do repertório de miRNA de uma espécie. De um total de 1.857.986 sequências únicas de 21 nucleotídeos, 1.405.134 (75,63%) mapearam no genoma. Destas, 470 (0,025%) tem algum ortólogo no miRBase com até 3 bases de diferença, sendo 95 destes conservados sem diferenças de base.

Dentre as sequências mapeadas candidatas a miRNAs, 178 (0,96%) tiveram estrutura secundária compatível com um precursor de miRNA (dados completos no Anexo II). Deste total, 82 foram identificadas na amostra de folha de *E. grandis* BRASUZ1, 55 na amostra de xilema de *E. grandis* BRASUZ1, 74 e 73 nas amostras de xilema de *E. globulus* A2 e C3, respectivamente. Dentre estes, 40 são isoformas de miRNAs conservados. Do total de 178 sequências, 163 tiveram, pelo menos, um alvo predito pelo psRNAtarget. O número total de transcritos preditos como alvo somou 690. Finalmente, 14 sequências foram validadas no microarranjo com intensidade de sinal significativamente maior em relação à média ( $p \leq 0,1$ ) e correspondem ao número de miRNAs identificados preenchendo aos quatro requisitos determinados para a anotação de um novo miRNA (Meyers, Axtell *et al.*, 2008).

Na montagem do genoma de *Populus trichocarpa*, modelo de espécies florestais, foi identificado computacionalmente, um número bastante próximo dos 178 preditos em *Eucalyptus*. Em *Populus*, foram identificados 169 genes de miRNAs representando 21 famílias (Tuskan, Difazio *et al.*, 2006).

Vale ressaltar que, no caso de *E. grandis*, a análise foi consideravelmente beneficiada pela disponibilidade da sequência de referência do genoma e pelo fato dos dados do sequenciamento de pequenos RNAs terem sido gerados a partir da mesma árvore que forneceu o genoma de referência. Além disso, a disponibilidade de um grande banco de transcritos desta mesma árvore, de outras árvores e de diversos tecidos também contribuiu, sobremaneira, para a identificação positiva de potenciais alvos.

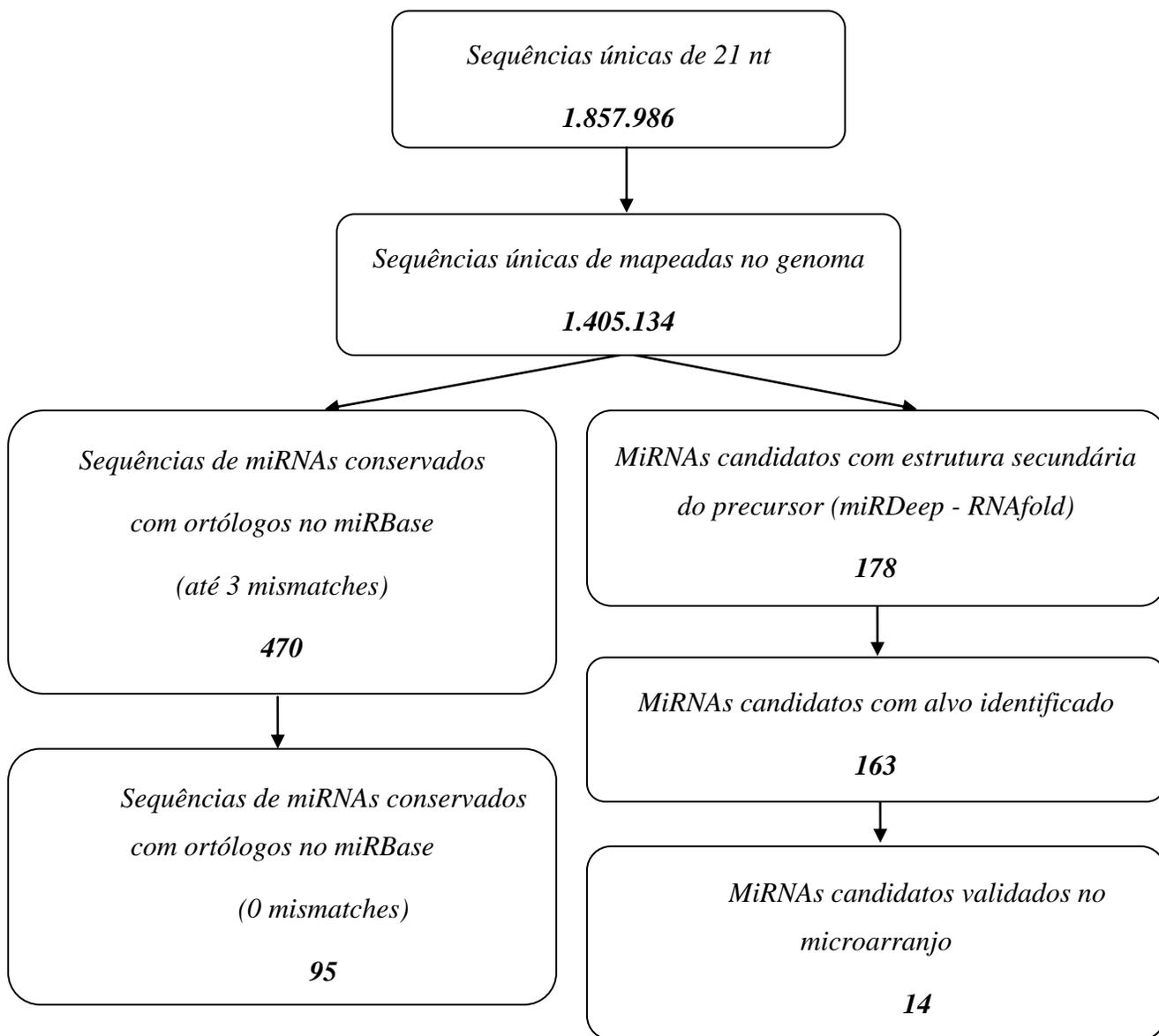


Figura 16: Fluxograma geral do experimento de descoberta de miRNA em Eucalyptus, destacando as contagens de sequências geradas inicialmente pelo sequenciamento de pequenos RNAs e os resultados quantitativos de cada etapa de análise, incluindo o mapeamento, análise de conservação, predição de estrutura e alvos e validação experimental por microarranjo.

## CONSERVAÇÃO COM *EUGENIA UNIFLORA*

Recentemente, o sequenciamento da fração de pequenos RNAs de *Eugenia uniflora* (pitanga) gerou a primeira publicação com descoberta em larga escala de miRNAs para uma espécie da família Myrtaceae (Guzman, Almerao *et al.*, 2012). Estes dados publicados permitiram uma análise comparativa dos dados de sequenciamento de *Eucalyptus* e *Eugenia* para avaliar a extensão da conservação de pequenos RNAs entre esses dois gêneros da família Myrtaceae. Vale destacar, entretanto, que o sequenciamento de *Eucalyptus* foi realizado em plataforma Illumina GAII enquanto os dados de *Eugenia* foram gerados em plataforma Illumina HiSeq2000, o que significa uma quantidade maior de dados gerados para a segunda espécie. Para efeito comparativo, o número de sequências únicas de *Eucalyptus* totalizou 1.857.986 e o de *Eugenia uniflora* foi pouco mais que o dobro, 4.010.705.

A análise das sequências de *Eugenia uniflora*, disponíveis no banco de dados GEO (*Gene Expression Omnibus*) hospedado no site do NCBI, mostrou que, do total de sequências únicas, 1.392.334 (34,7%) mapearam, com até duas diferenças de bases, no genoma de *E. grandis*. A observação do total de sequências mapeadas por tamanho, em nucleotídeos, das principais classes de pequenos RNAs mostra maior conservação de sequências de 21 nucleotídeos e menor conservação de sequências de 24 bases (Tabela 12). A mesma tendência pode ser observada considerando as sequências comuns entre as duas espécies por classe de tamanho. A porcentagem de sequências comuns de 21 nucleotídeos é 10 vezes maior considerando 100% de identidade e cerca de 7 vezes maior considerando isoformas com até uma base distinta.

Os dados revelam assim um elevado nível de conservação, especialmente, da fração de 21 nucleotídeos em grande parte, provavelmente, explicada pela pressão de seleção, principalmente para genes MIR conservados. O nível de conservação deve ser ainda maior por se tratar de espécies próximas, pertencentes a uma mesma família. Com isso, espera-se que, além dos miRNAs amplamente conservados em plantas, outros sejam conservados dentro da família Myrtaceae como foi anteriormente observado na análise de *northern blot* para algumas sequências de pequenos RNAs de 21 nucleotídeos (euc\_sRNA\_149582 e

euc\_sRNA\_75850), considerando também outras classes como siRNAs (Figura 12 – Validação experimental – *northern blot*).

Sugere-se que a menor conservação de siRNAs ocorra por sofrerem menor pressão seletiva para conservação de suas sequências que miRNAs, mostrando maior capacidade de adaptação. Este fenômeno provavelmente decorre do fato de siRNAs apresentarem mais comumente alvos relacionados, muitas vezes silenciando o próprio loco de origem, atuando em *cis*, muitos relacionados com sistemas como os de defesa, que requerem alta plasticidade e adaptabilidade, o que, em última instância, resulta em uma elevada taxa evolutiva.

Por outro lado, miRNAs, normalmente, silenciam alvos não relacionados e estão muito envolvidos em processos básicos como desenvolvimento e morfologia. Uma vez que o silenciamento depende da complementariedade a alvos heterólogos, estes pequenos RNAs estão mais sujeitos a pressão seletiva, mantendo, assim, maior conservação de sequências (Carthew e Sontheimer, 2009). A observação de uma maior conservação de pequenos RNAs de 22 nt (em sua maioria miRNAs), além dos sRNAs de 21 nt, entre taxa distintos, em comparação aos de 24 nt corrobora essa hipótese (Tabela 13).

Tabela 12: Número total de sequências únicas de *Eugenia uniflora* e de sequências de mapeadas no genoma de *E. grandis* por classe de tamanho em bases.

| <i>Número de bases</i> | <i>Total de sequências únicas</i> | <i>Número de sequências mapeadas</i> |
|------------------------|-----------------------------------|--------------------------------------|
| 21                     | 572.311                           | 352.691 (61,6%)                      |
| 22                     | 282.335                           | 121.025 (42,9%)                      |
| 24                     | 2.436.699                         | 514.683 (21,1%)                      |

Tabela 13: Número de sequências únicas por classe de tamanho, em nucleotídeos (nt), de *Eucalyptus* e *Eugenia* e número de sequências comuns entre ambos os taxa com 100% de identidade e com diferença de uma base.

|                            | <i>21nt</i>    | <i>22nt</i>    | <i>24nt</i>   |
|----------------------------|----------------|----------------|---------------|
| <i>Eucalyptus</i>          | 169.661        | 94.027         | 1.230.102     |
| <i>Eugenia</i>             | 572.311        | 282.335        | 2.436.699     |
| <i>Comum - 0 diferença</i> | 7.018 (4,1%)   | 4.742 (5,0%)   | 5.231 (0,4%)  |
| <i>Comum - 1 diferença</i> | 19.790 (11,7%) | 11.398 (12,1%) | 21.089 (1,7%) |

## CONCLUSÕES

O uso de tecnologias de sequenciamento em larga escala impulsionou a identificação de miRNAs conservados e a descoberta de miRNAs linhagem específicos que, possivelmente, seriam de difícil detecção via metodologias de clonagem e sequenciamento Sanger por serem expressos em menor quantidade. Como essas tecnologias se tornam cada vez mais acessíveis, tornou-se possível também a identificação de miRNAs expressos apenas em tecidos ou condições específicas. Contudo, a obtenção de milhares de sequências de pequenos RNAs por sequenciamento em larga escala gerou um desafio adicional considerável para uma correta classificação de miRNAs, outros pequenos RNAs e transcritos espúrios, parte derivados de degradação de outros RNAs (Kozomara e Griffiths-Jones, 2011).

Os números totais de sequências obtidas mostraram ainda a cautela e rigor analítico necessários para uma análise robusta de dados de sequenciamento de pequenos RNAs. Dentre os números obtidos, grande parte é representada por contaminantes como RNA ribossômico e de cloroplasto. Dentre os pequenos RNAs regulatórios, diferentes classes e subclasses estão representadas e os critérios para declarar uma sequência de pequeno RNA como efetivamente sendo um miRNA devem ser cuidadosamente observados.

A identificação de miRNAs conservados é facilitada pela busca por similaridade. Contudo, a busca por novos candidatos deve ser especialmente criteriosa. Os filtros usados com esse intuito seguiram os critérios recomendados por especialistas (Meyers, Axtell *et al.*, 2008) para a identificação de miRNAs: 1) mapeamento no genoma, 2) estrutura secundária do precursor, 3) alvos preditos, 4) validação experimental via microarranjo.

A busca por miRNAs conservados mostrou 95 isoformas absolutamente conservadas (sem nenhuma diferença de base) pertencente a vinte e cinco famílias. Incluindo novas isoformas com até três diferenças de bases, um total de 470 isoformas em 74 famílias foram identificadas. Para efeito de comparação, os dados de miRNA no miRBase de *A. thaliana*, que é a espécie modelo com sequência genômica amplamente anotada e onde a maior parte dos experimentos de descoberta e validação de miRNAs foi realizada, apresenta 338 sequências de miRNAs maduros em mais de 100 famílias.

Os potenciais novos miRNAs de *Eucalyptus* somam 178 sequências mapeadas no genoma e com estrutura secundária compatível. Destes, 163 tiveram alvos preditos e 14 foram validados no microarranjo e possuem alvos preditos, sendo potencialmente novos genes MIR de *Eucalyptus*.

A abordagem de descoberta e análise descrita neste trabalho se beneficiou consideravelmente da disponibilidade do genoma de referência de *E. grandis* e de tecido derivado da mesma árvore que forneceu o genoma de referência. Além da utilização desta amostra de *E. grandis*, foram estudadas duas árvores de *E. globulus* em vista da grande importância econômica desta espécie. A escolha por xilema como o tecido referência para a descoberta destes RNAs foi fundamentada no fato deste ser o tecido responsável pela formação da madeira, o principal produto do gênero *Eucalyptus* e alvo de interesse de manipulação via tecnologias de DNA recombinante. A análise de mapeamento das sequências de ambas as espécies confirmou a existência de variabilidade interespecífica de sequências de pequenos RNAs. Embora sejam espécies próximas, pertencentes ao mesmo gênero, o repertório de pequenos RNAs expressos apresentou sequências, aparentemente, espécie específicas e, portanto, de origem recente.

Uma análise de genômica comparativa de pequenos RNAs com a espécie *Eugenia uniflora*, pertencente à mesma família Myrtaceae, corrobora a alta conservação entre espécies relacionadas. Os dados mostram que a maior parte dos pequenos RNAs conservados refere-se aos de 21 nucleotídeos. Acredita-se que estes sejam, majoritariamente, miRNAs que, por suas características de atuação, sofrem maior pressão seletiva para a manutenção de suas sequências que as demais classes de pequenos RNAs o que explicaria assim a elevada conservação.

A descoberta de miRNAs novos e conservados em *Eucalyptus* contribui com a anotação do genoma de *E. grandis* adicionando informações sobre pequenos RNAs regulatórios e abrindo perspectivas de estudos experimentais pontuais para a elucidação de suas funções biológicas e do participação no controle da variação quantitativa em fenótipos de interesse industrial e adaptabilidade a estresses bióticos e abióticos. Adicionalmente, este foi o primeiro experimento para a criação de um microarranjo customizado de pequenos RNAs do gênero *Eucalyptus*. Esta ferramenta abre a perspectiva de realização de análises coletivas de pequenos RNAs em *Eucalyptus* e espécies relacionadas, abrindo as portas para investigação detalhada de seu papel funcional.

## **CAPÍTULO II**

*Identificação de pequenos RNAs em fase no genoma  
de Eucalyptus grandis*

## INTRODUÇÃO

Desde a recente descoberta da participação de pequenos RNAs não codificadores como elementos regulatórios, várias classes de sRNAs foram descritas. Diferentes classificações já foram propostas baseando-se na origem e em mecanismos de atuação. Na verdade, como bem colocado por Michael Axtell em sua recente revisão sobre classificação de pequenos RNAs em plantas, qualquer classificação sobre pequenos RNAs é uma construção intelectual que não reflete perfeitamente a realidade (Axtell, 2013). De maneira geral, miRNAs e siRNAs são consideradas as duas maiores e principais classes de pequenos RNAs em plantas. MiRNAs e siRNAs são, usualmente, considerados classes de pequenos RNAs não codificadores independentes, com mecanismos regulatórios e origens distintas.

A regulação pós-transcricional mediada por miRNAs, usualmente, ocorre, em última instância, por redução dos níveis da proteína traduzida pelo RNA mensageiro alvo, de forma que, quando o miRNA está ativo, o mRNA mensageiro alvo é silenciado. Contudo, estes elementos se mostram versáteis, podem participar de vias auto-regulatórias e são capazes de mecanismos mais sutis de regulação (Voinnet, 2009). Além disso, recentemente, foi mostrado que a biossíntese de siRNAs pode ser dependente da atuação de miRNAs (Yoshikawa, Peragine *et al.*, 2005; Talmor-Neiman, Stav *et al.*, 2006; Howell, Fahlgren *et al.*, 2007).

Recentemente, foi demonstrada a convergência entre as duas principais classes de pequenos RNAs em plantas pela observação da dependência da síntese de uma classe de siRNAs secundários induzida por miRNAs, em um mecanismo designado transitividade (*transitivity*), onde miRNAs desencadeiam a biossíntese de *trans-acting* siRNAs (tasiRNAs) ou RNAs em fase (*phased* RNAs) (Peragine, Yoshikawa *et al.*, 2004; Yoshikawa, Peragine *et al.*, 2005).

Este é mais um exemplo de quão ampla e complexa pode ser a rede de regulação da expressão gênica. Mecanismos e elementos, inicialmente, considerados paralelos e independentes são, na realidade, interdependentes e complementares, muitas vezes

apresentam pontos de redundância e podem fornecer indícios da história evolutiva de famílias gênicas.

A identificação desses novos elementos de regulação da expressão gênica representa mais um avanço na compreensão da atuação destes elementos e pode fornecer valiosas informações sobre a conservação e evolução desse mecanismo.

Neste capítulo o trabalho de descoberta e caracterização de miRNA em *Eucalyptus* foi expandido para a identificação de pequenos RNAs produzidos em fase no genoma de *Eucalyptus grandis*. A disponibilidade de dados de sequenciamento em larga escala de pequenos RNAs e da sequência de referência do genoma do mesmo indivíduo de *Eucalyptus grandis* permitiu o desenvolvimento desse tipo de estudo que envolve a identificação *in silico* de agrupamentos (*clusters*) de sequências e siRNA em fase no genoma.

## **REVISÃO BIBLIOGRÁFICA**

### ORIGEM DOS RNAs EM FASE

O mecanismo designado transitividade (*transitivity*), isto é, quando miRNAs desencadeiam a biossíntese de *trans-acting* siRNAs (tasiRNAs) ou *phased* RNAs (phasiRNAs), foi recentemente elucidado (Peragine, Yoshikawa *et al.*, 2004). Este não é um mecanismo padrão para todos os transcritos alvo de miRNAs, porém, alguns poucos mRNAs, designados transcritos TAS, são processados em fase para a produção de vários siRNAs secundários de 21 nt. SiRNAs secundários são aqueles cuja síntese depende de um outro pequeno RNA, usualmente um miRNA, atuando anteriormente para desencadear sua produção, dependente de RNA polimerase 6 dependente de RNA (RDR6).

A biossíntese de RNAs em fase envolve: 1) a clivagem de um mRNA alvo (TAS) por um sRNA inicial, usualmente um miRNA; 2) a estabilização do mRNA clivado por SGS3 (*supressor of gene silencing*); 3) a síntese de RNA fita dupla, a partir do fragmento 3' do mRNA alvo, pela RDR6; 4) este RNA fita dupla é processado em fase pela enzima

DCL4 em siRNAs de 21 nucleotídeos (Allen, Xie *et al.*, 2005) (Figura 17). Os tasiRNAs assim sintetizados silenciam genes alvo guiando proteínas argonautas (AGO) associadas a ele, como no mecanismo de silenciamento por miRNAs ou siRNAs. A incorporação dos tasiRNAs no complexo efetor AGO, também obedece à mesma regra observada para miRNAs, onde sequências com 5'uridina se associam, preferencialmente, à AGO1 (Mi, Cai *et al.*, 2008). A maquinaria envolvida na geração destes tasiRNAs é, então, distinta daquela envolvida na síntese de outros siRNAs e miRNAs e, consistentemente, dependente de RDR6, SGS3 e DCL4 (Peragine, Yoshikawa *et al.*, 2004; Yoshikawa, Peragine *et al.*, 2005; Talmor-Neiman, Stav *et al.*, 2006; Howell, Fahlgren *et al.*, 2007; Cuperus, Carbonell *et al.*, 2010).

