

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**Análise da diversidade molecular em *Araucaria  
angustifolia* a partir de sequências de microRNAs e do  
cloroplasto**

José Henrique Galdino

Porto Alegre

2018

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**Análise da diversidade molecular em *Araucaria  
angustifolia* a partir de sequências de microRNAs e do  
cloroplasto**

José Henrique Galdino

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular).


Orientador: Prof. Dr. Rogerio Margis

Coorientador: Dr. Frank Guzman Escudero

Porto Alegre, dezembro de 2018

## INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi realizado no Laboratório de Genômica e Populações de Plantas, Centro de Biotecnologia e Departamento de Biofísica da Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, com apoio financeiro do CNPq e da CAPES. O doutorando obteve bolsa de estudos do CNPq por um período de 48 meses.

 *No pain no gain*

## AGRADECIMENTOS

A Deus pelo dom da vida, pelas suas misericórdias que se renovam a cada manhã, por todas as oportunidades que tem me proporcionado e por me sustentar e guiar os meus passos nos momentos mais difíceis.

À minha mãe e minha irmã pelo amor incondicional, por todo o apoio, carinho e compreensão, por sempre estarem do meu lado e acreditarem em mim até mesmo quando eu não acreditei. Ao meu pai por ter fé em mim e sempre me apoiar nas minhas escolhas. Amo vocês!

À minha segunda mãe, tia Liege, a Robertinha, tio Tonho e tia Márcia por todo carinho e apoio.

A Tahís, Allen, Juninho e João Marcos por terem me recebido de portas abertas e com todo o carinho quando cheguei a Porto Alegre. Sou eternamente grato a vocês!

Ao meu orientador, professor Rogerio Margis, por atender a minha ligação da Bahia a quatro anos atrás e aceitar me orientar no doutorado. Pela sua orientação, confiança, paciência e compreensão. Por compartilhar um pouco da sua imensa experiência profissional comigo.

Ao meu coorientador, Frank, pela paciência e por me apoiar no desenvolvimento desse trabalho.

Aos meus colegas e parceiros queridos de laboratório, Isabel, Érika, Pábulo, Débora, Igor, Natália e Cordenonsi pela convivência maravilhosa, por todo o apoio e paciência, pela amizade, pela parceria em alguns dos melhores shows de metal que fui na minha vida, por me proporcionarem tantos momentos felizes e agradáveis.

A Ronei e Daniela, Helder e Cássia, Renato Bunge, Khym, Pri Yuyama, Nureyev e Maria, amigos irmãos do peito que me apoiaram incondicionalmente nessa fase. Sou eternamente grato a vocês!

A Laurinho, Jorge e Gabriela, Ferras, Botassoli, Breninho, Thomaz, Pedron, Raiane, Ana L. Paiva, Karen, Thais Lopes, Silvinha, Cida e Cris Todeschini pelo carinho, parceria e amizade.

A Elisa e sua família pelo carinho e respeito, por todo o apoio e parceria, por me proporcionarem momentos tão maravilhosos. Vocês moram no meu coração.

Ao PPGBM/UFRGS e ao CNPq pela oportunidade de estudo. À banca por aceitar contribuir com o meu trabalho.

Ao Elmo por toda a tranquilidade e serenidade, pela sua maestria em fazer o bem.

## SUMÁRIO

ABREVIATURAS.....	7
RESUMO.....	8
ABSTRACT.....	10
1. INTRODUÇÃO .....	12
1.1. <i>Araucaria angustifolia</i> (Bertol.) Kuntze.....	12
1.2. RNA-seq .....	14
1.2.1. Bioinformática aplicada nas análises de RNA-seq .....	16
1.2.2. Estudos transcriptômicos em plantas .....	18
1.3. microRNAs em plantas .....	20
1.3.1 Aspectos gerais .....	20
1.3.2. Biogênese e processamento .....	21
1.3.3. Montagem do complexo de silenciamento.....	23
1.3.4. Mecanismos de ação.....	24
1.3.5. Regulação da expressão gênica.....	24
1.3.6. Predição e análise de expressão de miRNAs .....	25
1.3.7. Predição computacional dos alvos.....	27
1.3.8. Aspectos evolutivos dos miRNAs de plantas .....	28
1.4. Genoma de cloroplasto cpDNA .....	30
2. OBJETIVOS.....	32
3 CAPITULO 1- Novel and conserved miRNAs among Brazilian pine and other gymnosperms.....	34
4 CAPÍTULO 2- <i>Araucaria angustifolia</i> chloroplast genome sequence and its relation to other Araucariaceae.....	53
5. DISCUSSÃO E CONSIDERAÇÕES FINAIS.....	60
6 REFERÊNCIAS .....	67

## **ABREVIATURAS**

AGO – argonauta

cDNA – DNA complementar

cpDNA – genoma de cloroplasto

DCL1 - DICER-LIKE1

EXPO5 - EXPORTIN5

miRNA – microRNA

mRNA – RNA mensageiro

NBS-LRR - nucleotide-binding site leucine-rich repeat

NCBI - National Center for Biotechnology Information

NGS - Next Generation Sequencing

Pre-miRNA – precursor de miRNA

Pri-miRNA – transcrito primário de miRNA

PTGS – post transcriptional gene silencing

RISC - RNA-induced silencing complex

RNA-seq – RNA sequencing

rRNA – RNA ribossômico

RT-qPCR – reverse transcriptase quantitative polymerase chain reaction

SNP - single nucleotide polymorphism

sRNA – small RNA

SSR – simple sequence repeat

tRNA – RNA de transferência

## RESUMO

*Araucaria angustifolia*, popularmente conhecida como Pinheiro do paran,  a principal confera endmica do Brasil com importncia comercial, destacando-se como a principal espcie madeireira da regio Sul do pas no sculo passado. Por apresentar madeira de excelente qualidade e alto valor comercial, esta espcie foi alvo de extrativismo predatrio durante dcadas, sofrendo um decrscimo populacional acentuado. Atualmente, *A. angustifolia*  classificada como espcie criticamente ameaada de extino. Esta espcie tem sido alvo de diversos estudos genticos, principalmente com foco em embriognese somtica. Recentemente, uma anlise transcriptmica em estgios iniciais de desenvolvimento foi realizada a partir de dados de RNA-seq, os quais encontram-se disponveis no NCBI. No presente estudo, esses dados foram utilizados em novas anlises a fim de se obter novos recursos genticos e genmicos em *A. angustifolia*.

No primeiro captulo, o transcriptoma do Pinheiro do paran foi analisado com foco na identificao de miRNAs. Este  o primeiro estudo baseado na identificao e anlise de pequenos RNAs em uma espcie da famlia Araucariaceae. A busca por miRNAs conservados no Pinheiro do paran resultou na identificao de 115 sequncias representativas de 30 famlias. Um total de 106 precursores de miRNAs foram preditos, dos quais, 41 correspondem a miRNAs conservados de 16 diferentes famlias, e 65 foram anotados como novos miRNAs. A comparao das sequncias de pre-miRNAs do pinheiro do paran com bibliotecas de pequenos RNAs de outras cinco espcies de gimnospermas indicou que 9 dos 65 novos miRNAs so conservados em gimnospermas, enquanto que 56 possivelmente so especficos do Pinheiro do paran, do gnero *Araucaria*, ou da famlia Araucariaceae. Em outra anlise comparativa, onde as sequncias dos novos pre-miRNAs do Pinheiro do paran foram alinhadas com unigenes do transcriptoma de *A. cunninghamii*, sete ortlogos foram identificados nessa espcie congnerica da Oceania. Uma srie de miRNAs identificados por anlises computacionais de predio foi validada experimentalmente atravs da tcnica de *stem-loop* RT-qPCR. Nessa anlise, os padres de expresso de miRNAs novos e conservados foram analisados em cinco tecidos distintos coletados de plantas com 90 dias de crescimento.

No segundo captulo, os dados pblicos de RNA-seq foram usados para determinar a sequncia completa do genoma de cloroplasto (cpDNA) de *A. angustifolia*. O cpDNA do Pinheiro do paran tem 146.203 pares de bases de comprimento, e abriga 122 genes, dos quais, 80 codificam



polipeptídios, 5 codificam RNAs ribossômicos e 37 codificam RNAs de transferência. As regiões codificantes totalizam 45,02%, sendo que 4,96% equivalem a rRNAs e tRNAs. Os genes normalmente localizados nas repetições invertidas (IRs) estão presentes em cópias únicas, com exceção dos genes *rrn5* e *tRNA-CAU*. A estrutura típica encontrada no cpDNA de vários grupos de plantas terrestres, LSC/IR/SSC/IR, não está presente no plastoma de *A. angustifolia*. Análises filogenéticas baseadas no método Bayesiano e Máxima Verossimilhança, com unanimidade e alto suporte estatístico, agruparam *A. angustifolia* com as espécies congênicas *A. heterophylla* e *A. columnaris*. Análises computacionais de predição de microssatélites indicaram que o cpDNA de *A. angustifolia* apresenta 100 SSRs, dos quais 14 correspondem a loci de tetranucleotídeos.

O presente estudo apresenta aspectos importantes em relação à composição, evolução e diversidade de miRNAs de *A. angustifolia*, uma espécie da família Araucariaceae criticamente ameaçada de extinção. Além de novos recursos genéticos resultantes da identificação de miRNAs, o presente trabalho também apresenta novos recursos genômicos em *A. angustifolia*, com a montagem e anotação do genoma de cloroplasto. O cpDNA de *A. angustifolia* apresenta importantes características que podem ser exploradas em diferentes aspectos, como estudos de diversidade populacional, estudos filogenéticos, entre outros.

Palavras-chave: *Araucaria angustifolia*, RNA-seq, miRNA, genoma de cloroplasto

## ABSTRACT

*Araucaria angustifolia*, commonly known as Brazilian pine, is the only conifer in Brazil with economic importance and was the most important wood species from south Brazil in the past century. Representing valuable source of high-quality wood, Brazilian pine was the target of extensive exploitation activities over decades, suffering massive population decrease. Currently, this conifer is classified as critically endangered species. Brazilian pine has been targeted by some genetic studies mainly with a focus on somatic embryogenesis. Recently, an RNA-seq data were used to perform a transcriptome comparative profile analysis of early development stages, and the sequences are available at NCBI repository. Once a high-throughput sequencing data is released, it can be used in a plethora of ways beyond the original purpose, and more information can be explored in the target organism. In the present study, public RNA-seq data in combination with bioinformatic and experimental analysis were used to identify novel genetic and genomic resources in *A. angustifolia*.

In chapter one, Brazilian pine has its transcriptome explored in respect to microRNAs, representing the first sRNA description in a member of the Araucariaceae family. The screening for conserved miRNAs in Brazilian pine revealed 115 sequences of 30 miRNA families. A total of 106 precursors sequences were predicted. 41 comprised conserved miRNAs from 16 families, whereas 65 were annotated as novel miRNAs. The comparison of Brazilian pine precursors with sRNA libraries of other five conifer species indicates that 9 out 65 novel miRNAs are conserved among gymnosperms, while 56 seem to be specific for Brazilian pine or restricted to Araucariaceae family. Another analysis comparing novel Brazilian pine miRNAs precursors and *Araucaria cunninghamii* RNA-seq data identified seven orthologs between both species. Mature miRNA identified by bioinformatic predictions were validated using stem-loop RT-qPCR assays. The expression pattern of conserved and novel miRNAs was analyzed in five different tissues of three-month-old *Araucaria* seedlings.

In chapter two, RNA-Seq data was used to determine the complete nucleotide sequence of the *A. angustifolia* chloroplast genome (cpDNA). The *A. angustifolia* cpDNA is 146,203 bp in length and contains 122 genes, including 80 protein-coding genes, 4 ribosomal RNA genes, and 36 tRNA genes. Coding regions comprise 45.02%; 4.96% correspond to rRNAs and tRNAs, and 50.02% of the genome encompasses non-coding regions. Genes found in inverted repeat (IR) are present as

single copy, with exceptions to *rrn5* and *trnI-AUG* loci. The typical LSC, SSC, IRa and IRb organization reported in several land-plant groups is not present in *A. angustifolia* cpDNA. Phylogenetic analyses using Bayesian and Maximum Likelihood methods clustered *A. angustifolia* in the Araucariaceae family, with *A. heterophylla* and *A. columnaris* as congeneric species. The screening of *A. angustifolia* cpDNA revealed 100 SSRs, 14 of them corresponding to tetrapolymer loci.

The present study provides important insights into the nature and composition of miRNAs in a critically endangered species from Araucariaceae, with valuable information on miRNA diversity and conservation in this taxon. Also, the *A. angustifolia* chloroplast genome assembled in the present study represents the first complete cpDNA genome sequence for this species and shows a set of features that could be further explored for population and phylogenetic studies within this group.

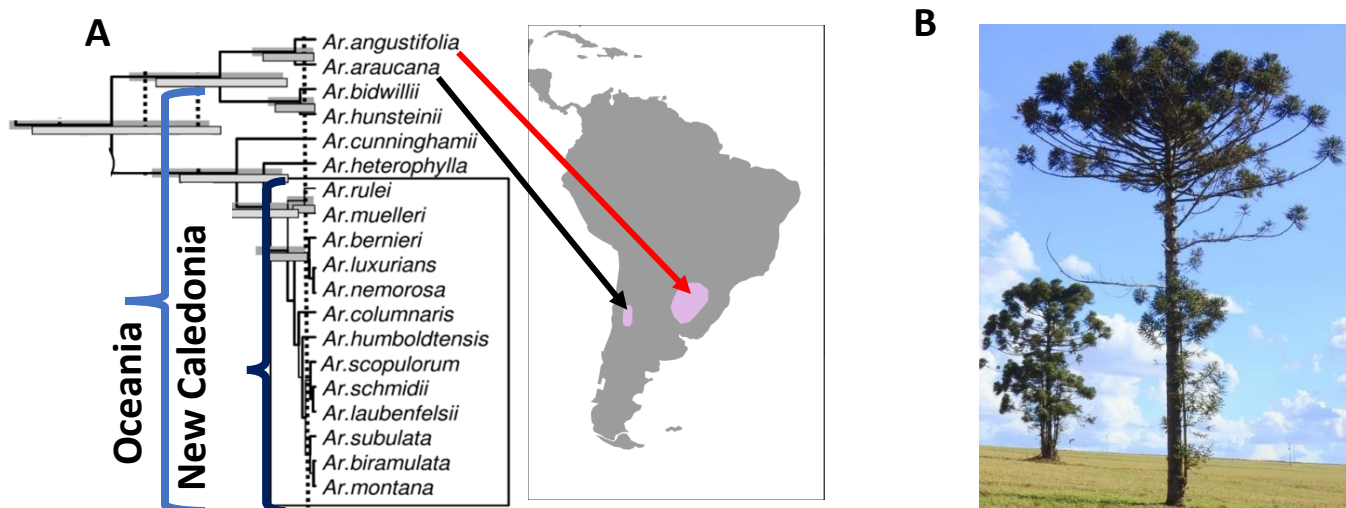
Keywords: *Araucaria angustifolia*, RNA-seq, miRNA, chloroplast genome

# 1. INTRODUÇÃO

## 1.1. *Araucaria angustifolia* (Bertol.) Kuntze

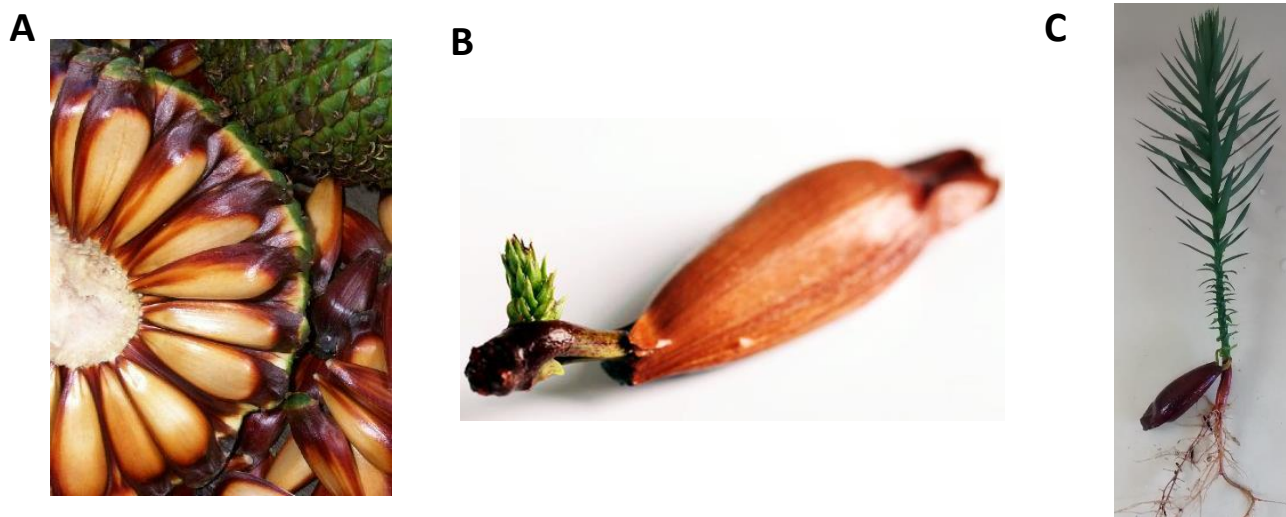
A família Araucariaceae abriga os gêneros *Wollemia*, *Agathis* e *Araucaria*, sendo este último o mais diversificado (Escapa and Catalano 2013). O gênero *Araucaria* abriga 19 espécies inteiramente distribuídas no hemisfério sul, com ocorrência na Austrália, Nova Zelândia, Nova Caledônia, Nova Guiné e na América do Sul, onde ocorrem as espécies *Araucaria araucana* e *Araucaria angustifolia* (Ruhsam et al. 2015) (Figura 1). *Araucaria angustifolia* (Bertol.) Kuntze, Pinheiro do Paraná ou Pinheiro brasileiro, é uma conífera dioica e perenifólia, com ocorrência predominante na região Sul do Brasil, com extensões e pequenas manchas no sudeste brasileiro e ao noroeste da Argentina e Paraguai (Reis et al. 2014).

Os espécimes de *A. angustifolia* são predominantemente dioicos, com ciclo de vida longo, atingindo a maturidade após 15 anos de crescimento, podendo atingir até 50 metros de altura (Auler et al. 2002; Paludo et al. 2016). Suas sementes são recalcitrantes (Figura 2). Os embriões acumulam alto teor de água e são dependentes desta condição para se desenvolverem (Agapito-Tenfen et al. 2011). Em situações de estresse a viabilidade pode ser comprometida, sobretudo quando o teor hídrico dos embriões atinge valores inferiores a 70% (Shibata et al. 2016).



**Figura 1.** Espécies do gênero *Araucaria* e suas áreas de ocorrência (A) adaptado de Kranitz et al., (2014); espécimes de *Araucaria angustifolia* (B).

Do ponto de vista ecológico, *A. angustifolia* desempenha importante papel em suas áreas de ocorrência (SOUZA 2007). É uma espécie que pode ser classificada como pioneira (Lowe et al. 2018) , comportando-se como espécie berçário no desenvolvimento de uma série de espécies epífitas, como bromélias e orquídeas (Longhi et al. 2009; Reis et al. 2014). Além disso, suas sementes ou pinhões são importantes fontes de alimento para a fauna silvestre durante as estações mais frias, quando diversas espécies de angiospermas não produzem frutos (Reis et al. 2014). Alguns animais estocadores de alimento como a cutia (*Dasyprocta spp.*) e a gralha-azul (*Cyanocorax caeruleus*) enterram os pinhões para se alimentar posteriormente, em períodos de baixas temperaturas (Reis et al. 2014). Por consequência, essa fauna estocadora atua como um importante agente dispersor de sementes (Longhi et al. 2009).



**Figura 2.** Sementes de *A. angustifolia* (A); pinhão germinando (B); planta jovem com 90 dias de desenvolvimento (C).

O Pinheiro do paran a   a principal con fera end mica do Brasil com import ncia comercial (Steiner et al. 2008). Suas sementes apresentam alto valor nutricional e s o tamb m adequadas para o consumo humano. Dessa forma, a produ o e venda de pinh es constitui uma importante atividade econ mica nas regi es de ocorr ncia, como, por exemplo, em grande parte do estado do Rio Grande do Sul. Al m da sua import ncia para o com rcio de alimentos, *A. angustifolia*   uma esp cie florestal que possui madeira de alta qualidade, com potencial para produ o de celulose, constru o civil, produ o de m veis, entre outros (Auler et al. 2002). Por conta disso, esta esp cie

foi massivamente explorada durante décadas, sofrendo considerável decréscimo populacional em virtualmente todas as áreas de ocorrência (Reis et al. 2014; Machado Mello and Peroni 2015; Tagliari and Peroni 2018). Atualmente estima-se que os remanescentes da mata de araucária representam menos de 10% da área original no território brasileiro, o que faz com que *A. angustifolia* seja classificada como espécie criticamente ameaçada de extinção (Thomas 2013).

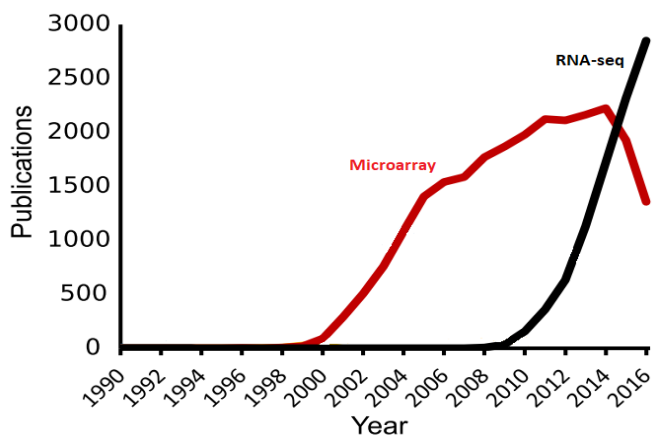
Levando em conta a condição recalcitrante, o longo ciclo reprodutivo e a potencial ameaça de extinção em *A. angustifolia*, uma série de estudos genéticos tem sido realizados nesta espécie. Os esforços concentram-se principalmente no estudo de variabilidade genética (Souza et al. 2009), estrutura e distância genética em populações (Auler et al. 2002) e embriogênese somática (Santos et al. 2002; Steiner et al. 2008; Santos et al. 2010; Agapito-Tenfen et al. 2011; Elbl et al. 2015). Porém, uma série de aspectos genéticos permanece inexplorada em *A. angustifolia*, como, por exemplo genoma nuclear, genoma organelar e biologia de pequenos RNAs.

## 1.2. RNA-seq

O termo transcriptoma refere-se ao conjunto completo de moléculas de RNA, expressas em um tecido ou tipo celular específico em um determinado momento (Weber 2015). Este conjunto abrange não só RNAs codificantes (mRNAs), como também os mais diversos tipos de RNAs não codificantes, como RNAs ribossômicos, RNAs de transferência, pequenos RNAs nucleares, pequenos RNAs nucleolares, microRNAs, RNAs longos não codificantes, RNAs circulares, dentre outros (Morgado and Johannes 2017). A complexidade do transcriptoma se acentua por conta de uma série de eventos relacionados ao controle da expressão gênica, os quais ocorrem nos níveis transcricional, pós-transcricional e na etapa de tradução (Hrdlickova et al. 2017). Ao nível de iniciação da transcrição, um gene pode dar origem a múltiplos transcritos com variações nos sítios de início da transcrição e posição dos primeiros éxons (McGettigan 2013). Transcritos de um mesmo gene também podem diferir entre si por conta de eventos de *splicing* alternativo e poliadenilação alternativa (McGettigan 2013). Além disso, transcritos processados de um mesmo gene podem apresentar sítios de início da tradução distintos e, portanto, codificar polipeptídeos diferentes (McGettigan 2013).

Os primeiros esforços para acessar perfis transcriptômicos ocorreram nos anos 90 com o desenvolvimento da técnica de microarranjos, à qual permaneceu por mais de uma década como uma das principais ferramentas para este tipo de análise (Weber 2015). No entanto, a técnica de microarranjos apresenta algumas limitações importantes, como a necessidade de conhecimento prévio de sequências genômicas da espécie alvo, a detecção dos padrões de expressão baseada em intensidade de fluorescência sujeita a ruídos, entre outros (Weber 2015).

O advento das plataformas NGS em meados da década de 2000 culminou no desenvolvimento do sequenciamento de RNA ou *RNA-sequencing* (RNA-seq) (McGettigan 2013). Esta tecnologia foi desenvolvida com a perspectiva de eliminar a maioria dos artefatos dos microarranjos, proporcionando a análise de transcriptomas com resolução ao nível de pares de bases, utilizando essencialmente os mesmos protocolos para qualquer espécie, e minimizando os ruídos de análise (McGettigan 2013) (Figura 3).



**Figura 3.** Surgimento e aplicação ao longo do tempo de duas das principais tecnologias utilizadas em estudos transcriptômicos: microarranjos e RNA-seq. Adaptado de Lowe et al., (2017).

O RNA-seq não depende do conhecimento prévio de sequências genômicas, ao contrário dos microarranjos, os quais são desenvolvidos a partir de sequências de referência. Enquanto os microarranjos indicam possíveis variações de expressão gênica a partir padrões de fluorescência, o RNA-seq proporciona medição direta da abundância de RNAs (Hrdlickova et al. 2017). O RNA-seq também permite simultaneamente a identificação e quantificação de sequências, inclusive transcritos com baixo padrão de expressão (Hrdlickova et al. 2017). Com esta técnica também é possível analisar a expressão de alelos específicos em indivíduos heterozigotos e a detecção de

polimorfismos entre indivíduos (Stein et al. 2011). Dessa forma, o RNA-seq proporcionou à comunidade científica estudar a expressão gênica minuciosamente e com uma profundidade nunca antes alcançada. Atualmente existem segmentos na pesquisa transcriptômica que surgiram apenas após o advento do RNA-seq, como, por exemplo, o estudo de longos RNAs intergênicos (lincRNAs) (Liu et al. 2015) e RNAs circulares (circRNAs) (Li et al. 2017a).

Em aspectos gerais, as reações de RNA-seq envolvem a preparação de bibliotecas de cDNA a partir de amostras de RNA (Hrdlickova et al. 2017). O protocolo de preparação destas bibliotecas pode variar de acordo com os tipos de RNA que se pretende analisar, o que faz com que as bibliotecas variem em tamanho, sequência, características estruturais e abundância (Hrdlickova et al. 2017). O sequenciamento de bibliotecas de RNAs poliadenilados é uma das aplicações mais comuns de RNA-seq. Em organismos eucarióticos, uma parcela substancial de RNAs codificantes (mRNAs) bem como RNAs não codificantes, como pri-miRNAs e lincRNAs, contém uma cauda poli-A (McGettigan 2013). Esta cauda é alvo de aplicações técnicas que promovem o enriquecimento de RNAs poli-A, os quais representam de 1 a 5% do RNA total de cada amostra (Buermans and den Dunnen 2014). Portanto, o seu enriquecimento com *beads* magnéticos acoplados a oligo dTs ou a partir de primers oligo dTs para transcrição reversa é uma etapa importante na preparação das bibliotecas (Weber 2015). No caso de RNAs não codificantes e sem cauda poli-A ou amostras parcialmente degradadas, a depleção de RNAs ribossômicos (rRNAs) aparece como uma importante estratégia, através de diferentes técnicas. Por exemplo, a degradação direcionada por sonda é uma técnica de degradação de RNAs ribossômicos que consiste na hibridização dos rRNAs com oligonucleotídeos de DNA e posterior digestão com RNase H (Weber 2015). Após a etapa de enriquecimento de RNAs poli-A e depleção de rRNAs, normalmente as amostras são submetidas à fragmentação e posterior transcrição reversa para formação das bibliotecas de cDNA (Weber 2015).

### **1.2.1. Bioinformática aplicada nas análises de RNA-seq**

As possibilidades de análise de dados de RNA-seq são tão variadas quanto as aplicações desta tecnologia (Conesa et al. 2016). Em um típico experimento, alguns dos principais procedimentos envolvem o controle de qualidade, o alinhamento dos *reads* com ou sem um genoma de referência, a obtenção de métricas em relação à abundância e expressão diferencial de transcritos (Conesa et al. 2016). O controle de qualidade é uma etapa crucial na eliminação de *reads* de baixa qualidade



e artefatos como sequências de adaptadores e sequências contaminantes de DNA nuclear e organelar (Han et al. 2015). Eventuais erros de sequenciamento também podem ser removidos e, em seguida, os *reads* pré-processados são agrupados em transcritos, os quais passam por outras etapas de controle de qualidade e contaminação ou pós-processamento (Han et al. 2015). O nível de expressão de cada transcrito pode ser estimado pela contagem de *reads* alinhados (Spies et al. 2017).

O agrupamento dos *reads* em transcritos consiste na montagem do transcriptoma (Schliesky et al. 2012). As estratégias de montagem variam em função da disponibilidade e uso de um genoma de referência e agrupam-se em três categorias: montagem com genoma de referência, montagem de novo ou a combinação das mesmas (Ward et al. 2012). Quando há um genoma de referência, os *reads* de RNA-seq podem ser alinhados ao mesmo com o uso de *splice-aware aligners* como TopHat (Trapnell et al. 2012), SliceMap (Barbier et al. 2018) e GSNAP (Wu and Nacu 2010). Em seguida os reads sobrepostos de cada *locus* são agrupados para construir um gráfico de todas as isoformas possíveis (Voshall and Moriyama 2018). Por fim, o gráfico é analisado para resolver as isoformas individuais (Voshall and Moriyama 2018).

Como mencionado anteriormente, a análise computacional de RNA-seq não depende necessariamente do conhecimento prévio de sequências ou sequências de referência. Na ausência de um genoma de referência, pode-se adotar estratégias de montagem *de novo* do transcriptoma (Martin and Wang 2011). Este tipo de montagem é executada com softwares que utilizam da redundância dos *reads* para encontrar sobreposição entre as leituras (Martin and Wang 2011). Programas como Trinity (Haas et al. 2013) usam o gráfico De Bruijn para reconstruir transcritos de uma ampla faixa de níveis de expressão e então processar a montagem de contigs e remover redundâncias (Voshall and Moriyama 2018). Comparada com a estratégia anterior, a montagem de novo tem uma série de vantagens, como por exemplo: não depende de um genoma de referência, o que representa, portanto, uma importante estratégia para o estudo de organismos sem genomas sequenciados; pode preencher lacunas em genomas de baixa cobertura, contribuindo, inclusive com a identificação de transcritos codificados por porções ausentes no rascunho do genoma; entre outros (Martin and Wang 2011).

### 1.2.2. Estudos transcriptômicos em plantas

O RNA-seq tem se mostrado como uma das ferramentas de maior resolução e versatilidade para estudos em biologia molecular vegetal. Essa ferramenta apresenta uma diversa gama de aplicações, com potencial para análises transcriptômicas detalhadas em tipos celulares e tipos de RNA específicos nas mais diversas situações e processos, como estresse, patologias e etapas de desenvolvimento (Conesa et al. 2016). As aplicações do RNA-seq incluem o sequenciamento e geração de dados genômicos em espécies ainda não estudadas, expandindo o universo de organismos-modelo; a compreensão detalhada de vias gênicas regulatórias e estruturais; o entendimento de como as plantas respondem fisiologicamente às mais variadas condições ambientais; um maior entendimento das relações entre os genes e seus respectivos produtos; e a fusão da transcriptômica com a genômica e metabolômica, integrando o universo das ferramentas ômicas (Martin et al. 2013).

Apesar de esforços recentes voltados para o sequenciamento genômico de várias espécies de plantas terrestres, bem como a disponibilização desses dados em repositórios especializados, como o Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>), eles representam uma discreta amostra da diversidade taxonômica vegetal (Martin et al. 2013). Usualmente, espécies vegetais de diversas divisões, como as gimnospermas, apresentam genomas consideravelmente longos, em decorrência de duplicações e/ou sequências altamente repetitivas (Pellicer et al. 2018). Nesse ponto, o sequenciamento *de novo* do transcriptoma aparece como uma importante estratégia para obtenção de recursos genéticos em espécies-não-modelo com considerável interesse biológico em função de aspectos evolutivos, ecológicos, agronômicos, comerciais, entre outros (Hrdlickova et al. 2017). A gimnosperma basal *Ginkgo biloba* tem sido alvo de diferentes estudos genéticos recentes baseados em RNA-seq, com enfoques que incluem metabolismo de flavonoides (Wu et al. 2018), análise de perfis transcriptômicos nos tipos sexuais (Du et al. 2016), perfis transcriptômicos em diferentes estágios de desenvolvimento das sementes (He et al. 2015), entre outros. Esforços similares foram feitos em diversas coníferas, como *Pinus tabulaeformis* (Yao et al. 2018), *Pinus bungeana* (Duan et al. 2017), *Platycladus orientalis* (Hu et al. 2016), *Chamaecyparis obtusa* (Sato et al. 2014), *Larix leptolepis* (Liu et al. 2016).