Sumarizando, miRNAs desencadeiam uma reação em cascata produzindo fitas duplas de RNAs, a partir de alguns transcritos específicos, e a clivagem destas em blocos consecutivos de 21 nucleotídeos dá origem aos siRNAs secundários, os tasiRNAs. Esta é uma subclasse onde, em teoria, o siRNA teria como alvo um transcrito não relacionado (em *trans*) criando uma complexa rede de regulação (Axtell, 2013).

Este mecanismo gera, muitas vezes, múltiplos siRNAs quase idênticos que tem como alvo um mesmo transcrito reforçando a intensidade de silenciamento e, ao mesmo tempo, garantindo a especificidade ao transcrito alvo (Zhai, Jeong *et al.*, 2011).

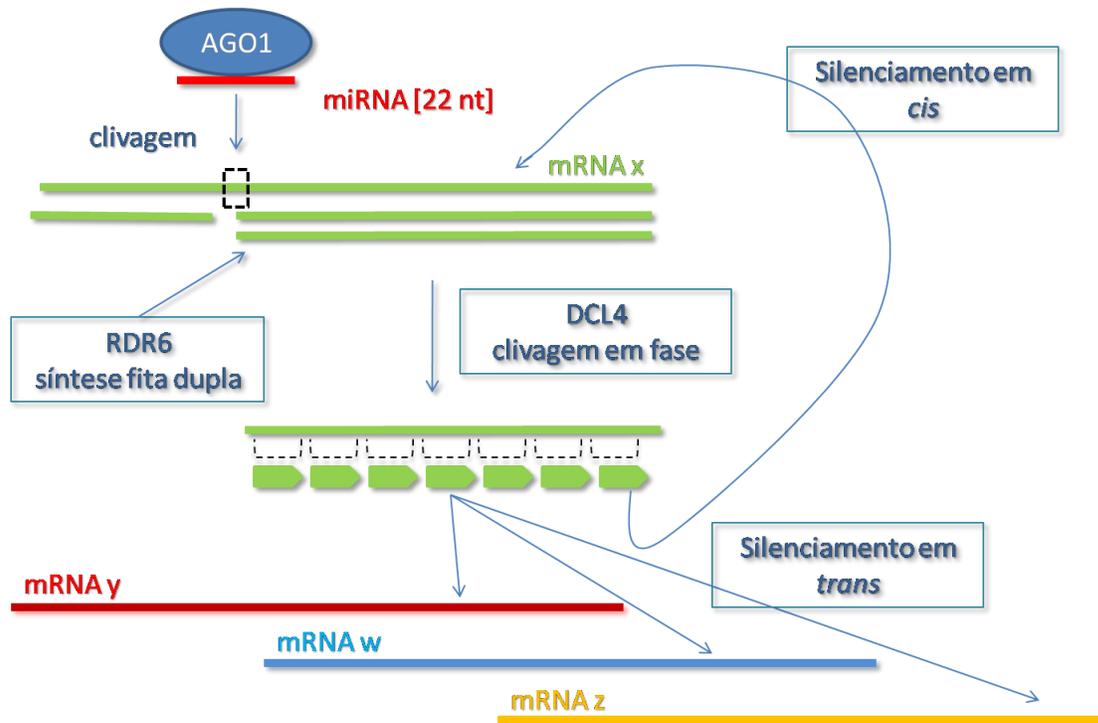


Figura 17: Representação esquemática da biossíntese de siRNAs secundários em fase a partir do processamento de um mRNA alvo de um miRNA (x) de 22 nt e do silenciamento em *cis* e em *trans* de genes alvo de siRNAs em fase.

## REDE DE REGULAÇÃO DOS RNAs EM FASE

Este mecanismo foi, originalmente, descrito em *Arabidopsis thaliana*. Quatro famílias de transcritos TAS foram descritas e validadas nesta espécie. TAS1, TAS2 e TAS4 são processados a partir de um único sítio de clivagem por um miRNA enquanto TAS3, a partir da clivagem em dois sítios do transcrito gerador (Allen e Howell, 2010). TAS1 e TAS2 são alvos de miR173 e seus tasiRNAs tem como alvo transcritos de proteínas com repetição pentatricopetídeo (PPR) (Allen, Xie *et al.*, 2005; Yoshikawa, Peragine *et al.*, 2005; Montgomery, Yoo *et al.*, 2008). Proteínas com essa repetição degenerada de 35 aminoácidos foram identificadas em um pequeno número de genes em genomas de eucariotos. No entanto em plantas, esta família sofreu grande expansão. A alta variabilidade observada em ortólogos desta família sugere que uma ativa dinâmica

evolutiva de seus membros com alta taxa de aparecimento/desaparecimento (*birth and death*), assim como ocorre em outras famílias gênicas expandidas, como os genes análogos de resistência, os *Resistance Gene Analogs* (Geddy e Brown, 2007). Apesar da função dos membros da família PPR não estar definida sugeriu-se que os tasiRNAs atuariam como elementos regulatórios tamponantes para contrabalançar a sua rápida expansão (Howell, Fahlgren *et al.*, 2007).

O miR828 regula o gene MYB113 e também o transcrito TAS4, cujos tasiRNAs também regulam MYB113 e outros membros desta família de fatores de transcrição (Rajagopalan, Vaucheret *et al.*, 2006). TAS4 está relacionado com regulação da biossíntese de flavonóides, antocianinas e fenilpropanóides, e, possivelmente, com senescência foliar (Allen, Xie *et al.*, 2005; Hsieh, Lin *et al.*, 2009; Luo, Mittal *et al.*, 2011).

TAS3 é alvo de miR390 e o processamento de tasiRNAs a partir deste transcrito obedece a um mecanismo distinto, onde dois sítios de ligação ao miRNA flanqueiam os tasiRNAs, contudo, somente o sítio 3' tem indícios de clivagem. Todos os TAS3 identificados geram, pelo menos, um tasiRNA maduro que silencia um dos fatores de resposta à auxina, ARF-2, ARF-3 ou ARF-4 (Allen, Xie *et al.*, 2005; Williams, Carles *et al.*, 2005; Axtell, Jan *et al.*, 2006; Luo, Mittal *et al.*, 2011) que influenciam a transição de fase vegetativa (da fase juvenil para adulta), crescimento lateral de raiz e morfologia e polaridade de folha (Peragine, Yoshikawa *et al.*, 2004; Fahlgren, Montgomery *et al.*, 2006).

As características que determinam se um mRNA alvo será processado para a síntese de tasiRNAs não está bem elucidada. Entretanto, o tamanho do miRNA inicial parece ser um dos pré-requisitos básicos para o desencadeamento de síntese de siRNAs secundários em fase. MiRNAs de 22 nucleotídeos, ao invés do tamanho padrão de 21 bases, com uma assimetria no duplex miRNA-miRNA\*, decorrente da diferença de tamanho entre estes, seriam necessários para a biossíntese dos tasiRNAs. É proposto que a conformação assumida pelo complexo AGO1-miRNA de 22 nt seja distinta daquela assumida pelo complexo AGO1-miRNA de 21 nt. Esta conformação promoveria, além da clivagem do mRNA alvo, o recrutamento das proteínas envolvidas na biossíntese de tasiRNAs – RDR6

e SGS3 (Manavella, Koenig *et al.*, 2012). A exceção a esse padrão é o miR390 que, com 21 nt, se liga à AGO7 e cliva TAS4 gerando tasiRNAs. Entretanto, destaca-se o fato de, neste caso, o mecanismo ser diferenciado, com dois sítios de ligação do miRNA no transcrito alvo (Axtell, Jan *et al.*, 2006; Montgomery, Howell *et al.*, 2008).

Apesar de, em geral, o miRNA gerador possuir 22 nt os tasiRNAs possuem um tamanho padrão de 21 bases o que, aparentemente, é crítico para sua funcionalidade, apesar de algum desvio na fase de síntese dos vários tasiRNAs ter sido observado (Howell, Fahlgren *et al.*, 2007).

Embora poucos tasiRNAs tenham sido identificados em plantas desde a descoberta destes em *A. thaliana* (Allen, Xie *et al.*, 2005), este parece ser um mecanismo de origem remota. A identificação de TAS3 e miR390 na briófito *Physcomitrella patens* indica um mecanismo altamente conservado de biossíntese de tasiRNAs em plantas (Axtell, Jan *et al.*, 2006; Talmor-Neiman, Stav *et al.*, 2006; Howell, Fahlgren *et al.*, 2007). MiR472 e os tasiRNAs secundários gerados por ele são conservados em *Populus* e em outras espécies cultivadas (Allen e Howell, 2010). A conservação de siRNAs secundários em angiospermas, mesmo em linhagens pouco relacionadas (Axtell, Jan *et al.*, 2006; Talmor-Neiman, Stav *et al.*, 2006), confirma estes elementos como uma classe robusta de pequenos RNAs regulatórios (Axtell, 2013). O par TAS3-miR390, altamente conservado, foi também identificado por análise *in silico* em *V. vinifera*. Contudo, os alvos de regulação pelos tasiRNAs de *V. vinifera* compreendem, além de genes envolvidos em desenvolvimento como proteínas envolvidas com senescência e fatores de resposta a auxina (ARFs – *Auxin Response Factors*), como previamente descrito em *A. thaliana* (Allen, Xie *et al.*, 2005), outros genes envolvidos em vias bioquímicas básicas – como transportadores de cálcio, enzimas envolvidas no metabolismo de celulose e ácidos graxos – e em genes de defesa em resposta à patógenos (Zhang, Li *et al.*, 2012). A identificação de novos tasiRNAs específicos de *Vitis vinifera*, *Oryza sativa* e *Brachipodium* indicam também a existência de siRNAs secundários de origem mais recente (Shivaprasad, Chen *et al.*, 2012).

Além da conservação observada, a grande variedade de alvos de regulação por tasiRNAs, envolvidos em diferentes processos biológicos, sugere que estes elementos regulatórios tenham papel fundamental no desenvolvimento de plantas (Zhang, Li *et al.*, 2012). Por exemplo, foi relatado que os tasiRNAs podem se movimentar entre células e, dessa forma, estabelecer gradientes de silenciamento que são necessários para a determinação da polaridade da folha por TAS3 (Chitwood, Nogueira *et al.*, 2009; Schwab, Maizel *et al.*, 2009).

## OS RNAs EM FASE E GENES DE RESISTÊNCIA A DOENÇAS

Transcritos TAS, como no caso de *Arabidopsis thaliana*, são mRNAs não codificadores. Entretanto, existem exemplos de transcritos codificadores de proteína gerando tasiRNAs. Entre estes estão membros da família de genes de resistência a doenças com a tríade de domínios protéicos TIR-NBS-LRR (*Toll-Interleukin receptor – nucleotide binding site – leucine rich repeat*), transcritos de proteínas *F-box*, envolvidas na sinalização por auxina, transcritos de transportadores cátion/hidrogênio e da família de proteínas PPR (Howell, Fahlgren *et al.*, 2007).

A identificação de tasiRNAs gerados a partir de genes de defesa e resistência a patógenos, particularmente genes da família NBS-LRR, também parece ser um mecanismo conservado já identificado em *Medicago truncatula*, batata e tomate (Zhai, Jeong *et al.*, 2011; Shivaprasad, Chen *et al.*, 2012). Em *M. truncatula*, 79 dos 112 locos identificados como *phasi* locos, ou geradores de transcritos TAS, são de genes NBS-LRR (Zhai, Jeong *et al.*, 2011). Em tomate, 58 dos 186 genes NBS-LRR tem sítios de ligação ao miR482, o agente desencadeador principal da produção de tasiRNAs em genes de resistência (Bonnet, He *et al.*, 2010).

A família miR482 tem como alvo genes NBS-LRR. Esta família é incomum pela presença predominante de miRNAs de 22 nt e pelo grande número de isoformas, totalizando 31 em espécies distintas. A família pode ainda ser expandida para incluir miR472 e miR2118. A família miR2118 apresenta o mesmo padrão de variação de

sequência e, acredita-se, que a superfamília miR482/miR2118 tenha o potencial de regular a expressão de vários genes NBS-LRR e atuem como reguladores fundamentais da resistência a doenças em tomate (Shivaprasad, Chen *et al.*, 2012). Uma outra observação é que a maior parte das 31 isoformas de mir482 é espécie específica, sendo provável que estas tenham origem bastante recente (Shivaprasad, Chen *et al.*, 2012).

Plantas apresentam um vastíssimo número de genes de receptores relacionados com imunidade inata. A expressão destes, embora ofereça proteção contra patógenos invasores, têm um alto custo para a planta pelo grande número de genes e o potencial de desencadear morte celular. Sugere-se, então, que mecanismos de silenciamento pós-transcricionais tenham co-evoluído para balancear o alto custo gerado pela expansão e diversificação dessas famílias de genes (Howell, Fahlgren *et al.*, 2007; Li, Pignatta *et al.*, 2012). Uma vez que bactérias e vírus possuem efetores para suprimir o silenciamento por miRNAs e siRNAs (Li e Ding, 2006; Navarro, Jay *et al.*, 2008), propõe-se que este mecanismo, além de favorecer a expansão do patógeno, contribua para o aumento da expressão de genes de resistência, criando um balanço entre resistência e susceptibilidade. Por tudo isso, sugere-se que miRNAs que atuam em genes de resistência sejam responsáveis por uma regulação fina da expressão destes ao longo da co-evolução de patógeno e hospedeiro (Li, Pignatta *et al.*, 2012).

A participação de silenciamento de RNA na expressão de genes de defesa abre um novo capítulo na interação entre patógeno e hospedeiro com mecanismos de defesa e contra-defesa. Em infecção por vírus e, em alguns casos, também por bactérias o silenciamento por RNA (RNAi) seria o primeiro mecanismo de defesa da planta. Os patógenos também possuem um arsenal para suplantar a defesa celular como no caso dos supressores virais (Voinnet, 2005a). O sistema coordenado por miR482 seria um exemplo de contra contra-defesa (Shivaprasad, Chen *et al.*, 2012). Um potencial novo sistema de defesa seria a existência de genes NBS-LRR truncados que funcionariam como sequestradores de miR482/miR2118, funcionando como “imitadores de alvo” (*target mimics*), proporcionando mais um nível de regulação fina da supressão de genes de defesa (Meyers, Kozik *et al.*, 2003).

Especula-se que a grande variação de sequência e níveis de expressão observados na superfamília miR482/miR2118 mostrem o balanço de custo e benefício em diferentes plantas dependendo do ambiente e de outros mecanismos de defesa (Shivaprasad, Chen *et al.*, 2012) que, em conjunto, constituem uma vasta e complexa rede.

Uma descoberta surpreendente foi a de que, em soja, SGS3, elemento fundamental da biossíntese de tasiRNAs, é também alvo de miR2118 (Zhai, Jeong *et al.*, 2011). O que, provavelmente, resultaria em um efeito tamponante da produção de tasiRNAs a partir de genes NBS-LRR.

PPR e ARF são regulados, independentemente, por siRNAs secundários e alguns miRNAs e, por isso, sugere-se que tasiRNAs reforcem ou coordenem a ação de outros siRNAs ou miRNAs (Shivaprasad, Chen *et al.*, 2012). Este tipo de grande rede de regulação envolvendo tasiRNAs, aparentemente, não está presente em *Arabidopsis thaliana* (Zhai, Jeong *et al.*, 2011).

Embora, por definição, tasiRNAs tenham como alvo transcritos não relacionados (*trans*), exemplos, como os citados anteriormente com genes de defesa NBS-LRR (Zhai, Jeong *et al.*, 2011; Shivaprasad, Chen *et al.*, 2012) e fatores de transcrição MYB, corroboram a existência de tasiRNAs com ação de silenciamento em *cis*.

A biossíntese de tasiRNAs parece ser um sistema eficaz de supressão em plantas, seja pelo estabelecimento de um gradiente de silenciamento, como no caso de ARFs, ou por regulação de grandes famílias gênicas, como no caso de PPR e NBS-LRR. Este pode ser um mecanismo de controle e amplificação de pequenos RNAs enquanto transforma um sinal de silenciamento célula específico (miRNA) em um sinal que se propaga criando um gradiente entre células (Allen e Howell, 2010).

A identificação *in silico* de tasiRNAs a partir do genoma é uma ferramenta poderosa (Chen, Li *et al.*, 2007; Howell, Fahlgren *et al.*, 2007; Zhai, Jeong *et al.*, 2011) uma vez que este é o único mecanismo conhecido de síntese de pequenos RNAs em fase (Axtell, 2013). Um dos critérios usados na predição *in silico* de grupos de siRNAs em fase é a identificação de, pelo menos, três siRNAs em 100 nucleotídeos (Johnson, Kasprzewska *et al.*, 2009). A disponibilidade de dados de sequenciamento em larga escala de pequenos

RNAs e de sequências de genomas de referência favorece o desenvolvimento desse tipo de estudo.

## **METODOLOGIA**

### SEQUENCIAMENTO

Os dados de sequenciamento de pequenos RNAs de folha e xilema de *E. grandis*, árvore BRASUZ1, e xilema de *E. globulus*, árvores A2 e C3, descrito no capítulo I foram utilizados na identificação de tasiRNAs.

### MAPEAMENTO E IDENTIFICAÇÃO DE siRNAs EM FASE

As sequências de pequenos RNAs de 21 bases foram mapeadas contra o genoma de referência de *E. grandis* (árvore BRASUZ 1) para identificação dos *locos* de siRNAs em fase. Para tanto, foi utilizado um algoritmo desenvolvido para identificar grupos de pequenos RNAs de 21 bases em fase conforme descrito anteriormente (Chen, Li *et al.*, 2007).

### PREDIÇÃO DE ALVOS

A predição de mRNAs alvo de siRNAs em fase foi realizada, como descrito no capítulo I, utilizando o programa psRNATarget contra o banco de transcritos de *E. grandis* do Phytozome. A função potencial desses tasiRNAs foi inferida pelos domínios protéicos presentes nos transcritos, identificados por meio da ferramenta BioMart (Phytozome).

## **RESULTADOS**

O sequenciamento de pequenos RNAs foi realizado multiplexando quatro amostras para a corrida em uma linha (*lane*) do sequenciador Illumina GA II. As sequências foram então mapeadas contra o genoma referência de *E. grandis*, conforme descrito no capítulo 1, e as leituras de 21 nucleotídeos foram usadas na busca por tasiRNAs. Foi utilizado um algoritmo desenvolvido para a identificação de pequenos RNAs em fase sem informação prévia de sítios de clivagem de miRNAs (Chen, Li *et al.*, 2007).

Foram identificados 173 agrupamentos de pequenos RNAs, considerando aqueles que se sobrepõem em uma mesma região genômica com fases de síntese distintas e em ambas as fitas. A partir destes agrupamentos, procedeu-se uma filtragem para identificar os grupos presentes dentro de regiões gênicas resultando em um total de 13 locos. Interessantemente, alguns dos pequenos RNAs de 21 nt que estão entre os mais abundantes nos dados de sequenciamento são justamente oriundos destes 13 locos.

Entre os 13 transcritos preditos como produtores de sRNAs em fase, três não tem função predita. São, potencialmente, não codificadores e se comportariam como transcritos TAS, inicialmente descritos em *A. thaliana* (Allen, Xie *et al.*, 2005; Williams, Carles *et al.*, 2005; Yoshikawa, Peragine *et al.*, 2005). Um deles é um fator de transcrição da família MYB, outro um fator de transcrição *X1-like* e sete são relacionados com resistência a doenças (Tabela 14). A predominância de locos de genes NBS-LRR (domínios NB-ARC e repetição *leucine rich*) de resistência a doenças entre os preditos como produtores de siRNAs em fase recapitula o que foi anteriormente observado em *Medicago truncatula* e tomate (Zhai, Jeong *et al.*, 2011; Shivaprasad, Chen *et al.*, 2012).

Em quatro desses transcritos, foi possível identificar sítios de ligação a pequenos RNAs de 22 nt descobertos na amostragem de sequenciamento, sugerindo que estes seriam os miRNAs desencadeadores da síntese de siRNAs em fase. Todos esses quatro genes são relacionados com imunidade inata e resistência a patógenos – Eucgr.C04268, Eucgr.G00363, Eucgr.G00368 e Eucgr.H03699 segundo nomenclatura do Phytozome (Tabela 14). Entretanto, nenhum desses candidatos a miRNAs de 22 nt tem um ortólogo no

miRBase e podem ser, potencialmente, linhagem-específicos e, portanto, de origem recente, acompanhando a dinâmica natureza evolutiva da família de genes NBS-LRR.