Além da análise transcriptômica global com foco em transcritos codificantes (mRNAs), o RNA-seq apresenta alto potencial para o estudo de pequenos RNAs não codificantes (sRNAs) (Morgado

and Johannes 2017). Desde a última década, o RNA-seq é utilizado massivamente na identificação e caracterização de tasiRNAs, phasiRNAs, hcRNAs e, principalmente, microRNAs (miRNAs), em virtualmente todos os grupos vegetais, em diferentes tecidos e sob as mais diversas condições ambientais e etapas de desenvolvimento (Axtell and Meyers 2018). A caracterização da atividade regulatória dos miRNAs frequentemente se deve a análises de expressão diferencial, sobretudo com a identificação de padrões de expressão tecido-específicos, como, por exemplo, a identificação de 46 miRNAs diferencialmente expressos e alguns com padrão tecido-específico em bibliotecas de pequenos RNAs provenientes de dois tipos de folha em *Ginkgo biloba* (Zhang et al. 2015b). As diferenças nos padrões de expressão espaço-temporal nos órgãos vegetais também fornecem importantes *insights* acerca do papel regulatório dos miRNAs, e uma série de estudos com dados de sRNA-seq foi realizada com esse propósito em diversas espécies de gimnospermas, como *Larix leptolepis* (Liu et al. 2016), *Sequoia sempervirens* (Chen et al. 2013), *Cunninghamia lanceolata* (Wan et al. 2012a), *Cryptomeria japonica* (Ujino-Ihara et al. 2018), *Picea balfouriana* (Li et al. 2017b), *Picea glauca* (Liu and El-Kassaby 2017), *Pinus tabulaeformis* (Yao et al. 2018), entre outras. Os miRNAs também podem apresentar expressão diferencial em tecidos submetidos às mais variadas condições de estresse. Uma série de miRNAs e outros sRNAs foram identificados com padrões de expressão diferencial em bibliotecas de sRNA-seq de indivíduos de várias espécies, submetidos à déficit de fosfato (Bai et al. 2018), tratamento com hidrogênio (Liu et al. 2016), estresse salino (Pegler et al. 2018), estresse hídrico (Pegler et al. 2018), estresse térmico (Zeng et al. 2018), entre outros.

Estudos recentes apontam o RNA-seq como uma importante ferramenta para análises genéticas em organelas (Smith 2013). As bibliotecas de RNA-seq de eucariotos frequentemente carregam uma considerável porção de sequências provenientes de transcritos de cloroplastos (Martin et al. 2013; Smith 2013; Conesa et al. 2016). Este enriquecimento ocorre principalmente por conta do grande número de cópias e alto padrão de expressão dos plastomas (Smith 2013). Além disso, os transcritos de cloroplastos são ricos em AT e frequentemente poliadenilados (Stern et al. 2010), o que favorece a sua seleção em protocolos de enriquecimento de bibliotecas de mRNAs Poly-A, ou até mesmo em protocolos de depleção de RNAs ribossômicos (Smith 2013). A disponibilidade em larga escala desses dados em repositórios públicos, como o NCBI, torna plausível a investigação de muitos aspectos transcricionais, evolutivos e funcionais em genômica organelar, como o processo de edição de RNA, identificação de SNPs, marcadores genéticos para estudos

populacionais e filogenéticos, marcadores com caráter taxonômico (*DNA barcoding*), entre outros (Yagi and Shiina 2014; Lee et al. 2014; Zhou et al. 2016; Rodrigues et al. 2017). Portanto, o que parece ser uma mera contaminação, na verdade é uma fonte praticamente inesgotável de informação, dado o volume de bibliotecas de RNA-seq de plantas disponível em bancos de dados públicos. Por fim, estudos recentes trazem evidências de que o genoma de cloroplasto de plantas terrestres, assim como o de algas, é transcrito em toda a sua extensão, principalmente em tecidos fotossintéticos, como folhas e caules em estágios iniciais (Smith 2013; Shi et al. 2016; Smith and Sanitá Lima 2016). Dessa forma, conforme apontam Shi et al. (2016), é possível montar genomas de cloroplasto a partir de dados de RNA-seq. Os mesmos autores confirmaram essa proposição com a montagem do plastoma de 12 espécies de angiospermas e duas espécies de briófitas usando como input dados de RNA-seq disponíveis no NCBI (Shi et al. 2016). Também utilizando dados de RNA-seq, Horvath et al. identificaram regiões promotoras, elementos repetitivos, marcadores genéticos, além da sequência completa do genoma de cloroplasto de *Euphorbia esula* (Horvath et al. 2018).

### **1.3. microRNAs em plantas**

#### **1.3.1 Aspectos gerais**

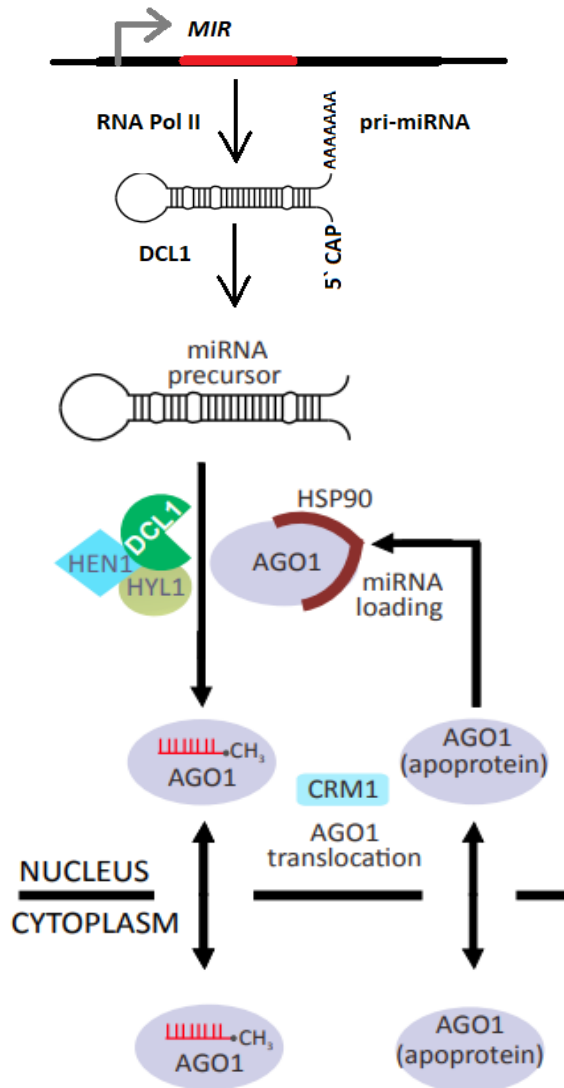
Os microRNAs (miRNAs) são uma classe de pequenos RNAs endógenos não codificantes com profundo impacto em uma série de processos biológicos (Moran et al. 2017). Os miRNAs são agentes regulatórios capazes de modular a expressão gênica agindo em sinergia com uma série de domínios protéicos que integram os complexos de silenciamento gênico. Os primeiros estudos com miRNAs se iniciaram no início da década de 90 com a identificação do miRNA *Lin-4* no verme nematódeo *Caenorhabditis elegans* (Lee et al. 1993). Neste trabalho, constatou-se a relação direta do miRNA *lin-4* com o controle pós-transcricional do gene *lin-14*, o qual apresenta atividade regulatória em etapas do desenvolvimento de *C. elegans* (Lee et al. 1993). Posteriormente, no início da década de 2000, estudos realizados na espécie modelo *Arabidopsis thaliana* constataram a existência de miRNAs em plantas (Reinhart 2002). Desde então, milhares de miRNAs foram identificados nos mais diversos grupos vegetais e os conhecimentos acerca da biologia dessa classe de pequenos RNAs vem se intensificando progressivamente.

### 1.3.2. Biogênese e processamento

A via canônica de biogênese dos miRNAs de plantas se inicia com a transcrição de genes *MIR* pela RNA polimerase II (Pol II) (Yu et al. 2017). Os genes *MIR* são similares aos genes codificantes de proteínas, apresentando elementos em comum, como uma região promotora com motivos TATA-box e sítios de ligação de fatores de transcrição (Yu et al. 2017). Os transcritos primários de miRNAs ou pri-miRNAs são RNAs com um CAP na extremidade 5' e cauda poli-A na porção 3'. Podem também conter íntrons e sofrer splicing canônico ou alternativo (Zielezinski et al. 2015). Os pri-miRNAs apresentam estrutura linear nas extremidades e na região central formam uma estrutura secundária em forma de grampo ou *hairpin*, que é crucial para as etapas de processamento pós transcricional (Zielezinski et al. 2015) (Figura 4). Os *hairpins* são reconhecidos pela família de enzimas endonucleases RNase III DICER-LIKE (DCL) (Budak and Akpinar 2015). O tamanho desta família varia em diferentes espécies de plantas. Por exemplo, a família DCL tem quatro membros em *Arabidopsis*, cinco membros no arroz (Margis et al. 2006) e três membros na briófito *Physcomitrella patens*. Em *Arabidopsis*, a DICER-LIKE1 (DCL1) é a principal responsável pela clivagem dos pri-miRNAs e consequente produção de *hairpins* menores conhecidos como precursores de miRNAs (pre-miRNAs) (Yu et al. 2017). Os pre-miRNAs são processados também por endonucleases DCL (principalmente a DCL1) para dar origem a pequenos RNAs de fita dupla compostos pelos miRNAs maduros 5p e 3p (Borges and Martienssen 2015).

Os pre-miRNAs de plantas são consideravelmente mais variados em tamanho e estrutura em comparação aos de metazoários (Miskiewicz et al. 2017). O seu processamento ocorre no núcleo por um complexo enzimático que contém a DCL1 (Miskiewicz et al. 2017). De acordo com os padrões de clivagem dos pre-miRNAs pela DCL, o processamento dos precursores pode ser classificado de diferentes formas, sendo as mais comuns *base-to-loop* e *loop-to-base* (Chorostecki et al. 2017). No primeiro grupo, os pre-miRNAs apresentam um dúplex (dsRNA) de 15 a 17 nucleotídeos abaixo do dúplex que contém os miRNAs maduros 5p e 3p e em direção à base (*lower stem*) (Chorostecki et al. 2017). Esta região de dsRNA é reconhecida pela maquinaria de processamento e especifica o primeiro sítio de corte da DCL1. O segundo sítio de corte situa-se imediatamente após a sequência do miRNA maduro, aproximadamente 21 nucleotídeos a partir do primeiro sítio de corte, e em direção ao loop (Chorostecki et al. 2017). Este tipo de processamento

se assemelha ao processamento dos miRNAs de animais. No segundo grupo, o pre-miRNA sofre a primeira clivagem abaixo do loop terminal e o processamento segue em direção à sua base (Chorostecki et al. 2017). Estes tipos de precursores apresentam o dúplex de reconhecimento do complexo de processamento acima do dúplex dos miRNAs maduros 5p e 3p. A clivagem dos pre-miRNAs produz um dúplex de 19 a 24 nucleotídeos que contém os microRNAs maduros (miRNAs) (Budak and Akpinar 2015). A maioria dos miRNAs de plantas apresentam comprimento de 20 a 22 nucleotídeos (nt), sendo 21nt o comprimento mais comum (Borges and Martienssen 2015).



**Figura 4.** Processo de biogênese de miRNAs em plantas. Adaptado de Bologna et al., (2018).

### 1.3.3. Montagem do complexo de silenciamento

Uma vez formados, os duplexes de miRNAs maduros são estabilizados via metilação das extremidades 3' pela metiltransferase HEN1, processo que ocorre no núcleo celular (Bologna et al. 2018). Alguns estudos sugerem a translocação deste dúplex para o citoplasma por intermédio da proteína EXPORTIN5 (EXPO5) e o subsequente carregamento de uma das fitas na proteína argonauta AGO1, para formar o complexo de silenciamento induzido por RNAs (RISC) (Bologna and Voinnet 2014). Porém, um estudo recente feito em *Arabidopsis* mostrou que os miRNAs em sua forma madura e metilada são carregados nas proteínas AGO ainda no núcleo e translocados para o citoplasma como complexos AGO:miRNA (Bologna et al. 2018) (Figura 4). Este processo é possível porque a proteína argonauta apresenta em sua porção N-terminal dois domínios, NES e NLS que sinalizam, respectivamente, o seu deslocamento do núcleo para o citoplasma e vice-versa, processo conhecido como *Nucleo-cytosolic Shuttling* (Bologna et al. 2018). Dessa forma, a proteína AGO1 recém-sintetizada e desprovida de miRNA, por meio do domínio sinalizador NLS se desloca para o núcleo, onde incorpora o miRNA maduro metilado na extremidade 3' e, pela interação do seu domínio NES com a proteína EXPO1 se desloca de volta para o citoplasma para desencadear os processos de silenciamento (PTGS) (Bologna et al. 2018).

A seleção das fitas do miRNA guia não ocorre de forma aleatória (Yu et al. 2017). Este processo depende dos fatores de processamento dos pre-miRNAs, do tipo de base na extremidade 5' e a estrutura do dúplex de miRNAs maduros (Chorostecki et al. 2017). A natureza da base na extremidade 5' direciona o carregamento da fita guia na proteína AGO. A maioria das fitas guias contém um resíduo de uracila em sua extremidade 5' e é incorporada pela AGO1 (Bologna and Voinnet 2014). Extremidades 5' compostas por resíduos de adenina e citosina sinalizam, respectivamente, o carregamento em AGO2 e AGO5 (Yu et al. 2017). A seleção das fitas guia em diferentes argonautas também ocorre em função dos *mismatches* e protuberâncias nos duplex miRNA5p/miRNA3p (Yu et al. 2017). Por exemplo, AGO2 favorece duplexes sem mismatches na região central, ao contrário da AGO1, que favorece *duplexes* com *mismatches* no centro (Yu et al. 2017).

#### 1.3.4. Mecanismos de ação

Os miRNAs de plantas regulam os genes alvo em nível pós transcricional por dois mecanismos principais: degradação dos transcritos e bloqueio da tradução (Borges and Martienssen 2015). A clivagem de transcritos guiada por miRNAs, também conhecida como *slicing*, ocorre em uma posição precisa no mRNA alvo (Borges and Martienssen 2015). O complexo AGO:miRNA interage com o mRNA alvo por complementariedade de bases miRNA/mRNA e, posteriormente, o domínio PIWI da argonauta catalisa a clivagem do transcrito. Em seguida, os fragmentos 5' e 3' do mRNA clivado são degradados por exonucleases (Borges and Martienssen 2015).

Em plantas a inibição da tradução mediada por miRNAs parece ser menos frequente em relação à clivagem dos alvos (Budak and Akpinar 2015). Os mecanismos moleculares característicos deste processo não são bem elucidados. Entretanto, análises *in vitro* indicaram que os miRNAs de plantas podem inibir o início da tradução ou bloquear o movimento dos ribossomos (Budak and Akpinar 2015). Alguns fatores requeridos na inibição da tradução mediada por miRNAs incluem a enzima KATANIN (KTN1), o componente do corpo de processamento (*P body*) VARICOSE (VCS), a proteína SUO e a proteína de membrana do retículo endoplasmático AMP1 (Yu et al. 2017). Mutações nestes genes afetam a repressão mediada por miRNAs ao nível de proteínas, o que sugere que a clivagem dos transcritos e a inibição da tradução são vias regulatórias distintas (Yu et al. 2017).

#### 1.3.5. Regulação da expressão gênica

Assim como em metazoários, os miRNAs desempenham papéis cruciais em diversos processos biológicos (Budak and Akpinar 2015). Os padrões de expressão dos miRNAs bem como o seu modo de ação são componentes fundamentais em várias etapas do desenvolvimento (Liu et al. 2018), como na transição da fase juvenil para a fase adulta (Xu et al. 2018), no desenvolvimento foliar (Yang et al. 2018), na iniciação e desenvolvimento de órgãos reprodutivos (D'Ario et al. 2017), no desenvolvimento de frutos (Hou et al. 2017), na formação e desenvolvimento do sistema radicular (Liu et al. 2018), entre outros. Diversos estudos com plantas modelo, como *Arabidopsis*, mostraram que mutantes deficientes na biogênese ou mecanismos de ação dos miRNAs exibem



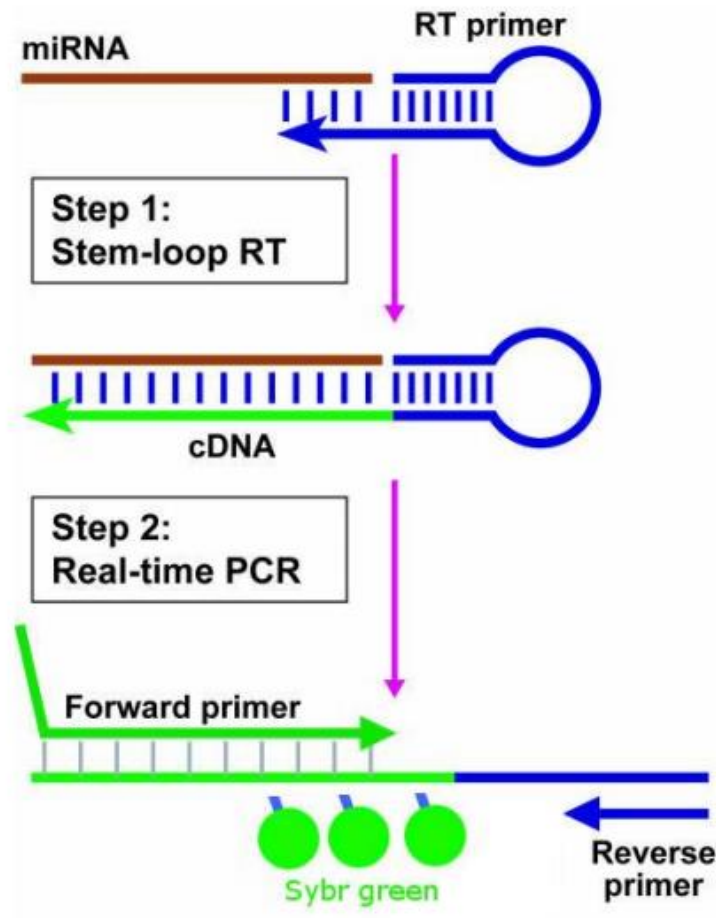
uma série de fenótipos (Yu et al. 2015). Por exemplo, em mutantes *dcl1* a consequência é a letalidade. Apenas mutantes hipomórficos, os quais apresentam função da proteína parcialmente afetada, completam as etapas do desenvolvimento e atingem o estágio de plântulas (Jacobsen et al. 1999). Os mutantes hipomórficos *ago1* apresentam adaxialização das folhas, presença de tricomas abaxiais em folhas jovens, baixa estatura da planta e diminuição no tamanho das flores (Yu et al. 2015).

### **1.3.6. Predição e análise de expressão de miRNAs**

Atualmente, os métodos mais utilizados na identificação de miRNAs dividem-se em duas categorias principais, predições computacionais e investigações experimentais (Axtell and Meyers 2018). As predições computacionais baseiam-se em algoritmos para identificar miRNAs candidatos a partir de dados de RNA-seq, sRNA-seq e sequenciamento genômico. Posteriormente, técnicas experimentais são aplicadas para validar os candidatos (Alptekin et al. 2017). Existem uma série de algoritmos especializados na predição de miRNAs como miRDeep (Friedländer et al. 2012), miRDeep-p (Yang and Li 2011), miRanalyzer (Hackenberg et al. 2009), e miR-PREFeR (Lei and Sun 2014). Uma análise comparativa entre estes softwares mostrou que o miR-PREFeR apresenta menor índice de falsos positivos e alto grau de acurácia e precisão na predição dos miRNAs (Lei and Sun 2014). O miR-PREFeR utiliza o padrão de expressão dos miRNAs, seguindo os critérios de anotação para miRNAs de plantas com o objetivo de realizar predições acuradas a partir de uma ou mais bibliotecas de sRNA-seq (Lei and Sun 2014). Este programa utiliza como dados de entrada um genoma ou transcriptoma de referência de determinada espécie, e um ou múltiplos arquivos para alinhamento dos *reads* de pequenos RNAs (sRNA-seq) da mesma espécie. A partir destes dados o programa realiza uma predição de pre-miRNAs a nível de sequência e estrutura secundária, bem como seus respectivos miRNAs maduros 5p e 3p (Lei and Sun 2014).

Atualmente, a principal técnica experimental para validação de miRNAs é a técnica de stem-loop RT-qPCR (Chai et al. 2015; Varkonyi-gasic and Varkonyi-gasic 2017; Pegler et al. 2018; He et al. 2018; Zhang et al. 2018). Uma vez que os miRNAs apresentam sequências curtas, esta técnica baseia-se no uso de iniciadores *stem-loop* como estratégia para aumentar o comprimento dos miRNAs maduros de modo que o cDNA produzido seja longo o suficiente para ser amplificado por PCR (Kramer 2011; Varkonyi-gasic and Varkonyi-gasic 2017). Os iniciadores *stem-loop* são

específicos para cada miRNA e são projetados para que sua porção 3' seja complementar aos últimos seis nucleotídeos da extremidade 3' do miRNA alvo (Kramer 2011). Assim, são sintetizados cDNAs correspondentes apenas ao miRNA-alvo para o qual o iniciador *stem-loop* foi utilizado na síntese. Na etapa de amplificação por PCR quantitativo utiliza-se um par de primers, sendo o primer *forward* específico para o miRNA alvo e o primer reverso com uma sequência universal (Kramer 2011) (Figura 5). A detecção dos níveis de transcritos pode ser feita a partir da fluorescência emitida por um agente intercalante como, por exemplo, SYBER® Green.



**Figura 5.** Stem-loop RT-qPCR para quantificação de miRNAs. Esta metodologia tem duas etapas principais, uma transcrição reversa (RT) seguida de uma PCR quantitativa em tempo real (qPCR). Adaptado de Chen et al., (2005).

As etapas subsequentes aos ensaios de RT-qPCR são a normalização e a análise de dados (Varkonyi-gasic and Varkonyi-gasic 2017). O objetivo da normalização é corrigir possíveis diferenças entre as amostras, de modo que a real expressão do miRNA-alvo seja medida (Joseph et al. 2018). A escolha do RNA normalizador é de suma importância e deve basear-se em transcritos que sejam afetados minimamente pelas diferentes condições biológicas (Joseph et al. 2018). É possível utilizar miRNAs com perfis de expressão mais estável para cumprir o papel de normalizadores para determinado conjunto de amostras (Kulcheski et al. 2010; Fausto et al. 2017). Programas como o geNorm (Vandesompele et al. 2002) e NormFinder (Andersen et al. 2004) avaliam a expressão de diferentes transcritos e apontam os RNAs com menor variação entre amostras distintas e indicam a melhor combinação de normalizadores para o cálculo da expressão relativa. A quantificação relativa é um dos principais métodos de análise de dados de RT-qPCR (Chen 2005). Este método baseia-se na representação dos valores de expressão do RNA-alvo em relação à expressão de um calibrador interno. A quantificação relativa normalmente é aplicada com o método comparativo  $2^{-\Delta\Delta CT}$  (Livak and Schmittgen 2001) .

### **1.3.7. Predição computacional dos alvos**

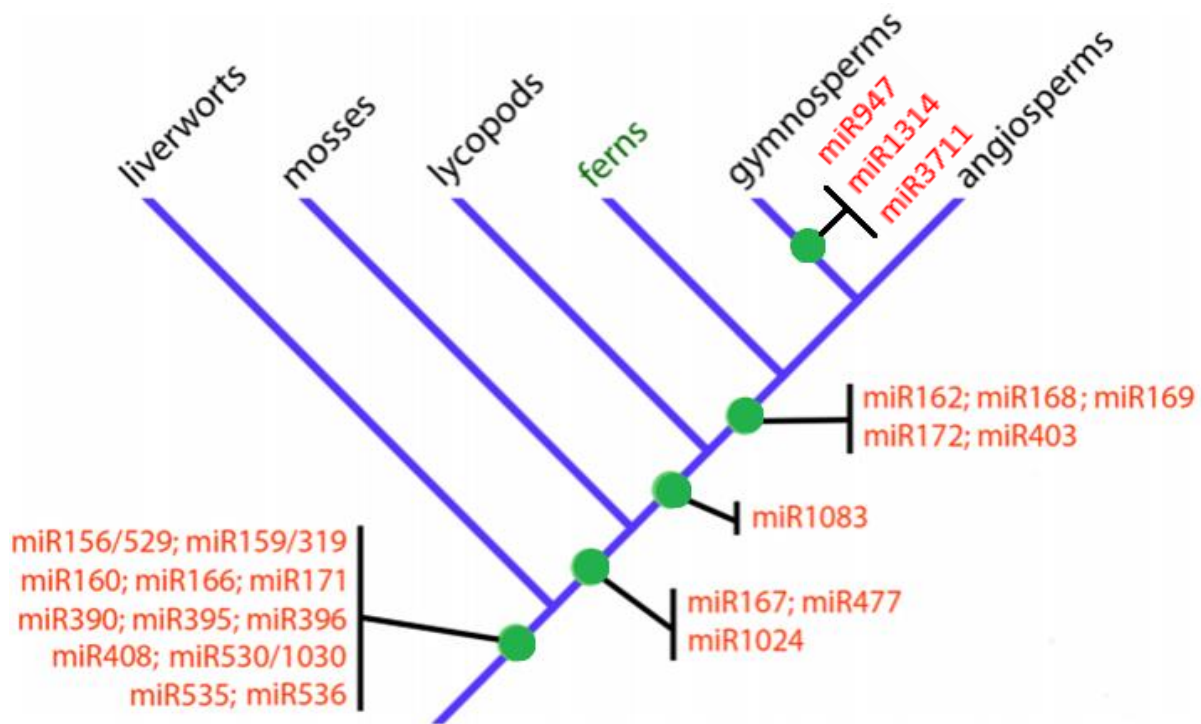
A compreensão da função dos miRNAs depende, em grande parte, da identificação de seus genes-alvo (Budak and Akpinar 2015; Borges and Martienssen 2015; Yu et al. 2017). A identificação de alvos de miRNAs pode ser feita com base em predições computacionais (*in silico*) (Srivastava et al. 2014; Das et al. 2015). Neste tipo de análise, a sequência do miRNA maduro é utilizada para buscar alvos em dados de genoma ou transcriptoma de determinado organismo (Srivastava et al. 2014). O psRNATarget (<http://plantgrn.noble.org/psRNATarget/>) é uma das principais plataformas para predição de alvos de miRNAs em plantas (Dai and Zhao 2011). Levando em conta o pareamento de bases entre a sequência do miRNA e do alvo, sobretudo em um trecho que abrange os nucleotídeos 2 ao 7 do miRNA (*seed region*), essa plataforma penaliza os *mismatches* e estima a energia requerida para modificar a estrutura secundária do mRNA e tornar acessível a sua região alvo (Dai and Zhao 2011). Além disso, o psRNATarget estima se a regulação miRNA-alvo ocorre pelo mecanismo de clivagem ou inibição da tradução (Dai and Zhao 2011).

### 1.3.8. Aspectos evolutivos dos miRNAs de plantas

Com o advento das tecnologias de sequenciamento de nova geração (NGS), o universo de pequenos RNAs tornou-se cada vez mais acessível, sobretudo no segmento dos miRNAs (Axtell and Meyers 2018). Desde a década passada, centenas ou talvez milhares de trabalhos com identificação e análise de miRNAs em plantas foram publicados. O acúmulo de dados de sequência e anotação de miRNAs culminou no desenvolvimento de bancos de dados específicos. O miRBase (<http://www.mirbase.org/cgi-bin/browse.pl>) atualmente é o principal repositório público especializado em miRNAs dos mais diversos organismos (Kozomara and Griffiths-Jones 2014). A versão mais atual do miRBase, *miRBase version 22 (v22)*, foi atualizada em março de 2018 e contém milhares de sequências de miRNAs de diferentes grupos vegetais, principalmente de angiospermas dicotiledôneas e monocotiledôneas-modelo como *Oryza sativa* e *Zea mays* (Kozomara and Griffiths-Jones 2014). A disponibilidade de sequências de miRNAs de diversos grupos vegetais trouxe à tona uma série de evidências e abordagens acerca da evolução desta classe de pequenos RNAs em plantas terrestres (Taylor et al. 2014). Assim como os genes codificantes de proteínas, os genes de miRNAs (genes *MIR*) também podem ser classificados em famílias gênicas (Moran et al. 2017). O número de genes *MIR* de espécies que sofreram duplicações genômicas excede consideravelmente o de espécies relacionadas que não passaram pelos mesmos processos (Shi et al. 2017). Na versão mais atual do miRBase constam 2.787 miRNAs anotados em *Glycine max* e 216 miRNAs anotados em *Phaseolus vulgaris*, duas espécies representantes da família Fabaceae; e 680 miRNAs anotados em *Brassica rapa* contra 80 em *Capsella rubella*, duas espécies da família Brassicaceae.

A identificação de miRNAs em diversos grupos vegetais também aponta questões relevantes em relação à conservação de miRNAs entre grupos mais próximos ou mais distantes, bem como a existência de miRNAs espécie-específicos (Qin et al. 2014). Axtell & Meyers estabeleceram o conceito “família de miRNAs conservados em plantas terrestres” à famílias de miRNAs anotados no miRBase em pelo menos duas das principais divisões taxonômicas de plantas terrestres: dicotiledôneas, monocotiledôneas, *Amborella*, gimnospermas, licófitas e briófitas (Axtell and Meyers 2018). Seguindo este critério, os mesmos autores identificaram 36 famílias de miRNAs conservados de plantas terrestres em acessos do miRBase, sendo nove delas presentes desde *P. patens* (briófitas) até o grupo das angiospermas: miR156, miR160, miR166, miR171, miR319,

miR390, miR477, miR529 e miR535 (Axtell and Meyers 2018). Isto indica que, possivelmente, esses miRNAs surgiram antes da linhagem ancestral das plantas terrestres e foram mantidos na maioria das linhagens derivadas deste ancestral comum (Axtell and Meyers 2018). Outras famílias de miRNAs foram identificadas em grupos menores ou em linhagens específicas, como miR1083, miR162, miR168, miR169, miR172 e miR403 em traqueófitas (Berruezo et al. 2017); e miR1311-miR1315, miR3709, miR3711, miR947, miR950 e miR951 em gimnospermas (Wan et al. 2012a; Wan et al. 2012b; Hao et al. 2012; Chen et al. 2013; Zhang et al. 2015b; Liu et al. 2016; Li et al. 2017b) (Figura 6). A grande maioria dos miRNAs identificados em diferentes grupos parece ser espécie-específica ou linhagem-específica (Moran et al. 2017). Por exemplo, mais de 10% dos genes *MIR* presentes em *Arabidopsis thaliana* e *Arabidopsis lyrata* são únicos em cada espécie, apesar do seu processo de especiação relativamente recente (Fahlgren et al. 2010).



**Figura 6.** Famílias de miRNAs conservados entre os diferentes grupos de plantas terrestres. Adaptado de Lowe et al., (2017).