Tabela 14: Anotação dos genes de *E. grandis* preditos como produtores de siRNAs secundários em fase. Nome do gene segundo o Phytozome, código (ID PFAM), nome do domínio protéico e descrição segundo PFAM e Panther.

| <i>Nome do gene</i> | <i>ID PFAM</i> | <i>Domínio protéico</i>                              | <i>Descrição PFAM</i>                         | <i>Descrição Panther</i>         |
|---------------------|----------------|--|---|----------------------------------|
| Eucgr.A01539        |                |  |   |                                  |
| Eucgr.B00373        | PF00249        | domínio de ligação a DNA Myb-like                    | domínio de ligação a DNA Myb-like             | relacionado a MYB                |
| Eucgr.B02765        |                |  |   |                                  |
| Eucgr.C02141        | PF01582        | domínio homólogo a receptor Toll/interleukin-1 (TIR) | resposta imune inata                          | -                                |
| Eucgr.C02142        | PF01582        | domínio homólogo a receptor Toll/interleukin-1 (TIR) | resposta imune inata                          | -                                |
| Eucgr.C02857        | PF07649        | domínio C1   | proteína com atividade de dissulfeto redutase | -                                |
| Eucgr.C03843        | PF01582        | domínio homólogo a receptor Toll/interleukin-1 (TIR) | resposta imune inata                          | -                                |
| Eucgr.C04268        | PF00931        | domínio NB-ARC                                       | apoptose                                      | proteína de resistência a doença |
| Eucgr.C04268        | PF01582        | domínio homólogo a receptor Toll/interleukin-1 (TIR) | resposta imune inata                          | proteína de resistência a doença |
| Eucgr.G00363        | PF00931        | domínio NB-ARC                                       | apoptose                                      | proteína de resistência a doença |
| Eucgr.G00368        | PF00931        | domínio NB-ARC                                       | apoptose                                      | proteína de resistência a doença |
| Eucgr.G00368        | PF01582        | domínio homólogo a receptor Toll/interleukin-1 (TIR) | resposta imune inata                          | proteína de resistência a doença |

| <i>Nome do gene</i> | <i>ID PFAM</i> | <i>Domínio protéico</i>                              | <i>Descrição PFAM</i>         | <i>Descrição Panther</i>                    |
|---------------------|----------------|--|-------------------------------|---|
| Eucgr.G01952        |                | -  | -                             | proteína com repetição leucine rich / F-BOX |
| Eucgr.H03699        | PF00931        | domínio NB-ARC                                       | apoptose                      | proteína com repetição <i>leucine rich</i>  |
| Eucgr.H03699        | PF01582        | domínio homólogo a receptor Toll/interleukin-1 (TIR) | resposta imune inata          | proteína com repetição <i>leucine rich</i>  |
| Eucgr.H03699        | PF00560        | repetição <i>leucine-rich</i> (LRR)                  | ligação a proteína            | proteína com repetição <i>leucine rich</i>  |
| Eucgr.H05056        | PF03470        | domínio XS <i>zinc finger</i>                        | domínio XS <i>zinc finger</i> | relacionado a ribonuclease P subunidade P38 |
| Eucgr.H05056        | PF03470        | domínio XS <i>zinc finger</i>                        | domínio XS <i>zinc finger</i> | fator de transcrição X1-like                |

Vale destacar que alguns transcritos alvos de candidatos a miRNAs de 22 bases são também alvos de siRNAs em fase (como o euc\_sRNA\_33215) sugerindo uma regulação integrada da expressão gênica por tasiRNAs e miRNAs como observado anteriormente (Howell, Fahlgren *et al.*, 2007; Shivaprasad, Chen *et al.*, 2012).

Um dos potenciais transcritos TAS preditos, sem anotação funcional (Eucgr.B02765), e não conservado em outras espécies, produz cerca de 20 sRNAs de 21 nt em fase identificados em todas as amostras do sequenciamento (Figura 18). O mais abundante desses (euc\_sRNA\_4546), tem, nos dados de sequenciamento, de 323 contagens em xilema de *E. globulus* A2 a 2.674 em folha de *E. grandis* (Figura 19). Quatro alvos foram preditos para esse siRNA secundário: um fator de liberação da tradução (Eucgr.F03826), uma transferase (Eucgr.A02205), um transcrito com o domínio LRR (*leucine rich repeat*) e, o mais intrigante, o mesmo transcrito TAS que origina os siRNAs secundários (Tabela 15). Este é mais um exemplo de um mecanismo de auto-regulação já observado em *Arabidopsis* e em *Medicago truncatula* (Allen, Xie *et al.*, 2005; Zhai, Jeong *et al.*, 2011). Este silenciamento por siRNAs secundários atuando em *cis* funcionaria como um mecanismo de retroregulação da biossíntese dos siRNAs em fase abrindo uma nova vertente de atuação destes pequenos RNAs silenciadores. Por isso, foi sugerido que os

então chamados tasiRNAs que, por definição atuam em *trans*, fossem denominados phasiRNAs (siRNAs em fase) (Zhai, Jeong *et al.*, 2011) (Figura 20).

Tabela 15: Transcritos preditos como alvo do siRNA secundário euc\_sRNA\_4546, derivado de Eucgr.B02765, pelo programa psRNATarget com a descrição dos domínios protéicos identificados segundo PFAM.

| <i>Nome do transcrito</i> | <i>ID PFAM</i> | <i>Domínio protéico</i>             | <i>Descrição PFAM</i>  |
|---------------------------|----------------|-------------------------------------|--|
| Eucgr.A02205.1            | PF02458        | família transferase                 | atividade de transferase, transferindo grupos acyl exceto amino-acyl |
| Eucgr.B02765.1            |                |                                     |  |
| Eucgr.C00993.1            | PF08263        | repetição <i>leucine-rich</i> (LRR) | domínio repetição <i>leucine-rich</i> N-terminal                     |
| Eucgr.C00993.1            | PF00560        | repetição <i>leucine-rich</i> (LRR) | ligação a proteína   |
| Eucgr.F03826.1            | PF00472        | domínio RF-1                        | atividade de fator de liberação da tradução                          |

### Eucgr.B02765 - TAS

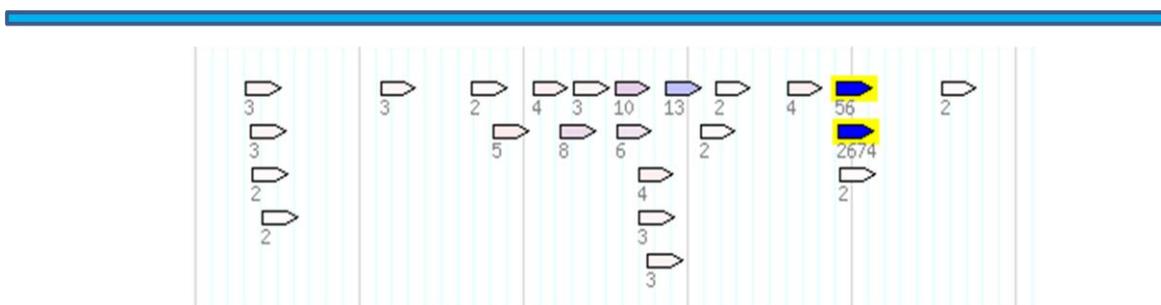


Figura 18: Visualização do mapeamento de pequenos RNAs de 21 nucleotídeos de folha de *E. grandis* BRASUZ1 em fase na região do gene Eucgr.B02765 do genoma referência de *E. grandis*. Em destaque (amarelo), os siRNAs mais abundantes no sequenciamento. Abaixo de cada siRNA, o número de seqüências obtidas no sequenciamento.

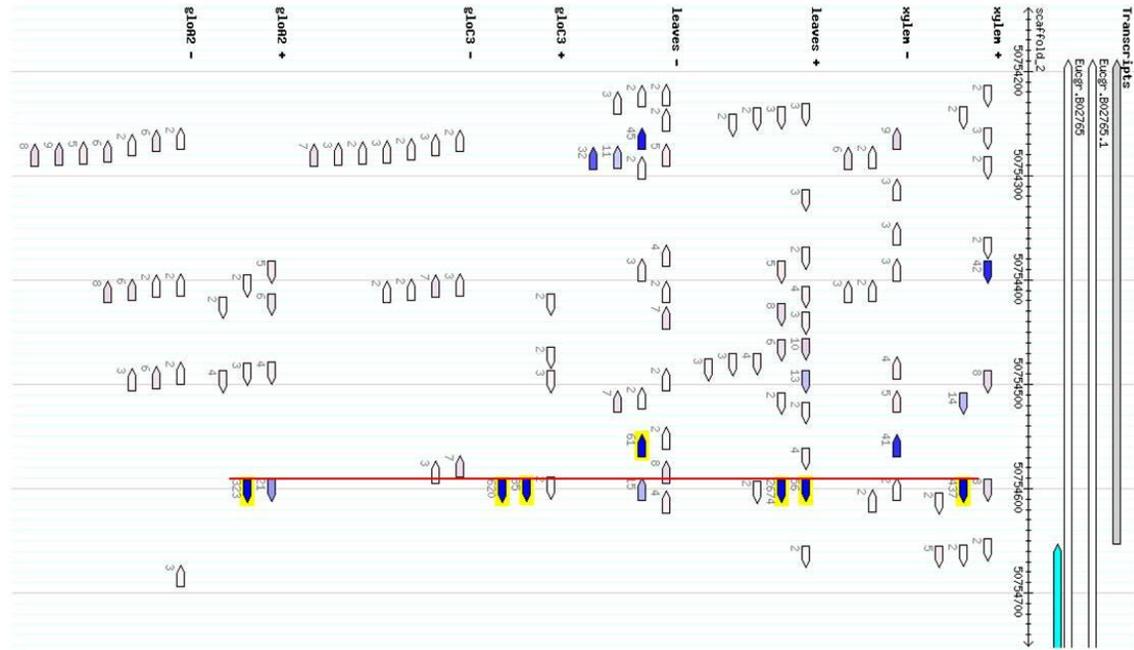


Figura 19: Mapeamento nas fitas senso (+) e antisenso (-) do genoma referência de *E. grandis*, na região do gene *Eucgr.B02765*, de pequenos RNAs de 21 nucleotídeos em fase oriundos do sequenciamento de quatro amostras: xilema (xylem) e folha (leaves) de *E. grandis* BRASUZ1 e xilema de *E. globulus* A2 e C3. Em destaque, os siRNAs mais abundantes no sequenciamento para cada uma das amostras. Abaixo de cada siRNA, o número de sequências obtidas no sequenciamento. O ponto inicial do *phasi RNA euc\_sRNA\_4546* foi marcado com linha vertical vermelha.

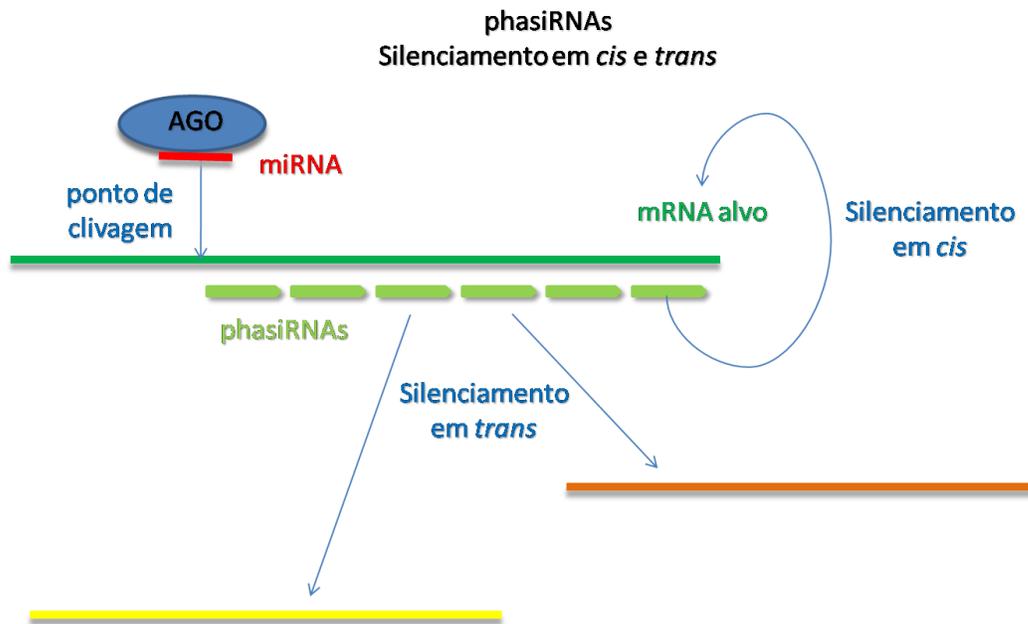


Figura 20: Representação esquemática da síntese de phasiRNAs e o silenciamento dirigido por estes em *cis* e em *trans*.

Outros exemplos de siRNAs em fase atuando em *cis* foram observados nos dados de *Eucalyptus*. O transcrito Eucgr.A01539, sem anotação funcional e específico para *Eucalyptus* de acordo com o Phytozome, possui um phasiRNA (euc\_sRNA\_30468) com 256 contagens na amostra de xilema de *E. grandis* atuando em *cis* no próprio transcrito primário. Interessante destacar que este phasiRNA não foi detectado em *E. globulus* C3 e nas demais amostras foi observado um número baixo de cópias – na amostra de xilema de *E. globulus* A2 (2 contagens) e folha de *E. grandis* (10 contagens).

Um fator de transcrição da família MYB (Eucgr.B00373) produz um siRNA em fase (euc\_sRNA\_766295) que tem como alvo o mesmo transcrito primário. Resultado similar foi observado em *Populus trichocarpa* (Tuskan, Difazio *et al.*, 2006). Em *Arabidopsis thaliana*, outro mecanismo de auto-regulação foi relatado envolvendo TAS4 e um grupo de fatores de transcrição MYB (PAP1, PAP2 e MYB113) que regulam a via de biossíntese de antocianina (Luo, Mittal *et al.*, 2011). Além do transcrito primário, dois outros alvos são preditos para o RNA em fase euc\_sRNA\_766295 (Tabela 16).

Tabela 16: Transcritos preditos como alvo do siRNA secundário euc\_sRNA\_766295, derivado de Eucgr.B00373, pelo programa psRNATarget com a descrição dos domínios protéicos identificados segundo PFAM.

| <i>Nome do transcrito</i> | <i>ID PFAM</i> | <i>Domínio protéico</i>                | <i>Descrição PFAM</i>              |
|---------------------------|----------------|--|------------------------------------|
| Eucgr.B00373.1            | PF00249        | domínio de ligação a DNA Myb-like      | domínio de ligação a DNA Myb-like  |
| Eucgr.B03556.1            | PF00076        | motivo de reconhecimento de RNA, RNP-1 | ligação a ácido nucleico           |
| Eucgr.B03556.1            | PF04059        | motivo 2 de reconhecimento de RNA      | motivo 2 de reconhecimento de RNA  |
| Eucgr.D00848.1            | PF01657        | Domínio sem função conhecida (DUF)     | Domínio sem função conhecida DUF26 |
| Eucgr.D00848.1            | PF07714        | tirosina kinase                        | fosforilação de proteína           |

Outros transcritos preditos como transcritos TAS são codificadores de proteínas. Um deles (Eucgr.C02141) codifica um gene de resistência com o domínio TIR (*Toll/Interleukin I like*). O mais abundante entre os siRNAs secundários derivados desse transcrito (euc\_sRNA\_33215) tem alta contagem nas amostras de xilema de *E. globulus* C3 e A2 (1.949 e 1.530, respectivamente) e folha de *E. grandis* (2.840) e, curiosamente, tem baixa contagem na amostra de xilema *E. grandis* (132). A análise de *Northern blot* do

siRNA secundário euc\_sRNA\_33215 em três espécies de *Eucalyptus* (*E. dunnii*, *E. urophylla* and *E. grandis*) mostra conservação no gênero mas, interessante, este parece não ser conservado em outras espécies de plantas e nem mesmo em duas espécies da família Myrtaceae testadas, *Corymbia citriodora* e *Eugenia uniflora* (Figura 21). Curiosamente, a expressão deste siRNA secundário parece ser, consistentemente, mais alta em folha do que em xilema das árvores adultas amostradas. Vários alvos foram preditos para este phasiRNA (Tabela 17). Dois são, supostamente, genes de resistência a doenças, envolvidos em imunidade inata, apoptose e com atividade de receptor transmembrana (Eucgr.E03126 and Eucgr.E03335).

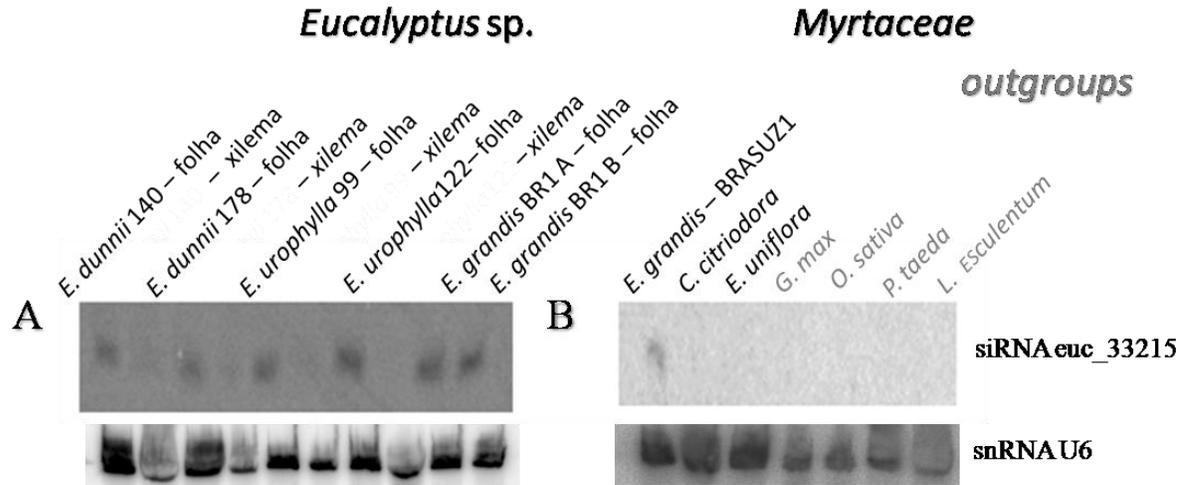


Figura 21: Northern blot do siRNA secundário euc\_sRNA\_33215. Em cada painel, da esquerda para a direita: (A) *E. dunnii* 140 – folha, *E. dunnii* 140 – xilema, *E. dunnii* 178 – folha, *E. dunnii* 178 – xilema, *E. urophylla* 99 – folha, *E. urophylla* 99 – xilema, *E. urophylla* 122 – folha, *E. urophylla* 122 – xilema, *E. grandis* BRASUZ1 a – folha e *E. grandis* BRASUZ1 b – folha; (B) Amostras de folha de *E. grandis* – BRASUZ1, *Corymbia citriodora*, *Eugenia uniflora*, *Glycine max*, *Oryza sativa*, *Pinus taeda* e *Solanum lycopersicum*. A hibridização do pequeno RNA nuclear snRNA U6 foi usada como controle de carregamento do gel.