#### 1.4. Genoma de cloroplasto (cpDNA)

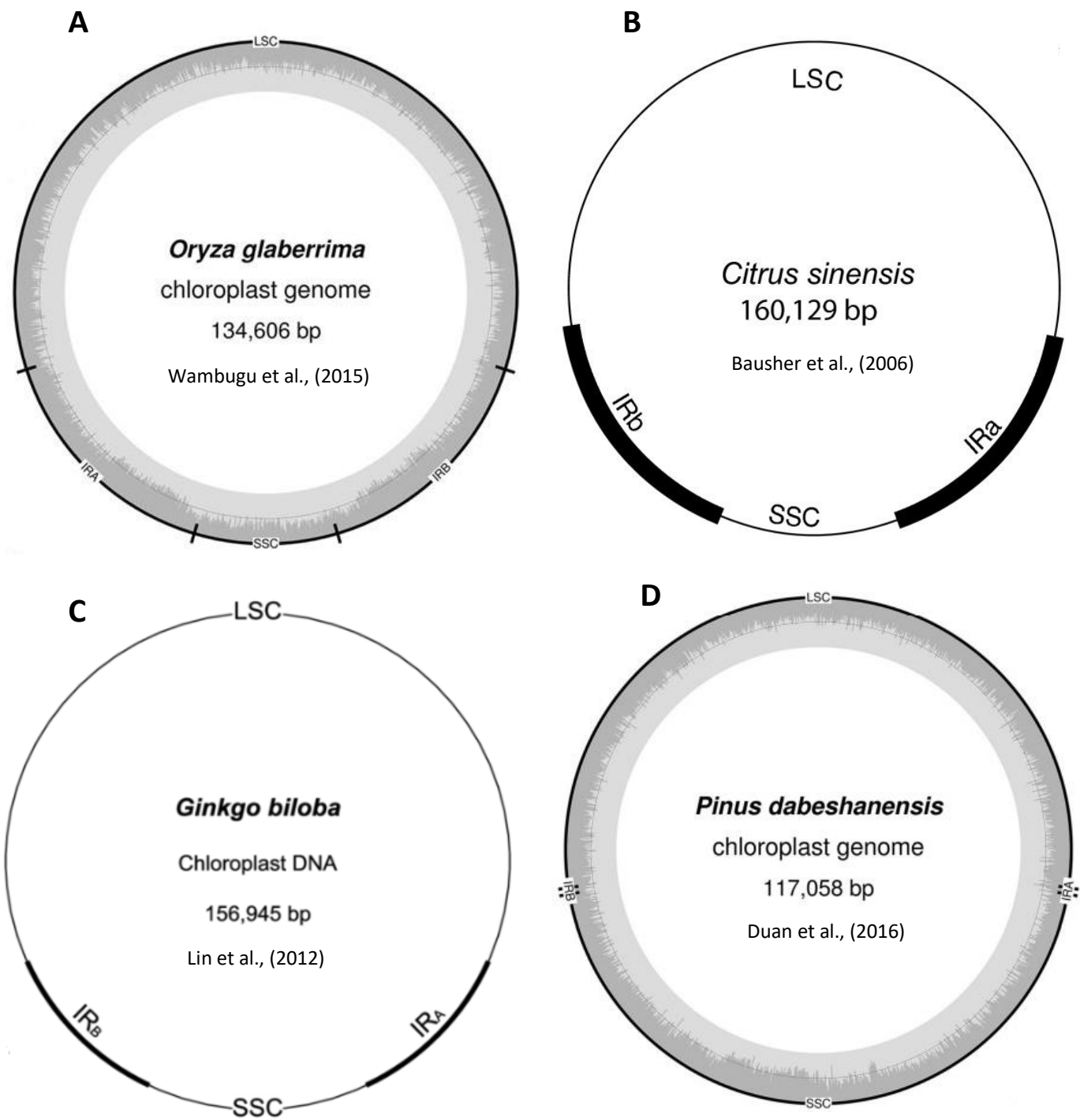
Os cloroplastos são centros metabólicos ativos que, através do processo fotossintético, convertem energia solar em carboidratos com liberação de oxigênio (Daniell et al. 2016). Além da fotossíntese, os cloroplastos apresentam outras funções vitais como a síntese de aminoácidos, nucleotídeos, ácidos graxos, hormônios vegetais e outros metabólitos; além da assimilação de enxofre e nitrogênio (Daniell et al. 2016). Estas funções são cruciais para as plantas, uma vez que os metabólitos sintetizados nos cloroplastos permitem uma resposta às mais variadas condições ambientais, incluindo fatores bióticos e abióticos (Sablok et al. 2016).

Os cloroplastos apresentam o seu próprio genoma, denominado cpDNA. O cpDNA codifica uma série de proteínas envolvidas na fotossíntese e em outras vias metabólicas (Stern et al. 2010). O primeiro genoma de cloroplasto a ser sequenciado foi o de *Nicotiana tabacum*, em 1986, com o método Sanger (Shinozaki et al. 1986). O advento das tecnologias NGS elevou drasticamente o patamar da genética e genômica vegetal e, desde então, centenas de genomas de cloroplasto foram sequenciados e os respectivos dados disponibilizados em repositórios como o NCBI (Smith 2013). O acúmulo destes dados tem contribuído com uma série de estudos em diversos grupos vegetais, proporcionando *insights* importantes como evidências de transferência horizontal de genes (Stegemann et al. 2012), rearranjos genéticos (Chen et al. 2015), identificação de SSRs (Hao et al. 2016), proposição de marcadores taxonômicos (*DNA barcoding*) (Yang et al. 2017), entre outros.

Os genomas de cloroplasto apresentam alto grau de conservação em estrutura e composição em diversos grupos de plantas terrestres (Daniell et al. 2016). A maioria dos cpDNAs apresenta uma estrutura de DNA circular dividida em quatro regiões: duas regiões repetidas invertidas (IRs) separando uma região longa de cópia única (LSC) e uma região curta de cópia única (SSC) (Daniell et al. 2016) (Figura 7). Os cpDNAs das plantas terrestres apresentam de 120 a 130 genes, os quais codificam componentes relacionados à fotossíntese, transcrição e síntese proteica (Sablok et al. 2016). Além de sequências codificantes, os cpDNAs abrigam sequências espaçadoras intergênicas não codificantes, às quais desempenham papéis regulatórios e exibem considerável grau de polimorfismo em tamanho e composição nucleotídica entre diferentes grupos (Coombe et al. 2016; Daniell et al. 2016; Horvath et al. 2018).

Apesar do alto grau de conservação estrutural em plantas terrestres, os genomas de cloroplasto podem variar consideravelmente em tamanho. Por exemplo, *Cathaya argyrophylla*, uma espécie de conífera pertencente à família Pinaceae, apresenta um genoma plastidial de 107 Kb, enquanto que espécies do gênero *Pelargonium*, família Geraniaceae, exibem um cpDNA de 218 Kb (Sablok et al. 2016). Esta variação provavelmente é consequência de rearranjos estruturais que ocorreram nos genomas de cloroplasto de certas linhagens de plantas terrestres, como na maioria das gimnospermas (Wu and Chaw 2014; Xu et al. 2015; Chaw et al. 2018). Entre estes rearranjos pode-se citar o encurtamento ou perda das regiões invertidas repetidas (IRs), bem como a perda de famílias gênicas inteiras (Wu et al. 2011; Wu and Chaw 2014; Chaw et al. 2018).

O genoma de cloroplasto de gimnospermas, principalmente em espécies coníferas, apresenta características distintas em relação ao cpDNA de angiospermas, sobretudo em relação às IRs (Chaw et al. 2018). Em angiospermas é comum que os plastomas contenham IRs com comprimento superior a 20 kb (Xu et al. 2015). A presença de longas repetições invertidas repetidas está diretamente ligada à estabilidade e manutenção da estrutura conservada dos genomas de cloroplasto (Xu et al. 2015). Porém, em gimnospermas o tamanho e até mesmo a presença das IRs variam entre os diferentes táxons. As linhagens mais basais de gimnospermas apresentam IRs com maior extensão, enquanto que linhagens mais derivadas apresentam IRs reduzidas ou ausentes (Figura 7). As regiões IR nas espécies basais *Ginkgo biloba* e *Cycas szechuanensis* apresentam um tamanho de 17 Kb e 25 Kb, respectivamente (Lin et al. 2012; Wang et al. 2018). Nas coníferas há uma tendência para redução ou perda das IRs (Wu and Chaw 2014). O cpDNA de pináceas contém IRs com comprimento inferior a 1 Kb, variando de 358 bp em *Pinus contorta* a 845 pb em *Pinus tabuliformes* (Asaf et al. 2018). Em contrapartida, o cpDNA de várias outras espécies de coníferas não apresenta IRs em sua estrutura, como ocorre em espécies do gênero *Agathis* (Chen et al. 2015), *Wollemia* (Yap et al. 2015) e *Podocarpus* (Vieira et al. 2014).



**Figura 7.** Diagramas indicando a estrutura (LSC/IRs/SSC) e tamanho do cpDNA de espécies representativas de diferentes grupos de plantas terrestres: *Oryza glaberrima* (Monocotiledôneas); *Citrus sinensis* (Dicotiledôneas); *Ginkgo biloba* (Ginkgophyta); *Pinus dabeshanensis* (Coniferophyta).



## 2. OBJETIVOS

O presente estudo tem como objetivo estudar a diversidade de miRNAs e o genoma de cloroplasto, dois componentes genéticos até então inexplorados em *Araucaria angustifolia*, que irão trazer informações relevantes da biologia de pequenos RNAs, regulação da expressão e expressão gênica diferencial, genoma organelar e aspectos evolutivos no gênero *Araucaria* e família Araucariaceae.

### Objetivos específicos

- ▶ Identificar miRNAs novos e conservados em araucaria;
- ▶ Caracterizar o padrão de expressão dos miRNAs novos e conservados em diferentes tecidos;
- ▶ Comparar as sequencias dos miRNAs preditos em *A. angustifolia* com dados de outras gimnospermas;
- ▶ Identificar alvos dos miRNAs preditos em *A. angustifolia*;
- ▶ Montar e anotar o genoma do cloroplasto (cpDNA);
- ▶ Comparar a estrutura e conteúdo gênico do cpDNA de *A. angustifolia* com o genoma do cloroplasto de outras espécies da família Araucariaceae;
- ▶ Realizar análises filogenéticas entre sequencias do cpDNA de *A. angustifolia* e ortólogos de outras gimnospermas com genoma plastidial disponível no NCBI;
- ▶ Identificar microssatélites (SSRs) no cpDNA de *A. angustifolia*.

### **3      CAPITULO 1- Novel and conserved miRNAs among Brazilian pine and other gymnosperms**

**Autores:** José Henrique Galdino<sup>1</sup>, Maria Eguiluz<sup>1</sup>, Frank Guzman<sup>2</sup>, Rogerio Margis<sup>1,2,3\*</sup>

<sup>1</sup>Programa de Pósgraduação e Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil.

<sup>2</sup>Programa de Pósgraduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil.

<sup>3</sup>Departamento de Biofísica, Instituto de Biociências, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil.

**Artigo publicado no periódico *Frontiers in Genetics* (2019)**



# Novel and Conserved miRNAs Among Brazilian Pine and Other Gymnosperms

José Henrique Galdino<sup>1</sup>, Maria Eguiluz<sup>1</sup>, Frank Guzman<sup>2</sup> and Rogerio Margis<sup>1,2,3\*</sup>

<sup>1</sup> Programa de Pós-graduação e Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, Brazil, <sup>2</sup> Programa de Pós-graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, Brazil, <sup>3</sup> Departamento de Biofísica, Instituto de Biociências, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, Brazil

The knowledge about plant miRNAs has increased exponentially, with thousands of miRNAs been reported in different plant taxa using high throughput sequencing technologies and bioinformatic tools. Nevertheless, several groups of plants remain unexplored, and the gap of knowledge about conifer miRNAs is considerable. There is no sequence or functional information available on miRNAs in Araucariaceae. This group is represented in Brazil by only one species, *Araucaria angustifolia*, an endangered species known as Brazilian pine. In the present study, Brazilian pine has its transcriptome explored with respect to small RNAs, representing the first description in a member of the Araucariaceae family. The screening for conserved miRNAs in Brazilian pine revealed 115 sequences of 30 miRNA families. A total of 106 precursors sequences were predicted. Forty one comprised conserved miRNAs from 16 families, whereas 65 were annotated as novel miRNAs. The comparison of Brazilian pine precursors with sRNA libraries of other five conifer species indicates that 9 out 65 novel miRNAs are conserved among gymnosperms, while 56 seems to be specific for Brazilian pine or restricted to Araucariaceae family. Analysis comparing novel Brazilian pine miRNAs precursors and *Araucaria cunninghamii* RNA-seq data identified seven orthologs between both species. Mature miRNA identified by bioinformatics predictions were validated using stem-loop RT-qPCR assays. The expression pattern of conserved and novel miRNAs was analyzed in five different tissues of 3-month-old *Araucaria* seedlings. The present study provides insights about the nature and composition of miRNAs in an Araucariaceae species, with valuable information on miRNAs diversity and conservation in this taxon.

**Keywords:** *Araucaria angustifolia*, Araucariaceae, microRNAs, non-coding RNAs, transcriptome

## OPEN ACCESS

### Edited by:

Rosane Garcia Collevatti,  
Universidade Federal de Goiás, Brazil

### Reviewed by:

André Luis Wendt Dos Santos,  
University of São Paulo, Brazil  
Evandro Novaes,  
Universidade Federal de Lavras, Brazil

### \*Correspondence:

Rogerio Margis  
rogerio.margis@ufrgs.br

### Specialty section:

This article was submitted to  
RNA,  
a section of the journal  
Frontiers in Genetics

**Received:** 12 December 2018

**Accepted:** 28 February 2019

**Published:** 22 March 2019

### Citation:

Galdino JH, Eguiluz M, Guzman F and Margis R (2019) Novel and Conserved miRNAs Among Brazilian Pine and Other Gymnosperms. *Front. Genet.* 10:222. doi: 10.3389/fgene.2019.00222

## INTRODUCTION

MicroRNAs (miRNAs) represent an important class of gene expression regulators (He and Hannon, 2004). These small nucleic acids correspond to 20–25 nt endogenous noncoding RNA sequences with considerable impact on virtually all biological processes (Budak and Akpinar, 2015). miRNA genes are transcribed as long precursor transcripts, called pri-miRNAs, which have the capacity to form a fold-back hairpin structure (Chorostecki et al., 2017). By dicer-like-1 enzyme (DL1) processing, precursors are cleaved and mature 5p/3p miRNA duplexes are produced (Zhang Y. et al., 2018). Usually, one of these mature miRNAs, is incorporated into RNA-induced silencing

complexes (RISCs) (Paroo et al., 2007), and by the action of Argonaute 1 proteins (AGO1), these complexes act over mRNA targets, directing their sequestration or degradation (Pratt and MacRae, 2009).

The first record of a miRNA was performed in 1993 in a study with *C. elegans* (Lee et al., 1993). Since then, a series of studies were carried out and thousands of miRNAs were described in several animal and plant taxa (Cui et al., 2017). Among land plants, the miRNA characterization approach was extensively propagated, mainly, in Angiosperm model species like *Oryza sativa* (Wang, 2004), *Zea mays* (Mica, 2006), *Panicum virgatum* (Xie et al., 2010), *Glycine max* (Severin et al., 2010) among others. In miRBase (Kozomara and Griffiths-Jones, 2014) it is possible to note that Angiosperm species miRNAs are remarkably predominant among Viridiplantae data. Even with predictions of novel miRNAs in some conifer species like *Picea abies* (Yakovlev et al., 2010), *Pinus taeda* (Lu et al., 2007), *Pinus densata* (Wan et al., 2012b), among others, the knowledge about the complexity of Gymnosperm miRNAs is very limited. Besides, in the field of small RNAs biology, a series of non-studied conifer species have great commercial and ecological importance, which means that a plethora of valuable genetic resources remains hidden in several taxa.

*Araucaria angustifolia* (Bertol.) Kuntze is the only endemic species of Gymnosperm with economic importance in Brazil (Steiner et al., 2008). This species, commonly named Brazilian pine, was the most important wood species from south Brazil in the past century (Santos et al., 2002). Representing a valuable source of seeds, wood, fiber and resin, Brazilian pine was the target of extensive exploitation over decades, suffering massive population decrease (Steiner et al., 2008). Brazilian pine seeds are recalcitrant, maintaining a high metabolism status during the storage (Steiner et al., 2008). Consequently, under normal conditions, the seeds have a short conservation period, with substantial decrease in water potential and viability reduction at 4 months after harvest (Araldi et al., 2015). This recalcitrant feature compromises the conservation of Brazilian pine seeds and, consequently, hampers recovery efforts for degraded populations (Longhi et al., 2009). Currently, this species is classified as critically endangered, according to the International Union of Conservation of Nature Red List of Threatened Species (Thomas, 2013).

Brazilian pine has been targeted by some genetic studies mainly with a focus on somatic embryogenesis (Santos et al., 2002, 2010; Steiner et al., 2008). Recently, an RNA-seq data were used to perform a transcriptome comparative profile analysis of early development stages (Elbl et al., 2015). However, there is no information about miRNAs in this species. In the present study, the Illumina technology was used for sequencing a Brazilian pine small RNA library. Using a bioinformatic approach, a series of conserved and putative novel miRNAs, including their stem-loop structure, sequences, and some potential targets were reported. Also, predicted miRNAs were compared with sRNA sequences from six different conifer species and with RNA-seq data from *Araucaria cunninghamii* to investigate the presence of these miRNAs in other conifer taxa. Finally, stem-loop RT-qPCR was applied to validate bioinformatics outputs and analyze

differential expression patterns of 12 conserved and 30 novel predicted miRNAs in five different tissues of 3-month *Araucaria* plants. The present data provide valuable information about Brazilian pine micro-RNA biology and will be very useful for future studies in this species as well as in Araucariaceae family.

## MATERIALS AND METHODS

### Plant Material

For small RNA library preparation and sequencing, fresh leaves were collected from an adult *Araucaria* tree situated at coordinate 29°51'52.3"S 50°53'51.9" in Rio Grande do Sul in Brazil.

For stem-loop RT-qPCR analysis, *Araucaria* seeds, obtained from the seasonal production, were used for germination and seedling production. Brazilian pine plants were grown under standard greenhouse conditions (Moreira-souza et al., 1994) until reaching 90 days (3-months). Then, samples of each replication were collected and frozen in an ultra-freezer at  $-80^{\circ}\text{C}$  for subsequent molecular analysis.

### Small RNA Isolation and Illumina Sequencing

Total RNA was extracted from Brazilian pine fresh leaves with Trizol reagent (Invitrogen, CA, USA), following the standard protocol. The quantification of isolated RNA was determined using Nanodrop (Nanodrop Technologies, Wilmington, DE, USA). RNA sample was sent to Fasteris SA (Plan-les-Ouates, Switzerland) for sequencing. Using the Illumina HiSeq2000 platform, one sRNA library was constructed and sequenced, comprising 28,376,092 single-end reads with a length of 50 bases (NCBI accession number SRR8599283). The sRNA library building follows a series of standard steps, briefly described as follows: gel purification of the RNA fragments ranging from 20 to 30 nt, ligation of the 3p and 5p adapters and followed by gel purification, cDNA synthesis and cDNA gel purification, and, finally, PCR amplification to generate a cDNA colony template library for deep sequencing.

### Bioinformatic Analysis of sRNA Library

The Illumina small RNA library was processed. First, poor-quality bases, with a Fastq value below 30, were removed and adapter sequences were trimmed using Sickle-Quality-Base-Trimming (<https://github.com/najoshi/sickle>) and Cutadapt (<https://cutadapt.readthedocs.io/en/stable/>), respectively. Second, reads with unknown nucleotides (containing one or more "N" bases) were removed with Prin-Seq script (Schmieder and Edwards, 2011). Third, sequences shorter than 18 and longer than 25 nucleotides were also excluded. Finally, Plant small RNAs derived from rRNAs, tRNAs, snRNAs, snoRNAs deposited at the tRNAdb (Jühling et al., 2009), SILVA rRNA (Jühling et al., 2009), and NONCODE v3.0 (Jühling et al., 2009) databanks as well as from Gymnosperm mtRNA and cpRNA deposited at NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/>) were used as references to align the reads by Bowtie (Langmead et al., 2009).

A set of 24 *A. angustifolia* mRNA-seq libraries were downloaded from NCBI Sequence Read Archive (SRA) under bioproject PRJNA240554 (Elbl et al., 2015). The libraries were

processed in order to cut poor-quality bases off the start and end of the reads, considering a fastq quality threshold of 30, and remove adapter sequences using Trim galore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)). Next, the complete transcriptome was assembled with Trinity (Haas et al., 2013) using default parameters, and the predicted contigs were used as reference sequences for pre-miRNA prediction and identification of potential miRNA targets.

## Identification of Conserved Mature miRNAs

To identify conserved mature miRNAs, all mature miRNA sequences of Viridiplantae from miRBase (version 22) (Kozomara and Griffiths-Jones, 2014) were downloaded and mapped against *A. angustifolia* clean small reads with bowtie I (Langmead et al., 2009), allowing no mismatches.

## Prediction of Conserved and Novel miRNA Precursors

miR-PREFeR (Lei and Sun, 2014) was used for prediction of miRNA precursor sequences. This pipeline uses SAM files obtained from the mapping between small RNAs and the complete transcriptome assembled with Bowtie (Langmead et al., 2009). The candidate precursors obtained were manually revised with Tablet software (Milne et al., 2009), and confirmed according to the anchoring patterns: correctly stem-loop secondary structures should harbor the mature miRNA sequence at one arm of the stem and the antisense miRNA sequence (miRNA\*), when detected, at the opposite arm.

The predicted precursor miRNAs were compared with miRBase stem-loop and mature sequences by BLASTn allowing no mismatches and classified into two categories, conserved precursors or putative novel precursors. The precursor stem-loop structures, as well as their minimal folding free energy (MFE), were analyzed using the annotation algorithm from the UEA sRNA toolkit (Moxon et al., 2008).

## Comparison Between Small RNA-Seq Data of Brazilian Pine and Other Conifer Species

Small RNA-seq libraries of another six-conifer species were downloaded from NCBI: *P. abies* (SRR824149; SRR824150) (Källman et al., 2013), *Ginkgo biloba* (SRR1658896, SRR1658901), *Cunninghamia lanceolata* (SRR066638) (Wan et al., 2012a), *Taxus mairei* (SRR797042) (Hao et al., 2012), and *Taxus wallichiana* (SRR1343578). The parameters for data cleaning and preprocessing were applied as described in the section Bioinformatic Analysis of sRNA Library. Next, all libraries were collapsed into a unique library, redundancy was removed, and reads were tagged with species code and read counts number. Using bowtie (Langmead et al., 2009), all libraries were mapped against Brazilian pine miRNA precursors, separately, allowing no mismatches. Using Tablet software (Milne et al., 2009), the anchoring patterns were visualized.

To investigate the presence of novel miRNAs predicted in *A. angustifolia* in another species of *Araucaria* genus,

RNA-seq libraries of *A. cunninghamii* were downloaded from GenBank (accession PRJNA277081) and this data was analyzed as follows. The libraries were processed, low-quality reads and adaptors were trimmed. The complete transcriptome was *de novo* assembled with Trinity. Sequences of novel miRNA precursors from *A. angustifolia* were blasted against *A. cumnigamia* unigenes. Matched sequences comprising a total extension or at least the region flanked by mature and antisense miRNAs were folded using UEA sRNA workbench (Stocks et al., 2012).

## Prediction of miRNA Targets

The prediction of target genes of the mature miRNAs from the conserved and novel pre-miRNAs was performed using psRNAtarget (Dai and Zhao, 2011) using *A. angustifolia* assembled unigenes. Default parameters and the expectation value of 3.0 were considered in this analysis. Blast2Go software (Conesa and Götz, 2008) was used to understand the functions of the putative target genes.

## Biological Confirmation of Predicted miRNAs by Stem-Loop RT-qPCR

To validate and analyze patterns of expression of Brazilian pine predicted miRNAs, Stem-loop RT-qPCR method was performed. Seeds were germinated, and seedlings were grown until reaching an age of 3-months. Then, RNA samples were isolated using the Trizol reagent (Invitrogen, CA, USA). Five different tissues: young leaf, old leaf, stem, main root, and secondary root, were analyzed using four biological replicates. RNA quality was evaluated using 1% agarose gel electrophoresis and Nanodrop. cDNA was obtained for 42 miRNAs based on the stem-loop method (Chen, 2005). Primers sequences for stem-loop cDNA synthesis and mature miRNA expression are in **Table S1**. RT-qPCR reactions were performed in a CFX 384 RealTime PCR System (Bio-Rad). PCR mixes were carried out in a final volume of 10  $\mu$ L, containing 5  $\mu$ L of diluted cDNA (1:100) and 5  $\mu$ L of reagents mix: 1X SYBR Green, 0.025 mM dNTP, 1X PCR buffer, 3 mM MgCl<sub>2</sub>, 0.25 U Platinum Taq DNA Polymerase (Invitrogen) and 200 nM of each reverse and forward primer. The RT-qPCR conditions were configured in this way: 94°C for 5 min, 40 cycles of 94°C for 15 s, 60°C for 10 and 25 s at 72°C. Melting curves were analyzed at the end of RT-qPCR runs to confirm the quality of amplified products. Samples were evaluated in four technical replicates. Using geNorm (<https://genorm.cmgg.be/>), normalizations for miRNA were performed with the Aang-miR171, Aang-nmiR009, and Aang-nmiR046, as the best combination of normalizers, following well-established criteria for miRNA RT-qPCR analysis (Kulcheski et al., 2010). To calculate the relative expression of miRNAs 2<sup>- $\Delta\Delta$ Ct</sup> method was used (Livak and Schmittgen, 2001). To carry out statistical analysis, ANOVA was applied using SAS software Version 9.4 (SAS Institute, Cary, NC, USA) and Duncan's multiple range test was performed to compare pairwise differences in expression, considering  $p < 0.05$ .

## RESULTS

### Diversity of Small RNA in *Araucaria angustifolia*

A total of 26,102,142 reads were obtained from the *A. angustifolia* small RNA library. This library was processed. Adapters, low-quality reads, base redundant reads, as well as reads longer than 25 and shorter than 18 nucleotides were removed. The clean library comprised 19,505,320 (74.73%) reads (Table 1), which were used for further analysis (Figure 1). The small RNA length distribution in Brazilian pine shown an interesting pattern. The length distribution and diversity of sRNAs are shown in the Figure 2A and Table S2. The highest abundance was observed in sequences with 21 and 24 nt, with 21 nt small RNA comprising more than 10 million sequences. The library

composition analysis showed that 14.45% of reads matched miRNAs, 19.20% matched rRNA, 1.52% matched tRNAs, 0.06% matched snRNAs, 0.69% matched snoRNAs, 7.35% matched mtRNA, 6.64% matched cpRNA, 13.16% matched transposons (TEs), and 36.93% matched other RNAs (Table 1).

### Identification of Conserved miRNAs in *Araucaria angustifolia*

All the Viridiplantae mature miRNAs deposited in miRBase were downloaded and mapped against *A. angustifolia* sRNA data with bowtie (Langmead et al., 2009). In this analysis, mismatches were not considered. As shown in Figure 2B and Table S3, 115 sequences matched miRNAs from 30 conserved families (miR156, miR159, miR160, miR164, miR165, miR166, miR167, miR168, miR169, miR171, miR319, miR390, miR394, miR395, miR396, miR397, miR398, miR399, miR403, miR408, miR529, miR535, miR858, miR894, miR947, miR1314, miR3711, miR4995, miR5139, miR5145, miR6300, and miR6478). The number of unique sequences per family as well as their read counts were highly variable, suggesting complex expression patterns of conserved miRNAs in *A. angustifolia* (Table S3).

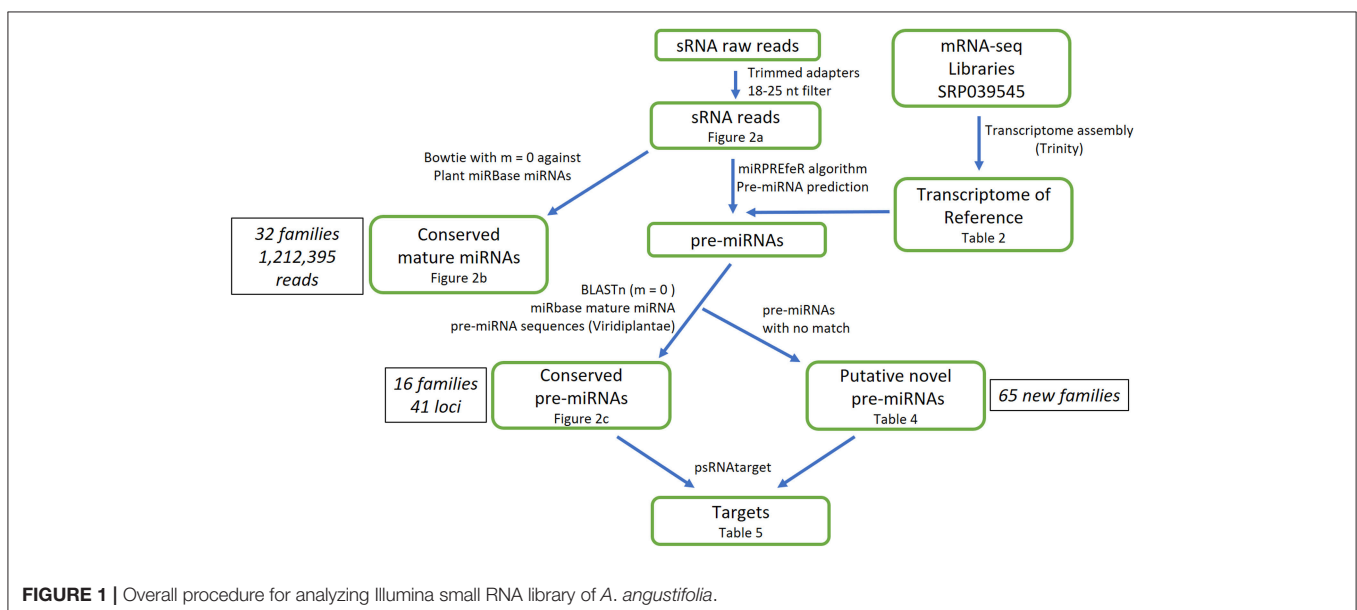
### Identification of Pre-miRNAs Hairpin Sequences in *Araucaria angustifolia*

Brazilian pine has no nuclear genome sequenced yet. Instead, mRNA-seq data are available in GenBank. Then, 24 libraries were downloaded and the complete transcriptome was *de novo* assembled with Trinity. The assembly features are shown in Table 2. The assembled Araucaria transcriptome comprised 360,259 transcripts with an average length of 673 nt. To identify *A. angustifolia* miRNA precursors, the reference transcriptome, as well as the sRNA library, were loaded into miR-PREFeR (Lei and Sun, 2014). This tool follows the criteria for plant miRNA annotation, using expression patterns of miRNAs to predict plant

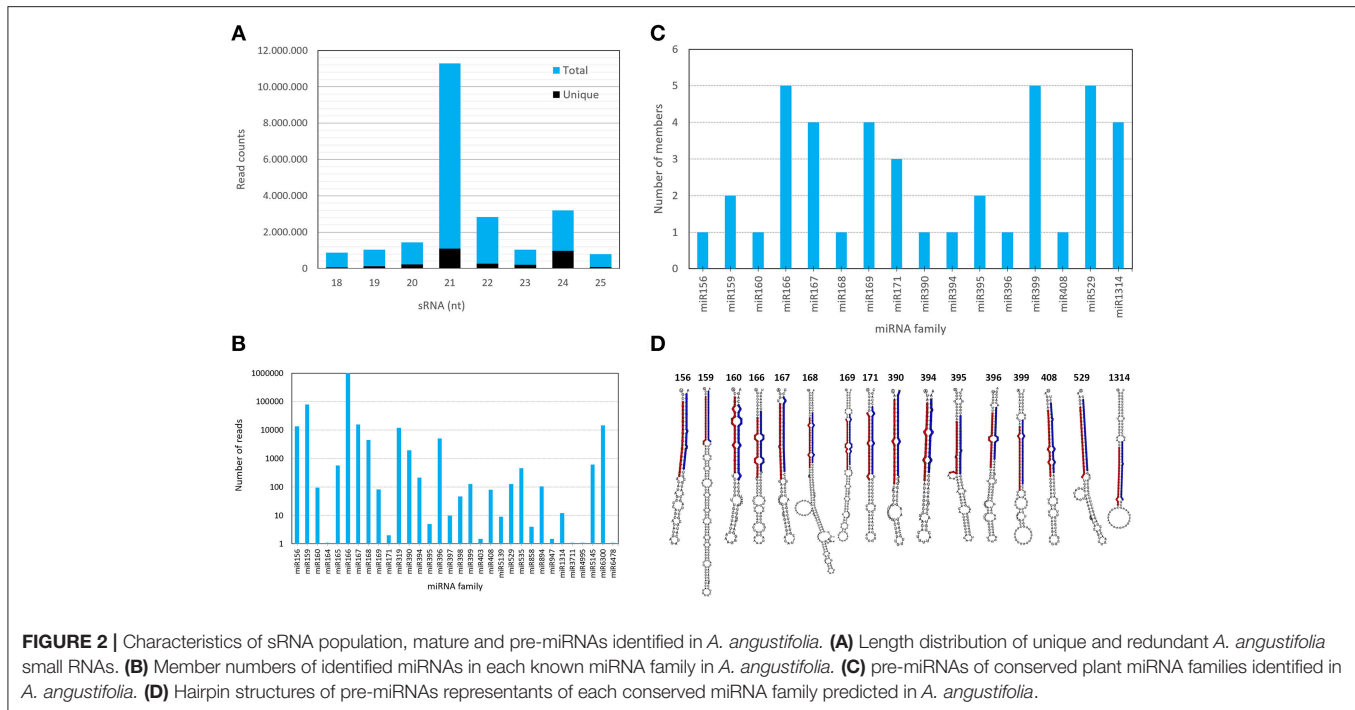
**TABLE 1** | Summary of data from *A. angustifolia* small RNA library sequencing.