Tabela 17: Transcritos preditos como alvo do siRNA secundário euc\_sRNA\_33215, derivado de Eucgr.C02141, pelo programa psRNATarget com a descrição dos domínios protéicos identificados segundo PFAM.

| <i>Nome do transcrito</i> | <i>ID PFAM</i>                | <i>Domínio protéico</i>  | <i>Descrição PFAM</i>                                  |
|---------------------------|-------------------------------|--|--|
| Eucgr.B00380.1            | PF00069<br>PF00433            | domínio proteína kinase<br>domínio proteína kinase C terminal  | fosforilação de proteína                               |
| Eucgr.B01155.1            | PF01738                       | família hidrolase dienelectona   | atividade de hidrolase                                 |
| Eucgr.D01467.1            | -                             | -  | -  |
| Eucgr.E01901.1            | PF02535                       | transportador <i>ZIP zinc</i>  | transporte de íons metálicos                           |
| Eucgr.E01907.1            | PF02535                       | transportador <i>ZIP zinc</i>  | transporte de íons metálicos                           |
| Eucgr.E03125.1            | PF00931<br>PF00560            | domínio NB-ARC<br>repetição <i>leucine-rich</i> (LRR)  | apoptose<br>ligação a proteína                         |
| Eucgr.E03126.1            | PF00931<br>PF00560<br>PF01582 | domínio NB-ARC<br>repetição <i>leucine-rich</i> (LRR)<br>domínio homólogo a receptor<br>Toll/interleukin-1 (TIR) | apoptose<br>ligação a proteína<br>resposta imune inata |
| Eucgr.E03335.1            | PF00931<br>PF01582            | domínio NB-ARC<br>domínio homólogo a receptor<br>Toll/interleukin-1 (TIR)  | apoptose<br>resposta imune inata                       |
| Eucgr.G02940.1            | PF03514                       | fator de transcrição família GRAS  | fator de transcrição<br>família GRAS                   |
| Eucgr.G03276.1            | PF00561                       | conformação alfa/beta hidrolase  | conformação alfa/beta<br>hidrolase                     |

| <i>Nome do transcrito</i> | <i>ID PFAM</i>     | <i>Domínio protéico</i>   | <i>Descrição PFAM</i>                              |
|---------------------------|--------------------|---|--|
| Eucgr.G03406.1            | PF03171            | superfamília oxigenase 2-oxoglutarato (2OG) e Fe(II)-dependente | atividade oxidoreductase                           |
| Eucgr.J01892.1            | PF00628<br>PF05641 | Domínio PHD (Plant Homeo Domain)<br>Domínio Agenet              | ligação a proteína<br>ligação a RNA                |
| Eucgr.K01839.1            | PF00069<br>PF01453 | domínio proteína kinase<br>lectina ligante de D-mannose         | fosforilação de<br>proteína<br>ligação a<br>açúcar |

O siRNA secundário acima descrito seria assim um exemplo de um mecanismo de amplificação de silenciamento a partir de um sRNA inicial, que tem como alvo um gene de resistência, produzindo siRNAs secundários em fase atuando em outros genes de resistência. Esses dados corroboram a hipótese anteriormente proposta de que siRNAs secundários produzidos em fase a partir de um gene de resistência e tendo como alvo outros genes de resistência representam um novo nível de defesa contra o ataque de patógenos (Zhai, Jeong *et al.*, 2011; Shivaprasad, Chen *et al.*, 2012). O patógeno silencia a primeira linha de defesa representada pelo sRNA inicial do hospedeiro, potencialmente um miRNA. Isto desencadearia a supressão do silenciamento, via siRNAs secundários, dos genes envolvidos no segundo nível de defesa dessa complexa rede, aumentando, assim, a expressão desses genes em última instância. Isto seria um processo que foi chamado de mecanismo de contra contra-defesa (Zhai, Jeong *et al.*, 2011). Em plantas infectadas de tomate, a supressão do silenciamento pode aumentar em duas a três vezes a expressão de determinados mRNAs de NBS-LRR. Contudo, como dezenas de mRNAs desses genes de resistência são alvos do miR482, que desencadeia a síntese de grande parte dos phasiRNA, o efeito em rede seria de um aumento da ordem de centenas de vezes (Shivaprasad, Chen *et al.*, 2012). Este siRNA secundário descoberto em *Eucalyptus* pode ser mais um exemplo de um sistema de amplificação do sinal de silenciamento que parece ser típico em famílias

gênicas que sofreram grande expansão como NBS-LRR e PPR (Howell, Fahlgren *et al.*, 2007; Shivaprasad, Chen *et al.*, 2012) (Figura 22).

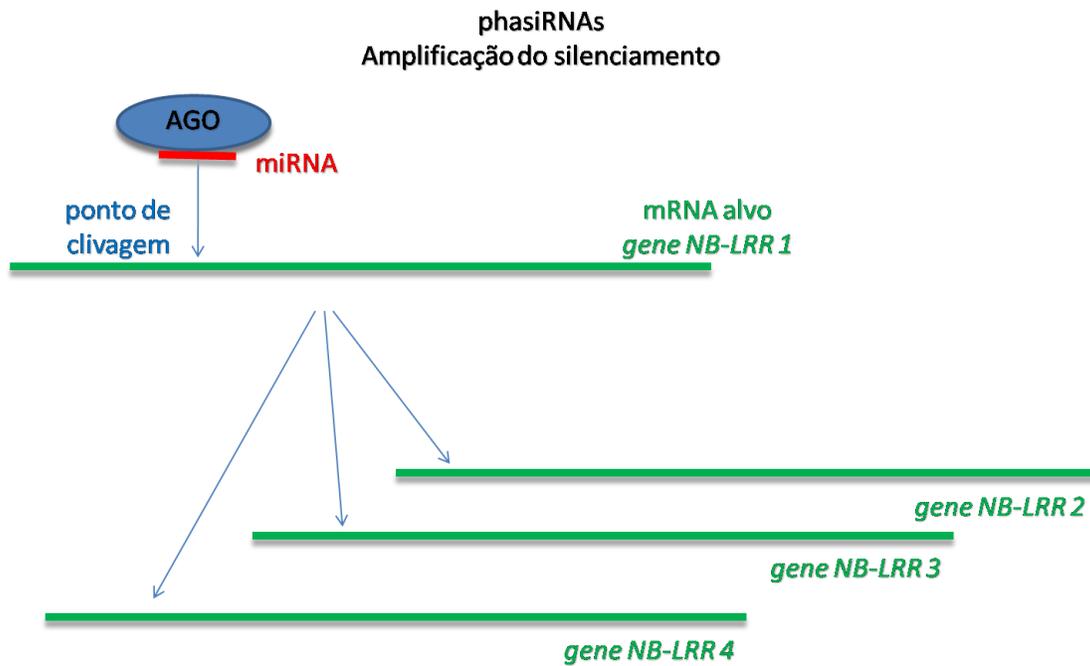


Figura 22: Representação esquemática da síntese de phasiRNAs e da amplificação do sinal de silenciamento dirigido a grandes famílias gênicas.

## CONCLUSÕES

A busca, *in silico*, por siRNAs produzidos em fase em regiões gênicas resultou na identificação de 13 transcritos. Uma busca por siRNAs em fase no genoma de *Populus trichocarpa* resultou em números similares, com 12 locos altamente significativos, quatro desses mapeando em genes NBS-LRR (Tuskan, Difazio *et al.*, 2006).

Vários siRNAs secundários, gerados a partir desses transcritos, apresentam alta contagem nos dados de sequenciamento. Algumas das sequências mais abundantes de 21 nt observadas no sequenciamento foram identificadas como siRNAs em fase como euc\_sRNA\_4546, euc\_sRNA\_766295 e euc\_sRNA\_33215. Este dado é importante no reconhecimento de que dados de sequenciamento em larga escala revelam diversas espécies de pequenos RNAs que agem em paralelo. Vale ressaltar que uma dissecção estrutural detalhada é necessária para a correta classificação das famílias de pequenos RNAs. A abundância não é um critério que possa ser diagnóstico de verdadeiros miRNAs, como demonstrado nos dados de sequenciamento de *Eucalyptus* que apontam uma grande amostragem de siRNAs de 21 nt entre as sequências mais abundantes.

A identificação do sRNA desencadeador em apenas quatro dos 13 transcritos preditos como produtores de siRNAs não invalida a predição destes transcritos que tem, no número de sRNAs de 21 nt mapeados, evidência suficiente para a declaração de um loco TAS. Existem exemplos na literatura onde nenhum dos sRNAs primários foi identificado nos transcritos primários (Howell, Fahlgren *et al.*, 2007).

A premissa inicial, quando da descoberta deste mecanismo, de que os siRNAs secundários produzidos em fase atuariam silenciando transcritos em *trans* mostrou-se incompleta. Este pode ser o padrão predominante mas não são raros os exemplos de atuação em *cis*, desencadeando uma cascata de auto-regulação de transcritos por siRNAs. Talvez a nomenclatura possa ser revista, em função disso, renomeando definitivamente os siRNAs secundários produzidos em fase como phasiRNAs. Uma parte dos phasiRNAs seria, de fato, tasiRNAs.

Interessantemente, em *Eucalyptus* foram encontradas todas as variantes de phasiRNAs reportadas na literatura para outras espécies. A identificação do mecanismo de síntese de siRNAs secundários em fase no gênero *Eucalyptus* corrobora a conservação deste sistema de silenciamento. A participação deste relevante mecanismo na manutenção do equilíbrio da expressão de genes que sofreram rápida expansão, de forma a manter a relação custo benefício da preservação um grande número de genes, parece ser mantida, especialmente, no caso de genes de resistência NBS-LRR.

Os resultados desta análise de siRNA em fase em *Eucalyptus* ressaltam novamente a importância da compreensão de mais essa via de silenciamento e da revelação de indicativos de que os vários mecanismos de regulação já descritos se integram de alguma forma. Como acontece em alguns transcritos regulados simultaneamente por miRNAs e phasiRNAs. Isto, mais uma vez, reafirma a complexidade da regulação da expressão gênica que envolve diversos níveis e elementos de silenciamento.

## **PERSPECTIVAS FUTURAS**

A elucidação crescente dos diversos níveis da regulação da expressão gênica e a consequente compreensão da complexidade da rede formada por todos os elementos regulatórios demonstra as razões pelas quais a busca de genes ou polimorfismos que, individualmente, possam explicar grande parte da variação fenotípica observada em uma característica complexa raramente tem sucesso.

Apesar do avanço representado pelo desenvolvimento de seleção assistida por marcadores (MAS – *marker assisted selection*), os alelos assim identificados representam apenas uma pequena parte da porção herdável dos fenótipos complexos. Uma parte significativa da origem destes caracteres permanece não elucidada, o que é a definição do conceito de *missing heritability*. Sugere-se que a determinação destes fenótipos complexos seja multifatorial e fatores como interação entre genes (epistasia), entre genes e ambiente e mudanças epigenéticas como o silenciamento transcricional por metilação de DNA, em parte controladas por miRNAs, possam contribuir para a determinação destes (Marian, 2012).

A recente descoberta de RNAs regulatórios adiciona um componente fundamental da regulação da expressão gênica. Por ser muito recente, a classificação destes elementos ainda é discutível e, provavelmente, será ainda revisitada. Assim como a nomenclatura atribuída aos novos sRNAs descobertos dentro de cada uma das classes e subclasses.

Esse trabalho representa um primeiro passo de descoberta em escala genômica ampla de elementos regulatórios no gênero *Eucalyptus*. Constitui ainda a identificação de elementos e mecanismos conservados e caracterização dos que são linhagem específicos.

A identificação de pequenos RNAs, incluindo miRNAs e siRNAs, em *Eucalyptus* abre a perspectiva para investigações pontuais de alguns desses elementos regulatórios na determinação de fenótipos de interesse. A inversão no padrão de intensidade da expressão dos miRNAs miR156 e miR172 na transição de fase vegetativa é um exemplo do que será mais detalhadamente estudado. A transição de fase vegetativa em *E. grandis* não é marcada por diferenças morfológicas claras como acontece em outras espécies, por

exemplo, *E. globulus*. A identificação do ponto de transição é uma característica de interesse para estudos de fisiologia e genética. A transição para a fase adulta em *Eucalyptus* tipicamente resulta na perda de totipotência celular e consequente impossibilidade de propagação vegetativa via estaquia e cultura de tecidos. No melhoramento, a possibilidade de acelerar a mudança da fase vegetativa para a fase adulta e consequente florescimento em prazo mais curto permite acelerar as gerações de recombinação e seleção recorrente.

A descoberta de mecanismos conservados como a amplificação do silenciamento em famílias de genes de resistência e a auto-regulação por phasiRNAs RNAs descritos neste trabalho, abre perspectivas interessantes de integrar este conhecimento em estudos da interação patógeno-hospedeiro, investigação da herança e busca de fontes de resistência. Os dados gerados em *Eucalyptus* poderão contribuir ainda para a elucidação da dinâmica desses elementos em espécies perenes em geral, bem como o entendimento da evolução destes e da co-evolução com grandes famílias gênicas.

A validação experimental da funcionalidade de cada um desses siRNAs secundários em fase constitui o próximo passo passível de alcance com a conclusão do sequenciamento do degradoma de *Eucalyptus grandis*, em andamento, que poderá confirmar o ponto de clivagem dos transcritos alvo inferidos *in silico*.

Várias aplicações das vias de silenciamento endógenas para repressão de determinados genes vem sendo desenvolvidas como a utilização de pequenos RNAs artificiais (Ossowski, Schwab *et al.*, 2008) e sequestradores de miRNAs, mímicos de mRNAs alvo em plantas (Chen, Jiang *et al.*, 2012). Esta tecnologia representa mais um caminho potencial para a manipulação de características complexas quando não houver variabilidade nativa na espécie alvo. Recentemente, por exemplo, um estudo conduzido pela empresa DuPont mostrou que é possível o uso de tasiRNAs artificiais cooptando transcritos TAS em *Arabidopsis* para silenciar a expressão de novos genes alvo. Substituindo tasiRNAs de TAS1c com sequências do gene FAD2 (*Fatty Acid Desaturation 2*), a atividade deste gene foi reduzida a níveis observados para alelo nulo (De La Luz Gutierrez-Nava, Aukerman *et al.*, 2008).

Em uma esfera mais ampla, pretende-se conduzir estudos epigenômicos que sejam, posteriormente, integrados aos dados mapeados de pequenos RNAs, uma vez que, parte desses RNAs é responsável pelo direcionamento de metilação DNA (metilação de DNA dependente de RNA).

Parece cada vez mais ser um consenso que o componente herdável da variação fenotípica vai além da sequência de DNA (Mosher e Melnyk, 2010). A variação epigenética também contribuiria fortemente para controle de fenótipos complexos e, portanto, para a evolução de mudanças fenotípicas relacionadas à adaptabilidade. Um estudo recente em linhagens epigenéticas recombinantes quase isogênicas (epiRILs) de *Arabidopsis thaliana*, onde as linhagens quase isogênicas se diferenciavam quase exclusivamente nos epialelos, mostrou variação epigenética herdável significativa em vários caracteres ecologicamente importantes e na plasticidade fenotípica em resposta a seca e condições nutricionais. Esta observação sugere que evolução a curto prazo, ou seja, plasticidade fenotípica, baseada exclusivamente em variação epigenética é um mecanismo importante (Zhang, Fischer *et al.*, 2013). Em espécies perenes com ciclos de vida da ordem de centenas de anos é bem provável que seus genomas tenham evoluído para se adaptar a mudanças climáticas expressivas por meio de mecanismos de plasticidade possivelmente envolvendo um importante componente de variação epigenética. Espécies de *Eucalyptus* representam um excelente modelo de espécie perene para estudos deste tipo tendo em vista a possibilidade de estudar os mesmos genótipos clonados e plantados em diferentes ambientes.

Um recente exemplo que mostra o impacto do controle epigenético sobre caracteres de interesse econômico é a regulação do amadurecimento do fruto de tomate por metilação. Foi mostrado que o processo de amadurecimento é acompanhado por uma extensa reprogramação do epigenoma. Cerca de 1% do genoma de 900 megabases do tomate é substancialmente alterada no processo, incluindo promotores de genes anteriormente descritos como pontos da chave do desencadeamento do amadurecimento (Zhong, Fei *et al.*, 2013).

Outros estudos também mostram os efeitos de epialelos em diferentes aspectos. Existe a sugestão de que o vigor híbrido observado em híbridos dos acessos C24 e Landsberg erecta de *Arabidopsis thaliana* seja decorrente da diminuição dos pequenos RNAs de 24 nt que mostram correlação com metilação de DNA e nível de expressão de genes (Groszmann, Greaves, Albertyn *et al.*, 2011). Resultados de experimentos de vernalização sugerem a participação do controle epigenético em diversos aspectos do desenvolvimento de plantas, especialmente, o florescimento (Groszmann, Greaves, Albert *et al.*, 2011). Novamente, a base molecular da heterose funcional em *Eucalyptus*, ou seja, a combinação de características favoráveis observada em híbridos interespecíficos ainda é obscura. Sua elucidação provavelmente passará pelo estudo dos componentes de epigenética e da interação entre epialelos das diferentes espécies.

## ANEXO I

Anotação das sequências de *Eucalyptus* ortólogas a miRNAs conservados depositados no miRBase. Código da sequência (ID), sequência, sequência de miRNA e espécie com maior similaridade no miRBase, número de bases distintas da sequência ortóloga do miRBase (*mismatches*) e extensão da similaridade em bases.

| ID       | Sequência              | miRBase    | Espécie                     | <i>mismatches</i> | Extensão da similaridade |
|----------|------------------------|------------|-----------------------------|-------------------|--------------------------|
| EUC_3918 | TATGTTTTCTTATTAACCAA   | miR156a    | <i>Brassica napus</i>       | 0                 | 21                       |
| EUC_2944 | TTGCAGAAGAGAATGGGACCT  | miR156d    | <i>Glycine max</i>          | 0                 | 21                       |
| EUC_3854 | GCCTGCTCCTGCATTTCTTGT  | miR156d    | <i>Glycine max</i>          | 0                 | 21                       |
| EUC_8565 | TGAAGGTGTTGAACGAGGTG   | miR156f-3p | <i>Oryza sativa</i>         | 2                 | 20                       |
| EUC_2420 | TTAGATTAGACGGACAAATATC | miR156f-3p | <i>Oryza sativa</i>         | 2                 | 22                       |
| EUC_4102 | TGATTTGAAATTTTTTGTGAT  | miR156k    | <i>Glycine max</i>          | 0                 | 21                       |
| EUC_9197 | TTCTTGAAGATATCCATTTA   | miR157d-3p | <i>Arabidopsis lyrata</i>   | 0                 | 20                       |
| EUC_2941 | CGGACTTAATTGCAGACTTGA  | miR159a    | <i>Arabidopsis thaliana</i> | 0                 | 21                       |
| EUC_4474 | CTCCCCAGCCAGTGCAGGCC   | miR159a    | <i>Arabidopsis thaliana</i> | 2                 | 21                       |
| EUC_5149 | TTGAATGCACCACAAGCTTGT  | miR159a    | <i>Arabidopsis thaliana</i> | 2                 | 21                       |
| EUC_4521 | CCCTGAACCCCTAATCCCTAA  | miR159a    | <i>Arabidopsis thaliana</i> | 2                 | 21                       |
| EUC_3796 | TGTCAGACAATAGTTTTTCT   | miR159a    | <i>Arabidopsis thaliana</i> | 2                 | 21                       |
| EUC_3814 | TGCTGCTAGAGCTATGAGTGC  | miR159a    | <i>Arabidopsis thaliana</i> | 1                 | 21                       |
| EUC_3815 | GTTGCCAACGTAGAGAAGCCC  | miR159a    | <i>Arabidopsis thaliana</i> | 2                 | 21                       |
| EUC_3009 | ATTGTTTTTTGAAGAGTTCGC  | miR159a    | <i>Arabidopsis thaliana</i> | 1                 | 21                       |
| EUC_3352 | TTGGGTCAAAGCATTAAATCG  | miR159a-5p | <i>Glycine max</i>          | 3                 | 21                       |
| EUC_4589 | CAAGGGATCAGATGCACTGGG  | miR159a.1  | <i>Oryza sativa</i>         | 1                 | 21                       |
| EUC_7780 | GGAATGGTGTCTGGCTCGAG   | miR159h-3p | <i>Zea mays</i>             | 1                 | 20                       |
| EUC_8901 | TGCATGCACTTAGAATTTCA   | miR159h-3p | <i>Zea mays</i>             | 1                 | 20                       |
| EUC_7342 | CAGTGTCTTCAATTCCTCA    | miR159h-3p | <i>Zea mays</i>             | 3                 | 20                       |
| EUC_3863 | TTGGACAAGGATTCTGTTGGA  | miR160a    | <i>Arabidopsis thaliana</i> | 1                 | 21                       |
| EUC_3861 | TTGGATAGATTAAGAAGGAAG  | miR160a    | <i>Arabidopsis thaliana</i> | 0                 | 21                       |
| EUC_8398 | CGCCTTCTTTCGCAGCTTCG   | miR160a    | <i>Arabidopsis thaliana</i> | 0                 | 20                       |
| EUC_4413 | CGAGCCTCGCCCAGCTGCTGA  | miR160f-5p | <i>Oryza sativa</i>         | 0                 | 21                       |
| EUC_3007 | TTAACTGTTTTATGTCATTGA  | miR160f-5p | <i>Oryza sativa</i>         | 0                 | 21                       |

|          |                       |         |                             |   |    |
|----------|-----------------------|---------|-----------------------------|---|----|
| EUC_7079 | TTCCCATCTTTTCTTGTTGG  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7074 | CCTTGACATCCATATTGAAC  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7076 | AACAAATATGAACCTATGAC  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_4060 | ACCGAGCCACGTGTCCTCTT  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_7068 | TTTGCTTGATTGTCACCAGA  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7198 | TTGAAAAGCCAGTGGCGCGA  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7196 | CACATCTTCCTCAACGATTG  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_3686 | CCAATGTCTCTCTCCATATTT | miR162a | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_7201 | ATCATTGTGAAGCAGAATTC  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7078 | TTTCTTGTTGGATTATACAA  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_4109 | CAATGTCATTAAGTATTTTG  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_7197 | CGGGTGACGGAGAATTAGGG  | miR162a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_4055 | TGGGATTTTACCTACCAACAT | miR162a | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_3816 | AGTTGCCAAAGTAGAGAAGCC | miR164a | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_3008 | AATTCTCGATGCCTGCTGCGA | miR164a | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_4424 | ACTCCACTTGGACAAACTAC  | miR166a | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_5303 | TAGTTAAGCGGTGTAAGGAAT | miR166a | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_7875 | TCAGAGATCTTGGAAAGGGAA | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7876 | ACAAGTGGAGAAGTTAGTTG  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_3237 | GTGCGTCATAAATTATCCTCA | miR166a | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_7877 | TTCGAGAGAAGTCATCAAGA  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7225 | TACAATCTAAATCCCTTAAC  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7224 | CCCCGACTGTCCCTGTTAAT  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_6554 | GTCATCGAGAAGAATCATTG  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_6555 | AGATGGAGGAAACTCACTTA  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_6552 | CTTGAGGTTCTTGAGATTGG  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_6553 | TGGTTCAAGTTGATAGGTCT  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7152 | TCGAGCTTGGAGCCACCCTC  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8881 | TCTACTTGAATTCATGCAAT  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_9033 | TTAGATGAGAAGACGAGACA  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_9034 | AATGAGGAAGGAAGCTAGAA  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_9035 | GATGAGGAATGAAGCTAGAA  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7643 | AGAAGAATGATCGTTCGTTT  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7436 | TTGGAAGATCTACATTGTGA  | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7437 | GAGGCCGAGGCGGGGTCTT   | miR166a | <i>Arabidopsis thaliana</i> | 1 | 20 |

|          |                        |            |                                |   |    |
|----------|------------------------|------------|--------------------------------|---|----|
| EUC_7649 | CAACGTGTCGGAATTCTCTA   | miR166a    | <i>Arabidopsis thaliana</i>    | 1 | 20 |
| EUC_7154 | AATTGCACACTCCACACGAG   | miR166a-5p | <i>Oryza sativa</i>            | 1 | 20 |
| EUC_7439 | GGAGTGGGAAGCCTCTGACAA  | miR166a-5p | <i>Oryza sativa</i>            | 3 | 20 |
| EUC_8506 | TTCTCGTCCTCCCCCGTCCC   | miR166c-5p | <i>Oryza sativa</i>            | 1 | 20 |
| EUC_6445 | AGCTTGGCTTGGTTCTGATA   | miR166d-5p | <i>Oryza sativa</i>            | 1 | 21 |
| EUC_8867 | ATCCTTCCCATTCTCATT     | miR166d-5p | <i>Oryza sativa</i>            | 1 | 20 |
| EUC_8882 | TTCTTGAATCAGATAAAGGC   | miR166d-5p | <i>Oryza sativa</i>            | 1 | 20 |
| EUC_8883 | TTCTCCAACGAGAGTTCCTT   | miR166d-5p | <i>Oryza sativa</i>            | 1 | 20 |
| EUC_4492 | GATGGTGTATGTGCATTGGTT  | miR166d-5p | <i>Oryza sativa</i>            | 0 | 21 |
| EUC_3351 | TACTTATGATCCACGTCATCG  | miR166e    | <i>Brachypodium distachyon</i> | 1 | 21 |
| EUC_4423 | CATTGGCAAGGGATTAGCAAG  | miR166e    | <i>Brachypodium distachyon</i> | 0 | 21 |
| EUC_3347 | CTAATTTATCATATGCTTTAT  | miR166f    | <i>Brassica napus</i>          | 0 | 21 |
| EUC_8027 | TCCAGGGGACTTTGAAGGTA   | miR166g-5p | <i>Zea mays</i>                | 1 | 20 |
| EUC_7700 | GTCTGTTGAAGAGACTTGAC   | miR166h-3p | <i>Glycine max</i>             | 2 | 20 |
| EUC_7644 | AGAAGAATGATCGCTCGTTT   | miR166h-3p | <i>Glycine max</i>             | 1 | 20 |
| EUC_4833 | TGCAGGATGATGGATGGCAGA  | miR166h-3p | <i>Glycine max</i>             | 1 | 21 |
| EUC_7878 | CCTCTTGTTTGCTGTTGGCA   | miR166h-3p | <i>Glycine max</i>             | 1 | 20 |
| EUC_7879 | TGCGAGGAGGATTTGTCCAA   | miR166h-3p | <i>Glycine max</i>             | 1 | 20 |
| EUC_3672 | GTTTTGCCGGAAATTGTGAG   | miR166h-3p | <i>Glycine max</i>             | 0 | 21 |
| EUC_7153 | TACAGTCTACTGGATCTCCA   | miR166h-3p | <i>Glycine max</i>             | 1 | 20 |
| EUC_7438 | TGACCTACTCAGAGACCTTG   | miR166h-3p | <i>Glycine max</i>             | 1 | 20 |
| EUC_3825 | CCGATTCTTCAATGTCCGTCG  | miR166j-5p | <i>Oryza sativa</i>            | 2 | 21 |
| EUC_3550 | TTGATGGTTCATGTTTTCTCG  | miR167a    | <i>Malus domestica</i>         | 2 | 21 |
| EUC_6587 | TTCGTATCCAGACGTGACGA   | miR167a    | <i>Citrus clementine</i>       | 0 | 20 |
| EUC_6586 | CTCTACCAAACGACCCATG    | miR167a    | <i>Citrus clementine</i>       | 1 | 20 |
| EUC_6589 | TAAAAGGCACCTTCCTCGAG   | miR167a    | <i>Malus domestica</i>         | 1 | 20 |
| EUC_6588 | TAATCCAAATTTACCTTAA    | miR167a    | <i>Malus domestica</i>         | 2 | 20 |
| EUC_2405 | AAAAAGAACTGAGATCCATGAC | miR167a    | <i>Brassica napus</i>          | 1 | 22 |
| EUC_3547 | CTTGTGAGTTCGCCAATAGAA  | miR167a    | <i>Citrus clementine</i>       | 1 | 21 |
| EUC_2531 | AGTTCTATTAAGTCCTTGCATG | miR167a    | <i>Citrus clementine</i>       | 1 | 22 |
| EUC_3549 | GCATGATTTGGTTCAGTTGAT  | miR167a    | <i>Citrus clementine</i>       | 1 | 21 |
| EUC_3967 | TCTGGAGTCTTTCTTTGGGCA  | miR167a    | <i>Malus domestica</i>         | 1 | 21 |
| EUC_2409 | ATTCCCAAACAACCCGACTCGC | miR167a    | <i>Citrus clementine</i>       | 0 | 22 |
| EUC_3777 | CGAGCACCATGAGAGATCGAG  | miR167a    | <i>Arabidopsis thaliana</i>    | 2 | 21 |
| EUC_4103 | AGAGCTCAAGTTACCAATAAG  | miR167c    | <i>Vitis vinifera</i>          | 1 | 21 |

|          |                        |            |                             |   |    |
|----------|------------------------|------------|-----------------------------|---|----|
| EUC_8888 | CAGAAAGGGTGAGTGTGGAA   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8889 | AGGAGAAGCAAAGTGAAGCC   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8886 | TCATCTAACGGAGGCGAAGA   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8887 | AAAGGGTGAGTGTGGAAGTG   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_2677 | CTTCTCTTTTTTCATTTTTCC  | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 22 |
| EUC_8897 | GTCCTTTCCAATCAAAAAC    | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8895 | AAATTGACTGGAGCTTGAAA   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8894 | CCGTAATTCATTGTTTTCTC   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8892 | CCGAAAGGAGAAACAAATTA   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8891 | CAAGCGAGTTTCTTCTGTCTG  | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8890 | AGAAGGAGAAGCAAAGTGAA   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8896 | TATTGAAGAGTTAAAGTTGA   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_2580 | TGCCAATTCAGTCCTTTCCGG  | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 22 |
| EUC_8893 | CTTCCACAGAACATAAATCA   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_6487 | CAAGTTGCTTAGCAGTCATG   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_6484 | ATAGATGGAATCTTTGCTGG   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_6485 | TGAGAGTTCATGTCTTTTGG   | miR167d    | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_4510 | TTCGCAGGCCCTCATCTATGA  | miR167d    | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_6502 | AGTCAGTCACTGAGGGCTAA   | miR167d-3p | <i>Arabidopsis lyrata</i>   | 3 | 21 |
| EUC_4511 | GTAGGTTACAAGTTCCTCCAA  | miR167d-3p | <i>Arabidopsis lyrata</i>   | 2 | 21 |
| EUC_8577 | CTATTAAGTCCTTGCATGAA   | miR167f-3p | <i>Populus trichocarpa</i>  | 1 | 20 |
| EUC_3966 | TTGCTGACCACTTCGCACGAA  | miR167f-3p | <i>Populus trichocarpa</i>  | 2 | 21 |
| EUC_4491 | TTTTGATCAAATCATGCATCT  | miR168a    | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_8952 | GGAGTTGCTCCTCATCATCC   | miR168a    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8953 | AATGTGTGCGACAATAATTG   | miR168a    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_4489 | TAATTTTTTTAAGGATAGAAA  | miR168a-3p | <i>Populus trichocarpa</i>  | 0 | 21 |
| EUC_8094 | CGGGGAATTGTAAGTGGCAG   | miR169a    | <i>Arabidopsis thaliana</i> | 0 | 20 |
| EUC_4081 | ATATGTCGTTTCGTTAGAAAC  | miR169b    | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_6537 | GAAGAACCGATAACCAAGAA   | miR169b    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7743 | TCGGACAGCTGTGTGAAAGG   | miR169b    | <i>Arabidopsis thaliana</i> | 0 | 20 |
| EUC_2557 | TCAGTACGGATGGGATGTCCGG | miR169b    | <i>Malus domestica</i>      | 1 | 22 |
| EUC_8559 | GCATCAACAGTCTATCATCA   | miR169n    | <i>Oryza sativa</i>         | 1 | 20 |
| EUC_8099 | CTTGGATGTGGTAGCCGTTT   | miR169n    | <i>Oryza sativa</i>         | 2 | 20 |
| EUC_6538 | TTGATTGTCCTTGATCTATC   | miR169r-3p | <i>Zea mays</i>             | 3 | 20 |
| EUC_9389 | GGCAAGTTGTCTTTGGCTAA   | miR169r-3p | <i>Zea mays</i>             | 2 | 20 |

|          |                         |            |                             |   |    |
|----------|-------------------------|------------|-----------------------------|---|----|
| EUC_3884 | TCTCATTGTTCTCGATGACGT   | miR169r-3p | <i>Zea mays</i>             | 3 | 21 |
| EUC_8547 | TCATTGTAGATGATTTTGGC    | miR171     | <i>Citrus trifoliata</i>    | 0 | 20 |
| EUC_4078 | CGAGCCTCGCCCAGCTGTTGG   | miR171b    | <i>Oryza sativa</i>         | 0 | 21 |
| EUC_8680 | GCACTTCTCCTTTCTGTTC     | miR171b    | <i>Arabidopsis thaliana</i> | 2 | 20 |
| EUC_3817 | TGTTGCCAAAGTAGAGAAGCC   | miR171b    | <i>Oryza sativa</i>         | 1 | 21 |
| EUC_8737 | CAAGTCTCCACAATCCACTT    | miR171b    | <i>Arabidopsis thaliana</i> | 2 | 20 |
| EUC_2989 | GTTTCCTCATTAGTCAACTTC   | miR171c    | <i>Aquilegia caerulea</i>   | 2 | 21 |
| EUC_4100 | TTCTTACAATGTCCACGGCTT   | miR171h-5p | <i>Populus trichocarpa</i>  | 1 | 21 |
| EUC_3371 | CTTCCTCGAGTTCATCCTTGT   | miR171h-5p | <i>Populus trichocarpa</i>  | 1 | 21 |
| EUC_8679 | GTCAAATAAGAGCCTGAAAT    | miR171i    | <i>Cucumis melo</i>         | 1 | 20 |
| EUC_3862 | TCATATATCAGACTCTATTGG   | miR172a    | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_6195 | CAACCTTGCAAAATCCGTCCA   | miR172a    | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_5094 | AGATCCACTTAGGTTATTATT   | miR172a    | <i>Arabidopsis thaliana</i> | 1 | 21 |
| EUC_7895 | CTGATCCTTCTTGACATAT     | miR172b    | <i>Nicotiana tabacum</i>    | 0 | 20 |
| EUC_7667 | CTTCGGTGGAGAAGTCAAAA    | miR172c    | <i>Arabidopsis thaliana</i> | 2 | 20 |
| EUC_7551 | CTCATGTCTCTACAGCAATT    | miR172c    | <i>Oryza sativa</i>         | 0 | 20 |
| EUC_3858 | TCCGACGTCGAGGTAGAGTGT   | miR172c    | <i>Arabidopsis thaliana</i> | 2 | 21 |
| EUC_5210 | AATTGTAGGTAAGAAGTATAT   | miR172d    | <i>Theobroma cacao</i>      | 0 | 21 |
| EUC_8051 | CAATGATCCTTCCGAAGGTT    | miR172d    | <i>Vitis vinifera</i>       | 1 | 20 |
| EUC_5211 | CCACCGTCTTAGTACAACACTAG | miR172d    | <i>Glycine max</i>          | 2 | 21 |
| EUC_8415 | TCGACGTATCTGTAGAATTT    | miR172d-5p | <i>Zea mays</i>             | 2 | 20 |
| EUC_5208 | AGTGTTCATCCCATGCTC      | miR172e    | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_9300 | GCTCAGGAGGGATAGCGCC     | miR172e    | <i>Arabidopsis thaliana</i> | 3 | 20 |
| EUC_3859 | AACAAAGTCAGACTTCAAGTA   | miR172g-3p | <i>Populus trichocarpa</i>  | 0 | 21 |
| EUC_5417 | TGGCGGCCAGCAGACCTTGTA   | miR172g-3p | <i>Populus trichocarpa</i>  | 0 | 21 |
| EUC_3819 | TGGGTGTCGTCCGAGCAGTTT   | miR172g-3p | <i>Populus trichocarpa</i>  | 1 | 21 |
| EUC_8068 | TGTAGGCTACGAATTGAGAC    | miR1862e   | <i>Oryza sativa</i>         | 2 | 20 |
| EUC_2538 | GGCATGGGTGGTATTTGCAAGA  | miR2089-5p | <i>Medicago truncatula</i>  | 3 | 22 |
| EUC_4074 | ATTTGAAATTTTTGTGATAC    | miR2111a   | <i>Populus trichocarpa</i>  | 0 | 21 |
| EUC_5466 | GACCCCTTGCCGAATCTGGCG   | miR2118b   | <i>Oryza sativa</i>         | 3 | 21 |
| EUC_7835 | CATCAACAACAACTTGCTG     | miR2623    | <i>Medicago truncatula</i>  | 3 | 20 |
| EUC_9165 | TTGAGATTGGATTGGGCTAC    | miR319a    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8059 | ACATCCGTCAAAGCTAGAAG    | miR319a    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7141 | AGAAGAACTGCTCGGAAAA     | miR319a    | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_3775 | CTACTCCGCCTCCCAAGAATT   | miR319a    | <i>Arabidopsis thaliana</i> | 3 | 21 |

|          |                         |             |                             |   |    |
|----------|-------------------------|-------------|-----------------------------|---|----|
| EUC_8873 | GCATGGCGTAGATCTCTTCG    | miR319b-5p  | <i>Arabidopsis lyrata</i>   | 2 | 20 |
| EUC_7550 | ATTGCCTTATTCTTGTTCTC    | miR319c     | <i>Arabidopsis thaliana</i> | 0 | 20 |
| EUC_8557 | CTGCAAGCATAGAGAGTTGG    | miR319c     | <i>Arabidopsis thaliana</i> | 0 | 20 |
| EUC_8280 | CTCAACTGGTCATGCATTCT    | miR319c     | <i>Arabidopsis thaliana</i> | 0 | 20 |
| EUC_8558 | CTGCAAGTTTTGAGAGTTGT    | miR319g     | <i>Glycine max</i>          | 0 | 20 |
| EUC_2417 | ATCCACATCATCGAGAAGACTG  | miR3627-5p  | <i>Vitis vinifera</i>       | 3 | 22 |
| EUC_7665 | AAAACCTCGTTCACATCAAAA   | miR3627-5p  | <i>Vitis vinifera</i>       | 1 | 20 |
| EUC_2370 | TCCAATGGCTCTCTCCATATTT  | miR3630-3p  | <i>Helianthus annuus</i>    | 3 | 22 |
| EUC_8774 | TTTCTGATTTTTGTGCTTGG    | miR3631b-3p | <i>Vitis vinifera</i>       | 3 | 20 |
| EUC_7309 | TATCATACAGTGCCATTTTCG   | miR3633a-3p | <i>Vitis vinifera</i>       | 2 | 20 |
| EUC_3843 | GAGAAAGAGTTGATTCGTCAA   | miR393a     | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_2449 | AAGCATAAACCAATTTCTGTTCT | miR393a     | <i>Arabidopsis thaliana</i> | 0 | 22 |
| EUC_3905 | TGTGTGTGTGTATATATAT     | miR393a-3p  | <i>Populus trichocarpa</i>  | 1 | 21 |
| EUC_3952 | TCATGCTCTTTCCAAGAATCT   | miR393b-3p  | <i>Zea mays</i>             | 2 | 21 |
| EUC_9299 | TATTGGCAAATCAACTCTT     | miR393d     | <i>Malus domestica</i>      | 1 | 20 |
| EUC_7647 | GAGAAGAATGATCGATTGTT    | miR393d     | <i>Malus domestica</i>      | 2 | 20 |
| EUC_8150 | TTTGCTGGAAGTATGGAAG     | miR394a     | <i>Vitis vinifera</i>       | 1 | 20 |
| EUC_7670 | AAAGTCTGGATGTTATAATT    | miR395a     | <i>Medicago truncatula</i>  | 1 | 20 |
| EUC_7559 | AGCATGCTTCTTTTTCAACG    | miR395a     | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_5045 | GGCTCTCATCACCATAGGTGA   | miR395a     | <i>Medicago truncatula</i>  | 0 | 21 |
| EUC_3614 | ATCTAGTGGATGGTGTCTGA    | miR395a     | <i>Arabidopsis thaliana</i> | 2 | 21 |
| EUC_7560 | ATAGTGAACAAGGAATGAAG    | miR395a     | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7144 | CCATATTTAGCCTTGGACGG    | miR395a     | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_4587 | TTCGGCTTTTTAGTAAACATG   | miR395a     | <i>Arabidopsis thaliana</i> | 0 | 21 |
| EUC_8098 | ATCAGGTCTCCAAGGTGAAC    | miR395a     | <i>Medicago truncatula</i>  | 0 | 20 |
| EUC_7145 | CCTTTGTGGGTTCTAGGTTA    | miR395a     | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_7143 | CAAAGGGCAGGGACGTAGTC    | miR395a     | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_8728 | TCGGACCAAGCTTCATTTCC    | miR395a     | <i>Arabidopsis thaliana</i> | 1 | 20 |
| EUC_4076 | CATCGGGCAATTTTGTAGGT    | miR395b     | <i>Oryza sativa</i>         | 0 | 21 |
| EUC_7557 | GTTGGTTCTGCATCCGTAGA    | miR395d-5p  | <i>Arabidopsis lyrata</i>   | 3 | 20 |
| EUC_7561 | AAGTCAAGGATAGTGAACAA    | miR395d-5p  | <i>Arabidopsis lyrata</i>   | 2 | 20 |
| EUC_3919 | ATAGCTTGTGGGCCATAATTA   | miR395d-5p  | <i>Arabidopsis lyrata</i>   | 3 | 21 |
| EUC_3894 | TTTTTTTTTCAGCGCCCGCAAT  | miR395d-5p  | <i>Arabidopsis lyrata</i>   | 3 | 21 |
| EUC_3965 | GTTTTAAAGTCAAATCGATCT   | miR395d-5p  | <i>Arabidopsis lyrata</i>   | 2 | 21 |
| EUC_3963 | CCAGATTCAACGGGGTTGACA   | miR395d-5p  | <i>Arabidopsis lyrata</i>   | 3 | 21 |