Type of small RNA	Number of reads	Percentage
Total*	26,102,142	100.00%
<18 nt	3,847,477	14.74%
>25 nt	2,749,345	10.53%
18–25 nt	19,505,320	74.73%
miRNA	2,819,381	10.80%
rRNA	3,744,202	14.34%
tRNA	297,176	1.14%
snRNA	12,395	0.05%
snoRNA	135,169	0.52%
mtRNA	1,433,049	5.49%
cpRNA	1,295,819	4.96%
retrotransposon	2,567,205	9.84%
Other sRNA	7,200,924	27.59%

\*Reads with length up to 44 nt.



**FIGURE 1** | Overall procedure for analyzing Illumina small RNA library of *A. angustifolia*.



**TABLE 2 |** Statistics on *Araucaria angustifolia* libraries and transcriptome *de novo* assembly.

Total reads	326,525,998
Number of transcripts	360,259
Median transcript length (nt)	256
Mean transcript length (nt)	673
Max transcript length (nt)	23,129
Number of transcripts >1 Kbases	77,861
N50	1,606

miRNAs from small RNA-Seq data (Lei and Sun, 2014). In this way, 106 miRNA precursors, were predicted (Data S1, S2). Using BLAST search, sequences of mature miRNA and antisense miRNA (miRNA\*) of each precursor was compared with mature and stem-loop Viridiplantae data from the miRBase platform, and mismatches were not considered. Following this stringent condition, 41 precursor sequences (pre-miRNAs) of 16 conserved miRNA families were reported (Table 3, Table S4, and Data S1). The most represented miRNA families were miR166, miR399, and miR529 with five members, followed by miR167 and miR1314, with four members (Figures 2C,D). In contrast, with only one member, miR156, miR160, miR168, miR390, miR394, miR396, and miR408 were the less represented miRNA families. The length of conserved pre-miRNAs ranged from 71 to 198 nt and the minimal folding free energy (MFE) ranged from  $-34.15$  to  $-99.70$  kcal/mol (Table 3 and Table S4). The anchoring patterns, as well as stem-loop structures of the 41 known precursors, are shown in the Data S1. The other 65 precursors were considered putative novel pre-miRNAs (Figure 3 and Data S2). The length of these sequences ranged from 61 to 422

nt and the MFE ranged from  $-335.72$  to  $-14.9$  kcal/mol with an average negative folding value of  $-55.9$  kcal/mol (Table 4 and Table S5). As occurred with conserved pre-miRNAs, all novel pre-miRNAs showed regular hairpin structures. Also, in 58 out of 65 (89.23%) novel pre-miRNAs were possible to detect the antisense miRNA (miRNA\*), which strongly support their prediction (Table 4, Table S5, and Data S2), indicating that these pre-miRNAs integrate the *A. angustifolia* miRNAome.

### Comparison Between *Araucaria angustifolia* Predicted Precursors and Small RNA Data From Different Gymnosperms

As a set of mature miRNA sequences of 65 precursors did not match miRBase data, the main online repository for all miRNA sequences and annotation (Kozomara and Griffiths-Jones, 2014), they could be classified as potential species-specific miRNAs. However, the miRBase platform has a restricted miRNA annotation in gymnosperms. There are only four conifer species of three genera with miRNA sequences annotated in this platform, *C. lanceolata*, *P. abies*, *P. densata*, and *P. taeda*. In addition, a series of studies with plant miRNAs, including conifer miRNAs (Chen et al., 2013; Zhang et al., 2013; Li et al., 2017a), was published and several novel plant miRNAs were proposed, but these data were not present in the currently miRBase version. Thus, it is possible that novel miRNAs proposed in different species and classified as species-specific miRNAs could be also present in other taxa.

To avoid overestimation of novel miRNAs, and to provide a comprehensive comparison between conifer miRNAs, sRNA-seq libraries of five conifer species, *P. abies*, *G. biloba*, *C. lanceolata*, *T.*

**TABLE 3** | pre-miRNAs and mature miRNAs identified in *A. angustifolia* matching miRNA families in other plant species.

Loci		Pre-miRNA			miRNA 5P			miRNA 3P		
miRNA family	miRNA locus	Length	Read counts	MFE	Sequence	Read count	Length	Sequence	Read count	Length
miR156	156	87	15,116	-48.5	TGACAGAAGAGAGTGAGCAC	13,367	20	GCTCACCATCTCTTTCTGTGTCAGC	1,583	23
miR159	159a	198	91,542	-71.8	CTTGGATTGAAGGGAGCTCC	77,010	20	AAGCTTCCTTCAGTCCAATCG	3	21
	159b	168	91,690	-91.2	AGCTCCCTTCGGTCCAATT	244	19	CTTGGATTGAAGGGAGCTCC	77,010	20
miR160	160	87	80	-47.1	TGCCTGGCTCCCTGTATGCCA	44	20	GTTGGCATAGAGGGAATCAAG	3	21
miR166	166a	71	1,008,365	-46.2	AAGGGGATTGCGGTCTGGCT	243	20	TCGGACCAGGCTTCATTCCCC	997,155	21
	166b	94	92,632	-42.6	GGACTGTTGTCTGGCTCGAAG	33	21	CCGGACCAGGCTTCATTCCCC	90,796	21
	166c	110	1,008,012	-54.5	GGAATGTTGTCTGGCTCGAGG	22	21	TCGGACCAGGCTTCATTCCCC	997,155	21
	166d	78	1,008,989	-47.7	GGAATGTTGTCTGGCTCGACT	780	21	TCGGACCAGGCTTCATTCCCC	997,155	21
	166e	96	1,063,754	-47.0	Non-detected	-	-	TCGGACCAGGCTTCATTCCCC	997,155	21
miR167	167a	120	13,848	-66.8	TGAAGCTGCCAGCATGATCTGA	11,024	22	AGATCATCTGGTAGCTTCAGC	580	21
	167b	117	4,785	-58.7	TGAAGCTGCCAGCATGATCTGG	2,611	22	TACCAGATCATGGTGGTGGCC	2	21
	167c*	91	4,788	-53.2	TGAAGCTGCCAGCATGATCTGG	2,611	22	AGGTCATCTGGCAGTTTCACC	6	21
	167d	109	4,783	-57.3	TGAAGCTGCCAGCATGATCTGG	2,611	22	TACCAGATCATGGTGGTGGCC	2	21
miR168	168	184	6,801	-74.0	TCGCTTGGTGCAGGTCGGGAA	4,436	21	CCCTGCTTGCATCAACTGAAT	332	21
miR169	169a	136	1,137	-99.7	AACAACCTGCGGCTATTCTA	1	21	GGCAAGTTGTTCTGGCTATG	918	21
	169b	131	195	-42.8	AAGCCAAGGATGAATTGCCGC	13	21	GGCAAGTTGTTCTGGCTACG	72	21
	169c	155	1,802	-72.7	AAGCCAAGGATGATTTGCCGG	938	21	GGCAAGTTGTTCTGGCTACG	72	21
	169d	158	2,886	-71.5	AAGCCAAGGATGATTTGCCGG	938	21	GGCAAGTTGTTCTGGCTATG	918	21
miR171	171a	94	491	-50.4	GGATATTGGAGCGTTCAACC	2	21	TTGAGCCGTGCCAATATCGCA	378	21
	171b	114	46	-63.1	GTGATGTTGGCTGGGCTCAAT	4	21	TGAGCCGTGCCAATATCACAA	16	21
	171c	84	36	-47.1	GTGATGTTGGCTGGGCTCAAT	4	21	TGAGCCGTGCCAATATCACAA	16	21
miR390	390a	83	1,998	-45.0	AAGCTCAGGAGGATAGCGCC	1,952	21	CGCTATCTATCTGAGCTTTT	13	23
miR394	394	80	611	-43.7	CTGGCATTCTGTCCACCTCC	312	21	AGGCGGACGGTATGCCAAGT	18	20
miR395	395a	112	2,712	-52.1	GTTCCCTCAACTACTTCAGAA	157	21	CTGAAGAGTTTGGGGAACTC	2,157	21
	395b	95	2,514	-36.5	Non-detected	-	21	CTGAAGAGTTTGGGGAACTC	2,157	21
miR396	396	120	5,246	-56.8	TTCCACAGCTTTCTTGAACCT	4,606	21	TTCAAGATTGCTGTGGGAAA	1	20
miR399	399a	114	136	-66.3	GGGGAGCTCTCCTTTGGCGGG	6	21	TGCCAAAGGAGAGTTGCCCTG	120	21
	399b	102	129	-64.0	GGGGGGCTCTCCTTTGGTGGG	2	21	TGCCAAAGGAGAGTTGCCCTG	120	21
	399c	111	134	-63.0	GGGGAGCTCTCCTTTGGCAGG	2	21	TGCCAAAGGAGAGTTGCCCTG	120	21
	399d	96	135	-53.2	GGGGAGCTCTCCTTTGGCGGG	6	21	TGCCAAAGGAGAGTTGCCCTG	120	21
	399e	112	134	-67.1	GGGGAGCTCTCCTTTGGCAGG	2	21	TGCCAAAGGAGAGTTGCCCTG	120	21
miR408	408	93	93	-55.6	GCCGGGAAGAGATAGCCCAT	1	20	TGCACTGCCTCTCCCTGGCTG	79	22
miR529	529a	115	542	-51.9	AGAAGAGAGAGAGCACAGCCT	357	21	AGGCTGTGCTCTCTCTCTTC	1	21
	529b	89	544	-51.4	AGAAGAGAGAGAGCACAGCCT	357	21	GCTGTGCTCTCTCTCTCTTC	2	21
	529c	119	540	-49.1	AGAAGAGAGAGAGCACAGCCT	357	21	Non-detected	-	-
	529d	104	540	-44.6	AGAAGAGAGAGAGCACAGCCT	357	21	Non-detected	-	-
	529e	103	463	-51.5	AGAAGAGAGAGAGTACAGCCC	35	21	GTTGTGCTCTCTCTCTCTTC	383	21
miR1314	1314a*	71	6,878	-34.1	CTCCTACATTTAGGGTCGCCG	1,188	21	TCGGCCTTGAATGTTAGGAGAG	4,523	22
	1314b*	107	6,904	-53.5	CTCCTACATTTAGGGTCGCCG	1,188	21	TCGGCCTTGAATGTTAGGAGAG	4,523	22
	1314c*	135	5,853	-60.2	CTTCTAAATTTAAGGTGCCG	560	21	TCGGCCTTGAATGTTAGGAGAG	4,523	22
	1314d*	124	5,921	-51.5	CTTCTAAATTTAAGGTGCCG	560	21	TCGGCCTTGAATGTTAGGAGAG	4,523	22

\*Precursor miRNAs also identified in *Araucaria cunninghamii*.

*mairei*, and *T. wallichiana*, were downloaded from GenBank (the accession codes were shown in Materials and Methods section). A comprehensive Phylogenetic relationship among this species is shown in Lu et al. (2014). The libraries were processed and only reads with length ranging from 18 to 25 nt remained. Then, using Bowtie (Langmead et al., 2009), the conifer sRNA libraries were mapped against the Brazilian pine miRNA precursors

in two different ways. First, the small conifer libraries were mapped separately allowing no mismatches. Second, all libraries, including Brazilian pine sRNA, were collapsed, organized into unique reads, and mapped allowing no mismatches, and the mapping was visualized with Tablet (Milne et al., 2009).

All precursor sequences of miR156, miR159, miR160, miR166, miR167, miR168, miR390, and miR396 families matched with





**TABLE 4** | Characteristics of novel pre-miRNAs and mature miRNAs identified in *A. angustifolia*.

Loci	Pre-nmiRNA			nmiRNA 5P			nmiRNA 3P		
	Length	Read count	MFE	Sequence	Read count	Length	Sequence	Read count	Length
Aang-nmiR001	93	303,638	-41.7	ACTGTGGGATGATGTCAAAAA	945	21	TTTGACATCACACCCGCGGTGA	294,457	22
Aang-nmiR002	78	299,835	-30.1	TATTCATCTATCACTGTGGAAA	4,092	22	TCCCCGGTGATTGATGAAGACA	231,531	22
Aang-nmiR003*	90	132,298	-40.4	CGTGGGCGACCGGGGAAAATT	51	21	TTTTCCCTGATCCGCCCATGCC	102,252	22
Aang-nmiR004	126	83,829	-94.3	TAGTAGACCAGACTCGCCATC	82,793	21	CGAGTCGGATCTACTACAACCT	243	22
Aang-nmiR005*	85	82,708	-38.0	TGACAGCCCGAAATCAGCGAGT	75,705	22	TGGCTGATTCGACTATCAAG	311	21
Aang-nmiR006	90	76,565	-36.8	GTGGTCGGCGAGAAGAATCC	37,665	20	Non-detected	-	-
Aang-nmiR007	142	73,017	-63.1	TGGGCTTACATGTCTGTCGATG	141	22	TCAGCAGACATGTAGCCAACC	71,657	22
Aang-nmiR008	117	62,299	-55.8	TCCCAAACATCGTCCAGAAATA	41,135	22	GTTTGGACGATGTTTGAATG	548	21
Aang-nmiR009	132	47,232	-65.3	TCTCGAACATCCTGCAGCCATT	44,752	22	TGGCTGCACGACTTCGAGATA	546	21
Aang-nmiR010	243	37,279	-93.6	CTGGTAAACAGATGGGGCACT	181	21	CGCCCATCTGATTACCGGTC	36,806	21
Aang-nmiR011	94	21,441	-30.8	TCACCGTGGACCGATGTAATA	354	21	TTACGTCAGGTCTCTGTGATT	17,626	22
Aang-nmiR012	90	18,386	-35.7	CCATCCGGCACTTGATGTCAAAA	39	22	TGACGTCAGGTCTCGATGGTT	16,623	22
Aang-nmiR013	94	13,462	-34.0	CCATTGAGCGCTTGGTGTCAAAA	7,534	22	TAACATCAGGCCCTCGATGATT	310	22
Aang-nmiR014	69	13,373	-14.9	TCTTGGATTATGGAAGACGAACC	8,615	24	AGGATGTTTTTCATTAATCAAGAAC	19	24
Aang-nmiR015*	64	13,041	-24.9	GGTCGTCACGGTCGGTCCGCC	3,356	21	Non-detected	-	-
Aang-nmiR016	67	12,997	-38.5	CGAGGAAATAATGTGAAGAAC	1,048	21	TCTTCACATCCTTTCTCGGA	7,684	21
Aang-nmiR017	81	11,945	-40.7	CGTGGGGCGTTGGACAAAACC	353	22	TTTTGCCAATACCTCCCATGCC	11,321	22
Aang-nmiR018	114	11,727	-40.7	CCGTATTCATTAACCATAGAG	328	21	CCTGTGGTTAATGAATACATCG	7,218	22
Aang-nmiR019	76	11,414	-35.3	CCATCGAGGCTTGACGTCAAAA	119	22	TTACGTCAGGTCTCTATGGTT	8,998	22
Aang-nmiR020	100	6,535	-45.4	AGGGCTGTCCGTGATTGGGCA	15	21	TCATACCCAATCACCGACAGC	3,981	21
Aang-nmiR021*	149	6,158	-61.0	Non-detected	-	-	TTTTTCCAATTCGCCCATGCC	5,945	21
Aang-nmiR022	140	5,009	-44.1	CCGTATTCATTAACCATAGAG	328	21	CCTATGATTAATGAATACATCG	4,168	22
Aang-nmiR023	99	4,579	-59.2	TGACTGTCTGGATGTATATC	4,464	21	TGTACATGCACGACGGTCACG	6	21
Aang-nmiR024	109	3,727	-55.0	CAGCCAAGAATGATTTGCCCGCC	565	23	GGCAGGTCATTCTTGGTGCT	1,089	20
Aang-nmiR025	177	3,525	-64.0	CCGCATCAGGTCTCCAAGGTG	3,447	21	Non-detected	-	-
Aang-nmiR026	88	3,344	-40.2	TGGATAGGAGGAGGATTCATG	1	21	TAAATCCTTCTGCTGTCCATA	3,062	21

(Continued)

TABLE 4 | Continued

Loci	Pre-nmiRNA			nmiRNA 5P			nmiRNA 3P		
	nmiRNA	Length	Read count	MFE	Sequence	Read count	Length	Sequence	Read count
Aang-nmiR027	75	2,601	-37.9	CCACCGTGGACCTCGGTGTGAA	10	21	TACAGTCAGGACCTCGGTGGTT	1,810	22
Aang-nmiR028	102	2,574	-49.3	TCCGGAGACGTCGGCGGGGGC	1,114	21	Non-detected	-	-
Aang-nmiR029	126	2,184	-54.9	AAAACCATTGACTATCAAAGA	47	22	TTTTGATAGCCAGTGGCAATC	1,364	21
Aang-nmiR030	127	1,798	-72.4	TGCGCCCTCGCGCGGGCC	74	19	GCGCTGGCCGGCGGGCTTTC	454	20
Aang-nmiR031	123	1,751	-65.2	ACCTCGCCAACAATCTCAGC	110	22	TGAGATTGTTGGAGAGTTTCG	847	21
Aang-nmiR032	104	1,283	-43.3	AGAAGAGAGAAAGCACATCCC	814	21	GTTGTGCTCTCTCTCTTCTTC	383	21
Aang-nmiR033*	176	1,238	-68.2	TTACCAGCCCCGCCATGCCTA	379	22	GGCGTTGCCGGTCTGGTAAAA	571	21
Aang-nmiR034	61	1,232	-18.3	GTCCTATTCGGTGGCCT	563	18	GAATAACGTGATAGGAGTCTG	12	21
Aang-nmiR035	127	1,203	-68.1	ATGCTTGTTATCTCTGTGCGGC	548	22	CCGCGCAGAAATAAAGCATG	19	21
Aang-nmiR036*	112	1,122	-46.0	CCTTGTTCCCTATTACTGGCA	932	21	TCAATAATAGGAACACAGGTT	133	22
Aang-nmiR037	122	855	-42.6	AGTCAACTCAAGTCTTTGAAA	14	21	TAAAGATTTGAGTTGTCCAA	666	21
Aang-nmiR038	65	830	-37.9	AGTGGGAGGAACGGGCAAAACT	96	23	TTTTCCGGCTCCTCCATTCC	662	22
Aang-nmiR039	112	777	-48.5	ATTGGACAACCTCAATCTTTGA	260	21	CTCAAGGACTTGAGCTGTCCAA	113	22
Aang-nmiR040	135	713	-55.4	CAGCAAGTGGAAACTAGAAT	11	21	TATTCAGTTCCTCACTTGCT	506	21
Aang-nmiR041	114	668	-45.3	TGGGCTTACATGTCTGTCGATG	141	22	TCAGCAGATATGTCAGCCAACC	379	22
Aang-nmiR042	141	664	-69.0	CACATTTTTAGTCTGAAACTG	317	21	TTCAGACTAAAGATGTGTATT	235	21
Aang-nmiR043	81	590	-47.3	TCCGAATTCGCGACGCTCCA	255	21	GGACCGTCGCTGAATTCGGAG	143	21
Aang-nmiR044*	104	559	-22.9	TTGCTGTCCATCAAAGAAGGC	377	21	Non-detected	-	-
Aang-nmiR045	143	465	-71.7	TTTCTGTGAACAAAATTTCAA	112	21	TTTGAAATTTGGTCATAGAG	38	21
Aang-nmiR046	86	442	-42.8	AGTGGGATGCGAGGATAAGACT	259	22	TCTTCTACGCTCCATTCC	172	22
Aang-nmiR047	131	392	-45.2	AGAATTGAAAACTTGCTTAT	126	21	AGGCATGTTTTCAATTCTGA	40	21
Aang-nmiR048	69	370	-27.0	CCATTGAGCACTTGTGTCAA	9	21	TTTTTTGACATCAGGCCCTC	206	21
Aang-nmiR049	89	369	-45.6	TGATAAGGCCCTAATGACACAA	269	21	GTGTTATTTGGCTTGTCAAT	27	21
Aang-nmiR050	82	343	-44.6	TTATTGAATACTGGTAAAAGG	12	21	TTACCAGTCTCAATGAGATC	279	21
Aang-nmiR051	198	330	-149.8	GTCTGCAGAGTGTATGGCCTG	272	21	CCTGCAGTCCAACATATACG	1	20
Aang-nmiR052	188	275	-67.6	TTCCAAAGCAGATAGATTGCCA	86	22	GCATCTGTCTGCTTCGGAATA	41	21

(Continued)

TABLE 4 | Continued

Loci	Pre-nmiRNA			nmiRNA 5P			nmiRNA 3P			
	nmiRNA	Length	Read count	MFE	Sequence	Read count	Length	Sequence	Read count	Length
Aang-nmiR053	94	231	-64.6	GCAATGAATCGGCTGAATCGC	141	21	Non-detected		-	-
Aang-nmiR054	99	223	-61.4	TGACCGTCGTGGATGTATATC	175	21	TGCACGACGGTCACGACTGCC	14	21	
Aang-nmiR055	67	218	-20.2	GTGGCCTATCGATCCTTTAG	33	20	CTAGAGGTGTCAGAAAAGTTAC	105	22	
Aang-nmiR056	100	180	-34.4	CTTGATGATGATAACCGTTGACG	15	23	CACGGTTTGTCTGAAAGAT	43	19	
Aang-nmiR057	135	176	-82.1	TGCTGAAATCGGTCGACTGA	78	21	GGTACGATCGATTTCCGTATA	70	21	
Aang-nmiR058	265	173	-102.0	TGGCATGACTTGCAAATTATG	9	21	CAATTTGTAAGGCCATGCTAAT	145	22	
Aang-nmiR059	422	147	-335.7	AGTGTCCAGCATTTCTCGTCT	4	21	TGAGAAATGCTGGACACTTCT	127	21	
Aang-nmiR060	98	147	-47.9	TGTTTCTACTGAGTTGGTTTCC	68	22	AACCTACTTAGTGAAAACATG	2	21	
Aang-nmiR061	83	123	-36.0	CGTGGGCGTCTTGACAAAGC	21	22	TTTTTCCAATGCCGCCCATGCC	91	22	
Aang-nmiR062	85	101	-49.6	CCCGTATTGAAGATCAACCCA	11	21	GGTTGATCTTCAATATGGCGC	59	21	
Aang-nmiR063	90	86	-54.8	TTTTGATTTTCAGTACGAATA	13	21	TTCGACTGAAAATCAAATC	8	21	
Aang-nmiR064	158	72	-87.6	GTTTTAACTCATGGATATGCA	42	21	CATATCCATGAGTTAAACCC	18	21	
Aang-nmiR065	113	46	-46.9	TTTATTGATTTGATGCTAATGA	3	22	CTTAACACCAGACTAATGAACA	12	22	

\*Correspond to pre-miRNAs also identified in *Araucaria cunninghamii*.

miRNAs (10 miRNA families) and 43 were targets of novel miRNAs (22 miRNA families). Detailed annotation outputs are shown in **Table 6**. Among the conserved miRNAs with predicted targets, Aang-miR156 was the only one that exhibited two targets, SBP (Squamosa promoter binding) and a gene encoding an exocyst complex component (EXO70A1-like). The other conserved miRNAs exhibited only one predicted potential target. For instance, Aang-miR159/transcription factor GAMYB, Aang-miR395/ATP sulfurylase 2 and Aang-miR1314/transcription factor ice 1. Among the novel miRNAs only eight exhibited only one predicted target (Aang-nmiR012, Aang-nmiR019, Aang-nmiR021, Aang-nmiR022, Aang-nmiR029, Aang-nmiR031, Aang-nmiR038, and Aang-nmiR052), the others showed two or more targets, for instance, Aang-nmiR006 exhibited four potential targets. The potential target genes regulated by the conserved miRNAs seem to display physiological functions, such as ATP-sulfurylase, regulation of gene expression (transcription factors GAMYB and ice1), RNA metabolism (U5 small ribonucleoprotein) and signaling cascades (ethylene receptor) (**Table 6**). Interestingly, a series of novel miRNAs (17 out of 22) are predicted to target genes related to disease resistance, like nucleotide binding site leucine-rich repeats (NBS-LRR). These results suggest that the conserved miRNAs are involved in a broad range of relatively conserved physiological functions, whereas most of the novel miRNAs,

possibly Araucariaceae-specific miRNAs, seem to be involved in disease resistance.

### Expression Profiles of Conserved and Novel miRNAs From *Araucaria angustifolia*

The stem-loop RT-qPCR method was used to validate and measure the expression of 12 conserved miRNAs (Aang-miR156, Aang-miR159, Aang-miR166, Aang-miR167, Aang-miR168, Aang-miR169, Aang-miR171, Aang-miR390, Aang-miR395, Aang-miR399, Aang-miR529, Aang-miR1314) and 30 novel miRNAs (Aang-nmiR001, Aang-nmiR002, Aang-nmiR003, Aang-nmiR004, Aang-nmiR005, Aang-nmiR007, Aang-nmiR008, Aang-nmiR009, Aang-nmiR011, Aang-nmiR012, Aang-nmiR016, Aang-nmiR017, Aang-nmiR018, Aang-nmiR019, Aang-nmiR021, Aang-nmiR023, Aang-nmiR025, Aang-nmiR026, Aang-nmiR027, Aang-nmiR029, Aang-nmiR038, Aang-nmiR040, Aang-nmiR044, Aang-nmiR046, Aang-nmiR049, Aang-nmiR051, Aang-nmiR054, Aang-nmiR057, Aang-nmiR059, Aang-nmiR061) in five tissues (young leaves, old leaves, stem, main root and secondary root) of 3-month-old plants (**Figure 5A**). The expression patterns of these miRNAs are illustrated in **Figures 5B,C, Datas S5--S7**.

Four conserved miRNAs showed high expression patterns in just one tissue compared to the others, Aang-miR166, and Aang-miR168 in the stem, and Aang-miR156 and Aang-miR159 in

**TABLE 5** | Mapping patterns of sRNAs from different gymnosperms against conserved and novel miRNA precursors predicted in *A. angustifolia*.

	miRNA precursor	<i>Cunninghamia lanceolata</i>	<i>Ginkgo biloba</i>	<i>Picea abies</i>	<i>Taxus mairei</i>	<i>Taxus wallichiana</i>	<i>Araucaria angustifolia</i>
Conserved miRNAs	Aang-miR156	283,795	4,702	230	21,967	1,574	15,118
	Aang-miR159ab	1	52	244	28	1	91,693
	Aang-miR160	38	269	270	48	2	80
	Aang-miR166abcde	4,983	719,138	433,596	21,426	4,574	1,063,754
	Aang-miR167abcd	33,472	5,437	955	779	45	13,848
	Aang-miR168	5,639	37,003	1,184	27,150	59,574	6,801
	Aang-miR169abcd	0	28	0	0	0	2,887
	Aang-miR171abc	17	10,214	89	3	4	491
	Aang-miR390a	395	8,922	1,974	164	52,965	1,999
	Aang-miR394	0	11,081	312	4	0	611
	Aang-miR395ab	0	2	0	2	0	2,722
	Aang-miR396	102	4,147	254	794	1,166	5,246
	Aang-miR399abcde	2	20	3	28	0	136
	Aang-miR408	325	47	0	190	22,972	94
	Aang-miR529abcde	0	77	58	174	11	1,086
	Aang-miR1314abcd	0	64,219	20	0	1	6,904
	Novel miRNAs	Aang-nmiR006	0	116	0	24	1,151
Aang-nmiR014		1,301	295	24	426	1,085	32,638
Aang-nmiR025		355	3,118	242	2,441	787	5,040
Aang-nmiR028		696	564	3	68	11,953	3,084
Aang-nmiR031		0	126,984	42,987	1,402	163	1,751
Aang-nmiR034		0	4	0	1,470	2,148	1,869
Aang-nmiR046		0	31	0	41	742	442
Aang-nmiR055		345	409	128	46	1,402	618
Aang-nmiR063		43	252	57	60	357	454

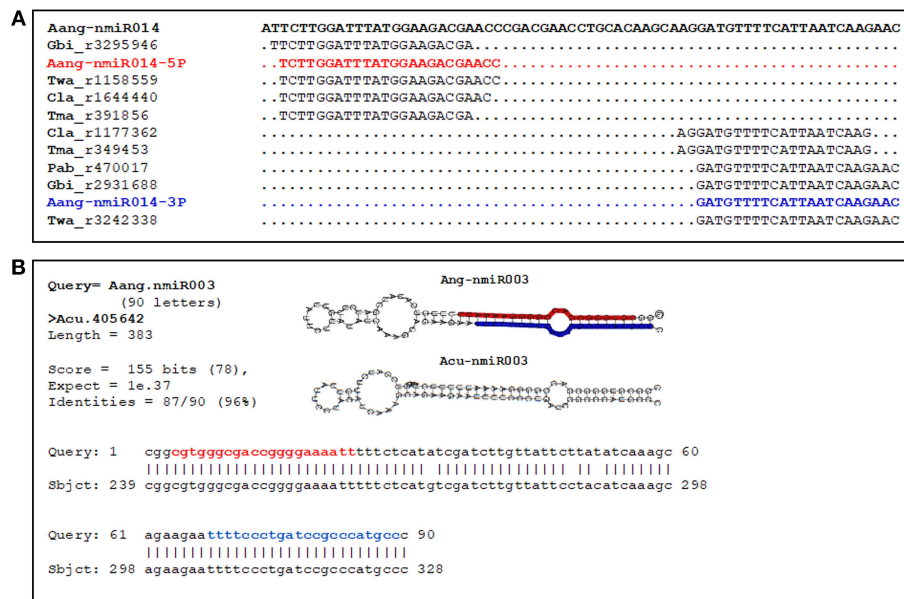
secondary root. The other conserved miRNAs showed different expression patterns (Figure 5B). For example, Aang-miR395 showed high levels in stem, followed by young and old leaves and low levels in main and secondary roots, Aang-miR167 showed highest expression levels in young leaves and secondary roots and Aang-miR399 showed highest levels in old leaves and secondary roots (Figure 5B).

The relative expression data of the novel miRNAs suggested highly complex expression patterns over the plant body (Figure 5C, Datas S6, S7). Among thirty novel miRNAs, twenty-seven exhibited some degree of variation in expression between the tissues (Figure 5C, Datas S6, S7). For example, Aang-nmiR038 was abundantly expressed in young and old leaves, moderately expressed in stem and main root and weakly in secondary roots (Datas S6, S7). The expression level of Aang-nmiR008 was approximately 3-fold higher in old leaves than in young leaves, 2-fold higher than secondary roots and slightly higher than stem and main root (Datas S6, S7). Three novel miRNAs, Aang-nmiR003, Aang-nmiR021, and Aang-nmiR061, had higher expression in the stem than in other tissues (Datas S6, S7). Aang-nmiR021 was barely detected in old leaves, moderately expressed in young leaves, main and secondary roots and strongly expressed in the stem with a predominant expression pattern in this tissue (Datas S6, S7). Aang-nmiR023,

Aang-nmiR044, and Aang-nmiR051 showed ubiquitous expression levels in the tissues with a slight increase in main root (Datas S6, S7). Fourteen novel miRNAs had higher expression in secondary roots than in other tissues: Aang-nmiR001, Aang-nmiR002, Aang-nmiR004, Aang-nmiR005, Aang-nmiR007, Aang-nmiR011, Aang-nmiR012, Aang-nmiR016, Aang-nmiR017, Aang-nmiR018, Aang-nmiR019, Aang-nmiR025, Aang-nmiR026, and Aang-nmiR049 (Figure 5, Datas S6, S7). In some cases, the expression levels were more than 10-fold higher in secondary roots than in other tissues. For example, Aang-nmiR016 was more than 50-fold higher in secondary roots than in young leaves and Aang-nmiR025 was 20-fold higher in secondary roots than in stem (Data S6). In contrast, some novel miRNAs were homogeneously expressed among the tissues, as illustrated by Aang-nmiR40 and Aang-nmiR044 (Figure 5, Datas S6, S7).

## DISCUSSION

High-throughput sequencing technologies represent a breakthrough in the molecular biology scientific world. Thousands of genome sequences, RNA-seq, and sRNA-seq projects have been released and a plethora of biological processes have been comprehensively analyzed. In the plant small RNA



**FIGURE 4** | Novel pre-miRNAs identified in *A. angustifolia* and present in other conifer species. **(A)** Locations of small RNAs of conifer-species mapped onto the Aang-nmiR014 precursor. The sequences corresponding to the *A. angustifolia* (Aan) most abundant 5p and 3p miRNAs are colored in blue and red, respectively. Cla, *Cunninghamia lanceolata*; Gbi, *Ginkgo biloba*; Pbi, *Picea abies*; Tma, *Taxus mairei*; Twa, *Taxus wallichiana*. **(B)** BLAST searching output showing the identification of precursor Aang-nmiR003 in *Araucaria cunninghamii*.

biology field, a series of miRNAs have been identified in a series of groups, but there is no genetic data available about miRNAs in any species (members) of Araucariaceae family. *Araucaria angustifolia* is the most important endemic conifer species in Brazil and has been used in a series of genetic studies (Auler et al., 2002; Souza et al., 2009; Elbl et al., 2015), but there is no available information about miRNAs in this species. In the present study, Illumina technology was used for deep sequencing of small RNA library to identify miRNAs in *A. angustifolia*.

The small RNA length distribution in *A. angustifolia* shows high abundance in sequences with 21 and 24 nt, (Figure 2B). Plant sRNAs are commonly reported in two principal size classes, 21 nucleotides and 24 nucleotides (Chávez Montes et al., 2014). This distribution pattern is well-documented in angiosperms (Guzman et al., 2012, 2013; Källman et al., 2013). However, non-angiosperm species comprise alternative ones (Chávez Montes et al., 2014). For example, in conifers, *P. abies* and *Pinus contorta* fail to produce significant numbers of 24-nt long small RNAs (Dolgosheina et al., 2008), whereas Chinese fir (Wan et al., 2012a) libraries were predominantly represented by this length class. The length enrichment toward 21 nt in the Brazilian pine sRNAome was also reported in other conifer-species (Dolgosheina et al., 2008; Chávez Montes et al., 2014; Zhang et al., 2015). The high abundance of 24 nt sRNAs, as well as the presence of 2.5 million sRNA, reads related to transposable elements (TEs) may reflect a myriad of sRNA types and functions, including events of transposition regulation. Liu and El-Kassaby pinpointed that a significant portion of 24 nt sRNAs may be related to TE silencing in *Picea glauca* during early developmental stages, and this expression decreases throughout the progression of phases (Liu

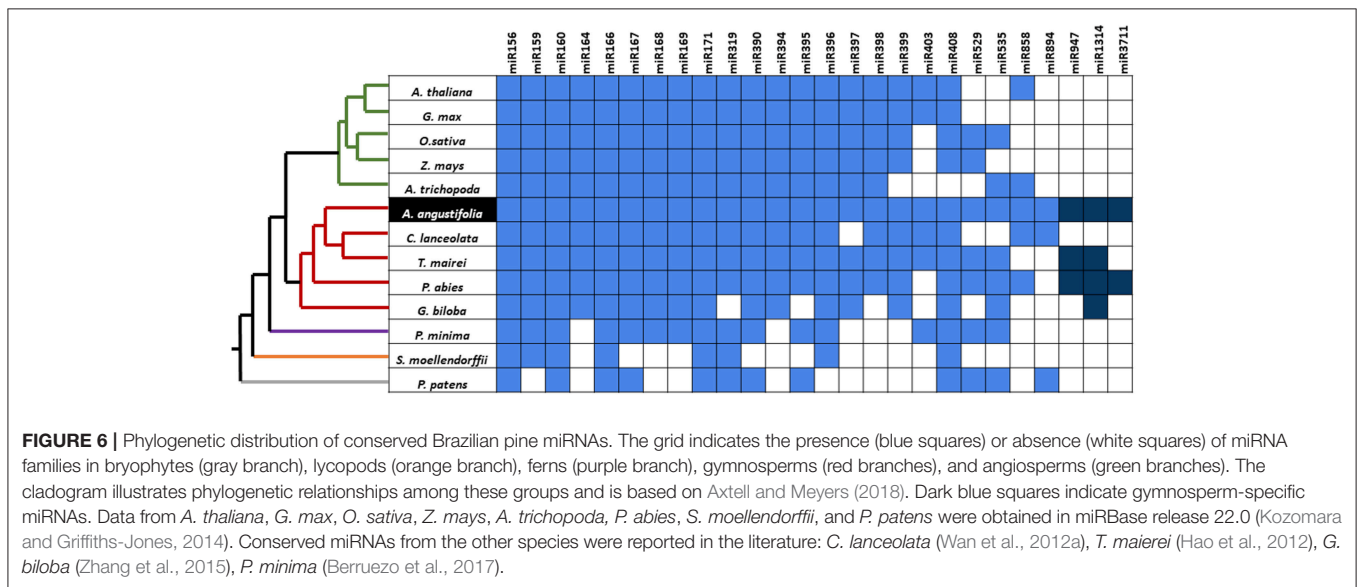
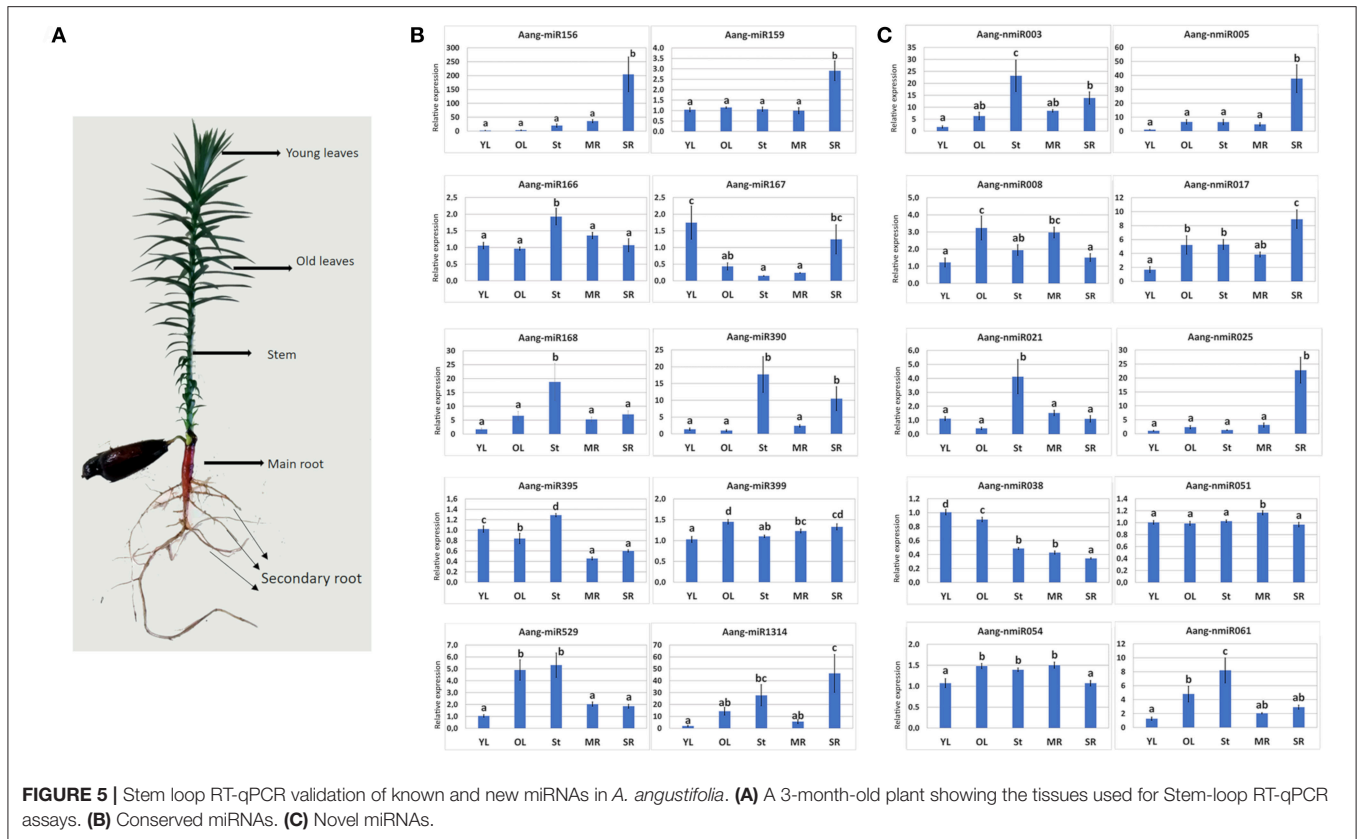
and El-Kassaby, 2017). Also, a small RNA class called hc-siRNAs, mostly 24 nts in length, are substantially numerous in sRNA libraries of land plants (Axtell and Meyers, 2018). Usually, these sRNAs are derived from intergenic, repetitive and transposon-related genomic regions (Axtell and Meyers, 2018). Therefore, the high diversity of 24 nt sequences, as well as the high number of sequences related to TEs suggests a role of these elements in Brazilian pine biology.

By comparing small RNA data from *A. angustifolia* with miRBase, 115 sequences representing 30 evolutionary plant miRNA families were identified (Figure 2C). The number of family members, as well as abundance, varied among families and sequences, respectively. Among conserved mature miRNA families identified in *A. angustifolia*, miR156, miR159, miR166, and miR167 have been documented as high abundant in several plant groups (Taylor et al., 2014). Other evolutionary conserved miRNA families were also identified in Brazilian pine. Among them, seven are conserved from Embryophytes (Berrueto et al., 2017), miR160, miR171, miR319, mir395, miR396, miR408, miR535, three are conserved from Tracheophytes (Berrueto et al., 2017), miR162, miR168, and miR403, and three conserved from Gymnosperms (Chávez Montes et al., 2014), miR947, miR1314, and miR3711 (Figure 6).

The miR403 family was reported in conifers and other vascular plant groups (Jagtap and Shivaprasad, 2014), with patterns that suggest a complex evolutionary history (Berrueto et al., 2017). For example, miRNA403 was reported in several dicots (Cuperus et al., 2011), but only in a few number of monocot species (Zhang et al., 2011). Beyond miR403, some interesting situations occur with miRNAs of other families

**TABLE 6** | Predicted putative targets of novel and conserved miRNAs in *A. angustifolia*.

miRNA	Target acc.	Expectation	Inhibition	Annotation
Aang-miR156	1931415	2.5	Cleavage	squamosa promoter-binding 20
	796092	3	Cleavage	exocyst complex component EXO70A1-like
Aang-miR159	977162	2.5	Cleavage	transcription factor GAMYB
Aang-miR166	1704393	2	Cleavage	homeodomain-leucine zipper transcription factor HB-3
Aang-miR167	1066318	2.5	Cleavage	zinc finger ZAT4-like
Aang-miR171	2163525	1.5	Cleavage	scarecrow 6
Aang-miR390	153786	2.5	Cleavage	ethylene receptor 2-like
Aang-miR395	968805	3	Cleavage	ATP sulfurylase 2
Aang-miR399	1917714	2.5	Cleavage	probable apyrase 6
Aang-miR529	1928068	2	Cleavage	U5 small nuclear ribonucleoprotein 40 kDa isoform X1
Aang-miR1314	1949852	2.5	Translation	transcription factor ice1
Aang-nmiR003	799146	2.5	Translation	diterpene synthase
Aang-nmiR005	2042481	2.5	Cleavage	disease resistance RPP13 4
	1950381	2.5	Translation	probable receptor kinase At1g49730
Aang-nmiR006	1063530	2.5	Cleavage	G-type lectin S-receptor-like serine threonine- kinase At2g19130
	1926825	0	Cleavage	disease resistance RGA3
	2436549	0	Cleavage	LRR and NB-ARC domain disease resistance
	1580639	1	Cleavage	NBS-LRR disease resistance
	2026350	1	Cleavage	disease resistance RPM1-like
Aang-nmiR009	225523	3	Cleavage	disease resistance RGA3
	967108	3	Cleavage	disease resistance RGA2-like isoform X1
	1681946	1.5	Cleavage	disease resistance RPM1-like
Aang-nmiR012	1601456	2	Cleavage	TMV resistance N-like
Aang-nmiR017	2007770	2.5	Cleavage	NBS-LRR
	2105514	2.5	Cleavage	disease resistance TAO1-like
	2105518	2.5	Cleavage	TMV resistance N-like
Aang-nmiR018	811798	2.5	Translation	disease resistance (TIR-NBS-LRR class)
	811797	2.5	Translation	TMV resistance N-like
	160015	2.5	Translation	disease resistance (TIR-NBS-LRR class)
Aang-nmiR019	1997048	2.5	Translation	TMV resistance N-like
Aang-nmiR021	2023865	1.5	Cleavage	disease resistance RPM1-like
Aang-nmiR022	2023865	1.5	Cleavage	disease resistance RPM1-like
Aang-nmiR025	869457	1.5	Cleavage	senescence-associated
	1863722	0	Cleavage	cytochrome P450 like TBP
Aang-nmiR027	169430	2.5	Cleavage	cyclin-T1-3-like isoform X1
	1942123	2.5	Cleavage	cyclin-T1-5
Aang-nmiR029	1580639	2.5	Translation	NBS-LRR disease resistance
Aang-nmiR031	2026350	2.5	Translation	disease resistance RPM1-like
Aang-nmiR038	1946613	1.5	Translation	disease resistance RPP13 4
Aang-nmiR039	1799491	2.5	Cleavage	disease resistance RGA1
	2509763	2.5	Cleavage	resistance family
Aang-nmiR051	2060073	0	Cleavage	probable xyloglucan endotransglucosylase hydrolase 10
	706255	1.5	Cleavage	probable xyloglucan endotransglucosylase hydrolase 32
Aang-nmiR052	984263	2.5	Translation	L-ascorbate oxidase homolog
Aang-nmiR054	1236307	3	Cleavage	L-ascorbate oxidase
Aang-nmiR059	2324731	2.5	Cleavage	cell division cycle 123 homolog
	2131961	2.5	Cleavage	target of AVRb operation1
	2174518	0	Cleavage	SUPPRESSOR OF npr1- CONSTITUTIVE 1-like
Aang-nmiR062	2151817	1.5	Cleavage	TMV resistance N-like isoform X2
	818026	0	Cleavage	TMV resistance N-like
	2126093	0	Cleavage	disease resistance TAO1-like
Aang-nmiR064	1820645	0	Cleavage	Disease resistance (TIR-NBS-LRR class) family
	1749835	2.5	Cleavage	disease resistance (TIR-NBS-LRR class)
	2031676	2.5	Cleavage	TMV resistance N-like



identified in Brazilian pine. miR894 has annotations only for the moss *Physcomitrella patens* (Arazi, 2012) in miRBase. Its absence in distant species like Arabidopsis and rice suggested a species-specific character. However, outside the miRBase platform, this family has been identified in some angiosperms, like moth orchid *Phalaenopsis aphrodite* (Chao et al., 2014) and citrus

(Tzarfati et al., 2013). miR6300, reported in *G. max* for the first time, has been identified in angiosperms (Biswas et al., 2016) but lack homolog sequences reported in other groups. miR5145 has annotation only for rice (Chen et al., 2011) and miR6478 has annotation only for *Populus trichocarpa* (Chen et al., 2011) in miRBase. Since mismatches were not allowed in



comparisons with miRBase data, this suggests that the miRNAs above-mentioned can integrate the miRNAome of Brazilian pine, and, for the first time, were identified in a gymnosperm species.

By using a well-established approach in the prediction of pre-miRNAs (Lei and Sun, 2014), 41 conserved and 65 potential novel pre-miRNAs were identified in *A. angustifolia* (Tables S4 and S5). Taking into account the main criteria for miRNA annotation (Taylor et al., 2014; Axtell and Meyers, 2018), i.e., a stable structure of stem-loop with high base complementarity between the arms (Velayudha Vimala Kumar et al., 2017) and evidence of mature and antisense miRNA with characteristic base-pairing between them (Gaspin et al., 2016), the novel and conserved *A. angustifolia* pre-miRNAs were consistently predicted (Datas S1, S2).

To exceed the miRBase blast-search in the novel and conserved miRNA identification and obtain more insights about the presence of pre-miRNAs predicted in *A. angustifolia* in different conifer species, including one of the same genera, some alternative approaches were applied. First, sRNA-seq of conifer species from different taxa, including the basal conifer *G. biloba* were mapped against pre-miRNAs predicted in *A. angustifolia*. Since mismatches were not considered, and the mapping patterns could be visualized, the output data showed that all predicted conserved pre-miRNAs were matched with small RNAs of different conifers, mainly in their sequences of 5P and 3P mature miRNAs. Interestingly, the same pattern was observed in 9 of the 65 novel pre-miRNAs predicted in *A. angustifolia* (Table 5). It is important to mention that pre-miRNAs were obtained from RNA-seq data of embryonic tissues, not representing neither the juvenile phase used in the RT-qPCR analysis, nor the mature leaf tissue used to obtain the small RNA seq data corresponding to mature miRNAs. So far, the amount pre-miRNAs detected in this study is certainly underestimated. These findings are particularly important since these miRNAs were considered novel because their mature and antisense miRNAs do not have potential ortholog sequences in the current version of the miRBase platform, even with data from four conifer species. In addition, the probable presence of these nine miRNAs in *G. biloba* suggests that they may evolved in early divergence times of gymnosperms. Extending this idea, another unusual analysis was applied with a similar purpose, but this time toward a related taxon. There is no data from miRNAs in the Araucariaceae family. This family has three genera, *Agathis*, *Araucaria*, and *Wollemia* with a primarily Southern Hemisphere distribution, with the clear majority of species endemic to Australia, New Zealand, New Guinea, and New Caledonia and just two species, *Araucaria araucana* and *A. angustifolia* endemic to South America (Escapa and Catalano, 2013). For the first time, a sRNA-seq data was analyzed and miRNAs were identified in a species representant from Araucariaceae. To extend this analysis in this family and in the genus *Araucaria*, RNA-seq data from *A. cunninghamii*, a related species endemic from Australia, was downloaded from GenBank. The complete transcriptome was assembled into unigenes, and a BLAST search between *A. angustifolia* pre-miRNAs and *A. cunninghamii* unigenes were applied. Again, some conserved pre-miRNAs predicted in *A. angustifolia* showed sequence identity or high similarity with

sequences from a different species, as could be illustrated by the identification of Aang-miR1314abcd family members in *A. cunninghamii* (Data S3). However, these analyses also indicated that seven novel pre-miRNAs identified in *A. angustifolia* showed potential orthologs in *A. cunninghamii* (Data S4). The sequence identity between these orthologs was high, reaching 96% between Aang-nmiR003 and Acun-nmiR003 (Figure 4B), and the mismatches appeared outside the mature miRNA sequences, mainly in the loop region. These patterns were already noted, but mainly in conserved miRNAs (Miskiewicz et al., 2017). For instance, Chorostecki and coworkers showed that plant miRNAs exhibit some evolutionary footprints that extend the mature miRNA sequences, comprising other conserved regions with structural determinants recognized during the biogenesis (Chorostecki et al., 2017). The putative novel orthologs in *A. angustifolia* and *A. cunninghamii* seem to follow the same pattern of evolutive fingerprints.

Several genes were predicted as potential targets for novel and conserved *A. angustifolia* miRNAs (Table 6). Among the conserved miRNA targets, functions related to energetic metabolism, signal transduction and gene expression control were found. The targets related to conserved miRNAs were reported in several taxa, miR171-scarecrow 6 (Zhu et al., 2015), miRNA395-ATP sulfurylase (Zhang et al., 2011; Jagadeeswaran et al., 2012), miR167-Zinc finger ZAT (Omidvar et al., 2015), miR166- HD-ZIPIII (Li et al., 2017b), miRNA159-GAMYB (Li et al., 2017a; Samad et al., 2017; Velayudha Vimala Kumar et al., 2017). This conservation of targets is consistent with the co-evolution of the miRNA-target pairs among plant lineages (Zhu et al., 2015).

Interestingly, a series of novel miRNAs are predicted to target NBS-LRR disease resistance genes. A strong association between the diversity NBS-LRRs and miRNAs was reported in a genome-wide study with 70 land plants (Zhang et al., 2016). There are evidence that NBS-LRR genes keep giving birth to new miRNAs targeting themselves in various plant lineages (Cui et al., 2017). Therefore, the novel miRNAs predicted in Brazilian pine seem to be more lineage-specific, given the high rate of evolution of their targets in response to the high evolution rate of pathogens.

Stem-loop RT-qPCR was applied to analyze expression patterns of conserved and novel miRNAs predicted in *A. angustifolia* (Figure 5, Datas S5–S7). Among the conserved miRNAs analyzed, Aang-nmiR166 showed highest expression levels on stem compared to the other tissues. This miRNA family has HD-ZIPIII transcription factors as potential targets, as reported in a series of studies and in the present work. Experimental analysis showed that the knock-down of miR166 promoted a substantial increase in expression levels of HD-ZIPIII genes OsHB3 and OsHB4 in stem, compared to other tissues in rice plants (Zhang J. et al., 2018). The high expression levels of this miRNA in the stem seem to be in accordance to the regulation of HD-ZIPIII transcription factors in the vascular tissues.

The expression of Aang-miR156 in secondary roots was 200-fold higher than in leaves, 10-fold higher than in stem and more than 5-fold higher than in main root (Figure 5B). These results corroborate other findings of the requirement of high

levels of miR156 expression for adventitious root formation in *Arabidopsis* and *Malus xiaojinensis* (Xu et al., 2016, 2017). Ang-miR159 also had higher expression in secondary roots than in other tissues (Figure 6). The target prediction analysis indicated GAMYB as a potential target for this miRNA (Table 6), a conserved miRNA/target association reported in several plant groups (Xie et al., 2010; Hao et al., 2012; Wan et al., 2012a; Karlova et al., 2013). However, in contrast with the miR156 pro-adventitious root formation in *Arabidopsis*, miR159 was reported as a posttranscriptional repressor of root growth in the same species (Xue et al., 2017). Once Aang-miR156 and Aang-miR159 were found to be high expressed in secondary roots of 3-month-old plants, these miRNAs seem to act synergistically as positive regulators of root growth in *A. angustifolia*.

Aang-miR171 was expressed homogeneously in all tissues, which suggests that this miRNA family plays important roles in different organs in *A. angustifolia*. A series of studies reported several phenotypes in miR171-overexpressing lineages in different species (Hai et al., 2018). Interestingly, a study with transgenic *Arabidopsis* plants overexpressing a miR171 mature sequence from the conifer species *P. densata* showed that these mutants exhibited larger leaves, shorter primary roots, higher plant height and early flowering stages (Hai et al., 2018).

Thirty novel miRNAs predicted in *A. angustifolia* were also biologically validated via stem-loop RT-qPCR (Figure 5, Datas S6, S7). These miRNAs exhibited extremely diverse expression patterns among young leaves, old leaves, stem, main root and secondary roots (Figure 5, Datas S6, S7). The association between RT-qPCR data and target prediction rises important clues about the function of these miRNAs in *A. angustifolia*. For instance, among the novel miRNAs targeting disease-resistant genes, Aang-nmiR038 (targeting RPP13) was high expressed in young and old leaves, Aang-nmiR003 (targeting RPP13) and Aang-nmiR021 (targeting RPM1-like) were high expressed in stem, and other three novel miRNAs were higher expressed in secondary roots, Aang-nmiR017 (targeting NBS-LRR, TAO and TMV), Aang-nmiR018 (targeting NBS-LRR class), Aang-nmiR019 (targeting TMV). These patterns and associations suggest that the novel pre-miRNAs predicted in *A. angustifolia* integrate a series pathways in this species.

## CONCLUSION

In the present study, a small-RNA library was constructed by high-throughput sequencing of *A. angustifolia* leaves with the aim of identifying miRNA precursors, mature miRNAs and miRNA targets in this species. Also, a series of conserved and novel miRNAs predicted in *A. angustifolia* was identified in RNA-seq data from different conifers, including the Australian

native congeneric species *A. cunninghamii*. This study provides the first report on the transcriptome-wide identification of miRNAs as well as the first view of the diversity, abundance and expression patterns of these small RNAs in Araucariaceae. Bioinformatics analysis suggests that Brazilian pine conserved and novel miRNAs might contribute to several physiological processes by targeting multiple targets and affecting different pathways. The novel lineage-specific miRNAs seem to be more involved with response to pathogens by targeting NBS-LRR resistance genes. Experimental analysis indicates that these miRNAs are expressed in different patterns through the plant body. This miRNA-target interaction remains to be further explored in order to achieve novel biological and evolutionary aspects in Brazilian pine and related species. It is possible that a series of miRNAs annotated in the present study can integrate the genetic pool of several non-studied conifers, including species from Araucariaceae and genus *Araucaria*. Therefore, these data represent valuable information for future genetic studies of miRNAs in Gymnosperms, by providing insights about biology, diversity, expression and evolution of these small RNAs. The upload of these data in miRBase will also be important for comparative analysis with other plant groups.

## AUTHOR CONTRIBUTIONS

RM, JG, and FG conceived and designed the study. FG conducted *in silico* analysis. JG and ME conducted the RT-qPCR experiments. JG and RM analyzed the data. JG and RM drafted the manuscript. All authors have read and approved the manuscript.

## FUNDING

RM is the recipient of a research fellowship 309030/2015-3 from Conselho Nacional de Pesquisa e Desenvolvimento Científico e Tecnológico, CNPq. JG and ME are the recipients of a Ph.D. fellowships from Conselho Nacional de Pesquisa e Desenvolvimento Científico e Tecnológico, CNPq. FG is the recipient of Post-Doctoral fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES. The present study was also partially supported through a grant from INCT-MCTIC.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00222/full#supplementary-material>

## REFERENCES

- Araldi, C. G., Maria, C., and Coelho, M. (2015). Establishment of post-harvest early-developmental categories for viability maintenance of *Araucaria angustifolia* seeds. *Acta Bot. Bras.* 29, 524–531. doi: 10.1590/0102-33062015abb0061
- Arazi, T. (2012). MicroRNAs in the moss *Physcomitrella patens*. *Plant Mol. Biol.* 80, 55–65. doi: 10.1007/s11103-011-9761-5

- Auler, N. M. F., dos Reis, M. S., Guerra, M. P., and Nodari, R. O. (2002). The genetics and conservation of *Araucaria angustifolia*: I. Genetic structure and diversity of natural populations by means of non-adaptive variation in the state of Santa Catarina, Brazil. *Genet. Mol. Biol.* 25, 329–338. doi: 10.1590/S1415-47572002000300014
- Axtell, M. J., and Meyers, B. C. (2018). Revisiting criteria for plant MicroRNA annotation in the era of big data. *Plant Cell* 30, 272–284. doi: 10.1105/tpc.17.00851

- Berrueto, F., de Souza, F. S. J., Picca, P. I., Nemirovsky, S. I., Martínez Tosar, L., Rivero, M., et al. (2017). Sequencing of small RNAs of the fern *Pleopeltis minima* (Polypodiaceae) offers insight into the evolution of the microRNA repertoire in land plants. *PLoS ONE* 12:e0177573. doi: 10.1371/journal.pone.0177573
- Biswas, S., Hazra, S., and Chattopadhyay, S. (2016). Identification of conserved miRNAs and their putative target genes in *Podophyllum hexandrum* (Himalayan Mayapple). *Plant Gene* 6, 82–89. doi: 10.1016/j.plgene.2016.04.002
- Budak, H., and Akpinar, B. A. (2015). Plant miRNAs: biogenesis, organization and origins. *Funct. Integr. Genomics* 15, 523–531. doi: 10.1007/s10142-015-0451-2
- Chao, Y.-T., Su, C.-L., Jean, W.-H., Chen, W.-C., Chang, Y.-C. A., and Shih, M.-C. (2014). Identification and characterization of the microRNA transcriptome of a moth orchid *Phalaenopsis aphrodite*. *Plant Mol. Biol.* 84, 529–548. doi: 10.1007/s11103-013-0150-0
- Chávez Montes, R. A., de Rosas-Cárdenas, F. F., De Paoli, E., Accerbi, M., Rymarquis, L. A., Mahalingam, G., et al. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat. Commun.* 5:3722. doi: 10.1038/ncomms4722
- Chen, C. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 33:e179. doi: 10.1093/nar/gni178
- Chen, C.-J., Liu, Q., Zhang, Y.-C., Qu, L.-H., Chen, Y.-Q., and Gautheret, D. (2011). Genome-wide discovery and analysis of microRNAs and other small RNAs from rice embryogenic callus. *RNA Biol.* 8, 538–547. doi: 10.4161/rna.8.3.15199
- Chen, Y. T., Shen, C. H., Lin, W. D., Chu, H. A., Huang, B. L., Kuo, C. I., et al. (2013). Small RNAs of *Sequoia sempervirens* during rejuvenation and phase change. *Plant Biol.* 15, 27–36. doi: 10.1111/j.1438-8677.2012.00622.x
- Chorostecki, U., Moro, B., Rojas, A. L. M., Debernardi, J. M., Schapire, A. L., Notredame, C., et al. (2017). Evolutionary footprints reveal insights into plant MicroRNA biogenesis. *Plant Cell* 29, 1248–1261. doi: 10.1105/tpc.17.00272
- Conesa, A., and Göttsch, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832
- Cui, J., You, C., and Chen, X. (2017). The evolution of microRNAs in plants. *Curr. Opin. Plant Biol.* 35, 61–67. doi: 10.1016/j.pbi.2016.11.006
- Cuperus, J. T., Fahlgren, N., and Carrington, J. C. (2011). Evolution and functional diversification of MIRNA genes. *Plant* 23, 431–442. doi: 10.1105/tpc.110.082784
- Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155–W159. doi: 10.1093/nar/gkr319
- Dolgoshina, E. V., Morin, R. D., Aksay, G., Sahinalp, S. C., Magrini, V., Mardis, E. R., et al. (2008). Conifers have a unique small RNA silencing signature. *RNA* 14, 1508–1515. doi: 10.1261/rna.1052008
- Elbl, P., Lira, B. S., Andrade, S. C. S., Jo, L., dos Santos, A. L. W., Coutinho, L. L., et al. (2015). Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, *Araucaria angustifolia* (Bertol.) Kuntze. *Plant Cell Tissue Organ Cult.* 120, 903–915. doi: 10.1007/s11240-014-0523-3
- Escapa, I. H., and Catalano, S. A. (2013). Phylogenetic analysis of *Araucariaceae*: integrating molecules, morphology, and fossils. *Int. J. Plant Sci.* 174, 1153–1170. doi: 10.1086/672369
- Gaspin, C., Rué, O., and Zytnicki, M. (2016). Ingredients for *in silico* miRNA identification and annotation. *JSM Biotechnol Bioeng.* 3:1071.
- Guzman, F., Almerão, M. P., Korbes, A. P., Christoff, A. P., Zanella, C. M., Bered, F., et al. (2013). Identification of potential miRNAs and their targets in *Vriesea carinata* (Poales, Bromeliaceae). *Plant Sci.* 210, 214–223. doi: 10.1016/j.plantsci.2013.05.013
- Guzman, F., Almerão, M. P., Körbes, A. P., Loss-Morais, G., and Margis, R. (2012). Identification of MicroRNAs from *Eugenia uniflora* by high-throughput sequencing and bioinformatics analysis. *PLoS ONE* 7:e49811. doi: 10.1371/journal.pone.0049811
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hai, B. Z., Qiu, Z. B., He, Y. Y., Yuan, M. M., and Li, Y. F. (2018). Characterization and primary functional analysis of *Pinus densata* miR171. *Biol. Plant.* 62, 318–324. doi: 10.1007/s10535-018-0774-7
- Hao, D.-C., Yang, L., Xiao, P.-G., and Liu, M. (2012). Identification of *Taxus* microRNAs and their targets with high-throughput sequencing and degradome analysis. *Physiol. Plant.* 146, 388–403. doi: 10.1111/j.1399-3054.2012.01668.x
- He, L., and Hannon, G. J. (2004). Correction: MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 631–631. doi: 10.1038/nrg1415
- Jagadeeswaran, G., Nimmakayala, P., Zheng, Y., Gowdu, K., Reddy, U. K., and Sunkar, R. (2012). Characterization of the small RNA component of leaves and fruits from four different cucurbit species. *BMC Genomics* 13:329. doi: 10.1186/1471-2164-13-329
- Jagtap, S., and Shivaprasad, P. V. (2014). Diversity, expression and mRNA targeting abilities of Argonaute-targeting miRNAs among selected vascular plants. *BMC Genomics* 15:1049. doi: 10.1186/1471-2164-15-1049
- Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., and Pütz, J. (2009). tRNADB 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* 37, 159–162. doi: 10.1093/nar/gkn772
- Källman, T., Chen, J., Gyllenstrand, N., and Lagercrantz, U. (2013). A significant fraction of 21-nucleotide small RNA originates from phased degradation of resistance genes in several perennial species. *Plant Physiol.* 162, 741–754. doi: 10.1104/pp.113.214643
- Karlova, R., Van Haarst, J. C., Maliepaard, C., Van De Geest, H., Bovy, A. G., Lammers, M., et al. (2013). Identification of microRNA targets in tomato fruit development using high-throughput sequencing and degradome analysis. *J. Exp. Bot.* 64, 1863–1878. doi: 10.1093/jxb/ert049
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Kulcheski, F. R., Marcelino-Guimaraes, F. C., Nepomuceno, A. L., Abdelnoor, R. V., and Margis, R. (2010). The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Anal. Biochem.* 406, 185–192. doi: 10.1016/j.ab.2010.07.020
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25
- Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y
- Lei, J., and Sun, Y. (2014). miR-PREFEr: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data. *Bioinformatics* 30, 2837–2839. doi: 10.1093/bioinformatics/btu380
- Li, Q., Deng, C., Xia, Y., Kong, L., Zhang, H., Zhang, S., et al. (2017a). Identification of novel miRNAs and miRNA expression profiling in embryogenic tissues of *Picea balfouriana* treated by 6-benzylaminopurine. *PLoS ONE* 12:e0176112. doi: 10.1371/journal.pone.0176112
- Li, X., Xie, X., Li, J., Cui, Y., Hou, Y., Zhai, L., et al. (2017b). Conservation and diversification of the miR166 family in soybean and potential roles of newly identified miR166s. *BMC Plant Biol.* 17:32. doi: 10.1186/s12870-017-0983-9
- Liu, Y., and El-Kassaby, Y. A. (2017). Landscape of fluid sets of hairpin-derived 21-/24-nt-long small RNAs at seed set uncovers special epigenetic features in *Picea glauca*. *Genome Biol. Evol.* 9, 82–92. doi: 10.1093/gbe/evw283
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-CRCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Longhi, S. J., Brena, D. A., Ribeiro, S. B., Gracioli, C. R., Longhi, R. V., and Mastella, T. (2009). Fatores ecológicos determinantes na ocorrência de *Araucaria angustifolia* e *Podocarpus lambertii*, na Floresta Ombrófila Mista da FLONA de São Francisco de Paula, RS, Brasil. *Ciência Rural* 40, 57–63. doi: 10.1590/S0103-84782009005000220
- Lu, S., Sun, Y. H., Amerson, H., and Chiang, V. L. (2007). MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. *Plant J.* 51, 1077–1098. doi: 10.1111/j.1365-313X.2007.03208.x
- Lu, Y., Ran, J.-H., Guo, D.-M., Yang, Z.-Y., and Wang, X.-Q. (2014). Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. *PLoS ONE* 9:e107679. doi: 10.1371/journal.pone.0107679
- Mica, E. (2006). Characterization of five microRNA families in maize. *J. Exp. Bot.* 57, 2601–2612. doi: 10.1093/jxb/erl013
- Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., et al. (2009). Tablet-next generation sequence assembly visualization. *Bioinformatics* 26, 401–402. doi: 10.1093/bioinformatics/btp666