|          |                        |            |                              |   |    |
|----------|------------------------|------------|------------------------------|---|----|
| EUC_4111 | CTTAATTCCTTCCTCAGCATT  | miR395d-5p | <i>Arabidopsis lyrata</i>    | 3 | 21 |
| EUC_7558 | TGCCAAAGTCATCTTCAATC   | miR395f    | <i>Oryza sativa</i>          | 0 | 20 |
| EUC_4460 | CAAGTTTGAGCTCAACAAGAT  | miR396a    | <i>Arabidopsis thaliana</i>  | 0 | 21 |
| EUC_3481 | GATTAGAAATAATTTGCTCAA  | miR396a    | <i>Hevea brasiliensis</i>    | 2 | 21 |
| EUC_3910 | CGATCGCATTCTATGATTGA   | miR396a    | <i>Arabidopsis thaliana</i>  | 0 | 21 |
| EUC_9340 | TTCCCACAGCTTTCTTGAGC   | miR396b-3p | <i>Arabidopsis lyrata</i>    | 0 | 20 |
| EUC_7125 | TAATCCGTTGCCTACAAAAT   | miR396b-3p | <i>Glycine max</i>           | 1 | 20 |
| EUC_6466 | CATTGTGTGTCCCATCATCA   | miR396c    | <i>Cynara cardunculus</i>    | 1 | 21 |
| EUC_9347 | TTCAAGCTAGCTGTGGGAAG   | miR396c-3p | <i>Oryza sativa</i>          | 3 | 20 |
| EUC_8561 | TGAAAGGGATAATGGTAGAA   | miR396e    | <i>Glycine max</i>           | 1 | 20 |
| EUC_8339 | TCGTTTAAAGGGGAAGAGGGC  | miR396g-3p | <i>Zea mays</i>              | 0 | 20 |
| EUC_7123 | TAAGAGAGAATTGAAGAGAC   | miR396j    | <i>Glycine max</i>           | 3 | 20 |
| EUC_5456 | GAGGCTGCAGAACCTGGGGTG  | miR397a    | <i>Arabidopsis thaliana</i>  | 1 | 21 |
| EUC_8057 | ATCGGACCAGGCTTCATTCC   | miR397a    | <i>Arabidopsis thaliana</i>  | 0 | 20 |
| EUC_5434 | AACCTCGTCATCTCGGAGAAT  | miR397a    | <i>Brassica napus</i>        | 2 | 21 |
| EUC_7474 | CAAACCTGACTACCAAGATGA  | miR398a    | <i>Arabidopsis thaliana</i>  | 2 | 20 |
| EUC_4070 | GTTGTTGGGTCGGCAATCGTG  | miR399a-5p | <i>Zea mays</i>              | 3 | 21 |
| EUC_4464 | GCGAGGCCAGGTGGCACCGC   | miR399d    | <i>Cucumis melo</i>          | 0 | 21 |
| EUC_6152 | TGCCTTGCCATCCATATTGAA  | miR399e    | <i>Theobroma cacao</i>       | 0 | 21 |
| EUC_3893 | GTCACATGGGTACAAGTTTGG  | miR399e    | <i>Theobroma cacao</i>       | 1 | 21 |
| EUC_3885 | CTGTTGGAGCTGATGTCACGC  | miR399e    | <i>Theobroma cacao</i>       | 0 | 21 |
| EUC_3926 | CATGAATTCTATATGTGCATG  | miR399g    | <i>Cucumis melo</i>          | 2 | 21 |
| EUC_4098 | CCCAGATCTGGGCGAGGGCTC  | miR399g-5p | <i>Zea mays</i>              | 3 | 21 |
| EUC_8735 | CCCCAATTTTTTTCTTACGA   | miR408     | <i>Arabidopsis thaliana</i>  | 1 | 20 |
| EUC_3704 | TCAATAGCCAAAGGAAGCACT  | miR408-5p  | <i>Oryza sativa</i>          | 2 | 21 |
| EUC_7411 | CCAATTTTCGTTCTCTGGTCA  | miR4234    | <i>Arabidopsis lyrata</i>    | 3 | 20 |
| EUC_3845 | GAATCCATCTTCTATAGTCAT  | miR4414b   | <i>Medicago truncatula</i>   | 1 | 21 |
| EUC_6406 | TACAGTTCCACAAGTGTCTT   | miR447b-3p | <i>Vitis vinifera</i>        | 3 | 21 |
| EUC_4901 | GTTCCATAATCAGTTTGGGA   | miR447b-3p | <i>Vitis vinifera</i>        | 2 | 21 |
| EUC_3709 | GGTGGTGAGTTGACAAAGAG   | miR472b    | <i>Populus trichocarpa</i>   | 3 | 21 |
| EUC_8868 | TCTTTAATTCTGTTTAGCAA   | miR477a    | <i>Nicotiana tabacum</i>     | 2 | 20 |
| EUC_3969 | TATCTAGCCTTTCCAAAACCTC | miR477a    | <i>Nicotiana tabacum</i>     | 3 | 21 |
| EUC_8611 | TTTCTACGACTGGAGGACTT   | miR477g-3p | <i>Physcomitrella patens</i> | 2 | 20 |
| EUC_5441 | GTTGGGCCAAGGTCGAGGCC   | miR482     | <i>Vitis vinifera</i>        | 1 | 21 |
| EUC_7795 | GCTCTTGAGGTAATGGGTTT   | miR482     | <i>Vitis vinifera</i>        | 3 | 20 |

|          |                         |            |                            |   |    |
|----------|-------------------------|------------|----------------------------|---|----|
| EUC_3707 | TCAATAGCCAAAAGAAGCCCT   | miR482     | <i>Gossypium raimondii</i> | 3 | 21 |
| EUC_3026 | GACTAGTTGAAGTAATGAGCC   | miR482     | <i>Gossypium raimondii</i> | 2 | 21 |
| EUC_4105 | GTATGACATGTCACTTATCTT   | miR482     | <i>Gossypium raimondii</i> | 3 | 21 |
| EUC_5447 | GTTGGTTGGCTTCTCTACAAT   | miR482     | <i>Gossypium raimondii</i> | 3 | 21 |
| EUC_3708 | TCAATAGCCAAACGAAGCCCT   | miR482     | <i>Gossypium raimondii</i> | 3 | 21 |
| EUC_2401 | GTCCGTACCAGTTATGAGTTGA  | miR482a    | <i>Pinus taeda</i>         | 3 | 22 |
| EUC_5168 | CGTGTCTTGGGGACGTCGAT    | miR482a-3p | <i>Glycine max</i>         | 2 | 21 |
| EUC_7348 | AAGGCTATAGACTCGTTGAA    | miR482a-3p | <i>Glycine max</i>         | 2 | 20 |
| EUC_5467 | TCACGCACATGAGGAGACTGA   | miR482a-3p | <i>Malus domestica</i>     | 2 | 21 |
| EUC_2616 | TATGCAGCATCATCAAGATTC   | miR482a-3p | <i>Glycine max</i>         | 3 | 22 |
| EUC_2617 | AGAAGTACAAGATGAGATTTT   | miR482a-3p | <i>Glycine max</i>         | 2 | 22 |
| EUC_5517 | TCCCAAGGCCGCCATTCCGA    | miR482a-3p | <i>Glycine max</i>         | 3 | 21 |
| EUC_3703 | CAGTCCGATTCCCGGGTGGAT   | miR482a-3p | <i>Glycine max</i>         | 2 | 21 |
| EUC_4347 | ACAGGTA AAAACCCATTTT    | miR482a-3p | <i>Malus domestica</i>     | 2 | 21 |
| EUC_8647 | CCGGCAAGTCATCCTTGGCT    | miR482a-5p | <i>Glycine max</i>         | 3 | 20 |
| EUC_2388 | ACTGATGATCATCGTCCGAGAA  | miR482b    | <i>Gossypium hirsutum</i>  | 3 | 22 |
| EUC_4008 | TGGAAAGCATGATCCAACCAT   | miR482b    | <i>Gossypium hirsutum</i>  | 3 | 21 |
| EUC_2570 | CCAAGCTCAGCCTCTCGATGG   | miR482b    | <i>Solanum tuberosum</i>   | 3 | 22 |
| EUC_2321 | TCGAACATCTCCAAGAAAGCAG  | miR482b-3p | <i>Nicotiana tabacum</i>   | 3 | 22 |
| EUC_7349 | TTCCGGTTCATCCCGCATCG    | miR482c    | <i>Pinus taeda</i>         | 2 | 20 |
| EUC_8054 | CTCGGCGGACATGAACAGAC    | miR482c    | <i>Solanum tuberosum</i>   | 3 | 20 |
| EUC_7311 | ATCCTTGTTCTTGACGATAT    | miR482c    | <i>Solanum tuberosum</i>   | 3 | 20 |
| EUC_7272 | CCCCTCGCGTATTTAAGTCG    | miR482e    | <i>Solanum tuberosum</i>   | 3 | 20 |
| EUC_2380 | ATGCTTTCGCAGTTGTTTCGTCT | miR482e    | <i>Solanum tuberosum</i>   | 2 | 22 |
| EUC_7276 | GAGTCACTATTGAAGGCATG    | miR482e    | <i>Solanum tuberosum</i>   | 3 | 20 |
| EUC_7277 | GCAGCCCACAATATTGATCA    | miR482e    | <i>Solanum tuberosum</i>   | 3 | 20 |
| EUC_7275 | GAAGATGACCTTAGCAAGAG    | miR482e    | <i>Solanum tuberosum</i>   | 3 | 20 |
| EUC_2434 | TCTCCTAAGCTCTTTCAAAGAA  | miR482e    | <i>Solanum tuberosum</i>   | 3 | 22 |
| EUC_5454 | GGATGGGATGTCGGGAGAATG   | miR482e    | <i>Solanum tuberosum</i>   | 3 | 21 |
| EUC_5518 | TGGGCGGTCATGTTGGGACCA   | miR482e    | <i>Solanum tuberosum</i>   | 3 | 21 |
| EUC_5516 | AGTTGTCTTTCCATACCTCC    | miR482e    | <i>Solanum tuberosum</i>   | 3 | 21 |
| EUC_7278 | TGACGATCATGTGAAGGAGA    | miR482e    | <i>Solanum tuberosum</i>   | 1 | 20 |
| EUC_5455 | AGTACGGATGGGATGTCGGGA   | miR482e    | <i>Solanum tuberosum</i>   | 3 | 21 |
| EUC_5448 | CGAAATGCTAGACTGTTTGGT   | miR482e    | <i>Solanum tuberosum</i>   | 2 | 21 |
| EUC_7274 | TTCTCTTCAAGTCCATGA      | miR482e    | <i>Solanum tuberosum</i>   | 3 | 20 |

|          |                         |             |                              |   |    |
|----------|-------------------------|-------------|------------------------------|---|----|
| EUC_7540 | CTTGAAGCGCAAATTCTCCG    | miR482e     | <i>Solanum tuberosum</i>     | 2 | 20 |
| EUC_2470 | AGATCACGCTGGCAGCTTCATC  | miR482e     | <i>Solanum tuberosum</i>     | 3 | 22 |
| EUC_8053 | AAGAGCTTCGCCATATCCTT    | miR482e     | <i>Solanum tuberosum</i>     | 2 | 20 |
| EUC_2469 | TCTTGCCAATACCACCCATGCC  | miR482e     | <i>Solanum tuberosum</i>     | 3 | 22 |
| EUC_5453 | CTTCAAGTAATGATGACGAGT   | miR482e     | <i>Solanum tuberosum</i>     | 3 | 21 |
| EUC_7340 | TCAACTTCAACTCTTCAATT    | miR482e     | <i>Solanum tuberosum</i>     | 3 | 20 |
| EUC_2455 | AGAACTTGTTTAACTTGCATAT  | miR482e     | <i>Solanum tuberosum</i>     | 2 | 22 |
| EUC_2294 | ATGATAAAAATGTCCGATCTAGT | miR482e     | <i>Solanum tuberosum</i>     | 3 | 22 |
| EUC_6982 | TTGCCAGCCGTCTTCAAAT     | miR482e     | <i>Solanum tuberosum</i>     | 2 | 20 |
| EUC_2433 | TATCTCTTTGATGGTTGGAGCT  | miR482e     | <i>Solanum tuberosum</i>     | 3 | 22 |
| EUC_2435 | CATCTCCTAAGCTCTTTCAAAG  | miR482e     | <i>Solanum tuberosum</i>     | 3 | 22 |
| EUC_7273 | TGCATCCCTTCCAAAAGTCG    | miR482e     | <i>Solanum tuberosum</i>     | 3 | 20 |
| EUC_6981 | CTTACGCAAAGATCTGCTC     | miR482e     | <i>Solanum tuberosum</i>     | 1 | 20 |
| EUC_6452 | TCAAGAATAAATCCAGCCGT    | miR5248     | <i>Medicago truncatula</i>   | 3 | 21 |
| EUC_6691 | TGAATTCATCTCCGGCATAT    | miR5261     | <i>Medicago truncatula</i>   | 3 | 20 |
| EUC_6768 | CCATCAGAACTCCGAAGTTA    | miR5261     | <i>Medicago truncatula</i>   | 3 | 20 |
| EUC_6738 | AAAGAAGAGGCCGGCAAGGT    | miR5261     | <i>Medicago truncatula</i>   | 2 | 20 |
| EUC_6729 | CGGAGCAACAACCTGATCGAA   | miR5261     | <i>Medicago truncatula</i>   | 3 | 20 |
| EUC_6753 | CTCCCGAAACAAAGCTTATC    | miR5261     | <i>Medicago truncatula</i>   | 3 | 20 |
| EUC_8267 | TAAAAAAGAGAGAGGTGATA    | miR5261     | <i>Medicago truncatula</i>   | 1 | 20 |
| EUC_5588 | CGACTTCGATACTGATGAGTT   | miR5261     | <i>Medicago truncatula</i>   | 3 | 21 |
| EUC_3864 | ATCATTCATTCAATCAACTA    | miR530a     | <i>Citrus sinensis</i>       | 1 | 21 |
| EUC_2940 | TCAGCACTAGGGTCCCCAATG   | miR535a     | <i>Physcomitrella patens</i> | 0 | 21 |
| EUC_8013 | AAAGGGATAATGGTAGAAAG    | miR5565g-3p | <i>Sorghum bicolor</i>       | 3 | 20 |
| EUC_8669 | CAGGCATGCCTTCAAGAGAG    | miR6214     | <i>Hordeum vulgare</i>       | 2 | 20 |
| EUC_8650 | TTAAGAATGTTTTGACGGAA    | miR6478     | <i>Populus trichocarpa</i>   | 0 | 20 |
| EUC_3800 | CCTCGGTTTGATTTGATTCAT   | miR827      | <i>Malus domestica</i>       | 1 | 21 |
| EUC_6524 | GAGGCTCAAAGTTGTTCGG     | miR858      | <i>Malus domestica</i>       | 0 | 20 |
| EUC_6570 | CGAGCTCCCCACGATGTTAG    | miR858      | <i>Malus domestica</i>       | 1 | 20 |
| EUC_4228 | ATTTTGGGATGAACAAAATAG   | miR858a     | <i>Arabidopsis thaliana</i>  | 1 | 21 |
| EUC_3687 | TCCAATGTCTCTCTCCCTATT   | miR858b     | <i>Arabidopsis thaliana</i>  | 0 | 21 |
| EUC_3803 | GATCCGAAGACAACCTCTTGA   | miR858b     | <i>Arabidopsis thaliana</i>  | 0 | 21 |
| EUC_5481 | TCAACTTCTTGCAATCCAAAT   | miR858b     | <i>Arabidopsis thaliana</i>  | 2 | 21 |

## **ANEXO II**

Anotação das sequências de pequenos RNAs de *Eucalyptus* mapeadas pelo pacote de programas miRDeep2 no genoma de referência de *Eucalyptus grandis* com estrutura secundária predita a partir da sequência genômica flanqueante pelo programa RNAfold. Cromossomo onde a sequência foi mapeada, coordenada da posição da sequência no cromossomo (início e fim), fita senso (+) ou antisense (-) e sequência obtida no smRNA-Seq.

| <i>Cromossomo</i> | <i>Início</i> | <i>Fim</i> | <i>Fita</i> | <i>Sequência</i>               |
|-------------------|---------------|------------|-------------|--------------------------------|
| scaffold_1        | 5317735       | 5317802    | -           | G TTCAGTTTCTTATGATT CAGCTTGAC  |
| scaffold_1        | 29793496      | 29793572   | +           | A GGTGTCATACTAAATCTGACTATTT    |
| scaffold_1        | 8964013       | 8964065    | +           | T GGATATTTCAACACTTGTTGATC      |
| scaffold_1        | 26041774      | 26041837   | +           | C CTTATTGTTTTAAAGCGTGGTAC      |
| scaffold_1        | 11695211      | 11695302   | -           | A TACATAAGCTCAAACCAACTGGG      |
| scaffold_1        | 11978313      | 11978358   | -           | C GGGTGGGCGGTTCCCGGC           |
| scaffold_2        | 37526275      | 37526348   | -           | C AATTTACATGAACTCTTGATTTTGGTAG |
| scaffold_2        | 8901361       | 8901433    | -           | A CTGTTTTGGCTAGCGGGTCGGTGCAAAT |
| scaffold_2        | 30644960      | 30645028   | +           | A AGTTTTTCATCATTGAAGCTCAATAG   |
| scaffold_2        | 45657154      | 45657208   | +           | C AATTTTCATGGTTTTATATGAGATGG   |
| scaffold_2        | 43664039      | 43664107   | -           | G CCATAAATATAAAAGCAACATTGA     |
| scaffold_2        | 26919578      | 26919653   | -           | A CGTTGGCCTATGAGATTGTTGGA      |
| scaffold_2        | 5713770       | 5713832    | +           | C TTGAAGAATGGTTGCCAAACACC      |
| scaffold_2        | 6397973       | 6398027    | -           | G GTTCAGTTCTCGGGTAGATTCAT      |
| scaffold_2        | 36162697      | 36162757   | +           | T CATATTTGATTCATAACCTGTCG      |
| scaffold_2        | 36163955      | 36164015   | +           | T CATATTTGATCCATCACTTGTTG      |
| scaffold_2        | 60516084      | 60516140   | +           | T GC TTTGTGCTTTCAGCATT TTTA    |
| scaffold_2        | 46674871      | 46674940   | -           | T TGACCAGCCATCAGCGAGGCTCG      |
| scaffold_2        | 50779484      | 50779539   | +           | T ATCAACTCTCAACCAGTTTTTT       |
| scaffold_2        | 50114530      | 50114593   | -           | C TAATCCTCAGGTTGTGGATTAT       |
| scaffold_2        | 44579347      | 44579420   | +           | A ACTGCTGATCGTCCTTAGCT         |
| scaffold_2        | 50114529      | 50114594   | -           | A TCCTCAGGTTGTGGATTATC         |
| scaffold_2        | 21568113      | 21568173   | +           | A AATGAATATGTAGCAAGATG         |
| scaffold_2        | 50757583      | 50757638   | -           | T GGAGCGGCTGTCTTCACTTT         |
| scaffold_2        | 44568478      | 44568549   | -           | C TGCTGATCGTTCTTAGCTGA         |
| scaffold_2        | 42520479      | 42520565   | +           | T TTCCACGACTTTTTTCCTT          |
| scaffold_2        | 44181267      | 44181328   | -           | C GACGGTGAACGAATGGTT           |
| scaffold_2        | 46900630      | 46900715   | +           | C ACTATAAGAGCTAGCCAT           |

|            |          |          |   |                              |
|------------|----------|----------|---|------------------------------|
| scaffold_2 | 15501504 | 15501553 | - | GGGCTTACACCTACTCACA          |
| scaffold_3 | 17384720 | 17384773 | - | AGGGCGGTGCTAAGCTATGACCAATTG  |
| scaffold_3 | 59392853 | 59392910 | - | AACTCCTATTTTCGATGTGTCTTCTAT  |
| scaffold_3 | 59725906 | 59725963 | - | GACTCCTATTTTCGATGTGTCTTCTAT  |
| scaffold_3 | 46124276 | 46124348 | - | TTGACTACAATAGTTGTGAGGAGC     |
| scaffold_3 | 56270845 | 56270898 | + | AAATATTTTCTGGTGTTTGGATTG     |
| scaffold_3 | 32477079 | 32477159 | + | ACCGTTACGAATTGCTTGAAACCG     |
| scaffold_3 | 77857061 | 77857120 | - | TTTCATGGGTTTTTCGTGGAATG      |
| scaffold_3 | 33774400 | 33774485 | - | GCTCATTTCTCTTTCCGTCAAGC      |
| scaffold_3 | 63101314 | 63101367 | + | CACCAGATTCGGTGAGCTCGCG       |
| scaffold_3 | 62002495 | 62002571 | + | TTCTGGTAAATCACTGATCCGA       |
| scaffold_3 | 26337656 | 26337740 | + | GTGAAACCTATAAAGAACCTGG       |
| scaffold_3 | 16901740 | 16901830 | - | AGCAAGGGGCCTAACGCCCTCG       |
| scaffold_3 | 9674215  | 9674291  | - | AGGTCGATTCAATGAACCATGT       |
| scaffold_3 | 3925073  | 3925143  | - | GTACGTAAGTAAAATATTTT         |
| scaffold_3 | 4901554  | 4901642  | - | TACTTCTTACCTACAATTCTG        |
| scaffold_3 | 63013301 | 63013377 | + | GGGCAGTGACTTATCAGAAAA        |
| scaffold_3 | 75497983 | 75498049 | + | TCCACGCATCAATGTATTGGG        |
| scaffold_3 | 22949943 | 22950026 | + | AGCCACGGCGAGGCTCGGCCT        |
| scaffold_3 | 75497983 | 75498048 | + | CTCCACGCATCAATGTATTGG        |
| scaffold_3 | 37351603 | 37351650 | - | CTTTTGTCTCTCCAATGAA          |
| scaffold_3 | 1676342  | 1676426  | - | CGATTTTCTTCGAGCCACT          |
| scaffold_4 | 40072989 | 40073068 | - | GAATCGACTGGTTTGGTTCGGTTCTT   |
| scaffold_4 | 2856070  | 2856121  | + | GCGTTTTGTCAGTGATCACAAAAGA    |
| scaffold_4 | 1945023  | 1945103  | - | CTCACAAGTGTGTATTTTGTATTTT    |
| scaffold_4 | 10342362 | 10342428 | + | TATAAGGGGGTAGTAATGGTTCTTG    |
| scaffold_4 | 32822189 | 32822255 | + | GGGTCTATTCTCCAACTCCAGTTG     |
| scaffold_4 | 15167020 | 15167088 | + | CCCATTTCAAGTCATCAACTAGTGT    |
| scaffold_4 | 4346107  | 4346174  | + | TCAACAATGTATATCAAGCATATT     |
| scaffold_4 | 32900065 | 32900125 | + | TCTATTTTTTGTGTTGTTCACTA      |
| scaffold_4 | 35712130 | 35712215 | - | AGCATGGTGGTAAAGGCATCT        |
| scaffold_4 | 10840520 | 10840594 | - | CTGGTAAACCAGGGTCTTGT         |
| scaffold_4 | 23961816 | 23961893 | + | GAGCAAGGTACTAACTTGG          |
| scaffold_5 | 2852183  | 2852250  | + | GTTCCATGTCCATTGGCATGGCTGTTAC |
| scaffold_5 | 17245094 | 17245186 | - | TTCGGTTCCGAGTTACTTCTCTGGA    |
| scaffold_5 | 2855160  | 2855226  | + | TACTTAAAGAGGAAAGCTACTTGT     |
| scaffold_5 | 18557831 | 18557881 | - | TATGCTTTATTATACATAAAGAA      |
| scaffold_5 | 6693118  | 6693185  | - | CTATGAACTCTTGATTTTGGTAG      |
| scaffold_5 | 8969774  | 8969845  | + | TCTCTCCCTCAAAGACTTCCGA       |
| scaffold_5 | 6697174  | 6697239  | - | GTGTAGGGCTGTCATGGGATAG       |
| scaffold_5 | 57892783 | 57892850 | - | GCCCGTTTTGGTTCAATTTTGG       |

|            |          |          |   |                              |
|------------|----------|----------|---|------------------------------|
| scaffold_5 | 8969774  | 8969845  | + | GGAGGTCTTTGAGGTGAGTCC        |
| scaffold_5 | 1765810  | 1765855  | - | TTCTGTGATTATTTTCGTA          |
| scaffold_6 | 49705890 | 49705951 | - | CAATTATGTGCTGGCTGTTTGATCTGT  |
| scaffold_6 | 36752308 | 36752365 | + | ACCTCATATAATAGTGTATCTTTTT    |
| scaffold_6 | 34632585 | 34632665 | + | TCCCATCAGTTTGGTCGATCATGTC    |
| scaffold_6 | 35263408 | 35263498 | + | CTGGTTAGGCTGGGCGAGGCTCGAC    |
| scaffold_6 | 52395801 | 52395858 | + | GTCTATGGTCGCAAAGGCCAGATTT    |
| scaffold_6 | 32988777 | 32988835 | - | CTCCGATTTGGGTTTCATCTCTATA    |
| scaffold_6 | 34888415 | 34888474 | + | TACTCATCCGTACTTTAATCTTA      |
| scaffold_6 | 41845398 | 41845449 | + | ATTGTTATCAAACGGGTTTCTTT      |
| scaffold_6 | 49869821 | 49869904 | - | GCTCAGGACAATGCCATTAC         |
| scaffold_6 | 46457329 | 46457419 | - | TCGCGGCCAGGGCTCCAGATC        |
| scaffold_6 | 44575129 | 44575214 | + | CATGGTATTTTGTACCTCCG         |
| scaffold_7 | 50327614 | 50327699 | - | CCCCGAGAATGTGTAAGCGTGGTTATTT |
| scaffold_7 | 21948032 | 21948097 | - | TCGAGAGTATCGGAGAGTATCGAAC    |
| scaffold_7 | 6629417  | 6629499  | + | CCAGCACTGCCAATTGCTGGAGGTA    |
| scaffold_7 | 52336538 | 52336606 | - | CACTTATTTTCAGATATACATTCA     |
| scaffold_7 | 19481917 | 19481972 | + | CCAATTTTCTATTCAAATGTTCT      |
| scaffold_7 | 40119641 | 40119697 | - | ATGATCAACGGCTGTGATGGACTC     |
| scaffold_7 | 25658926 | 25659002 | + | GCCACCACGTCAGCGATTTTCAAC     |
| scaffold_7 | 35709030 | 35709106 | - | TTTAATATTGCTTTTATCATTCAA     |
| scaffold_7 | 7116505  | 7116568  | + | ATTTGAGTAACCGAACCTTTCTC      |
| scaffold_7 | 13245381 | 13245458 | + | AGGTCGATTCAATAGAACCATGT      |
| scaffold_7 | 6542066  | 6542138  | - | AGAATTTGGTGGAGCTGGAGAC       |
| scaffold_7 | 33275771 | 33275832 | + | TAGCAACTGTGCTAGGGACCGT       |
| scaffold_7 | 44038972 | 44039057 | - | TGTTTATGTTTAAAAATTGTTA       |
| scaffold_7 | 27629283 | 27629350 | + | GCCATTTGGTTCAACTTTGG         |
| scaffold_7 | 18424906 | 18424950 | + | TGATATGCCACGGAAGAATAG        |
| scaffold_7 | 33275052 | 33275123 | + | GCTGGTTGGGCGACTAAGGAG        |
| scaffold_7 | 35441197 | 35441255 | + | ACTTCTGACGGACGTGGGAAA        |
| scaffold_7 | 35441198 | 35441255 | + | CTTCTGACGGACGTGGGAAAG        |
| scaffold_7 | 18342016 | 18342088 | - | GACGATAGTTCATGCACGAAT        |
| scaffold_7 | 41617841 | 41617898 | + | GGTTTGGTTTCAAGATTAATT        |
| scaffold_7 | 23809792 | 23809871 | - | AAGTCCTTAGTCAGAGGTG          |
| scaffold_8 | 31126504 | 31126554 | + | GTCGTTGTCAGTGTTGTCGCTTCG     |
| scaffold_8 | 31133935 | 31134009 | + | GTAATTC AACATGTTATTGTCATTG   |
| scaffold_8 | 49196511 | 49196574 | + | GTTTCGAAATTGAAATGGCTTGCC     |
| scaffold_8 | 38204273 | 38204339 | + | TTGACGTTGTTCAATGCTATAAAA     |
| scaffold_8 | 17410401 | 17410490 | - | CCCCATGGGGTCTCAATCAATCTC     |
| scaffold_8 | 8730809  | 8730857  | - | AATAGAGTTGATCCTCGGTCTTA      |
| scaffold_8 | 69574908 | 69574983 | + | GCATTCGGCCGAATGCTAAAAGC      |

|             |          |          |   |                             |
|-------------|----------|----------|---|-----------------------------|
| scaffold_8  | 7443115  | 7443186  | + | AGATCACGCTGGCAGCTTCATC      |
| scaffold_8  | 1973507  | 1973571  | + | TCCTACTTAGCCGCCAATCAA       |
| scaffold_8  | 5335410  | 5335487  | - | CCCTTGCAAGTTTATCCGGTG       |
| scaffold_8  | 36363532 | 36363612 | + | ATGTCAAGCTTAAAATCAAAT       |
| scaffold_8  | 32371749 | 32371829 | + | ATCTCAAGCTCAAATCAAAT        |
| scaffold_8  | 53906658 | 53906740 | - | TGCTGCAAGTACTCAATGCC        |
| scaffold_8  | 19256241 | 19256320 | - | AATGCATGAATCAGGTCCATG       |
| scaffold_8  | 59285796 | 59285842 | + | TGGGTCCGCATGACCTCAGG        |
| scaffold_9  | 28910581 | 28910654 | + | TCTTTGTTTCTTTCGTTTGGTTGTTG  |
| scaffold_9  | 8685865  | 8685936  | - | CTTGCATCAAATCAGCAGTGGTCTTC  |
| scaffold_9  | 20015643 | 20015724 | - | TCCTTGCTTCTCCCTCATGCAAAGC   |
| scaffold_9  | 19201822 | 19201892 | - | AATCCCAGAGACATTCAATAAAT     |
| scaffold_9  | 23230509 | 23230566 | + | CATGAAATAGAGTGAACGAGCCC     |
| scaffold_9  | 578357   | 578415   | + | ACCTTGGGCTTTTAGCCTTCTAAA    |
| scaffold_9  | 38159437 | 38159505 | + | GCAAAGACTTCTTCAATAAATGC     |
| scaffold_9  | 28998996 | 28999082 | + | CCGTCTCTCCTGCGACAAGATG      |
| scaffold_9  | 19977754 | 19977836 | - | TGGCTTCTCCCTCACGCAAAGC      |
| scaffold_9  | 891672   | 891747   | + | CAATGTCGTCCTTGATCTTGA       |
| scaffold_9  | 23843047 | 23843109 | - | TCAAGAATTTGACTTTCAAAG       |
| scaffold_9  | 26196287 | 26196368 | + | TACTCCAGGCTACTTTCAATA       |
| scaffold_9  | 22353844 | 22353900 | + | TGCCGGATCGTACGAAGTTAG       |
| scaffold_9  | 28988680 | 28988747 | + | TGCCATCCTTCTACGACATGA       |
| scaffold_9  | 33470517 | 33470600 | - | TGTCGTTTTTTTTGCAGTCAA       |
| scaffold_9  | 364030   | 364074   | - | CAGAGGCACCGAATTAGAAT        |
| scaffold_10 | 24405493 | 24405580 | - | TGGGTCGGTCCTTGATCCTGGAATTCC |
| scaffold_10 | 28846438 | 28846508 | + | CCAATATTAACAAGTTTCTATTCTTT  |
| scaffold_10 | 10953566 | 10953659 | + | TTCACAATATTCCTCTCTTTTCTCC   |
| scaffold_10 | 17647583 | 17647649 | + | CGCGCTCGCCGATCTGGTGAGTCTC   |
| scaffold_10 | 10795415 | 10795496 | + | AGGAACGTGAACTTACTTGAATC     |
| scaffold_10 | 19921359 | 19921442 | - | AATCAAATCCTAAACTTTTCAA      |
| scaffold_10 | 19349778 | 19349829 | - | GTCCTAGGCCTCGACCTTGG        |
| scaffold_10 | 12142684 | 12142773 | + | TTTTAAGGTCTTGCTAAGGAC       |
| scaffold_11 | 29739765 | 29739832 | + | GTAGGATATGGTGAGGTCACAAAGTTG |
| scaffold_11 | 19890922 | 19890989 | + | TTGACTTATTCCAAATTCCAAAGTTG  |
| scaffold_11 | 40625339 | 40625414 | - | CGGGTTAGTCGGAACCATTGATCA    |
| scaffold_11 | 40625339 | 40625416 | - | ACCGGGTTAGTCGGAACCATTGAT    |
| scaffold_11 | 40625335 | 40625416 | - | ATGGTTGTGATTGACTCGGTCCAT    |
| scaffold_11 | 42245591 | 42245653 | - | CGAGAAAGTTGAAAGCCCAAGGCA    |
| scaffold_11 | 23723532 | 23723619 | - | TCTTCTCTCCACAATGTTGAA       |
| scaffold_11 | 17541956 | 17542042 | - | TCTTCTCTCCGCAATGTTCAA       |
| scaffold_11 | 43275663 | 43275749 | + | TCTTCCATTTTCCCTCCTCTT       |

|             |          |          |   |                       |
|-------------|----------|----------|---|-----------------------|
| scaffold_11 | 42264074 | 42264150 | + | GATCATGCTGGTTGCTTCAAC |
| scaffold_11 | 42711502 | 42711549 | - | TGAGGTCACGCAGACCTCAGG |

---

## **BIBLIOGRAFIA**

Addo-Quaye, C., T. W. Eshoo, *et al.* Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol, v.18, n.10, May 20, p.758-62. 2008.

Ahuja, M. R. e D. B. Neale. Evolution of Genome Size in Conifers. . Silvae Genetica, v.54, n.3, p.126-137. 2005.

Allen, E. e M. D. Howell. miRNAs in the biogenesis of trans-acting siRNAs in higher plants. Semin Cell Dev Biol, v.21, n.8, Oct, p.798-804. 2010.

Allen, E., Z. Xie, *et al.* microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell, v.121, n.2, Apr 22, p.207-21. 2005.

Allmer, J. e M. Yousef. Computational methods for ab initio detection of microRNAs. Front Genet, v.3, p.209. 2012.

Alonso-Peral, M. M., J. Li, *et al.* The microRNA159-regulated GAMYB-like genes inhibit growth and promote programmed cell death in Arabidopsis. Plant Physiol, v.154, n.2, Oct, p.757-71. 2010.

Altschul, S. F., T. L. Madden, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, v.25, n.17, Sep 1, p.3389-402. 1997.

Amarilyo, G. e A. La Cava. miRNA in systemic lupus erythematosus. Clin Immunol, v.144, n.1, Jul, p.26-31. 2012.

Ambros, V., R. C. Lee, *et al.* MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr Biol, v.13, n.10, May 13, p.807-18. 2003.

Anokye-Danso, F., M. Snitow, *et al.* How microRNAs facilitate reprogramming to pluripotency. J Cell Sci, Oct 17.

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

Arif, M. A., I. Fattash, *et al.* DICER-LIKE3 activity in *Physcomitrella patens* DICER-LIKE4 mutants causes severe developmental dysfunction and sterility. Mol Plant, v.5, n.6, Nov, p.1281-94. 2012.

Assis, T. F. D. Production and use of Eucalyptus hybrids for industrial purposes. Proceedings of the QFRI/CRC Workshop on Hybrid Breeding and Genetics of Forest Trees. Brisbane, Australia: Department of Primary Industries, 2000. 63–74 p.

Aukerman, M. J. e H. Sakai. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell, v.15, n.11, Nov, p.2730-41. 2003.

Axtell, M. J. Classification and Comparison of Small RNAs from Plants. Annu Rev Plant Biol, Jan 16. 2013.

Axtell, M. J. e D. P. Bartel. Antiquity of MicroRNAs and Their Targets in Land Plants. Plant Cell, v.17, n.6, p.1658-1673. 2005.

Axtell, M. J. e J. L. Bowman. Evolution of plant microRNAs and their targets. Trends Plant Sci, v.13, n.7, Jul, p.343-9. 2008.

Axtell, M. J., C. Jan, *et al.* A two-hit trigger for siRNA biogenesis in plants. Cell, v.127, n.3, Nov 3, p.565-77. 2006.

Axtell, M. J., J. A. Snyder, *et al.* Common functions for diverse small RNAs of land plants. Plant Cell, v.19, n.6, p.1750-1769. 2007.

Axtell, M. J., J. O. Westholm, *et al.* Vive la difference: biogenesis and evolution of microRNAs in plants and animals. Genome Biol, v.12, n.4, p.221. 2011.

Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, v.116, n.2, p.281-297. 2004.

Bartel, D. P. e C. Z. Chen. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nature Reviews Genetics, v.5, n.5, p.396-400. 2004.

Baulcombe, D. RNA silencing in plants. Nature, v.431, n.7006, Sep 16, p.356-363. 2004.

Baumberger, N. e D. C. Baulcombe. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A, v.102, n.33, Aug 16, p.11928-33. 2005.

Bedon, F., J. Grima-Pettenati, *et al.* Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). BMC Plant Biol, v.7, p.17. 2007.

Bolstad, B. M., R. A. Irizarry, *et al.* A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, v.19, n.2, Jan 22, p.185-93. 2003.

Bonnet, E., Y. He, *et al.* TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. Bioinformatics, v.26, n.12, Jun 15, p.1566-8. 2010.

Breakfield, N. W., D. L. Corcoran, *et al.* High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. Genome Res, v.22, n.1, Jan, p.163-76. 2011.

Carthew, R. W. e E. J. Sontheimer. Origins and Mechanisms of miRNAs and siRNAs. Cell, v.136, n.4, Feb 20, p.642-55. 2009.

Chang, S., J. Puryear, *et al.* A Simple and Efficient Method for Isolating RNA from Pine Trees. Plant Molecular Biology Reporter, v.11, n.2, p.113-116. 1993.

Chen, H., S. Jiang, *et al.* Improving panicle exertion of rice cytoplasmic male sterile line by combination of artificial microRNA and artificial target mimic. Plant Biotechnol J, Nov 17. 2012.

Chen, H. M., Y. H. Li, *et al.* Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. Proc Natl Acad Sci U S A, v.104, n.9, Feb 27, p.3318-23. 2007.

Chen, X. Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol, v.25, p.21-44. 2009.

\_\_\_\_\_. Small RNAs in development - insights from plants. Curr Opin Genet Dev, v.22, n.4, Aug, p.361-7. 2012.

Chinnusamy, V. e J. K. Zhu. RNA-directed DNA methylation and demethylation in plants. Sci China C Life Sci, v.52, n.4, Apr, p.331-43. 2009.

Chitwood, D. H., F. T. Nogueira, *et al.* Pattern formation via small RNA mobility. Genes Dev, v.23, n.5, Mar 1, p.549-54. 2009.

Cleveland, W. S. e S. J. Devlin. Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting. . Journal of the American Statistical Association v.83, n.403, p.596–610. 1988.

Corsini, L. R., G. Bronte, *et al.* The role of microRNAs in cancer: diagnostic and prognostic biomarkers and targets of therapies. Expert Opin Ther Targets, v.16 Suppl 2, Apr, p.S103-9. 2012.

Cuperus, J. T., A. Carbonell, *et al.* Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol, v.17, n.8, 2010, p.997-1003. 2010.

Cuperus, J. T., N. Fahlgren, *et al.* Evolution and functional diversification of MIRNA genes. Plant Cell, v.23, n.2, Feb, p.431-42. 2011.

D'alessandra, Y., G. Pompilio, *et al.* MicroRNAs and myocardial infarction. Curr Opin Cardiol, v.27, n.3, May, p.228-35. 2012.

Dai, X. e P. X. Zhao. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res, v.39, n.Web Server issue, Jul, p.W155-9. 2011.

De La Luz Gutierrez-Nava, M., M. J. Aukerman, *et al.* Artificial trans-acting siRNAs confer consistent and effective gene silencing. Plant Physiol, v.147, n.2, Jun, p.543-51. 2008.

Deluc, L., F. Barrieu, *et al.* Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. Plant Physiol, v.140, n.2, Feb, p.499-511. 2006.

Ding, D., L. Zhang, *et al.* Differential expression of miRNAs in response to salt stress in maize roots. Ann Bot, v.103, n.1, Jan, p.29-38. 2009.

Dolgosheina, E. V., R. D. Morin, *et al.* Conifers have a unique small RNA silencing signature. RNA, v.14, n.8, Aug, p.1508-15. 2008.

Du, J., E. Miura, *et al.* The Populus Class III HD ZIP transcription factor POPCORONA affects cell differentiation during secondary growth of woody stems. PLoS One, v.6, n.2, p.e17458. 2011.

Dunoyer, P., G. Schott, *et al.* Small RNA duplexes function as mobile silencing signals between plant cells. Science, v.328, n.5980, May 14, p.912-6. 2010.

Eldem, V., U. Celikkol Akcay, *et al.* Genome-wide identification of miRNAs responsive to drought in peach (*Prunus persica*) by high-throughput deep sequencing. PLoS One, v.7, n.12, p.e50298. 2012.

Fahlgren, N., M. D. Howell, *et al.* High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE, v.2, n.2, p.e219. 2007.

Fahlgren, N., S. Jogdeo, *et al.* MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. Plant Cell, v.22, n.4, Apr, p.1074-89. 2010.

Fahlgren, N., T. A. Montgomery, *et al.* Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. Curr Biol, v.16, n.9, May 9, p.939-44. 2006.

Fao. GLOBAL WOOD AND WOOD PRODUCTS FLOW. FAO. Shangai. 2007

Fire, A., S. Xu, *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature, v.391, n.6669, Feb 19, p.806-11. 1998.