- Miskiewicz, J., Tomczyk, K., Mickiewicz, A., Sarzynska, J., and Szachniuk, M. (2017). Bioinformatics study of structural patterns in plant MicroRNA precursors. *Biomed Res. Int.* 2017, 1–8. doi: 10.1155/2017/6783010
- Moreira-souza, M., Jurandy, E., and Nogueira, B. (1994). Practical Method For germination of *Araucaria angustifolia* (Bert.) O. Ktze. seeds. *Sci. Agric.* 60, 389–391.
- Moxon, S., Schwach, F., Dalmay, T., MacLean, D., Studholme, D. J., and Moulton, V. (2008). A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 24, 2252–2253. doi: 10.1093/bioinformatics/btn428
- Mutum, R. D., Kumar, S., Balyan, S., Kansal, S., Mathur, S., and Raghuvanshi, S. (2016). Identification of novel miRNAs from drought tolerant rice variety Nagina 22. *Sci. Rep.* 6:30786. doi: 10.1038/srep30786
- Omidvar, V., Mohorianu, I., Dalmay, T., and Fellner, M. (2015). Identification of miRNAs with potential roles in regulation of anther development and male-sterility in 7B-1 male-sterile tomato mutant. *BMC Genomics* 16:878. doi: 10.1186/s12864-015-2077-0
- Paroo, Z., Liu, Q., and Wang, X. (2007). Biochemical mechanisms of the RNA-induced silencing complex. *Cell Res.* 17, 187–194. doi: 10.1038/sj.cr.7310148
- Pratt, A. J., and MacRae, I. J. (2009). The RNA-induced silencing complex: a versatile gene-silencing machine. *J. Biol. Chem.* 284, 17897–17901. doi: 10.1074/jbc.R900012200
- Samad, A. F. A., Sajad, M., Nazaruddin, N., Fauzi, I. A., Murad, A. M. A., Zainal, Z., et al. (2017). MicroRNA and transcription factor: key players in plant regulatory network. *Front. Plant Sci.* 8:565. doi: 10.3389/fpls.2017.00565
- Santos, A. L. W., dos Silveira, V., Steiner, N., Maraschin, M., and Guerra, M. P. (2010). Biochemical and morphological changes during the growth kinetics of *Araucaria angustifolia* suspension cultures. *Braz. Arch. Biol. Technol.* 53, 497–504. doi: 10.1590/S1516-89132010000300001
- Santos, A. L. W., dos Silveira, V., Steiner, N., Vidor, M., and Guerra, M. P. (2002). Somatic embryogenesis in parana pine (*Araucaria angustifolia* (Bert.) O. Kuntze). *Braz. Arch. Biol. Technol.* 45, 97–106. doi: 10.1590/S1516-89132002000100015
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026
- Severin, A. J., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160
- Souza, M. I. F., de Salgueiro, F., Carnavale-Bottino, M., Félix, D. B., Alves-Ferreira, M., Bittencourt, J. V. M., et al. (2009). Patterns of genetic diversity in southern and southeastern *Araucaria angustifolia* (Bert.) O. Kuntze relic populations. *Genet. Mol. Biol.* 32, 546–556. doi: 10.1590/S1415-47572009005000052
- Steiner, N., Catarina, C. S., Balbuena, T. S., and Guerra, M. P. (2008). *Araucaria angustifolia* biotechnology. *Funct. Plant Sci. Biotechnol.* 2, 20–28.
- Stocks, M. B., Moxon, S., Mapleson, D., Woolfenden, H. C., Mohorianu, I., Folkes, L., et al. (2012). The UEA sRNA workbench : a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* 28, 2059–2061. doi: 10.1093/bioinformatics/bts311
- Taylor, R. S., Tarver, J. E., Hiscock, S. J., and Donoghue, P. C. J. (2014). Evolutionary history of plant microRNAs. *Trends Plant Sci.* 19, 175–182. doi: 10.1016/j.tplants.2013.11.008
- Thomas, P. (2013). *Araucaria angustifolia*. *The IUCN Red List of Threatened Species 2013:e.T32975A2829141*. doi: 10.2305/IUCN.UK.2013-1.RLTS.T32975A2829141.en
- Tzarfati, R., Sela, I., and Goldschmidt, E. E. (2013). Graft-induced changes in MicroRNA expression patterns in citrus leaf petioles. *Open Plant Sci. J.* 7, 17–23. doi: 10.2174/1874294701307010017
- Velayudha Vimala Kumar, K., Srikakulam, N., Padbhanabhan, P., and Pandi, G. (2017). Deciphering microRNAs and their associated hairpin precursors in a non-model plant, *Abelmoschus esculentus*. *Noncoding RNA* 3:19. doi: 10.3390/ncrna3020019
- Wan, L.-C., Wang, F., Guo, X., Lu, S., Qiu, Z., Zhao, Y., et al. (2012a). Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. *BMC Plant Biol.* 12:146. doi: 10.1186/1471-2229-12-146
- Wan, L. C., Zhang, H., Lu, S., Zhang, L., Qiu, Z., Zhao, Y., et al. (2012b). Transcriptome-wide identification and characterization of miRNAs from *Pinus densata*. *BMC Genomics* 13:132. doi: 10.1186/1471-2164-13-132
- Wang, J.-F. (2004). Identification of 20 microRNAs from *Oryza sativa*. *Nucleic Acids Res.* 32, 1688–1695. doi: 10.1093/nar/gkh332
- Xie, F., Frazier, T. P., and Zhang, B. (2010). Identification and characterization of microRNAs and their targets in the bioenergy plant switchgrass (*Panicum virgatum*). *Planta* 232, 417–434. doi: 10.1007/s00425-010-1182-1
- Xu, M., Hu, T., Zhao, J., Park, M. Y., Earley, K. W., Wu, G., et al. (2016). Developmental functions of miR156-regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes in *Arabidopsis thaliana*. *PLoS Genet.* 12:e1006263. doi: 10.1371/journal.pgen.1006263
- Xu, X., Li, X., Hu, X., Wu, T., Wang, Y., Xu, X., et al. (2017). High miR156 expression is required for auxin-induced adventitious root formation via MxSPL26 independent of PINs and ARFs in *Malus xiaojinensis*. *Front. Plant Sci.* 8:1059. doi: 10.3389/fpls.2017.01059
- Xue, T., Liu, Z., Dai, X., and Xiang, F. (2017). Primary root growth in *Arabidopsis thaliana* is inhibited by the miR159 mediated repression of MYB33, MYB65 and MYB101. *Plant Sci.* 262, 182–189. doi: 10.1016/j.plantsci.2017.06.008
- Yakovlev, I. A., Fossdal, C. G., and Johnsen, Ø. (2010). MicroRNAs, the epigenetic memory and climatic adaptation in Norway spruce. *New Phytol.* 187, 1154–1169. doi: 10.1111/j.1469-8137.2010.03341.x
- Zhang, J., Wu, T., Li, L., Han, S., Li, X., Zhang, S., et al. (2013). Dynamic expression of small RNA populations in larch (*Larix leptolepis*). *Planta* 237, 89–101. doi: 10.1007/s00425-012-1753-4
- Zhang, J., Zhang, H., Srivastava, A. K., Pan, Y., Bai, J., Fang, J., et al. (2018). Knockdown of rice MicroRNA166 confers drought resistance by causing leaf rolling and altering stem xylem development. *Plant Physiol.* 176, 2082–2094. doi: 10.1104/pp.17.01432
- Zhang, L., Zheng, Y., Jagadeeswaran, G., Li, Y., Gowdu, K., and Sunkar, R. (2011). Identification and temporal expression analysis of conserved and novel microRNAs in Sorghum. *Genomics* 98, 460–468. doi: 10.1016/j.ygeno.2011.08.005
- Zhang, Q., Li, J., Sang, Y., Xing, S., Wu, Q., and Liu, X. (2015). Identification and characterization of MicroRNAs in *Ginkgo biloba* var. *epiphylla* Mak. *PLoS ONE* 10:e0127184. doi: 10.1371/journal.pone.0127184
- Zhang, Y., Wang, Y., Gao, X., Liu, C., and Gai, S. (2018). Identification and characterization of microRNAs in tree peony during chilling induced dormancy release by high-throughput sequencing. *Sci. Rep.* 8:4537. doi: 10.1038/s41598-018-22415-5
- Zhang, Y., Xia, R., Kuang, H., and Meyers, B. C. (2016). The diversification of plant NBS-LRR defense genes directs the evolution of MicroRNAs that target them. *Mol. Biol. Evol.* 33, 2692–2705. doi: 10.1093/molbev/msw154
- Zhu, X., Leng, X., Sun, X., Mu, Q., Wang, B., Li, X., et al. (2015). Discovery of conservation and diversification of genes by phylogenetic analysis based on global genomes. *Plant Genome* 8, 1–11. doi: 10.3835/plantgenome2014.10.0076

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Galdino, Eguiluz, Guzman and Margis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

#### **4      CAPÍTULO 2- *Araucaria angustifolia* chloroplast genome sequence and its relation to other Araucariaceae**

**Autores:** José Henrique S. G. Brandão<sup>1</sup>, Nureyev F. Rodrigues<sup>2</sup>, Maria Eguiluz<sup>1</sup>, Frank Guzman<sup>2</sup>, and Rogerio Margis<sup>1,2,3\*</sup>

<sup>1</sup>PPGBM, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

<sup>2</sup>PPGBCM, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

<sup>3</sup>Departamento de Biofísica, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil

**Manuscrito publicado no periódico *Genetics and Molecular Biology* (2019)**



## *Araucaria angustifolia* chloroplast genome sequence and its relation to other Araucariaceae

José Henrique S. G. Brandão<sup>1</sup>, Nureyev F. Rodrigues<sup>2</sup>, Maria Eguiluz<sup>1</sup>, Frank Guzman<sup>2</sup> and Rogerio Margis<sup>1,2,3</sup> 

<sup>1</sup>PPGBM, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.

<sup>2</sup>PPGBCM, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.

<sup>3</sup>Departamento de Biofísica, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.

### Abstract

*Araucaria angustifolia* is endemic to southern Brazil. Known as Brazilian pine, *A. angustifolia* is the only native conifer species with economic and social relevance in this country. Due to massive exploitation, it has suffered a significant population decline and currently is classified as critically endangered. This encouraged the scientific community to investigate genetic features in Brazilian pine to increase resources for management and preservation. In this work, RNA-Seq data was used to determine the complete nucleotide sequence of the *A. angustifolia* chloroplast genome (cpDNA). The cpDNA is 146,203 bp in length and contains 122 genes, including 80 protein-coding genes, 5 ribosomal RNA genes, and 37 tRNA genes. Coding regions comprise 45.02%, 4.96% correspond to rRNAs and tRNAs, and 50.02% of the genome encompasses non-coding regions. Genes found in the inverted repeat (IR) are present as single copy, with exception of the *rnm5* and *trnI*-CAU loci. The typical LSC, SSC, IRa and IRb organization reported in several land-plant groups is not present in *A. angustifolia* cpDNA. Phylogenetic analyses using Bayesian and Maximum Likelihood methods clustered *A. angustifolia* in the Araucariaceae family, with *A. heterophylla* and *A. columnaris* as congeneric species. The screening of *A. angustifolia* cpDNA revealed 100 SSRs, 14 of them corresponding to tetrapolymer loci.

**Keywords:** Brazilian pine, plastid, genome, cpDNA, conservation.

Received: July 18, 2018; Accepted: November 28, 2018.

*Araucaria angustifolia* (Bertol.) Kuntze syn. *Columbea angustifolia*, also known as Brazilian pine, is a native Brazilian tree of the class Pinopsida, order Pinales, and family Araucariaceae. It is widely distributed in the southern and southeastern areas of Brazil but also occurs in limited areas from Argentina and Paraguay (de Souza *et al.*, 2009). *A. angustifolia* is a dioecious wind-pollinated species with a mixed seed dispersion by barochoric autochory and birds (Lowe *et al.*, 2018). The seeds maintain elevated levels of water and active metabolic rates at the mature stage, resulting in a rapid loss of viability (Astarita *et al.*, 2004). Due to seed recalcitrance to storage, conservation strategies are restricted mainly to propagation by embryonic cultures (Steiner *et al.*, 2008).

*A. angustifolia* is one of the most important trees in its region of natural occurrence due to its relevant ecological, economic, and social functions. Its seeds are rich in starch, proteins, and flavonoids, having a high nutritional value during the winter season. As results of its social and economic relevance, *A. angustifolia* went through an indiscriminate exploitation and a substantial population decline, having been categorized as a critically endangered species in the International Union for the Conservation of Nature and Natural Resources (IUCN), Red List of Threatened Species (Thomas, 2013). The taxonomic classification of the genus *Araucaria* is well resolved (Stefenon *et al.*, 2006). It comprises 19 species with an interesting distribution worldwide. The species are distributed only in tropical and subtropical zones of the Southern hemisphere (Stefenon *et al.*, 2006). Seventeen species are present in Oceania, 13 of which are endemic to the small archipelago New Caledonia (Lu *et al.*, 2014). The two remaining species, *Arau-*

Send correspondence to Rogerio Margis, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS) PO Box 15005, 91501-970 Porto Alegre, RS, Brazil. E-mail: [rogerio.margis@ufrgs.br](mailto:rogerio.margis@ufrgs.br).

*caria araucana*, and *Araucaria angustifolia*, are present in southern South America (Lu *et al.*, 2014).

Brazilian pine has been targeted by genetic studies that mainly focused on somatic embryogenesis, with the purpose of developing technologies for the conservation and genetic improvement of this species. One of these studies has generated RNA-seq data from early stage tissues and the libraries are available in the NCBI database (Elbl *et al.*, 2015). Once high-throughput sequencing data is generated, it can be used in a plethora of ways beyond the original purpose, and relevant information can be further explored from the targeted organism. In the present study, the RNA-seq data composed by 24 libraries (Elbl *et al.*, 2015) was used as input to perform the complete assembly and annotation of the *A. angustifolia* chloroplast (cp) genome. The paired-end sequence reads were filtered against 58 Pinidae cp genomes (Table S1) using BWA software with two mismatches allowed (Li and Durbin, 2009). The reads were used for an assembly *de novo* with ABySS software (Simpson *et al.*, 2009). The cp genome scaffolds were orientated using cp genome sequences of *Araucaria heterophylla* (NC\_026450.1) using BLASTN (Camacho *et al.*, 2009). A gap region relative to an intergenic region was filled in after Sanger sequencing using the primers F: 5' ACCGTGAGGGTTCAAGTCC and R: 3' GTGGCACG AGGATTTTCAGT. For this purpose, total DNA was extracted by the CTAB method from young leaves of an *A. angustifolia* tree. DNA quality was evaluated by electrophoresis in a 1% agarose gel, and quantity was determined using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Genes were annotated using GeSeq (Tillich *et al.*, 2017) and BLAST similarity searches. Transfer RNAs (tRNAs) were predicted using the Aragorn program (Laslett and Canback, 2004) implemented in the GeSeq program and confirmed by comparison with the appropriate homologs in *A. heterophylla*. The circular cp genome map was drawn using the online program OGDRAW (Lohse *et al.*, 2013).

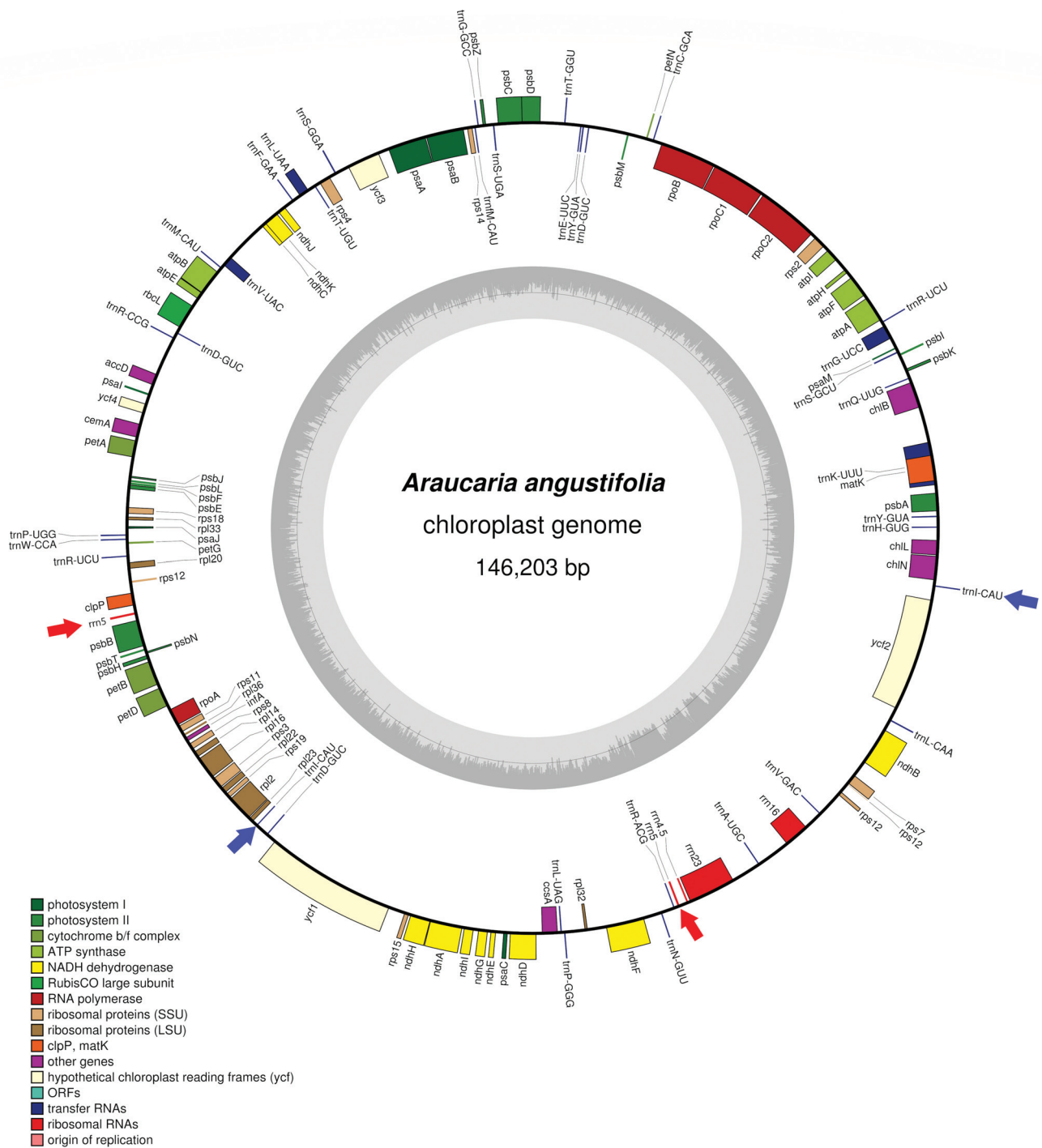
To determine the phylogenetic relationships of *A. angustifolia* in the Pinidae division and corroborate with the Brazilian pine plastid genome sequence, a set of 73 cp protein-coding sequences (Table S2) from 18 conifer species, 17 belonging to Pinidae (Table S3) and *Ginkgo biloba* serving as outgroup were used. Nucleotide sequences were aligned separately using MUSCLE available in MEGA version 6.0 (Tamura *et al.*, 2013). Alignments were concatenated and nucleotide positions of each gene were specified and a Bayesian tree was generated using MrBayes version 3.2.6 (Ronquist *et al.*, 2012), with the JC evolutionary model as determined by MODELTEST version 3.7 (Posada and Crandall, 1998), and 10,000,000 generations sampled every 100 generations. The first 25% of trees were discarded as burn-in to produce a consensus phylogram, with posterior probability (PP) values for each node. Maximum Likelihood (ML) analysis was also applied, using RaxML

(Stamatakis, 2014) program and the ML tree was compared to Bayesian topology. The phylogenetic trees were rooted and visualized using FigTree software (<http://tree.bio.ed.ac.uk/software/figtree/>).

Simple sequence repeats (SSRs) were detected using MISA perl script, available at (<http://pgrc.ipk-gatersleben.de/misa/>), with thresholds of 10 repeat units for mononucleotide SSRs, five repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs. The interruption threshold among SSRs, which indicates the maximum difference between two SSRs was 50 base pairs. A total of 229,914,266 high quality Illumina paired-end reads from the *A. angustifolia* transcriptome libraries generated via the HiScanSQ platform and available at NCBI Sequence Read Archive (SRA) under accession number SRP039545 were filtered against Pinidae cp genomes. The 2,107,993 obtained reads were *de novo* assembled into non-redundant contigs and singletons covering about 99.65% of the cp genome (minimum coverage of 23 reads, maximum coverage of 1,780 reads). Two final large scaffolds were obtained and joined into a single cp circular genome after the use of Sanger sequencing. The complete cp genome of *A. angustifolia* is 146,203 bp in size and was submitted to GenBank (accession number: MH599004). This size is similar to those found in other *Araucaria* species (Ruhsam *et al.*, 2015).

In *A. angustifolia*, duplicated genes present in inverted repeat regions IRa and IRb are found as a single copy, with exception of a sequence of 513 bp corresponding to the *rrn5* gene, placed between *clpP* and *psbB*, suggesting a recombination event, and the two inverted copies from *tRNA-CAU* gene (Figure 1). The same pattern was described in other species of the Araucariaceae family, *Agathis dammara* and *Wollemia nobilis*, which lack canonical IRs and harbor double inverted copies of *rrn5* and *tRNA-CAU* in their plastomes (Wu and Chaw, 2014; Yap *et al.*, 2015). A reduced size of IRa and IRb was described in *Pinus taeda* L. (Asaf *et al.*, 2018). Loss of the IR was also reported in the chloroplast genomes of some species of Pinaceae and Cupressophytes (Wu *et al.*, 2011).

The coding sequences comprise 45.02%, 4.96% correspond to rRNAs and tRNAs, and 50.02% of the genome comprises non-coding regions, including introns, pseudogenes and intergenic spacers (Table 1). In general, all genomic features showed similarity in size, structure, and gene abundance with other *Araucaria* species (Table S4) (Ruhsam *et al.*, 2015). The genome contains 122 genes in total, which includes 120 single-copy genes corresponding to 80 protein-coding genes, 36 transfer RNA (tRNA) genes and four ribosomal genes (rRNA) (Figure 1, Table 1). The cp genome has 14 intron-containing genes: 9 protein-coding genes, one pseudogene, and four tRNA genes. The *rps12* gene, a trans-spliced gene entirely located in the LSC region, and the *yef3* gene contain two introns each; the other



**Figure 1** - Gene map of the *Araucaria angustifolia* chloroplast genome. Genes drawn inside the circle are transcribed counterclockwise and those out-sides are clockwise. Genes belonging to different functional groups are indicated by different tonalities. The darker gray in the inner circle corresponds to the GC content, while the lighter gray corresponds to the AT content. Red and blue arrows indicate, respectively, double inverted copies of the *rrm5* and *trnI-CAU* genes.

genes have only one intron each. The *trnK-UUU* intron has 2,407 bp, with the largest intron encompassing also the *matK* gene, a common feature of land-plants chloroplast genomes.

A phylogenetic analysis was performed to evaluate the position of *A. angustifolia* in the Araucariaceae family

and subclass Pinidae, and 73 protein-coding genes from other 18 conifers were used for this purpose. The final alignment reached 55,435 nucleotides. These species were intentionally sampled to comprise the main representative taxa of the subclass Pinidae, while *Ginkgo biloba* was used as outgroup. The Bayesian analysis resulted in a consistent



**Table 1** - Summary of the *Araucaria angustifolia* chloroplast genome features.

Feature	<i>Araucaria angustifolia</i>
Total cpDNA size	146,203 bp
Protein coding regions (%)	45.02%
rRNA and tRNA (%)	4.96%
Introns (% total)	8.3%
Intergenic sequences and pseudogenes (%)	41.72%
Number of genes	122
Number of protein coding genes	80
Number of tRNA genes	36
Number of rRNA genes	4
Number of duplicated genes	2
Pseudogenes	1
GC content (%)	36.54%

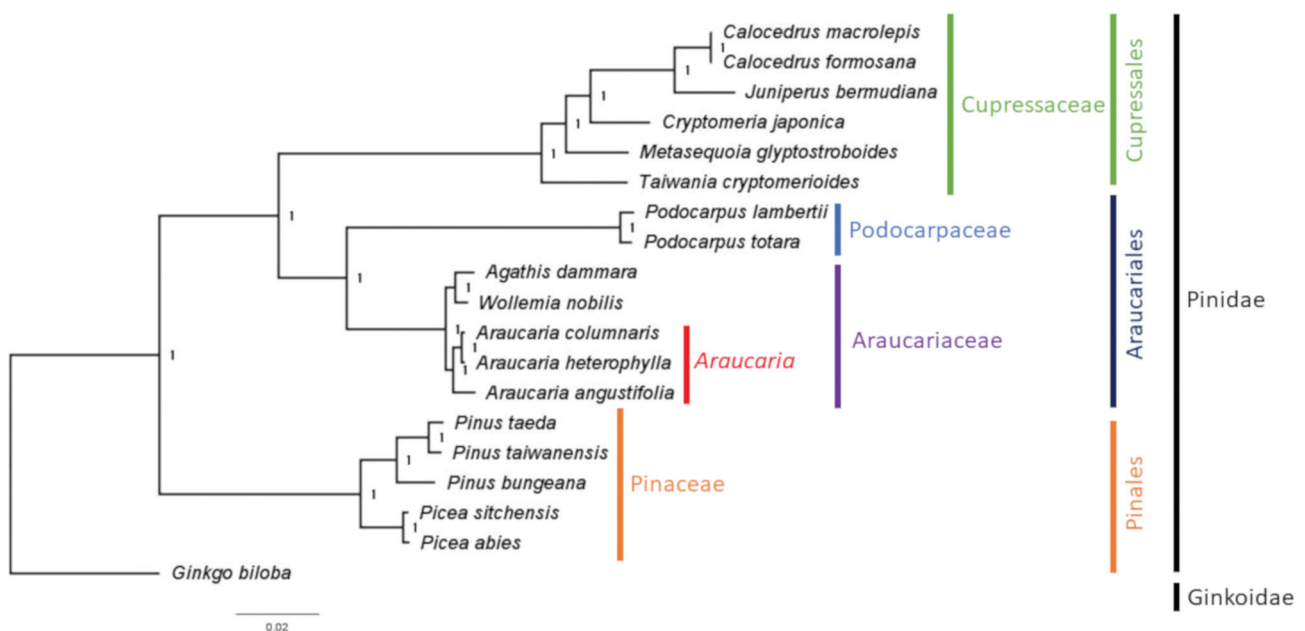
phylogenetic relationship of *A. angustifolia* and the 18 conifer represented species (Figure 2). Within Pinidae we found, two branches, the first one represented by the family Pinaceae, order Pinales, and the other by the families Cupressaceae, Podocarpaceae, Araucariaceae comprising Conifer I and Conifer II, respectively (Lu *et al.*, 2014). Within Conifer II, the genera and species distribution is clear and well supported among Araucariaceae and Podocarpaceae, comprising the order Araucariales and Cupressaceae and comprising the order Cupressales (Lu *et al.*, 2014). Within Araucariaceae, *Araucaria angustifolia*, *Araucaria heterophylla*, and *Araucaria columnaris* were grouped together in the genus *Araucaria*, which grouped

with *Agathis dammara* and *Wollemia nobilis*. In the monophyletic *Araucaria* clade, *Araucaria columnaris* and *Araucaria heterophylla*, representing endemic species from New Caledonia and Norfolk Island, Australia, respectively, are grouped together, and another basal branch corresponds to *Araucaria angustifolia*, endemic from South America (Lu *et al.*, 2014). The strongly supported topology within the genus *Araucaria*, family Araucariaceae, and among the other taxa (Cupressaceae, Podocarpaceae, and Pinaceae) is congruent with a series of phylogenetic studies (Lu *et al.*, 2014). The ML analysis corroborated the Bayesian approach (Figure S1), which strongly reinforces the importance of cpDNA for phylogenetic inference.

Using the MISA perl script, 100 simple sequence repeats (SSRs) were detected in *A. angustifolia* cpDNA. The 53 homopolymers A/T and 24 dipolymers AT were the most common SSRs, while 14 different tetrapolymers and a single hexapolymer were also found (Table S5). SSR pentapolymers were not present in the cpDNA. The present *A. angustifolia* chloroplast genome is the first complete cpDNA sequence for this species and shows a set of features that could be further explored for population and phylogenetic studies within this group. Moreover, the present study increases the genetic and genomic resources available in *Araucaria* and shows that, as reported in bryophytes and angiosperms (Shi *et al.*, 2016), the plastome sequence can be straightforwardly assembled from transcriptome data generated for conifers.

## Acknowledgments

This study was carried out with financial support from the Conselho Nacional de Desenvolvimento Cien-

**Figure 2** - Phylogenetic tree of 18 species of Pinidae based on 73 cp protein-coding genes generated by Bayesian method. Numbers above each node are posterior probability values. Family and order are also indicated. *Ginkgo biloba* was used as outgroup.

tífico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Finance code 001) and Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS).

### Conflict of Interests

The authors declare that they do not have any conflict of interest

### Author contributions

JHSGB and ME performed RNA extraction, PCR reactions and other experiments; NFR, FG, ME, JHSGB and RM performed bioinformatics and statistical analyses of data; JHSGB and RM wrote the manuscript. All authors read and contributed to the final version of the manuscript; RM coordinate the project and findings.

### References

- Asaf S, Khan AL, Khan MA, Shahzad R, Lubna, Kang SM, Al-Harrasi A, Al-Rawahi A and Lee I-J (2018) Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *PLoS One* 13:e0192966.
- Astarita LV, Floh EIS and Handro W (2004) Free amino acid, protein and water content changes associated with seed development in *Araucaria angustifolia*. *Biol Plant* 47:53–59.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
- de Souza MIF, Salgueiro F, Carnavale-Bottino M, Félix DB, Alves-Ferreira M, Bittencourt JVM and Margis R (2009) Patterns of genetic diversity in southern and southeastern *Araucaria angustifolia* (Bert.) O. Kuntze relict populations. *Genet Mol Biol* 32:546–556.
- Elbl P, Lira BS, Andrade SCS, Jo L, dos Santos ALW, Coutinho LL, Floh EIS and Rossi M (2015) Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, *Araucaria angustifolia* (Bertol.) Kuntze. *Plant Cell Tissue Organ Cult* 120:903–915.
- Laslett D and Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16.
- Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Lohse M, Drechsel O, Kahlau S and Bock R (2013) OrganelleGenomeDRAW — a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41:W575–W581.
- Lowe AJ, Breed MF, Caron H, Colpaert N, Dick C, Finegan B, Gardner M, Gheysen G, Gribel R, Harris JBC *et al.* (2018) Standardized genetic diversity-life history correlates for improved genetic resource management of Neotropical trees. *Divers Distrib* 24:730–741.
- Lu Y, Ran JH, Guo DM, Yang ZY and Wang XQ (2014) Phylogeny and divergence times of Gymnosperms inferred from single-copy nuclear genes. *PLoS One* 9:e107679.
- Posada D and Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA and Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61: 539–542.
- Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, Mei W, Thomas PI, Gardner MF, Ennos RA *et al.* (2015) Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour* 15:1067–1078.
- Shi C, Wang S, Xia EH, Jiang JJ, Zeng FC and Gao LZ (2016) Full transcription of the chloroplast genome in photosynthetic eukaryotes. *Sci Rep* 6:30135.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM and Birol I (2009) ABySS: A parallel assembler for short read sequence data. *Genome Res* 19:1117–1123.
- Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Stefenon VM, Gailing O and Finkeldey R (2006) Phylogenetic relationship within genus *Araucaria* (Araucariaceae) assessed by means of AFLP fingerprints. *Silvae Genet* 55:45–52.
- Steiner N, Catarina CS, Balbuena TS and Guerra MP (2008) *Araucaria angustifolia* Biotechnology. *Funct Plant Sci Biotechnol* 2:20–28.
- Tamura K, Stecher G, Peterson D, Filipowski A and Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729.
- Thomas P (2013) *Araucaria angustifolia*. IUCN Red List Threat Species 2013 eT32975A2829141.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R and Greiner S (2017) GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45:W6–W11.
- Wu CS and Chaw SM (2014) Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): Evolution towards shorter intergenic spacers. *Plant Biotechnol J* 12:344–353.
- Wu CS, Wang YN, Hsu CY, Lin CP and Chaw SM (2011) Loss of different inverted repeat copies from the chloroplast genomes of pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol Evol* 3:1284–1295.
- Yap JYS, Rohner T, Greenfield A, Van Der Merwe M, McPherson H, Glenn W, Kornfeld G, Marendy E, Pan AYH, Wilton A *et al.* (2015) Complete chloroplast genome of the Wollemi Pine (*Wollemia nobilis*): Structure and evolution. *PLoS One* 10:e0128126.

### Internet Resources

IUCN, Red List of Threatened Species, <http://www.iucnredlist.org/> (accessed May 27, 2018)

### Supplementary Material

The following online material is available for this article:

Figure S1 - Phylogenetic tree of 18 species of Pinidae based on 73 cp protein-coding genes generated using Maximum Likelihood.

Table S1 - List of 58 Pinidae complete chloroplast genomes used for mapping plastid reads from *Araucaria angustifolia*.

Table S2 - List of 73 chloroplast protein-coding genes used in the phylogenetic analysis.

Table S3 - List of 18 plastome sequences of Pinidae included in the Bayesian and Maximum Likelihood phylogenetic analysis.

Table S4 - Comparison of cpDNA characteristics in different species from Araucariaceae

Table S5 - List of simple sequence repeats (SSRs) identified in the *Araucaria angustifolia* chloroplast genome.

*Associate Editor: Ana Tereza R. Vasconcelos*

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.

## 5. DISCUSSÃO E CONSIDERAÇÕES FINAIS

A espécie alvo do presente trabalho, *Araucaria angustifolia*, é uma conífera arbórea endêmica do território brasileiro, com alta relevância evolutiva, ecológica, econômica e histórico-cultural. Essa espécie atualmente enfrenta grave risco de extinção e inspira cuidados e iniciativas de manejo e conservação. Algumas características intrínsecas como, por exemplo, sementes sazonais, recalcitrantes e intolerantes ao estoque, e longo ciclo reprodutivo, dificultam iniciativas de reflorestamento, manejo e preservação de *A. angustifolia*. Essa problemática despertou o interesse da comunidade científica em investigar essa espécie sob diversos aspectos. Do ponto de vista genético, diversos estudos foram desenvolvidos com foco principalmente em diversidade populacional e embriogênese somática. Os dados gerados por esses trabalhos apontam uma série de questões relevantes, como a caracterização da estrutura genética de populações naturais fragmentadas (Auler et al. 2002), modificações bioquímicas e parâmetros fisiológicos durante a embriogênese zigótica (Astarita et al. 2003b; Astarita et al. 2003a; Silveira et al. 2008), modificações morfofisiológicas durante a cinética de crescimento de amostras em suspensão celular (Santos et al. 2010), caracterização proteômica em etapas da embriogênese somática (Steiner et al. 2008) e análise transcriptômica comparativa em etapas iniciais do desenvolvimento de embriões somáticos (Elbl et al. 2015). Porém, assim como em muitas espécies da família Araucariaceae, vários aspectos genéticos permanecem inexplorados em *A. angustifolia*. Dada a sua importância econômica, ecológica e cultural, além da sua grave situação de risco de extinção, é imprescindível que novos esforços sejam feitos em pesquisa básica a fim acessar novos ramos da biologia dessa espécie.

O presente trabalho, dividido em dois artigos, traz novos dados com relação a aspectos genéticos de *A. angustifolia*. O primeiro artigo focou na predição de miRNAs, comparação dos mesmos com o banco de dados do miRBase, comparação com dados de RNA-seq da espécie congênica *A. cunninghamii* e sRNA-seq de outras seis espécies de gimnospermas; análise de expressão diferencial dos miRNAs preditos via RT-qPCR, em diferentes tecidos; e predição de alvos dos miRNAs novos e conservados. A partir do mesmo input de RNA-seq, o segundo artigo focou-se na montagem e anotação do genoma de cloroplasto (cpDNA) de *A. angustifolia*.

O primeiro artigo realizou uma caracterização e análise comparativa de miRNAs em *A. angustifolia*. Nesse estudo, uma biblioteca de sRNA-seq de folhas e 24 bibliotecas públicas de RNA-seq, provenientes de estágios iniciais de embriões somáticos, foram utilizadas para a predição de precursores de miRNAs (pre-miRNAs) e seus respectivos miRNAs maduros. Essa análise foi feita com o software miR-PREFeR, o qual apresenta alta acurácia e baixa frequência de falsos positivos na predição de miRNAs (Lei and Sun 2014). Um estudo recente mostrou que uma parcela significativa dos miRNAs de plantas depositados no miRBase é questionável em relação a vários critérios de anotação, podendo representar falsos positivos (Axtell and Meyers 2018). A metodologia adotada no presente trabalho foi executada em concordância com critérios atuais de anotação de miRNAs de plantas (Taylor et al. 2014; Axtell and Meyers 2018).

As sequências de miRNAs maduros foram comparadas com os dados do miRBase via BLAST. Nessa análise *mismatches* não foram considerados. Seguindo esses critérios, 41 miRNAs conservados de 16 famílias diferentes foram identificados. Destes Aang-miR156, Aang-miR159, Aang-miR160, Aang-miR166, Aang-miR171, Aang-miR390, Aang-miR195, Aang-miR396 e Aang-miR529 aparecem em todos os grupos de plantas terrestres (Taylor et al. 2014), Aang-miR168 e Aang-miR169 são específicos de traqueófitas (Berruezo et al. 2017) e miR1314 ocorre apenas em gimnospermas (Chávez Montes et al. 2014). Sessenta e cinco miRNAs preditos não apresentaram identidade de sequência com os dados do miRBase e, portanto, foram classificados como possíveis novos miRNAs em *A. angustifolia* (Aang-nmiRs). Este é o padrão clássico adotado para identificação de miRNAs conservados e novos nos mais diversos grupos vegetais (Wang 2004; Mica 2006; Chen et al. 2011; Guzman et al. 2012; Hao et al. 2012; Zhang et al. 2013; Chao et al. 2014; Bologna and Voinnet 2014; Chai et al. 2015; Zhang et al. 2015b; Velayudha Vimala Kumar et al. 2017; Berruezo et al. 2017; Li et al. 2017c; He et al. 2018). Porém, conforme apontado por (Axtell and Meyers 2018), o banco de dados do miRBase levou mais de três anos sem receber atualizações, e mesmo após o seu *upgrade* para a versão 22 em março deste ano, uma série de miRNAs de vários grupos não consta nesse repositório. Essa lacuna é proeminente em gimnospermas. O miRBase contém acessos de miRNAs de apenas quatro espécies, *Picea abies*, *Pinus densata*, *Pinus taeda* e *Cunninghamia lanceolata*. Porém, existem dados públicos de miRNAs anotados em várias outras gimnospermas, como *Ginkgo biloba* (Zhang et al. 2015b), *Taxus mairei* (Hao et al. 2012), *Taxus wallichiana* (Zhang et al. 2015a), *Sequoia sempervirens* (Chen et al. 2013), *Larix leptolepis* (Liu et al. 2016), *Picea balfouriana* (Li et al. 2017b), *Picea*

*glauca* (Liu and El-Kassaby 2017), entre outros. Além das comparações com o miRBase, bibliotecas de pequenos RNAs de outras cinco espécies de gimnospermas foram alinhadas com as sequências dos 106 pre-miRNAs de *A. angustifolia*. Esta análise indicou que os miRNAs conservados preditos em *A. angustifolia* apresentam sequências correspondentes em *Cunninghamia lanceolata*, *Taxus mairei*, *Taxus wallichiana*, *Picea abies* e *Ginkgo biloba*. O mesmo ocorreu com 9 dos 65 miRNAs novos, Aang-nmiR006, Aang-nmiR014, Aang-nmiR025, Aang-nmiR028, Aang-nmiR031, Aang-nmiR034, Aang-nmiR046, Aang-nmiR055 e Aang-nmiR063. Isso indica que, possivelmente, esses miRNAs também apresentam ortólogos em outras espécies de gimnospermas, e surgiram nas linhagens que originaram esse grupo.

O presente trabalho é pioneiro na identificação de miRNAs em uma espécie da família Araucariaceae e, conseqüentemente, do gênero *Araucaria*. Este gênero apresenta 19 espécies inteiramente distribuídas no hemisfério sul, com 17 espécies presentes na Oceania e duas espécies na América do sul. Os estudos genéticos no gênero concentram-se principalmente em aspectos filogenéticos e filogeográficos (Auler et al. 2002; Escapa and Catalano 2013; Lu et al. 2014; Ruhsam et al. 2015). Portanto, a identificação de miRNAs em *A. angustifolia* representa uma importante iniciativa para estudos relacionando os mecanismos de regulação da expressão gênica mediada por pequenos RNAs no gênero *Araucaria* e incentiva novas ações investigativas neste aspecto. A partir de buscas no NCBI por dados de sequenciamento de espécies do gênero *Araucaria*, foram localizadas bibliotecas de RNA-seq de *A. cunninghamii*, uma araucária endêmica da Austrália. Essas bibliotecas foram baixadas do NCBI e usadas como *input* para montagem do transcriptoma de referência, o qual teoricamente contém sequências provenientes de transcritos de genes *MIR*. Em seguida, as sequências dos 106 pre-miRNAs de *A. angustifolia* foram comparadas com o transcriptoma de *A. cunninghamii* via BLAST, e apenas sequências de *A. cunninghamii* alinhadas em toda a extensão dos pre-miRNAs de *A. angustifolia* foram consideradas. Por fim, os pre-miRNAs candidatos de *A. cunninghamii* (Acu pre-miRNAs) foram submetidas a análises de *foldng* para visualização das estruturas em grampo (*hairpins*). Dessa forma, precursores da família miR1314, específica de coníferas, e sete novos precursores identificados em *A. angustifolia* foram também identificados em *A. cunninghamii*, e classificados como possíveis ortólogos. A identidade de sequência entre os precursores de *A. angustifolia* e os seus possíveis ortólogos em *A. cunninghamii* foi alta, atingindo 96% entre Aang-nmiR003 e Acu-nmiR003, com mismatches ocorrendo principalmente na região do loop e fora das sequências dos

miRNAs maduros. Estudos recentes apontam padrões similares em diversos miRNAs conservados, quando suas sequências são comparados entre diferentes grupos de Angiospermas (Miskiewicz et al. 2017; Chorostecki et al. 2017).

O alvos preditos para miRNAs novos e conservados foram categorizados em processos biológicos e funções moleculares por análises computacionais de ontologia gênica (Conesa and Götz 2008). Diversos alvos de miRNAs conservados em *A. angustifolia* foram identificados em outros grupos vegetais sob a mesma combinação miRNA-alvo. Exemplos incluem miR171- scarecrow 6 (Zhu et al. 2015), miRNA395- ATP sulfurylase (Jagadeeswaran et al. 2012), miR167-Zinc finger ZAT (Omidvar et al. 2015), miR166- HD-ZIPIII (Li et al. 2017c) e miRNA159-GAMYB (Velayudha Vimala Kumar et al. 2017). Uma série de alvos de miRNAs novos identificados em *A. angustifolia* correspondem a genes de resistência que codificam proteínas com domínio NBS-LRR. Evidências recentes apontam o papel evolutivo dos genes NBS-LRR em direcionar o surgimento de novos miRNAs a partir de eventos de duplicação invertida (González et al. 2015; Zhang et al. 2016). Dessa forma, os novos miRNAs, por apresentarem regiões complementares, assumem um papel regulatório sobre os próprios genes parentais (González et al. 2015; Zhang et al. 2016).

Os miRNAs estão envolvidos em uma série de processos metabólicos em diferentes etapas do desenvolvimento (Chávez Montes et al. 2014; Samad et al. 2017). A análise de expressão diferencial de miRNAs pode trazer importantes insights acerca das funções regulatórias desses pequenos RNAs em diferentes condições, períodos de desenvolvimento e tipos de tecido (Yakovlev et al. 2010; Kantar et al. 2011; Zhang et al. 2011; Bai et al. 2018; Pegler et al. 2018). Entre os miRNAs preditos em *A. angustifolia*, 12 miRNAs conservados e 30 miRNAs novos foram selecionados para análise de expressão diferencial em cinco tecidos de plantas jovens, utilizando-se a técnica de stem-loop RT-qPCR. Entre os miRNAs conservados diferencialmente expressos, Aang-miR167 foi mais expresso em folhas jovens, Aang-miR399 foi mais expresso em folhas maduras, Aang-miR166, Aang-miR168, Aang-miR390 e Aang-miR395 foram mais expressos no caule; Aang-miR156, Aang-miR159 e Aang-miR1314 foram mais expressos nas raízes adventícias. Entre os miRNAs novos diferencialmente expressos, 14 miRNAs foram mais expressos nas raízes adventícias, com padrões mais acentuados em Aang-nmiR005 e Aang-nmiR025; Aang-nmiR051 foi mais expresso na raiz principal; Aang-nmiR003, Aang-nmiR021 e Aang-nmiR061 foram mais expressos no caule; e Aang-nmiR029 e Aang-nmiR038 foram mais

expressos nas folhas. O padrão tecido-específico dos miRNAs tem sido comprovado em várias espécies (Mica 2006; Lu et al. 2007; Chen et al. 2011; Wan et al. 2012a; Zhang et al. 2015b; Xu et al. 2016; Liu et al. 2016; Li et al. 2017b; Rodrigues et al. 2017).

Uma série de pequenos RNAs conservados em embriófitas, traqueófitas e gimnospermas estão presentes em *A. angustifolia* e regulando alvos também conservados nos respectivos grupos. Os pre-miRNAs conservados de *A. angustifolia* apresentam tamanho e estrutura secundária similar aos encontrados em outras espécies. A análise de expressão diferencial via RT-qPCR mostrou que os miRNAs novos e conservados de *A. angustifolia* apresentam padrões variados de expressão os tecidos investigados, indicando a complexidade das vias regulatórias por parte desses pequenos RNAs.

No segundo trabalho, as bibliotecas de RNA-seq obtidas do NCBI e utilizadas nas análises de predição de miRNAs e alvos foram utilizadas para montagem e anotação do genoma de cloroplasto de *A. angustifolia*. A montagem de genoma do cloroplasto a partir de dados de RNA-seq foi documentada em diferentes espécies de algas e plantas terrestres, incluindo briófitas e angiospermas (Shi et al. 2016; Horvath et al. 2018). Algumas características peculiares do genoma de cloroplasto como elevado número de cópias por célula (Golczyk et al. 2014; Morley and Nielsen 2016), transcrição completa do cpDNA em RNAs, incluindo regiões espaçadoras intergênicas (Shi et al. 2016), níveis de expressão elevados (Yagi and Shiina 2014; Zoschke and Bock 2018), e alta frequência de AT (Fang et al. 2015; Yap et al. 2015; Chen et al. 2015; Coombe et al. 2016; Zoschke and Bock 2018) favorecem o seu enriquecimento durante a preparação de bibliotecas de RNA-seq, frequentemente correspondendo a mais de 20% do total de *reads* (Smith 2013). Dessa forma, com a disponibilidade de 64GB de *reads* derivados do sequenciamento Illumina de bibliotecas de RNA-seq de amostras de *A. angustifolia*, este estudo baseou-se nas características supracitadas do cpDNA para montar e anotar o genoma cloroplastidial de *A. angustifolia* a partir de dados transcriptômicos.

Desde a última década, uma série de genomas de cloroplasto de espécies de gimnospermas tem sido estudada (Hirao et al. 2008; Lin et al. 2012; Fang et al. 2015; Yap et al. 2015; Chen et al. 2015; Chen et al. 2015; Coombe et al. 2016; Wang et al. 2018). Na família Araucariaceae os plastomas de *Wollemia nobilis* e *Agathis tortara* foram sequenciados e anotados (Vieira et al. 2014; Yap et al. 2015). No gênero *Araucaria*, 15 espécies da Oceania foram utilizadas em estudos



filogenéticos com base em suas sequências completas de cpDNA (Ruhsam et al. 2015), porém não constam publicações descritivas desses plastomas. Apenas sequências completas e anotações do cpDNA de *A. heterophylla* (NC\_026450) e *A. columnaris* (KM678417) estão disponibilizadas no NCBI. Portanto, a montagem e anotação do genoma cloroplastidial de *A. angustifolia* representa uma importante contribuição em recursos genéticos e genômicos na espécie, no gênero *Araucaria* e na família Araucariaceae.

As análises bioinformáticas aplicadas na montagem do cpDNA de *A. angustifolia* incluem: *download* das bibliotecas *paired-end* de RNA-seq do NCBI; filtragem dos reads utilizando-se 58 genomas completos de diferentes espécies de coníferas (subclasse Pinidae); montagem *de novo* do cpDNA com software ABySS utilizando como input os *reads* filtrados provenientes de transcritos de cpDNA. Seguindo essa estratégia, o número total de reads de alta qualidade derivados do cpDNA foi de 2.107.993. Os *contigs* montados a partir desses *reads* apresentaram valores de cobertura entre 23 e 1.780, e foram suficientes para cobrir 99,65% do cpDNA de *Araucaria*. O sequenciamento Sanger foi utilizado para preencher um *gap* referente a uma região intergênica.

O cpDNA completo de *A. angustifolia* apresenta 146.203 pares de bases, tamanho semelhante aos encontrados em outras espécies do gênero (Ruhsam et al. 2015) e da família Araucariaceae (Wu and Chaw 2014; Yap et al. 2015). A estrutura típica encontrada em cpDNAs de plantas terrestres, cópia única longa (LSC) e cópia única curta (SSC) separadas por cópias invertidas repetidas (IRA e IRb) não existe no cpDNA de *A. angustifolia*. Ao invés disso, apenas dois genes apresentam duplicações invertidas repetidas, *rrn5* e *tRNA-CAU*. O mesmo padrão foi descrito no cpDNA de espécies da família Araucariaceae, *Agathis damara* (Wu and Chaw 2014) e *Wollemia nobilis* (Yap et al. 2015). A ausência ou encurtamento das IRs foi atribuída em genomas plastidiais de diversas gimnospermas (Hirao et al. 2008; Wu et al. 2011; Vieira et al. 2014; Fang et al. 2015; Li et al. 2015; Yap et al. 2015; Chen et al. 2015; Asaf et al. 2018). Além do comprimento e estrutura, o cpDNA de *A. angustifolia* é similar aos de outras espécies do gênero e da família também em relação ao número de genes. *A. angustifolia* contém 122 genes, dos quais 80 codificam proteínas, 5 codificam RNAs ribossomais e 37 codificam RNAs de transferência. Esses números são, respectivamente, 124, 82, 5 e 37 em *Araucaria columnaris* (KM678417); 124, 82, 5 e 37 em *Araucaria heterophylla* (NC\_026450.1); 122, 82, 5 e 35 em *Wollemia nobilis* (Yap et al. 2015); e 123, 82, 5 e 36 em *Agathis damara* (Wu and Chaw 2014).

Duas abordagens foram empregadas para avaliar a posição filogenética de *A. angustifolia* em relação a 18 espécies representantes de diferentes grupos de gimnospermas, incluindo família Araucariaceae e gênero *Araucaria*, com base no alinhamento concatenado de 73 genes codificantes de proteínas, totalizando 55.453 nucleotídeos. Estas abordagens, análise bayesiana e máxima verossimilhança, produziram topologias iguais e com alto suporte estatístico. Em ambas as árvores *A. angustifolia* apareceu agrupada com *A. heterophylla* e *A. columnaris*, compondo o clado monofilético referente ao gênero *Araucaria*, o qual, por sua vez, agrupou-se com *Wollemia nobilis* e *Agathis damara*, compondo o clado monofilético referente à família Araucariaceae. Os demais agrupamentos foram congruentes com diversos estudos filogenéticos baseados em sequências de cpDNA com representantes dos mesmos grupos (Wu and Chaw 2014; Yi et al. 2015; Chen et al. 2015; Sudianto et al. 2016; Tao et al. 2016).

Por fim, análises bioinformáticas foram empregadas na identificação de microssatélites no cpDNA de *A. angustifolia*. Os microssatélites ou *simple sequence repeats* (SSR) são sequências repetitivas de um a seis pares de bases amplamente distribuídas no genoma (Vieira et al. 2016). Os SSRs geralmente são co-dominantes e apresentam alto grau de polimorfismos (Choudhary and Trivedi 2010). Por conta disso são excelentes marcadores moleculares muito utilizados em estudos de diversidade populacional (Dean et al. 2015), mapeamento genético (Mahfooz et al. 2016), entre outros. Com base na análise de predição de SSRs, 100 microssatélites foram identificados no cpDNA de *A. angustifolia*, dos quais 53 são mononucleotídeos A/T, 24 são dinucleotídeos AT, e os demais compreendem tetranucleotídeos e hexanucleotídeos. Esses dados são consistentes com o alto percentual de SSRs formados por curtas repetições Poli-A e Poli-T nos genomas de cloroplasto de vários grupos (Kuang et al. 2011; Vieira et al. 2014; Ozyigit et al. 2015; Yap et al. 2015; Chen et al. 2015). Os dados obtidos nesse trabalho com relação ao cpDNA de *A. angustifolia* apresentam subsídios que podem ser explorados em estudos de diversidade populacional, filogenéticos, entre outros.

## 6 REFERÊNCIAS

Agapito-Tenfen SZ, Steiner N, Guerra MP and Nodari RO (2011) Patterns of polyembryony and frequency of surviving multiple embryos of the Brazilian pine *Araucaria angustifolia*. *Aust J Bot* 59:749–755.

Alptekin B, Akpinar BA and Budak H (2017) A Comprehensive Prescription for Plant miRNA Identification. *Front Plant Sci* 7:1–28. doi: 10.3389/fpls.2016.02058

Andersen CL, Jensen JL and Ørntoft TF (2004) Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. *Cancer Res* 64:5245–5250. doi: 10.1158/0008-5472.CAN-04-0496

Asaf S, Khan AL, Khan MA, Shahzad R, Lubna, Kang SM, Al-Harrasi A, Al-Rawahi A and Lee I-J (2018) Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *PLoS One* 13:e0192966. doi: 10.1371/journal.pone.0192966

Astarita LV, Handro W and Floh EIS (2003a) Changes in polyamines content associated with zygotic embryogenesis in the Brazilian pine, *Araucaria angustifolia* (Bert.) O. Ktze. *Rev Bras Botânica* 26:163–168. doi: 10.1590/S0100-84042003000200003

Astarita L V., Floh EIS and Handro W (2003b) Changes in IAA, tryptophan and activity of soluble peroxidase associated with zygotic embryogenesis in *Araucaria angustifolia* (Brazilian pine). *Plant Growth Regul* 39:113–118. doi: 10.1023/A:1022542618945

Auler NMF, Reis MS dos, Guerra MP and Nodari RO (2002) The genetics and conservation of *Araucaria angustifolia*: I. Genetic structure and diversity of natural populations by means of non-adaptive variation in the state of Santa Catarina, Brazil. *Genet Mol Biol* 25:329–338. doi: 10.1590/S1415-47572002000300014

Axtell MJ and Meyers BC (2018) Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. *Plant Cell* 30:272–284. doi: 10.1105/tpc.17.00851

Bai Q, Wang X, Chen X, Shi G, Liu Z, Guo C and Xiao K (2018) Wheat miRNA TaemiR408 Acts as an Essential Mediator in Plant Tolerance to Pi Deprivation and Salt Stress via Modulating

Stress-Associated Physiological Processes. *Front Plant Sci* 9:1–17. doi: 10.3389/fpls.2018.00499

Barbier M, Bottelbergs A, Nuydens R, Ebneith A and De Vos WH (2018) SliceMap: an algorithm for automated brain region annotation. *Bioinformatics* 34:718–720. doi: 10.1093/bioinformatics/btx658

Behjati S and Tarpey PS (2013) What is next generation sequencing? *Arch Dis Child - Educ Pract Ed* 98:236–238. doi: 10.1136/archdischild-2013-304340

Berruezo F, de Souza FSJ, Picca PI, Nemirovsky SI, Martínez Tosar L, Rivero M, Mentaberry AN and Zelada AM (2017) Sequencing of small RNAs of the fern *Pleopeltis minima* (Polypodiaceae) offers insight into the evolution of the microRNA repertoire in land plants. *PLoS One* 12:e0177573. doi: 10.1371/journal.pone.0177573

Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL and Trees E (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin Microbiol Infect* 24:335–341. doi: 10.1016/j.cmi.2017.10.013

Bologna NG, Iselin R, Abriata LA, Sarazin A, Pumplin N, Jay F, Grentzinger T, Dal Peraro M and Voinnet O (2018) Nucleo-cytosolic Shuttling of ARGONAUTE1 Prompts a Revised Model of the Plant MicroRNA Pathway. *Mol Cell* 69:709–719.e5. doi: 10.1016/j.molcel.2018.01.007

Bologna NG and Voinnet O (2014) The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis. *Annu Rev Plant Biol* 65:473–503. doi: 10.1146/annurev-arplant-050213-035728

Borges F and Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* 16:727–741. doi: 10.1038/nrm4085

Budak H and Akpinar BA (2015) Plant miRNAs: biogenesis, organization and origins. *Funct Integr Genomics* 15:523–531. doi: 10.1007/s10142-015-0451-2

Buermans HPJ and den Dunnen JT (2014) Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta - Mol Basis Dis* 1842:1932–1941. doi: 10.1016/j.bbadis.2014.06.015

Chai J, Feng R, Shi H, Ren M, Zhang Y and Wang J (2015) Bioinformatic Identification and

Expression Analysis of Banana MicroRNAs and Their Targets. PLoS One 10:e0123083. doi: 10.1371/journal.pone.0123083

Chao Y-T, Su C-L, Jean W-H, Chen W-C, Chang Y-CA and Shih M-C (2014) Identification and characterization of the microRNA transcriptome of a moth orchid *Phalaenopsis aphrodite*. Plant Mol Biol 84:529–548. doi: 10.1007/s11103-013-0150-0

Chávez Montes R a, Rosas-Cárdenas de FF, De Paoli E, Accerbi M, Rymarquis L a, Mahalingam G, Marsch-Martínez N, Meyers BC, Green PJ and de Folter S (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun 5:3722. doi: 10.1038/ncomms4722

Chaw S-M, Wu C-S and Sudianto E (2018) Evolution of Gymnosperm Plastid Genomes. Advances in Botanical Research, 1st ed. Elsevier Ltd., pp 195–222

Chen C-J, Liu Q, Zhang Y-C, Qu L-H, Chen Y-Q and Gautheret D (2011) Genome-wide discovery and analysis of microRNAs and other small RNAs from rice embryogenic callus. RNA Biol 8:538–547. doi: 10.4161/rna.8.3.15199

Chen C (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33:e179–e179. doi: 10.1093/nar/gni178

Chen J, Hao Z, Xu H, Yang L, Liu G, Sheng Y, Zheng C, Zheng W, Cheng T and Shi J (2015) The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. Front Plant Sci 6:1–11. doi: 10.3389/fpls.2015.00447

Chen YT, Shen CH, Lin WD, Chu HA, Huang BL, Kuo CI, Yeh KW, Huang LC and Chang IF (2013) Small RNAs of *Sequoia sempervirens* during rejuvenation and phase change. Plant Biol 15:27–36. doi: 10.1111/j.1438-8677.2012.00622.x

Chorostecki U, Moro B, Rojas ALM, Debernardi JM, Schapire AL, Notredame C and Palatnik J (2017) Evolutionary Footprints Reveal Insights into Plant MicroRNA Biogenesis. Plant Cell 29:tpc.00272.2017. doi: 10.1105/tpc.17.00272

Choudhary OP and Trivedi S (2010) Microsatellite or Simple Sequence Repeat ( SSR ) instability depends on repeat characteristics during replication and repair. J Cell Mol Biol 8:21–34.