Friedlander, M. R., S. D. Mackowiak, *et al.* miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res, v.40, n.1, Jan, p.37-52. 2011.

Frizzi, A. e S. Huang. Tapping RNA silencing pathways for plant biotechnology. Plant Biotechnol J, v.8, n.6, Aug, p.655-77. 2010.

Gaiotto, F. A., M. Bramucci, *et al.* Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* s.t. Blake with dominant RAPD and AFLP markers. . Theoretical and Applied Genetics v.95, p.842–849. 1997.

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

Geddy, R. e G. G. Brown. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genomics, v.8, p.130. 2007.

Govaerts, R., M. Sobral, *et al.* World checklist of Myrtaceae. Kew: Kew Publishing, Royal Botanic Gardens. 2008

Grattapaglia, D. Integrating genomics into *Eucalyptus* breeding. Genetics and Molecular Research, v. 3, n.3, p.369-379. 2004.

Grattapaglia, D. e M. Kirst. *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytologist, v.179, n.4, p.911-929. 2008.

Grattapaglia, D., R. E. Vaillancour, *et al.* Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. Tree Genetics & Genomes, v.8, n.3, p.463-508. 2012.

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

Griffiths-Jones, S., H. K. Saini, *et al.* miRBase: tools for microRNA genomics. Nucleic Acids Res, v.36, n.Database issue, Jan, p.D154-8. 2008.

Groszmann, M., I. K. Greaves, *et al.* Epigenetics in plants-vernalisation and hybrid vigour. Biochim Biophys Acta, v.1809, n.8, Aug, p.427-37. 2011.

\_\_\_\_\_. Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. Proc Natl Acad Sci U S A, v.108, n.6, Feb 8, p.2617-22. 2011.

Gursansky, N. R., I. R. Searle, *et al.* Mobile microRNAs hit the target. Traffic, v.12, n.11, Nov, p.1475-82. 2011.

Guzman, F., M. P. Almerao, *et al.* Identification of microRNAs from *Eugenia uniflora* by high-throughput sequencing and bioinformatics analysis. PLoS One, v.7, n.11, p.e49811. 2012.

Hofacker, I. L. Vienna RNA secondary structure server. Nucleic Acids Res, v.31, n.13, Jul 1, p.3429-31. 2003.

Howell, M. D., N. Fahlgren, *et al.* Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell, v.19, n.3, Mar, p.926-42. 2007.

Hsieh, L. C., S. I. Lin, *et al.* Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. Plant Physiol, v.151, n.4, Dec, p.2120-32. 2009.

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

Johnson, C., A. Kasprzewska, *et al.* Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res, v.19, n.8, Aug, p.1429-40. 2009.

Jones-Rhoades, M. W. Conservation and divergence in plant microRNAs. Plant Mol Biol, v.80, n.1, Sep, p.3-16. 2012.

Jones-Rhoades, M. W., D. P. Bartel, *et al.* MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol, v.57, p.19-53. 2006.

Jung, J. H. e C. M. Park. MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. Planta, v.225, n.6, May, p.1327-38. 2007.

Khraiwesh, B., M. A. Arif, *et al.* Transcriptional control of gene expression by microRNAs. Cell, v.140, n.1, Jan 8, p.111-22. 2010.

Khraiwesh, B., J. K. Zhu, *et al.* Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta, v.1819, n.2, Feb, p.137-48. 2011.

Kidner, C. A. e R. A. Martienssen. The developmental role of microRNA in plants. Curr Opin Plant Biol, v.8, n.1, p.38-44. 2005.

Ko, J. H., C. Prassinis, *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, v.169, n.3, p.469-78. 2006.

\_\_\_\_\_. 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, v.169, n.3, p.469-478. 2006.

Kozomara, A. e S. Griffiths-Jones. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res, v.39, n.Database issue, Jan, p.D152-7. 2011.

Kruszka, K., M. Pieczynski, *et al.* Role of microRNAs and other sRNAs of plants in their changing environments. J Plant Physiol, v.169, n.16, Nov 1, p.1664-72. 2012.

Kulcheski, F. R., L. F. De Oliveira, *et al.* Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics, v.12, p.307. 2011.

Kunej, T., I. Godnic, *et al.* Cross talk between microRNA and coding cancer genes. Cancer J, v.18, n.3, May-Jun, p.223-31. 2012.

Ladiges, P. Y., F. Udovicic, *et al.* Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. Journal of Biogeography, v.30, n.7, Jul, p.989-998. 2003.

Lee, R. C., R. L. Feinbaum, *et al.* The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell, v.75, n.5, p.843-854. 1993.

Lertpanyasampatha, M., L. Gao, *et al.* Genome-wide analysis of microRNAs in rubber tree (*Hevea brasiliensis* L.) using high-throughput sequencing. Planta, v.236, n.2, Aug, p.437-45. 2012.

Li, F. e S. W. Ding. Virus counterdefense: diverse strategies for evading the RNA-silencing immunity. Annu Rev Microbiol, v.60, p.503-31. 2006.

Li, F., D. Pignatta, *et al.* MicroRNA regulation of plant innate immune receptors. Proc Natl Acad Sci U S A, v.109, n.5, Jan 31, p.1790-5. 2012.

Li, H. e R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, v.25, n.14, Jul 15, p.1754-60. 2009.

Lu, C., S. S. Tej, *et al.* Elucidation of the small RNA component of the transcriptome. Science, v.309, n.5740, p.1567-1569. 2005.

Lu, S., Y. H. Sun, *et al.* MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. Plant J, v.51, n.6, Sep, p.1077-98. 2007.

\_\_\_\_\_. Novel and mechanical stress-responsive MicroRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. Plant Cell, v.17, n.8, p.2186-2203. 2005.

Luo, Q. J., A. Mittal, *et al.* An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in Arabidopsis. Plant Mol Biol, v.80, n.1, Sep, p.117-29. 2011.

Lynd, L. R., M. S. Laser, *et al.* How biotech can transform biofuels. Nature Biotechnology, v.26, n.2, p.169-172. 2008.

Mallory, A. C., T. Elmayan, *et al.* MicroRNA maturation and action--the expanding roles of ARGONAUTES. Curr Opin Plant Biol, v.11, n.5, Oct, p.560-6. 2008.

Mallory, A. C. e H. Vaucheret. Functions of microRNAs and related small RNAs in plants. Nature Genetics, v.38, p. Suppl:S31-36. 2006.

Manavella, P. A., D. Koenig, *et al.* Plant secondary siRNA production determined by microRNA-duplex structure. Proc Natl Acad Sci U S A, v.109, n.7, Feb 14, p.2461-6. 2012.

Marian, A. J. Elements of 'missing heritability'. Curr Opin Cardiol, v.27, n.3, May, p.197-201. 2012.

Martinez, G., J. Forment, *et al.* High-throughput sequencing, characterization and detection of new and conserved cucumber miRNAs. PLoS One, v.6, n.5, p.e19523. 2011.

Matranga, C. e P. D. Zamore. Small silencing RNAs. Current Biology, v.17, n.18, p.789-793. 2007.

Mecchia, M. A., J. M. Debernardi, *et al.* MicroRNA miR396 and RDR6 synergistically regulate leaf development. Mech Dev, Aug 4. 2012.

Meyers, B. C., M. J. Axtell, *et al.* Criteria for annotation of plant MicroRNAs. Plant Cell, v.20, n.12, Dec, p.3186-90. 2008.

Meyers, B. C., A. Kozik, *et al.* Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell, v.15, n.4, Apr, p.809-34. 2003.

Mi, S., T. Cai, *et al.* Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell, v.133, n.1, Apr 4, p.116-27. 2008.

Molnar, A., C. W. Melnyk, *et al.* Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. Science, v.328, n.5980, May 14, p.872-5. 2010.

Montgomery, T. A., M. D. Howell, *et al.* Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell, v.133, n.1, Apr 4, p.128-41. 2008.

Montgomery, T. A., S. J. Yoo, *et al.* AGO1-miR173 complex initiates phased siRNA formation in plants. Proc Natl Acad Sci U S A, v.105, n.51, Dec 23, p.20055-62. 2008.

Moran, G., C. B. J, *et al.* Reduction in levels of inbreeding in a seed orchard of Eucalyptus regnans F. Muell, compared with natural populations. Silvae Genetica v.38, p.32–36. 1989.

Morin, R. D., G. Aksay, *et al.* Comparative analysis of the small RNA transcriptomes of Pinus contorta and Oryza sativa. Genome Res, v.18, n.4, Apr, p.571-84. 2008.

Mosher, R. A. e C. W. Melnyk. siRNAs and DNA methylation: seedy epigenetics. Trends Plant Sci, v.15, n.4, Apr, p.204-10. 2010.

Nag, A., S. King, *et al.* miR319a targeting of TCP4 is critical for petal growth and development in Arabidopsis. Proc Natl Acad Sci U S A, v.106, n.52, Dec 29, p.22534-9. 2009.

Napoli, C., C. Lemieux, *et al.* Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell, v.2, n.4, Apr, p.279-289. 1990.

Navarro, L., F. Jay, *et al.* Suppression of the microRNA pathway by bacterial effector proteins. Science, v.321, n.5891, Aug 15, p.964-7. 2008.

Nikitina, E. G., L. N. Urazova, *et al.* MicroRNAs and human cancer. Exp Oncol, v.34, n.1, p.2-8. 2012.

Nikovics, K., T. Blein, *et al.* The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell, v.18, n.11, Nov, p.2929-45. 2006.

Ong, S. S. e R. Wickneswari. Expression profile of small RNAs in *Acacia mangium* secondary xylem tissue with contrasting lignin content - potential regulatory sequences in monolignol biosynthetic pathway. BMC Genomics, v.12 Suppl 3, Nov 30, p.S13. 2012.

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

Ossowski, S., R. Schwab, *et al.* Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J, v.53, n.4, Feb, p.674-90. 2008.

Papageorgiou, N., D. Tousoulis, *et al.* The role of microRNAs in cardiovascular disease. Curr Med Chem, v.19, n.16, p.2605-10. 2012.

Park, W., J. Li, *et al.* CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr Biol, v.12, n.17, Sep 3, p.1484-95. 2002.

Patzlaff, A., S. Mcinnis, *et al.* Characterisation of a pine MYB that regulates lignification. Plant J, v.36, n.6, Dec, p.743-54. 2003.

Peragine, A., M. Yoshikawa, *et al.* SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. Genes Dev, v.18, n.19, Oct 1, p.2368-79. 2004.

Pfaff, N., T. Moritz, *et al.* miRNAs involved in the generation, maintenance, and differentiation of pluripotent cells. J Mol Med (Berl), v.90, n.7, Jul, p.747-52.

Potts, B. M. e M. Savva, Eds. Self-incompatibility in Eucalyptus. Pollination '88. Melbourne: University of Melbourne, p.165-170, Pollination '88ed. 1988.

Prüfer, K., U. Stenzel, *et al.* PatMan: rapid alignment of short sequences to large databases. Bioinformatics, v.24, n.13, Jul 1, p.1530-1. 2008.

Pryor, L. D. e L. A. S. Johnson. A classification of the eucalypts. Australia: Australian National University Press. 1971

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

Rivals, E., C. Bruyere, *et al.* Formation of the *Arabidopsis* pentatricopeptide repeat family. Plant Physiol, v.141, n.3, Jul, p.825-39. 2006.

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

Romano, N. e G. Macino. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. Mol Microbiol, v.6, n.22, Nov, p.3343-53. 1992.

Schreiber, A. W., B. J. Shi, *et al.* Discovery of barley miRNAs through deep sequencing of short reads. BMC Genomics, v.12, p.129. 2011.

Schuler, M. A. e D. Werck-Reichhart. Functional genomics of P450s. Annu Rev Plant Biol, v.54, p.629-67. 2003.

Schwab, R., A. Maizel, *et al.* Endogenous TasiRNAs mediate non-cell autonomous effects on gene regulation in *Arabidopsis thaliana*. PLoS One, v.4, n.6, p.e5980. 2009.

Schwab, R., J. F. Palatnik, *et al.* Specific effects of microRNAs on the plant transcriptome. Development Cell, v.8, n.4, p.517-527. 2005.

Shivaprasad, P. V., H. M. Chen, *et al.* A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell, v.24, n.3, Mar, p.859-74. 2012.

Silvicultura, S. B. D. Fatos e Números do Brasil Florestal. Sociedade Brasileira de Silvicultura. 2003

Song, C., J. Fang, *et al.* Identification and characterization of 27 conserved microRNAs in citrus. Planta, v.230, n.4, Sep, p.671-85. 2009.

Song, C., C. Wang, *et al.* Deep sequencing discovery of novel and conserved microRNAs in trifoliolate orange (*Citrus trifoliata*). BMC Genomics, v.11, p.431. 2010.

Song, C., M. Yu, *et al.* Validation and characterization of *Citrus sinensis* microRNAs and their target genes. BMC Res Notes, v.5, p.235. 2012.

Song, Q. X., Y. F. Liu, *et al.* Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol, v.11, p.5. 2011.

Stracke, R., H. Ishihara, *et al.* Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J, v.50, n.4, May, p.660-77. 2007.

Sun, Y. H., R. Shi, *et al.* MicroRNAs in trees. Plant Mol Biol, v.80, n.1, Sep, p.37-53. 2011.

Sunkar, R., T. Girke, *et al.* Cloning and Characterization of MicroRNAs from Rice. Plant Cell, v.17, n.5, p.1397-1411. 2005.

Talmor-Neiman, M., R. Stav, *et al.* Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. Plant J, v.48, n.4, Nov, p.511-21. 2006.

Tang, G. Plant microRNAs: an insight into their gene structures and evolution. Semin Cell Dev Biol, v.21, n.8, Oct, p.782-9. 2010.

Tarver, J. E., P. C. Donoghue, *et al.* Is evolutionary history repeatedly rewritten in light of new fossil discoveries? Proc Biol Sci, v.278, n.1705, Feb 22, p.599-604. 2010.

Tempel, S. e F. Tahi. A fast ab-initio method for predicting miRNA precursors in genomes. Nucleic Acids Res, v.40, n.11, Jun, p.e80. 2012.

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

Tuskan, G. A., S. Difazio, *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science, v.313, n.5793, Sep 15, p.1596-604. 2006.

Varkonyi-Gasic, E., R. H. Lough, *et al.* Kiwifruit floral gene APETALA2 is alternatively spliced and accumulates in aberrant indeterminate flowers in the absence of miR172. Plant Mol Biol, v.78, n.4-5, Mar, p.417-29. 2012.

Voinnet, O. Induction and suppression of RNA silencing: insights from viral infections. Nat Rev Genet, v.6, n.3, Mar, p.206-20. 2005a.

\_\_\_\_\_. Non-cell autonomous RNA silencing. FEBS Lett, v.579, n.26, Oct 31, p.5858-71. 2005b.

\_\_\_\_\_. Origin, biogenesis and activity of plant microRNAs. Cell, v.136, n.4, Feb 20, p.669-87. 2009.

Wan, L. C., F. Wang, *et al.* Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. BMC Plant Biol, v.12, p.146. 2012.

Wan, L. C., H. Zhang, *et al.* Transcriptome-wide identification and characterization of miRNAs from *Pinus densata*. BMC Genomics, v.13, p.132. 2012.

Wang, J. W., M. Y. Park, *et al.* miRNA control of vegetative phase change in trees. PLoS Genet, v.7, n.2, Feb, p.e1002012. 2011.

Wang, J. W., R. Schwab, *et al.* Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in *Arabidopsis thaliana*. Plant Cell, v.20, n.5, May, p.1231-43. 2008.

Wang, L., X. Gu, *et al.* miR396-targeted AtGRF transcription factors are required for coordination of cell division and differentiation during leaf development in *Arabidopsis*. J Exp Bot, v.62, n.2, Jan, p.761-73. 2011.

Wang, L., H. Liu, *et al.* Identification and characterization of maize microRNAs involved in the very early stage of seed germination. BMC Genomics, v.12, p.154. 2011.

Wang, T., L. Chen, *et al.* Identification of drought-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. BMC Genomics, v.12, p.367. 2011.

Wang, X. J., J. L. Reyes, *et al.* Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol, v.5, n.9, p.R65. 2004.

Wang, Z. J., J. Q. Huang, *et al.* Discovery and profiling of novel and conserved microRNAs during flower development in *Carya cathayensis* via deep sequencing. Planta, v.236, n.2, Aug, p.613-21. 2012.

Wei, L. Q., L. F. Yan, *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, v.12, n.6, p.R53. 2011.

Williams, L., C. C. Carles, *et al.* A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the *Arabidopsis* ARF2, ARF3, and ARF4 genes. Proc Natl Acad Sci U S A, v.102, n.27, Jul 5, p.9703-8. 2005.

Wilson, P. G. Myrtaceae. In: K. Kubitzki (Ed.). The Families and Genera of Vascular Plants Springer, v.X: Sapindales, Cucurbitales, Myrtaceae, 2011. Myrtaceae, p.212-271

Wong, M. M., C. H. Cannon, *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, v.12, p.342. 2011.

Wu, G., M. Y. Park, *et al.* The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. Cell, v.138, n.4, Aug 21, p.750-9. 2009.

Wu, G. e R. S. Poethig. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. Development, v.133, n.18, Sep, p.3539-47. 2006.

Wu, L., H. Zhou, *et al.* DNA methylation mediated by a microRNA pathway. Mol Cell, v.38, n.3, May 14, p.465-75. 2010.

Xia, R., H. Zhu, *et al.* Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol, v.13, n.6, p.R47. 2012.

Xuan, P., M. Guo, *et al.* PlantMiRNAPred: efficient classification of real and pseudo plant pre-miRNAs. Bioinformatics, v.27, n.10, May 15, p.1368-76. 2011.

Yakovlev, I. A., C. G. Fossdal, *et al.* MicroRNAs, the epigenetic memory and climatic adaptation in Norway spruce. New Phytol, v.187, n.4, Sep, p.1154-69. 2010.

Yao, Y. e Q. Sun. Exploration of small non coding RNAs in wheat (*Triticum aestivum* L.). Plant Mol Biol, v.80, n.1, Sep, p.67-73. 2011.

Yoshikawa, M., A. Peragine, *et al.* A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev, v.19, n.18, Sep 15, p.2164-75. 2005.

Zen, K. e C. Y. Zhang. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med Res Rev, v.32, n.2, Mar, p.326-48. 2012.

Zhai, J., D. H. Jeong, *et al.* MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev, v.25, n.23, Dec 1, p.2540-53. 2011.

Zhang, B., X. Pan, *et al.* Plant microRNA: a small regulatory molecule with big impact. Developmental Biology, v.289, n.1, p.3-16. 2005.

Zhang, C., G. Li, *et al.* Identification of trans-acting siRNAs and their regulatory cascades in grapevine. Bioinformatics, v.28, n.20, Oct 15, p.2561-8. 2012.

Zhang, L., Y. Zheng, *et al.* Identification and temporal expression analysis of conserved and novel microRNAs in Sorghum. Genomics, v.98, n.6, Dec, p.460-8. 2011.

Zhang, X., Z. Zou, *et al.* Over-expression of sly-miR156a in tomato results in multiple vegetative and reproductive trait alterations and partial phenocopy of the sft mutant. FEBS Lett, v.585, n.2, Jan 21, p.435-9. 2010.

Zhang, Y., Y. Bai, *et al.* Bioinformatics prediction of miRNAs in the *Prunus persica* genome with validation of their precise sequences by miR-RACE. J Plant Physiol, v.170, n.1, Jan 1, p.80-92. 2012.

Zhang, Y. Y., M. Fischer, *et al.* Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytol, v.197, n.1, Jan, p.314-22. 2013.

Zhong, R., E. A. Richardson, *et al.* The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. Plant Cell, v.19, n.9, Sep, p.2776-92. 2007.

Zhong, S., Z. Fei, *et al.* Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnol, v.31, n.2, Jan 27, p.154-9. 2013.

Zhou, X., Q. Zhu, *et al.* MicroRNA profiling using microParaflo microfluidic array technology. Methods Mol Biol, v.822, p.153-82. 2012.

Zhou, Z. S., J. B. Song, *et al.* Genome-wide identification of *Brassica napus* microRNAs and their targets in response to cadmium. J Exp Bot, v.63, n.12, Jul, p.4597-613. 2012.

Zhu, C., Y. Ding, *et al.* MiR398 and plant stress responses. Physiol Plant, v.143, n.1, Sep, p.1-9. 2011.

Zhu, H., F. Hu, *et al.* *Arabidopsis* Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell, v.145, n.2, Apr 15, p.242-56. 2011.

Zhu, H., R. Xia, *et al.* Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. BMC Plant Biol, v.12, p.149. 2012.

Zhu, Q. H. e C. A. Helliwell. Regulation of flowering time and floral patterning by miR172. J Exp Bot, v.62, n.2, Jan, p.487-95. 2011.