- Conesa A and Götz S (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int J Plant Genomics* 2008:1–12. doi: 10.1155/2008/619832
- Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczęśniak MW, Gaffney DJ, Elo LL, Zhang X et al. (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13. doi: 10.1186/s13059-016-0881-8
- Coombe L, Warren RL, Jackman SD, Yang C, Vandervalk BP, Moore RA, Pleasance S, Coope RJ, Bohlmann J, Holt RA et al. (2016) Assembly of the Complete Sitka Spruce Chloroplast Genome Using 10X Genomics' GemCode Sequencing Data. *PLoS One* 11:e0163059. doi: 10.1371/journal.pone.0163059
- D'Ario M, Griffiths-Jones S and Kim M (2017) Small RNAs: Big Impact on Plant Development. *Trends Plant Sci* 22:1056–1068. doi: 10.1016/j.tplants.2017.09.009
- Dai X and Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39:W155–W159. doi: 10.1093/nar/gkr319
- Daniell H, Lin C-S, Yu M and Chang W-J (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:134. doi: 10.1186/s13059-016-1004-2
- Das A, Chaudhury S, Kalita MC and Mondal TK (2015) In silico identification, characterization and expression analysis of miRNAs in *Cannabis sativa* L. *Plant Gene* 2:17–24. doi: 10.1016/j.plgene.2015.03.003
- Dean D, Wadl PA, Hadziabdic D, Klingeman WE, Ownley BH, Rinehart TA, Dattilo AJ, Scheffler B and Trigiano RN (2015) Analysis of genetic diversity and population structure for the native tree *Viburnum rufidulum* occurring in Kentucky and Tennessee. *J Am Soc Hortic Sci* 140:523–531.
- Du S, Sang Y, Liu X, Xing S, Li J, Tang H and Sun L (2016) Transcriptome Profile Analysis from Different Sex Types of *Ginkgo biloba* L. *Front Plant Sci* 7:1–10. doi: 10.3389/fpls.2016.00871
- Duan D, Jia Y, Yang J and Li Z-H (2017) Comparative Transcriptome Analysis of Male and Female Conelets and Development of Microsatellite Markers in *Pinus bungeana*, an Endemic Conifer in China. *Genes (Basel)* 8:393. doi: 10.3390/genes8120393

- Elbl P, Lira BS, Andrade SCS, Jo L, dos Santos ALW, Coutinho LL, Floh EIS and Rossi M (2015) Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, *Araucaria angustifolia* (Bertol.) Kuntze. *Plant Cell, Tissue Organ Cult* 120:903–915. doi: 10.1007/s11240-014-0523-3
- Escapa IH and Catalano SA (2013) Phylogenetic Analysis of Araucariaceae: Integrating Molecules, Morphology, and Fossils. *Int J Plant Sci* 174:1153–1170. doi: 10.1086/672369
- Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, Laubinger S, Smith LM, Dasenko M, Givan SA, Weigel D et al. (2010) MicroRNA Gene Evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* 22:1074–1089. doi: 10.1105/tpc.110.073999
- Fang M-F, Wang Y-J, Zu Y-M, Dong W-L, Wang R-N, Deng T-T and Li Z-H (2015) The complete chloroplast genome of the Taiwan red pine *Pinus taiwanensis* (Pinaceae). *Mitochondrial DNA* 27:1–2. doi: 10.3109/19401736.2015.1046169
- Fausto AKS, Silva T da F, Romanel E and Vaslin MFS (2017) microRNAs as reference genes for quantitative PCR in cotton. *PLoS One* 12:e0174722. doi: 10.1371/journal.pone.0174722
- Friedländer MR, Mackowiak SD, Li N, Chen W and Rajewsky N (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* 40:37–52. doi: 10.1093/nar/gkr688
- Golczyk H, Greiner S, Wanner G, Weihe A, Bock R, Börner T and Herrmann RG (2014) Chloroplast DNA in Mature and Senescing Leaves: A Reappraisal. *Plant Cell* 26:847–854. doi: 10.1105/tpc.113.117465
- González VM, Müller S, Baulcombe D and Puigdomènech P (2015) Evolution of NBS-LRR Gene Copies among Dicot Plants and its Regulation by Members of the miR482/2118 Superfamily of miRNAs. *Mol Plant* 8:329–331. doi: 10.1016/j.molp.2014.11.013
- Guzman F, Almerão MP, Körbes AP, Loss-Morais G and Margis R (2012) Identification of MicroRNAs from *Eugenia uniflora* by High-Throughput Sequencing and Bioinformatics Analysis. *PLoS One* 7:e49811. doi: 10.1371/journal.pone.0049811
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the

Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. doi: 10.1038/nprot.2013.084

Hackenberg M, Sturm M, Langenberger D, Falcón-Pérez JM and Aransay AM (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res* 37:W68–W76. doi: 10.1093/nar/gkp347

Han Y, Gao S, Muegge K, Zhang W and Zhou B (2015) Advanced applications of RNA sequencing and challenges. *Bioinform Biol Insights* 9:29–46. doi: 10.4137/BBI.S28991

Hao D-C, Yang L, Xiao P-G and Liu M (2012) Identification of *Taxus* microRNAs and their targets with high-throughput sequencing and degradome analysis. *Physiol Plant* 146:388–403. doi: 10.1111/j.1399-3054.2012.01668.x

Hao Z, Cheng T, Zheng R, Xu H, Zhou Y, Li M, Lu F, Dong Y, Liu X, Chen J et al. (2016) The Complete Chloroplast Genome Sequence of a Relict Conifer *Glyptostrobus pensilis*: Comparative Analysis and Insights into Dynamics of Chloroplast Genome Rearrangement in Cupressophytes and Pinaceae. *PLoS One* 11:e0161809. doi: 10.1371/journal.pone.0161809

He B, Gu Y, Xu M, Wang J, Cao F and Xu L (2015) Transcriptome analysis of *Ginkgo biloba* kernels. *Front Plant Sci* 6:1–10. doi: 10.3389/fpls.2015.00819

He X, Shenkute AG, Wang W and Xu S (2018) Characterization of Conserved and Novel microRNAs in *Lilium lancifolium* Thunb. by High-Throughput Sequencing. *Sci Rep* 8:2880. doi: 10.1038/s41598-018-21193-4

Hirao T, Watanabe A, Kurita M, Kondo T and Takata K (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biol* 8:70. doi: 10.1186/1471-2229-8-70

Horvath DP, Patel S, Dođramaci M, Chao WS, Anderson J V., Foley ME, Scheffler B, Lazo G, Dorn K, Yan C et al. (2018) Gene Space and Transcriptome Assemblies of Leafy Spurge (*Euphorbia esula*) Identify Promoter Sequences, Repetitive Elements, High-Quality Markers, and a Full-Length Chloroplast Genome. *Weed Sci* 66:355–367. doi: 10.1017/wsc.2018.2

Hou Y, Zhai L, Li X, Xue Y, Wang J, Yang P, Cao C, Li H, Cui Y and Bian S (2017) Comparative



Analysis of Fruit Ripening-Related miRNAs and Their Targets in Blueberry Using Small RNA and Degradome Sequencing. *Int J Mol Sci* 18:2767. doi: 10.3390/ijms18122767

Hrdlickova R, Toloue M and Tian B (2017) RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip Rev RNA* 8:e1364. doi: 10.1002/wrna.1364

Hu X-G, Liu H, Jin Y, Sun Y-Q, Li Y, Zhao W, El-Kassaby YA, Wang X-R and Mao J-F (2016) De Novo Transcriptome Assembly and Characterization for the Widespread and Stress-Tolerant Conifer *Platycladus orientalis*. *PLoS One* 11:e0148985. doi: 10.1371/journal.pone.0148985

Jacobsen SE, Running MP and Meyerowitz EM (1999) Disruption of an RNA helicase/RNase III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* 126:5231–43.

Jagadeeswaran G, Nimmakayala P, Zheng Y, Gowdu K, Reddy UK and Sunkar R (2012) Characterization of the small RNA component of leaves and fruits from four different cucurbit species. *BMC Genomics* 13:329. doi: 10.1186/1471-2164-13-329

Joseph JT, Poolakkalody NJ and Shah JM (2018) Plant reference genes for development and stress response studies. *J Biosci* 43:173–187. doi: 10.1007/s12038-017-9728-z

Kantar M, Lucas SJ and Budak H (2011) miRNA expression patterns of *Triticum dicoccoides* in response to shock drought stress. *Planta* 233:471–484. doi: 10.1007/s00425-010-1309-4

Kchouk M, Gibrat JF and Elloumi M (2017) Generations of Sequencing Technologies: From First to Next Generation. *Biol Med*. doi: 10.4172/0974-8369.1000395

Kozomara A and Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42:D68–D73. doi: 10.1093/nar/gkt1181

Kramer MF (2011) Stem-Loop RT-qPCR for miRNAs. *Curr Protoc Mol Biol* 95:15.10.1-15.10.15. doi: 10.1002/0471142727.mb1510s95

Kuang D-Y, Wu H, Wang Y-L, Gao L-M, Zhang S-Z and Lu L (2011) Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54:663–673. doi: 10.1139/g11-026

Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV and Margis R (2010)

The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Anal Biochem* 406:185–192. doi: 10.1016/j.ab.2010.07.020

Lee J, Kang Y, Shin SC, Park H and Lee H (2014) Combined Analysis of the Chloroplast Genome and Transcriptome of the Antarctic Vascular Plant *Deschampsia antarctica* Desv. *PLoS One* 9:e92501. doi: 10.1371/journal.pone.0092501

Lee RC, Feinbaum RL and Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–854. doi: 10.1016/0092-8674(93)90529-Y

Lei J and Sun Y (2014) miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data. *Bioinformatics* 30:2837–2839. doi: 10.1093/bioinformatics/btu380

Li Q-F, Zhang Y-C, Chen Y-Q and Yu Y (2017a) Circular RNAs roll into the regulatory network of plants. *Biochem Biophys Res Commun* 488:382–386. doi: 10.1016/j.bbrc.2017.05.061

Li Q, Deng C, Xia Y, Kong L, Zhang H, Zhang S and Wang J (2017b) Identification of novel miRNAs and miRNA expression profiling in embryogenic tissues of *Picea balfouriana* treated by 6-benzylaminopurine. *PLoS One* 12:e0176112. doi: 10.1371/journal.pone.0176112

Li X, Xie X, Li J, Cui Y, Hou Y, Zhai L, Wang X, Fu Y, Liu R and Bian S (2017c) Conservation and diversification of the miR166 family in soybean and potential roles of newly identified miR166s. *BMC Plant Biol* 17:32. doi: 10.1186/s12870-017-0983-9

Li Z-H, Zhu J, Yang Y-X, Yang J, He J-W and Zhao G-F (2015) The complete plastid genome of Bunge's pine *Pinus bungeana* (Pinaceae). *Mitochondrial DNA* 27:1–2. doi: 10.3109/19401736.2015.1060465

Lin C-P, Wu C-S, Huang Y-Y and Chaw S-M (2012) The Complete Chloroplast Genome of *Ginkgo biloba* Reveals the Mechanism of Inverted Repeat Contraction. *Genome Biol Evol* 4:374–381. doi: 10.1093/gbe/evs021

Liu H, Yu H, Tang G and Huang T (2018) Small but powerful: function of microRNAs in plant development. *Plant Cell Rep* 37:515–528. doi: 10.1007/s00299-017-2246-5

Liu J, Wang H and Chua N-H (2015) Long noncoding RNA transcriptome of plants. *Plant Biotechnol J* 13:319–328. doi: 10.1111/pbi.12336

Liu Y and El-Kassaby YA (2017) Landscape of fluid sets of hairpin-derived 21-/24-nt-long small RNAs at seed set uncovers special epigenetic features in *Picea glauca*. *Genome Biol Evol* 9:82–92. doi: 10.1093/gbe/evw283

Liu Y, Han S, Ding X, Li X, Zhang L, Li W, Xu H, Li Z and Qi L (2016) Transcriptome Analysis of mRNA and miRNA in Somatic Embryos of *Larix leptolepis* Subjected to Hydrogen Treatment. *Int J Mol Sci* 17:1951. doi: 10.3390/ijms17111951

Livak KJ and Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* 25:402–408. doi: 10.1006/meth.2001.1262

Longhi SJ, Brena DA, Ribeiro SB, Gracioli CR, Longhi RV and Mastella T (2009) Fatores ecológicos determinantes na ocorrência de *Araucaria angustifolia* e *Podocarpus lambertii*, na Floresta Ombrófila Mista da FLONA de São Francisco de Paula, RS, Brasil. *Ciência Rural* 40:57–63. doi: 10.1590/S0103-84782009005000220

Lowe AJ, Breed MF, Caron H, Colpaert N, Dick C, Finegan B, Gardner M, Gheysen G, Gribel R, Harris JBC et al. (2018) Standardized genetic diversity-life history correlates for improved genetic resource management of Neotropical trees. *Divers Distrib* 24:730–741. doi: 10.1111/ddi.12716

Lu S, Sun YH, Amerson H and Chiang VL (2007) MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. *Plant J* 51:1077–1098. doi: 10.1111/j.1365-313X.2007.03208.x

Lu Y, Ran JH, Guo DM, Yang ZY and Wang XQ (2014) Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. *PLoS One*. doi: 10.1371/journal.pone.0107679

Machado Mello AJ and Peroni N (2015) Cultural landscapes of the *Araucaria* Forests in the northern plateau of Santa Catarina, Brazil. *J Ethnobiol Ethnomed* 11:51. doi: 10.1186/s13002-015-0039-x

Mahfooz S, Singh SP, Rakh R, Bhattacharya A, Mishra N, Singh PC, Chauhan PS, Nautiyal CS and Mishra A (2016) A Comprehensive Characterization of Simple Sequence Repeats in the

Sequenced *Trichoderma* Genomes Provides Valuable Resources for Marker Development. *Front Microbiol* 7:1–12. doi: 10.3389/fmicb.2016.00575

Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ and Waterhouse PM (2006) The evolution and diversification of Dicers in plants. *FEBS Lett* 580:2442–2450. doi: 10.1016/j.febslet.2006.03.072

Martin JA and Wang Z (2011) Next-generation transcriptome assembly. *Nat Rev Genet* 12:671–682. doi: 10.1038/nrg3068

Martin LBB, Fei Z, Giovannoni JJ and Rose JKC (2013) Catalyzing plant science research with RNA-seq. *Front Plant Sci* 4:1–10. doi: 10.3389/fpls.2013.00066

McGettigan PA (2013) Transcriptomics in the RNA-seq era. *Curr Opin Chem Biol* 17:4–11. doi: 10.1016/j.cbpa.2012.12.008

Mica E (2006) Characterization of five microRNA families in maize. *J Exp Bot* 57:2601–2612. doi: 10.1093/jxb/erl013

Miskiewicz J, Tomczyk K, Mickiewicz A, Sarzynska J and Szachniuk M (2017) Bioinformatics Study of Structural Patterns in Plant MicroRNA Precursors. *Biomed Res Int* 2017:1–8. doi: 10.1155/2017/6783010

Moran Y, Agron M, Praher D and Technau U (2017) The evolutionary origin of plant and animal microRNAs. *Nat Ecol Evol* 1:0027. doi: 10.1038/s41559-016-0027

Morgado L and Johannes F (2017) Computational tools for plant small RNA detection and categorization. *Brief Bioinform* 4–10. doi: 10.1093/bib/bbx136

Morley SA and Nielsen BL (2016) Chloroplast DNA Copy Number Changes during Plant Development in Organelle DNA Polymerase Mutants. *Front Plant Sci* 7:1–10. doi: 10.3389/fpls.2016.00057

Omidvar V, Mohorianu I, Dalmay T and Fellner M (2015) Identification of miRNAs with potential roles in regulation of anther development and male-sterility in 7B-1 male-sterile tomato mutant. *BMC Genomics* 16:878. doi: 10.1186/s12864-015-2077-0

Ozyigit II, Dogan I and Filiz E (2015) In silico analysis of simple sequence repeats (SSRs) in

chloroplast genomes of *Glycine* species. *Plant Omics* 8:24–29.

Paludo GF, Duarte RI, Bernardi AP, Mantovani A and Reis MS dos (2016) THE SIZE OF *Araucaria angustifolia* (Bertol.) Kuntze ENTERING INTO REPRODUCTIVE STAGES AS A BASIS FOR SEED MANAGEMENT PROJECTS. *Rev Árvore* 40:695–705. doi: 10.1590/0100-67622016000400013

Pegler J, Grof C and Eamens A (2018) Profiling of the Differential Abundance of Drought and Salt Stress-Responsive MicroRNAs Across Grass Crop and Genetic Model Plant Species. *Agronomy* 8:118. doi: 10.3390/agronomy8070118

Pellicer J, Hidalgo O, Dodsworth S and Leitch IJ (2018) Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes (Basel)* 9:88. doi: 10.3390/genes9020088

Qin Z, Li C, Mao L and Wu L (2014) Novel insights from non-conserved microRNAs in plants. *Front Plant Sci* 5:1–10. doi: 10.3389/fpls.2014.00586

Reinhart BJ (2002) MicroRNAs in plants. *Genes Dev* 16:1616–1626. doi: 10.1101/gad.1004402

Reis MS dos, Ladio A and Peroni N (2014) Landscapes with *Araucaria* in South America: evidence for a cultural dimension. *Ecol Soc* 19:art43. doi: 10.5751/ES-06163-190243

Rodrigues NF, Christoff AP, da Fonseca GC, Kulcheski FR and Margis R (2017) Unveiling Chloroplast RNA Editing Events Using Next Generation Small RNA Sequencing Data. *Front Plant Sci*. doi: 10.3389/fpls.2017.01686

Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, Mei W, Thomas PI, Gardner MF, Ennos RA et al. (2015) Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour* 15:1067–1078. doi: 10.1111/1755-0998.12375

Sablok G, Mudunuri SB, Edwards D and Ralph PJ (2016) Chloroplast genomics: Expanding resources for an evolutionary conserved miniature molecule with enigmatic applications. *Curr Plant Biol* 7–8:34–38. doi: 10.1016/j.cpb.2016.12.004

Samad AFA, Sajad M, Nazaruddin N, Fauzi IA, Murad AMA, Zainal Z and Ismail I (2017) MicroRNA and Transcription Factor: Key Players in Plant Regulatory Network. *Front Plant Sci*

8:1–18. doi: 10.3389/fpls.2017.00565

Santos ALW dos, Silveira V, Steiner N, Maraschin M and Guerra MP (2010) Biochemical and morphological changes during the growth kinetics of *Araucaria angustifolia* suspension cultures. *Brazilian Arch Biol Technol* 53:497–504. doi: 10.1590/S1516-89132010000300001

Santos ALW dos, Silveira V, Steiner N, Vidor M and Guerra MP (2002) Somatic Embryogenesis in Parana Pine (*Araucaria angustifolia* (Bert.) O. Kuntze). *Brazilian Arch Biol Technol* 45:97–106. doi: 10.1590/S1516-89132002000100015

Sato S, Yoshida M, Hiraide H, Ihara K and Yamamoto H (2014) Transcriptome Analysis of Reaction Wood in Gymnosperms by Next-Generation Sequencing. *Am J Plant Sci* 05:2785–2798. doi: 10.4236/ajps.2014.518295

Schliesky S, Gowik U, Weber APM and Bräutigam A (2012) RNA-Seq Assembly – Are We There Yet? *Front Plant Sci* 3:1–12. doi: 10.3389/fpls.2012.00220

Shi C, Wang S, Xia E-H, Jiang J-J, Zeng F-C and Gao L-Z (2016) Full transcription of the chloroplast genome in photosynthetic eukaryotes. *Sci Rep* 6:30135. doi: 10.1038/srep30135

Shi T, Wang K and Yang P (2017) The evolution of plant microRNAs: insights from a basal eudicot sacred lotus. *Plant J* 89:442–457. doi: 10.1111/tpj.13394

Shibata M, Coelho CMM, Araldi CG, Adan N and Peroni N (2016) Physiological and physical quality of local *Araucaria angustifolia* seed variety. *Acta Sci Agron* 38:249. doi: 10.4025/actasciagron.v38i2.27976

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5:2043–2049. doi: 10.1002/j.1460-2075.1986.tb04464.x

Silveira V, Santa-Catarina C, Balbuena TS, Moraes FMS, Ricart CAO, Sousa M V., Guerra MP, Handro W and Floh EIS (2008) Endogenous abscisic acid and protein contents during seed development of *Araucaria angustifolia*. *Biol Plant* 52:101–104. doi: 10.1007/s10535-008-0018-3

Slatko BE, Gardner AF and Ausubel FM (2018) Overview of Next-Generation Sequencing

Technologies. *Curr Protoc Mol Biol* 122:e59. doi: 10.1002/cpmb.59

Smith DR (2013) RNA-Seq data: a goldmine for organelle research. *Brief Funct Genomics* 12:454–456. doi: 10.1093/bfgp/els066

Smith DR and Sanitá Lima M (2016) Unraveling chloroplast transcriptomes with ChloroSeq, an organelle RNA-Seq bioinformatics pipeline. *Brief Bioinform* 18:bbw088. doi: 10.1093/bib/bbw088

SOUZA AF (2007) Ecological interpretation of multiple population size structures in trees: The case of *Araucaria angustifolia* in South America. *Austral Ecol* 32:524–533. doi: 10.1111/j.1442-9993.2007.01724.x

Souza MIF de, Salgueiro F, Carnavale-Bottino M, Félix DB, Alves-Ferreira M, Bittencourt JVM and Margis R (2009) Patterns of genetic diversity in southern and southeastern *Araucaria angustifolia* (Bert.) O. Kuntze relict populations. *Genet Mol Biol* 32:546–556. doi: 10.1590/S1415-47572009005000052

Spies D, Renz PF, Beyer TA and Ciaudo C (2017) Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief Bioinform* 1–11. doi: 10.1093/bib/bbx115

Srivastava PK, Moturu T, Pandey P, Baldwin IT and Pandey SP (2014) A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics* 15:348. doi: 10.1186/1471-2164-15-348

Stegemann S, Keuthe M, Greiner S and Bock R (2012) Horizontal transfer of chloroplast genomes between plant species. *Proc Natl Acad Sci* 109:2434–2438. doi: 10.1073/pnas.1114076109

Stein S, Bahrami-samani E and Xing Y (2011) RNA Processing. doi: 10.5772/832

Steiner N, Claudete S, Andrade JBR, Balbuena TS, Guerra MP, Handro W, Floh EIS and Silveira V (2008) *Araucaria angustifolia* Biotechnology. *Funct Plant Sci Biotechnol* 2:20–28.

Stern DB, Goldschmidt-Clermont M and Hanson MR (2010) Chloroplast RNA Metabolism. *Annu Rev Plant Biol* 61:125–155. doi: 10.1146/annurev-arplant-042809-112242

Sudianto E, Wu C-S, Lin C-P and Chaw S-M (2016) Revisiting the Plastid Phylogenomics of

Pinaceae with Two Complete Plastomes of *Pseudolarix* and *Tsuga*. *Genome Biol Evol* 8:1804–1811. doi: 10.1093/gbe/evw106

Tagliari MM and Peroni N (2018) Local varieties of *Araucaria angustifolia* ( Bertol .) Kuntze ( Pinales : Araucariaceae ) in southern Brazil : A brief discussion about landscape domestication. *Biotemas* 31:59–68. doi: doi.org/10.5007/2175-7925.2018v31n3p59

Tao K, Gao L, Li J, Chen S, Su Y and Wang T (2016) The complete chloroplast genome of *Torreya fargesii* (Taxaceae). *Mitochondrial DNA Part A* 27:3512–3513. doi: 10.3109/19401736.2015.1074195

Taylor RS, Tarver JE, Hiscock SJ and Donoghue PCJ (2014) Evolutionary history of plant microRNAs. *Trends Plant Sci* 19:175–182. doi: 10.1016/j.tplants.2013.11.008

Thomas P (2013) *Araucaria angustifolia*. IUCN Red List Threat Species 2013 eT32975A2829141. doi: 10.2305/IUCN.UK.2013-1.RLTS.T32975A2829141.en

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL and Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–578. doi: 10.1038/nprot.2012.016

Ujino-Ihara T, Ueno S, Uchiyama K and Futamura N (2018) Comprehensive analysis of small RNAs expressed in developing male strobili of *Cryptomeria japonica*. *PLoS One* 13:e0193665. doi: 10.1371/journal.pone.0193665

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A and Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034. doi: 10.1007/s00603-018-1496-z

Varkonyi-gasic E and Varkonyi-gasic E (2017) *Plant Epigenetics*. doi: 10.1007/978-3-319-55520-1

Velayudha Vimala Kumar K, Srikakulam N, Padbhanabhan P and Pandi G (2017) Deciphering microRNAs and Their Associated Hairpin Precursors in a Non-Model Plant, *Abelmoschus esculentus*. *Non-Coding RNA* 3:19. doi: 10.3390/ncrna3020019

Vieira L do N, Faoro H, Rogalski M, Fraga HP de F, Cardoso RLA, de Souza EM, de Oliveira



- Pedrosa F, Nodari RO and Guerra MP (2014) The Complete Chloroplast Genome Sequence of *Podocarpus lambertii*: Genome Structure, Evolutionary Aspects, Gene Content and SSR Detection. *PLoS One* 9:e90618. doi: 10.1371/journal.pone.0090618
- Vieira MLC, Santini L, Diniz AL and Munhoz C de F (2016) Microsatellite markers: What they mean and why they are so useful. *Genet Mol Biol* 39:312–328. doi: 10.1590/1678-4685-GMB-2016-0027
- Voshall A and Moriyama EN (2018) Next-Generation Transcriptome Assembly: Strategies and Performance Analysis. *Bioinforma Era Post Genomics Big Data*. doi: 10.5772/intechopen.73497
- Wan L-C, Wang F, Guo X, Lu S, Qiu Z, Zhao Y, Zhang H and Lin J (2012a) Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. *BMC Plant Biol* 12:146. doi: 10.1186/1471-2229-12-146
- Wan LC, Zhang H, Lu S, Zhang L, Qiu Z, Zhao Y, Zeng QY and Lin J (2012b) Transcriptome-wide identification and characterization of miRNAs from *Pinus densata*. *BMC Genomics*. doi: 10.1186/1471-2164-13-132
- Wang J-F (2004) Identification of 20 microRNAs from *Oryza sativa*. *Nucleic Acids Res* 32:1688–1695. doi: 10.1093/nar/gkh332
- Wang Y, Wang Z, Zhang L and Yang X (2018) The complete chloroplast genome of *Cycas Szechuanensis*, an extremely endangered species. *Mitochondrial DNA Part B* 3:974–975. doi: 10.1080/23802359.2018.1507635
- Ward JA, Ponnala L and Weber CA (2012) Strategies for transcriptome analysis in nonmodel plants. *Am J Bot* 99:267–276. doi: 10.3732/ajb.1100334
- Weber APM (2015) Discovering new biology through RNA-Seq. *Plant Physiol* 169:pp.01081.2015. doi: 10.1104/pp.15.01081
- Wu C-S and Chaw S-M (2014) Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnol J* 12:344–353. doi: 10.1111/pbi.12141
- Wu CS, Wang YN, Hsu CY, Lin CP and Chaw SM (2011) Loss of different inverted repeat copies

from the chloroplast genomes of pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol Evol* 3:1284–1295. doi: 10.1093/gbe/evr095

Wu TD and Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873–881. doi: 10.1093/bioinformatics/btq057

Wu Y, Guo J, Zhou Q, Xin Y, Wang G and Xu L (2018) De novo transcriptome analysis revealed genes involved in flavonoid biosynthesis, transport and regulation in *Ginkgo biloba*. *Ind Crops Prod* 124:226–235. doi: 10.1016/j.indcrop.2018.07.060

Xu J-H, Liu Q, Hu W, Wang T, Xue Q and Messing J (2015) Dynamics of chloroplast genomes in green plants. *Genomics* 106:221–231. doi: 10.1016/j.ygeno.2015.07.004

Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, Yang L and Poethig RS (2016) Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in *Arabidopsis thaliana*. *PLoS Genet* 12:1–29. doi: 10.1371/journal.pgen.1006263

Xu Y, Zhang L and Wu G (2018) Epigenetic Regulation of Juvenile-to-Adult Transition in Plants. *Front Plant Sci* 9:1–8. doi: 10.3389/fpls.2018.01048

Yagi Y and Shiina T (2014) Recent advances in the study of chloroplast gene expression and its evolution. *Front Plant Sci* 5:1–7. doi: 10.3389/fpls.2014.00061

Yakovlev IA, Fossdal CG and Johnsen Ø (2010) MicroRNAs, the epigenetic memory and climatic adaptation in Norway spruce. *New Phytol* 187:1154–1169. doi: 10.1111/j.1469-8137.2010.03341.x

Yang J, Vázquez L, Chen X, Li H, Zhang H, Liu Z and Zhao G (2017) Development of Chloroplast and Nuclear DNA Markers for Chinese Oaks (*Quercus* Subgenus *Quercus*) and Assessment of Their Utility as DNA Barcodes. *Front Plant Sci* 8:1–17. doi: 10.3389/fpls.2017.00816

Yang T, Wang Y, Teotia S, Zhang Z and Tang G (2018) The Making of Leaves: How Small RNA Networks Modulate Leaf Development. *Front Plant Sci* 9:1–7. doi: 10.3389/fpls.2018.00824

Yang X and Li L (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 27:2614–2615. doi: 10.1093/bioinformatics/btr430

Yao Y, Han R, Gong Z, Zheng C and Zhao Y (2018) RNA-seq Analysis Reveals Gene Expression

Profiling of Female Fertile and Sterile Ovules of *Pinus Tabulaeformis* Carr. during Free Nuclear Mitosis of the Female Gametophyte. *Int J Mol Sci* 19:2246. doi: 10.3390/ijms19082246

Yap J-YS, Rohner T, Greenfield A, Van Der Merwe M, McPherson H, Glenn W, Kornfeld G, Marendy E, Pan AYH, Wilton A et al. (2015) Complete Chloroplast Genome of the Wollemi Pine (*Wollemia nobilis*): Structure and Evolution. *PLoS One* 10:e0128126. doi: 10.1371/journal.pone.0128126

Yi D-K, Yang JC, So S, Joo M, Kim D-K, Shin CH, Lee Y-M and Choi K (2015) The complete plastid genome sequence of *Abies koreana* (Pinaceae: Abietoideae). *Mitochondrial DNA* 27:1–3. doi: 10.3109/19401736.2015.1025260

Yu A, Saudemont B, Bouteiller N, Elvira-Matelot E, Lepère G, Parent J-S, Morel J-B, Cao J, Elmayer T and Vaucheret H (2015) Second-Site Mutagenesis of a Hypomorphic *argonaute1* Allele Identifies SUPERKILLER3 as an Endogenous Suppressor of Transgene Posttranscriptional Gene Silencing. *Plant Physiol* 169:1266–1274. doi: 10.1104/pp.15.00585

Yu Y, Jia T and Chen X (2017) The ‘how’ and ‘where’ of plant microRNAs. *New Phytol* 216:1002–1017. doi: 10.1111/nph.14834

Zeng X, Xu Y, Jiang J, Zhang F, Ma L, Wu D, Wang Y and Sun W (2018) Identification of cold stress responsive microRNAs in two winter turnip rape (*Brassica rapa* L.) by high throughput sequencing. *BMC Plant Biol* 18:52. doi: 10.1186/s12870-018-1242-4

Zhang L, Zheng Y, Jagadeeswaran G, Li Y, Gowdu K and Sunkar R (2011) Identification and temporal expression analysis of conserved and novel microRNAs in *Sorghum*. *Genomics* 98:460–468. doi: 10.1016/j.ygeno.2011.08.005

Zhang M, Dong Y, Nie L, Lu M, Fu C and Yu L (2015a) High-throughput sequencing reveals miRNA effects on the primary and secondary production properties in long-term subcultured *Taxus* cells. *Front Plant Sci* 6:604. doi: 10.3389/fpls.2015.00604

Zhang Q, Li J, Sang Y, Xing S, Wu Q and Liu X (2015b) Identification and Characterization of MicroRNAs in *Ginkgo biloba* var. *epiphylla* Mak. *PLoS One* 10:e0127184. doi: 10.1371/journal.pone.0127184

Zhang Q, Zhao C, Li M, Sun W, Liu Y, Xia H, Sun M, Li A, Li C, Zhao S et al. (2013) Genome-

wide identification of *Thellungiella salsuginea* microRNAs with putative roles in the salt stress response. *BMC Plant Biol* 13:180. doi: 10.1186/1471-2229-13-180

Zhang Y, Wang Y, Gao X, Liu C and Gai S (2018) Identification and characterization of microRNAs in tree peony during chilling induced dormancy release by high-throughput sequencing. *Sci Rep* 8:4537. doi: 10.1038/s41598-018-22415-5

Zhang Y, Xia R, Kuang H and Meyers BC (2016) The Diversification of Plant NBS-LRR Defense Genes Directs the Evolution of MicroRNAs That Target Them. *Mol Biol Evol* 33:2692–2705. doi: 10.1093/molbev/msw154

Zhou T, Chen C, Wei Y, Chang Y, Bai G, Li Z, Kanwal N and Zhao G (2016) Comparative Transcriptome and Chloroplast Genome Analyses of Two Related Dipteronia Species. *Front Plant Sci* 7:1–13. doi: 10.3389/fpls.2016.01512

Zhu X, Leng X, Sun X, Mu Q, Wang B, Li X, Wang C and Fang J (2015) Discovery of Conservation and Diversification of Genes by Phylogenetic Analysis based on Global Genomes. *Plant Genome* 8:1–11. doi: 10.3835/plantgenome2014.10.0076

Zielezinski A, Dolata J, Alaba S, Kruszka K, Pacak A, Swida-Barteczka A, Knop K, Stepien A, Bielewicz D, Pietrykowska H et al. (2015) mirEX 2.0 - an integrated environment for expression profiling of plant microRNAs. *BMC Plant Biol* 15:144. doi: 10.1186/s12870-015-0533-2

Zoschke R and Bock R (2018) Chloroplast Translation: Structural and Functional Organization, Operational Control, and Regulation. *Plant Cell* 30:745–770. doi: 10.1105/tpc.18.00016